THE MATHEMATICAL THEORY OF HUYGENS' PRINCIPLE

THE MATHEMATICAL THEORY OF HUYGENS' PRINCIPLE

BY

BEVAN B. BAKER

FORMERLY PROFESSOR OF MATHEMATICS AT THE ROYAL HOLLOWAY COLLEGE IN THE UNIVERSITY OF LONDON

AND

E. T. COPSON

PROFESSOR OF MATHEMATICS AT UNIVERSITY COLLEGE, DUNDEE IN THE UNIVERSITY OF ST. ANDREWS

SECOND EDITION

OXFORD
AT THE CLARENDON PRESS
1950

Oxford University Press, Amen House, London E.C.4
GLASGOW NEW YORK TORONTO MELBOURNE WELLINGTON
BOMBAY CALCUTTA MADRAS CAPE TOWN
Geoffrey Cumberlege, Publisher to the University

FIRST EDITION 1939

Reprinted lithographically at the University Press, Oxford, from corrected sheets of the first edition.

PREFACE

STIMULATED by a course of post-graduate lectures on the Partial Differential Equations of Mathematical Physics which Professor E. T. Whittaker gave sixteen years ago in the Mathematical Institute of Edinburgh University, one of the authors of the present work (B. B. B.) planned a comprehensive treatise covering the whole of this field. Unfortunately, ill health and pressure of other duties have, so far, prevented the completion of this scheme. In the meantime the subject has been treated from different points of view by Bateman (1932), Courant and Hilbert (1924 and 1938), and Webster (1927).

During the same period there have been great advances in mathematical physics, especially in the various developments of quantum mechanics. As these new theories are still developing rapidly, it would perhaps be unwise to attempt at the present juncture another general treatise on the mathematics of physics: and, after much consideration, we have decided to abandon the original plan and to replace it by the publication of a number of monographs, each complete in itself, on various special topics not adequately treated in existing books.

The present monograph deals with the mathematical theory of Huygens' principle in optics and its application to the theory of diffraction. No attempt is made to give a complete account of the various methods of solving special diffraction problems. We are concerned only with the general theory of the solution of the partial differential equations governing the propagation of light and we discuss some of the simpler diffraction problems merely as illustrative examples. For an account of the more technical developments of the theory of diffraction we refer the reader to the excellent articles by von Laue and Epstein in the *Encyklopädie der mathematischen Wissenschaften* (Band V, 3. Teil) and that of Wolfsohn in the *Handbuch der Physik* (Band XX—'Licht als Wellenbewegung').

The standard of knowledge of pure and applied mathematics expected of the reader is roughly that of the undergraduate who has completed the compulsory part of an honours course and is about to take up some specialized study.

We wish to express here our great indebtedness to Professor Whittaker for the original stimulus which led us to this work and for his continued interest, encouragement, and advice. We also desire to thank the Delegates of the Clarendon Press for undertaking this book and their Staff for their unfailing skill in printing it.

B. B. B.
E. T. C.

PREFACE TO THE SECOND EDITION

ADVANTAGE has been taken of the preparation of a new edition of this monograph to add a chapter on the application of the theory of integral equations to problems of diffraction by a plane screen. This method goes back in principle to the work of Lord Rayleigh in the second volume of his book on 'The Theory of Sound', and has gained importance during the war in the use of diffraction theory in radio problems. It is hoped that this chapter will prove a useful introduction to the growing literature of the subject.

The first four chapters of the book are virtually unchanged, apart from the correction of minor errors and misprints and the addition of references to more recent work.

My thanks are due to the Delegates of the Clarendon Press for publishing a second edition and to their Staff who continue to maintain their reputation for fine printing.

June 1949 E. T. C.

CONTENTS

I

THE ANALYTICAL REPRESENTATION OF HUYGENS' PRINCIPLE

§ 1. The principle of Huygens

§ 1.1. Huygens' geometrical theory of wave-propagation in optics

In his *Traité de la Lumière*, published in 1690, Huygens discussed the process of the propagation of light by the aid of a new principle, which has since been generally known by his name. At that time† light was regarded as a disturbance in a medium, the aether, of much the same general character as sound in air. Huygens supposed that, at any instant $t = t_0$, a point-source of light generates a disturbance which is propagated into the surrounding medium as an isolated spherical wave‡ which expands with a large constant velocity, the velocity of light. This initial disturbance develops into the disturbance at the subsequent instant $t = t_1$ through a succession of states at the intermediate instants, and a knowledge of the state at any intermediate instant $t = t'$ suffices to determine the state at the instant $t = t_1$. Thus, if we regard each element of the isolated wave at the instant $t = t'$ as the centre of a new disturbance, the actual effect at the instant $t = t_1$ is the resultant of all these secondary effects, the actual wave the envelope of the secondary waves.

The principle, stated in this form, is somewhat vague, and so, before we go on to discuss the difficulties of Huygens' geometrical theory of wave-propagation in optics, it is convenient to follow Hadamard|| and analyse the principle in the form of a syllogism.

A. (Major premiss.) In order to determine the effect at the instant $t = t_1$ of a luminous phenomenon caused by a given disturbance at the initial instant $t = t_0$, we may calculate the state at some intermediate instant $t = t'$ and from that deduce the state at the instant $t = t_1$.

B. (Minor premiss.) If at the instant $t = t_0$, or, more precisely,

† For the history of the development of the theories of light, see E. T. Whittaker, *History of the Theories of Aether and Electricity* (Dublin, 1910).

‡ An isolated spherical wave, or pulse as it is sometimes called, is a disturbance of the medium localized on the surface of a sphere.

|| See Hadamard's lecture 'Le Principe de Huygens', *Bull. de la Soc. math. de France*, **52** (1924), 610–40, or his book *Lectures on Cauchy's Problem* (Yale, 1923), 53–6.

in the short interval $t_0-\epsilon \leqslant t \leqslant t_0$, we produce a luminous disturbance localized in the immediate neighbourhood of a point O, the effect at the subsequent instant $t = t'$ is localized in a very thin spherical shell with centre O and radius $c(t'-t_0)$, where c is the velocity of light.

C. (Conclusion.) In order to calculate the effect at the instant $t = t_1$ due to a luminous disturbance localized at O at the instant $t = t_0$, we may replace the initial disturbance by a suitable system of luminous disturbances taking place at the intermediate instant $t = t'$ and distributed over the surface of the sphere with centre O and radius $c(t'-t_0)$.

When analysed in this way, Huygens' principle is seen to involve three propositions, and different authors have attached the name 'Huygens' principle' indiscriminately to any one of the three. In the present book we are concerned with proposition C and its generalizations involving luminous disturbances which cannot be generated by the superposition of spherical waves.

Proposition A would probably be accepted as immediately obvious: it is nothing other than the principle of determinism which runs all through classical mathematical physics. Nevertheless, the fact that the solutions of the differential equations which govern the propagation of light do satisfy this proposition is of considerable interest, for it leads to certain remarkable relations connecting these solutions.†

Proposition A is what the philosophers would describe as one of the laws of thought—its contrary is inconceivable. Proposition C is a physical law capable of very wide generalization. Proposition B, on the other hand, has a very special character, since it is a property peculiar to certain special types of luminous phenomena. It states that an isolated spherical light wave has clean-cut propagation; for if such a wave is due to a disturbance localized at the origin and acting only during the very short interval $t_0-\epsilon \leqslant t \leqslant t_0$, the effect at a distance cT is null until the instant $t = t_0+T-\epsilon$ and is null again after the instant $t = t_0+T$; an isolated spherical wave leaves no residual after-effect. We cannot generalize proposition B to cover, for example, two-dimensional wave-motions; for, in two dimensions, an initial disturbance always gives rise to a residual after-effect.

† Hadamard, *Bull. de la Soc. math. de France*, **52** (1924), 241–78; *Acta Math.* **49** (1926), 203–44; *Journal de Math.* **8** (1929), 197–228. See also E. Hille, *Functional Analysis and Semi-Groups*, (New York, 1948), Ch. XX.

To sum up, although the premisses A and B do imply the conclusion C, an argument of this type is incapable of generalization owing to the great restrictions under which the proposition B holds. We shall prove a general form of proposition C, and then show that B is a consequence of C for the special type of spherical wave considered by Huygens.

Throughout the book we are concerned only with the generalizations of C which are, in effect, governed by the partial differential equation of wave-motions. For the still wider generalizations concerning other differential equations we refer the reader to Hadamard's Yale lectures already cited.

§ 1.2. The difficulties of Huygens' theory

In applying his geometrical theory of wave-propagation Huygens encountered certain difficulties which he was able to overcome only by making special *ad hoc* hypotheses. In the first place, he found that he could account for the rectilinear propagation of light only by assuming that a secondary wave has no effect except at the point where it touches its envelope. Secondly, the envelope of the secondary spherical waves consists of two sheets, one on each side of the surface on which the secondary sources of disturbance lie. It would seem, therefore, that one of Huygens' isolated waves would be propagated not only forwards but also backwards. To get over this difficulty Huygens had to assume that only one sheet of the envelope is to be considered.

If we wish to avoid making this last assumption, we must give up the purely geometrical theory and have recourse to analysis. To illustrate this, let us consider plane waves of sound of small amplitude. We shall prove in § 3.2 that an initial disturbance is actually propagated in both directions unless certain conditions are satisfied. Only when the initial values q_0 and s_0 of the velocity and condensation are connected by one or other of the relations $q_0 \pm cs_0 = 0$ do we get a plane wave which is propagated in a definite direction.†

A similar conclusion holds for electromagnetic waves in a vacuum: in a progressive plane electromagnetic wave-motion whose wave fronts are perpendicular to the vector $\mathbf{n}$ and move in the direction

† See a letter from Fresnel to Poisson (*Œuvres complètes de Fresnel*, **2**, 227: quoted by Poincaré, *Theorie math. de la Lumière*, **1** (1889), 81; Croze, *Annales de Physique*, **5** (1926), 371–439 (380–1)) on this point. For a similar result concerning general sound waves, see Love, *Proc. London Math. Soc.* (2), **1** (1903), 37–62 (54).

of $\mathbf{n}$, the electric and magnetic vectors $\mathbf{d}$ and $\mathbf{h}$ must be equal in magnitude and $\mathbf{n}$, $\mathbf{d}$, and $\mathbf{h}$ must form a set of right-handed orthogonal axes.

§ 2. Huygens' construction as a contact-transformation

§ 2.1. The definition of a contact-transformation

A surface element at a point P of space is specified by the coordinates (x, y, z) of P and the direction cosines (l, m, n) of the normal to the element. Let a transformation from the set of variables (x, y, z, l, m, n) to the set (x', y', z', l', m', n') be regarded as turning this surface element into another surface element at P' (x', y', z'), the normal to which has direction cosines (l', m', n'). There are ∞^2 surface elements through any given point P; to each corresponds in general a surface element through a different point P', since the coordinates (x', y', z') depend on (l, m, n) as well as on the coordinates of P.

From the equations which define (x', y', z', l', m', n') in terms of (x, y, z, l, m, n) it may be possible to eliminate completely the direction cosines so as to obtain one or more relations between the coordinates of P and P'.

There are three cases to be considered:

(*a*) There may be only a single relation

$$\Omega(x, y, z, x', y', z') = 0.$$

When (x, y, z) are given, this equation, regarded as an equation in (x', y', z'), represents a surface. Thus each point P, or, more precisely, the set of surface elements through P, is transformed into a surface Ω_P.

(*b*) There may be two relations

$$\Omega_1(x, y, z, x', y', z') = 0, \qquad \Omega_2(x, y, z, x', y', z') = 0.$$

Then each point P is transformed into a curve.

(*c*) There may be three relations, in which case each point P is transformed into a point P'.

Let us restrict our attention to transformations of the type (*a*), and consider the effect of applying such a transformation to the surface elements of a given surface S. It may happen that, no matter what surface S is chosen, the transformed surface elements build up another surface S'; if this is so, the transformation is called a *contact-transformation*.† It can be shown that, in this case, the surface S' is the envelope of the surfaces Ω_P corresponding to the individual points P of S.

A contact-transformation of this type is precisely the sort of transformation which appears in Huygens' geometrical construction. For each point of the wave-front S at the instant t is transformed into a sphere of radius $c(t'-t)$, and the wave-front S' at the instant t' is the envelope of these spheres. The equations of the transformation are

$$\begin{aligned} x' &= x+c(t'-t)l, \qquad & l' &= l, \\ y' &= y+c(t'-t)m, \qquad & m' &= m, \\ z' &= z+c(t'-t)n, \qquad & n' &= n. \end{aligned}$$

§ 2.2. The analogy with dynamical systems

In the modern theory of general dynamics, contact-transformations play an important part, since the history of any dynamical system may be regarded as the gradual self-unfolding of a contact-transformation. Consider a dynamical system in which $(q_1, q_2,..., q_n)$ are the generalized coordinates specifying the state at the instant t; let $(p_1, p_2,..., p_n)$ be the corresponding generalized momenta. Then if $(q_1', q_2',..., q_n')$ and $(p_1', p_2',..., p_n')$ are the values of the coordinates and momenta at the instant t', the transformation of the set of variables $(q_1, q_2,..., q_n;\ p_1, p_2,..., p_n)$ into the corresponding set of accented variables is a contact-transformation.‡

Placing this result beside the statement that the simple geometrical construction suggested by Huygens is a contact-transformation, it is natural to conjecture that the general analytical expression of Huygens' principle ought to involve a contact-transformation, generalizing the result just enunciated for finite systems to dynamical systems with an infinite number of degrees of freedom.

Actually Huygens' ideas have not been developed on these lines. The reason for this is that, in the practical applications of Huygens' principle, radiation, the disturbance in the medium which we may for convenience call the aether, generally proceeds from sources.

† The name is due to S. Lie. It was suggested by the evident fact that, if two surfaces are in contact, so also are their transforms by a contact-transformation.

‡ For the theory of the application of contact-transformations to dynamics, see E. T. Whittaker, *Analytical Dynamics* (Cambridge, 1917), Ch. XI et seq.; Prange, *Encyc. der math. Wissenschaften*, Band IV, 1_2, Heft 4.

These sources are, from the dynamical point of view, singularities at which energy is introduced into the aether; the existence of these singularities prevents us from developing Huygens' principle in the way which would be natural if we were dealing with a self-contained conservative dynamical system.

In particular, the solution of a dynamical problem with a finite number of degrees of freedom is a solution valid for all values of t, both subsequent and antecedent to the instant t_0, whereas in radiation problems we cannot trace the radiation back beyond the instant when it issued from the source. The formulae which will be useful in the practical applications of Huygens' principle will generally be formulae which are valid only at instants t subsequent to some initial instant t_0; the results obtained by substituting values of t less than t_0 need bear no relation whatever to the actual phenomenon.

§ 3. The propagation of sound waves in air

§ 3.1. The differential equations of sound waves of small amplitude

Huygens' principle in its crudest form takes no account whatever of the phenomenon of polarization, although this phenomenon was discovered by Huygens himself in his experimental work on Iceland spar. Until the time of Young and Fresnel, light was regarded as a disturbance in a medium analogous to that of sound in air. We know now that the propagation of light is of an entirely different character from that of sound. To specify a light wave, we need to know the three components of the 'light-vector', whereas a sound wave is specified by a single quantity, the scalar velocity potential. There is, then, no precise analogy between the propagation of sound and the propagation of light.

We now proceed to consider the problem of expressing in an analytical form the principle of Huygens for a scalar phenomenon, namely, the propagation of sound waves of small amplitude. This will serve as an introduction to the vector form of Huygens' principle, which is based on the electromagnetic theory of light. It will, moreover, provide a justification for Huygens' principle in optics as that subject was understood in the days before Young and Fresnel.

We begin with a brief sketch of the theory of the propagation of sound waves of small amplitude. We denote the velocity of the medium at the point (x, y, z) at the instant t by $\mathbf{q}$, where q^2 is negligible. We suppose that the motion is irrotational, so that $\mathbf{q}$ is derived from a velocity potential u by the equation

$$\mathbf{q} = -\operatorname{grad} u, \tag{3.11}$$

where $\operatorname{grad} u$ denotes the vector with components $(\partial u/\partial x, \partial u/\partial y, \partial u/\partial z)$. A knowledge of $\mathbf{q}$ alone does not specify the state of the medium at (x, y, z) at the instant t; it is necessary to know in addition the pressure p and the density ρ of the medium. Actually it is more convenient to consider, instead of ρ, the *condensation* s, defined by

$$\rho = \rho_0(1+s),$$

where ρ_0 is the density in the undisturbed state. In dealing with sound waves of small amplitude, the square of s is negligible.

The motion is governed by two equations, a kinematical and a dynamical equation. The kinematical equation is the *equation of continuity*†

$$\dot{\rho}+\operatorname{div}(\rho\mathbf{q}) = 0.$$

For sound waves of small amplitude this reduces to

$$\operatorname{div}\mathbf{q} = -\dot{s}. \tag{3.12}$$

The dynamical equation is the expression in vector form of *Euler's Dynamical Equations*,‡ namely

$$\rho\dot{\mathbf{q}}+(\rho\mathbf{q}\cdot\nabla)\mathbf{q} = -\operatorname{grad} p.$$

It being assumed that p is a function of ρ alone, this equation simplifies to

$$\rho_0\,\dot{\mathbf{q}} = -\left(\frac{dp}{d\rho}\right)_0 \operatorname{grad}\rho,$$

where the suffix 0 denotes the value in the undisturbed state. From this we have

$$\dot{\mathbf{q}} = -c^2 \operatorname{grad} s, \tag{3.13}$$

where $c^2 = (dp/d\rho)_0$. The constant c has the dimensions of a velocity.

From equations (3.11), (3.12), and (3.13) it follows that

$$\frac{\partial s}{\partial t} = \nabla^2 u, \qquad c^2 s = \frac{\partial u}{\partial t}, \tag{3.14}$$

where ∇^2 is Laplace's operator‖

$$\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}.$$

† See, for example, Ramsey's *Hydrodynamics* (1920), 5 (1).

‡ Ibid. 17 (1). The symbol ∇ denotes the vector operator $\mathbf{i}\frac{\partial}{\partial x}+\mathbf{j}\frac{\partial}{\partial y}+\mathbf{k}\frac{\partial}{\partial z}$; the full-stop denotes the scalar product.

‖ It is the square of the vector operator ∇ which occurs in the preceding footnote.

Hence the velocity potential u satisfies the equation

$$\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}.$$

This is called the *equation of wave-motions*, any solution of it a *wave-function*. Evidently s and each component of $\mathbf{q}$ are wave-functions.

§ 3.2. Plane waves of sound

A wave function u which depends only on t and one of the Cartesian coordinates, x say, satisfies the equation

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2},$$

which is called the one-dimensional wave-equation. Its general solution is

$$u = f(ct-x)+F(ct+x), \tag{3.21}$$

where f and F denote arbitrary functions.

The particular solution

$$u = f(ct-x) \tag{3.22}$$

is the velocity potential of a disturbance which is propagated parallel to the axis of x with velocity c; for the disturbance at the instant t at a point of coordinate x is the same as that at the initial instant $t = 0$ at the point of coordinate $x-ct$. In other words, the expression (3.22) is the velocity potential of a plane sound wave which is propagated without change of type parallel to Ox with velocity c. The constant c is then called the *velocity of sound*. In virtue of equations (3.11) and (3.14), the particle velocity q (which is parallel to Ox) and the condensation s are given by

$$q = cs = f'(ct-x);$$

thus c is not the velocity of the constituent particles of air but is the velocity of propagation of the disturbance as a whole.

In the same way, $F(ct+x)$ is the velocity potential of a plane sound wave which is propagated parallel to Ox with velocity $-c$. The general motion with velocity potential (3.21) is the resultant of superposing two plane waves travelling parallel to Ox with velocities $\pm c$.

The arbitrary functions in the velocity potential

$$\phi = f(ct-x)+F(ct+x)$$

of the general plane wave of sound can be determined if we are given the values q_0 and s_0 of the velocity and condensation at the initial instant $t = 0$ for all values of x. For we have

$$q_0 = f'(-x) - F'(x), \qquad cs_0 = f'(-x) + F'(x),$$

and so

$$2f'(-x) = q_0 + cs_0, \qquad 2F'(x) = -q_0 + cs_0. \tag{3.23}$$

These equations determine f and F apart from an unimportant additive constant.

From (3.23) we see that, if q_0 and s_0 have general values, neither f nor F is identically zero and so the wave is propagated in both directions. Only when one of the functions $q_0 \pm cs_0$ is identically zero do we get a wave of sound which is propagated in one direction.

§ 3.3. Isotropic spherical waves of sound

The equation of wave-motions

$$\nabla^2 u = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2}$$

expressed in terms of spherical polar coordinates (r, θ, ϕ) has the form

$$\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\cot\theta}{r^2}\frac{\partial u}{\partial \theta} + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 u}{\partial \phi^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}.$$

In particular, if the wave-function u depends only on r and t, it satisfies the equation

$$\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}. \tag{3.31}$$

A solution of this equation is the velocity potential of sound waves of small amplitude in which the disturbance at any given instant is the same at all points of any sphere whose centre is at the origin; that is, a solution of (3.31) is the velocity potential of isotropic spherical waves.

Since (3.31) can be written in the form

$$\frac{\partial^2}{\partial r^2}(ru) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}(ru),$$

the function ru satisfies the equation of plane waves, and so the velocity potential of isotropic spherical waves is

$$u = \{f(ct-r) + F(ct+r)\}/r, \tag{3.32}$$

where f and F are arbitrary functions.

The particular wave-function

$$u = \frac{1}{r}f(ct-r) \tag{3.33}$$

is the velocity potential of isotropic spherical waves diverging from O. The constant c is still called the velocity of sound, since it is the velocity with which the disturbance as a whole is propagated. But the actual mode of propagation differs from that of plane waves in that there is a change of type.

The law of propagation of the condensation s is simple. For, since

$$s = \frac{1}{cr}f'(ct-r), \tag{3.34}$$

the quantity rs is propagated outwards without change of type with constant velocity c: hence s diminishes like $1/r$ as the disturbance diverges from O.

The particle velocity $\mathbf{q}$ is directed radially outwards and is of magnitude

$$q = \frac{1}{r^2}f(ct-r)+\frac{1}{r}f'(ct-r). \tag{3.35}$$

Thus q is in general the sum of two terms: the first,

$$\frac{1}{r^2}f(ct-r), \tag{3.36}$$

predominates when r is small; the second,

$$\frac{1}{r}f'(ct-r), \tag{3.37}$$

predominates when r is large. Thus when an isotropic expanding spherical wave diverges from O, the particle velocity does not undergo a mere attenuation but actually changes its type.

It follows from (3.34) and (3.37) that $q = cs$ for a diverging isotropic spherical wave of large radius, just as for a progressive plane wave. We should then expect that a progressive plane wave and a diverging isotropic spherical wave of large radius would have very similar characters; but this is not the case, as we now show.

Let us consider an isotropic diverging disturbance which is confined to a spherical shell: inside and outside the shell the condensation and the velocity are zero. From (3.34) and (3.35) we have

$$r^2(q-cs) = f(ct-r),$$

so that $f(ct-r)$ vanishes everywhere except in the shell. Hence, by (3.34),

$$\int_a^b sr\,dr = \int_a^b \frac{1}{c}f'(ct-r)\,dr = \left[-\frac{1}{c}f(ct-r)\right]_a^b$$

and this vanishes if a and b are respectively less than and greater than the inner and outer radii of the shell. From this it follows that s cannot be of one sign throughout the shell; in other words *a spherical wave of positive condensation cannot exist alone.*† This does not hold for a progressive plane wave, since the velocity potential is constant but not necessarily zero in regions in which the velocity and condensation both vanish.

§ 3.4. Simple and double sources

If r denotes distance from a fixed point P', the expression

$$u = \frac{1}{r}f\left(t-\frac{r}{c}\right) \tag{3.41}$$

is the velocity potential of an isotropic spherical wave-motion in which the waves expand from the centre P'. Evidently u and also the velocity and condensation are infinite at P', which is therefore a singular point of the wave-motion. Moreover, air is flowing across a sphere of centre P' and radius r at the rate

$$4\pi r^2 q = 4\pi f\left(t-\frac{r}{c}\right)+\frac{4\pi r}{c}f'\left(t-\frac{r}{c}\right),$$

and this tends to $4\pi f(t)$ as $r \to 0$. Hence the motion is characterized by the fact that air is being introduced at P' at the rate‡ $4\pi f(t)$. For this reason we say that (3.41) is the velocity potential of a *simple source* at P' and we call $f(t)$ the strength of the source.

More general wave-functions can be constructed by adding together the velocity potentials due to several different simple sources. Let us consider, for example, a simple source of strength $f(t)$ at P' (x', y', z') and another of strength $-f(t)$ at the adjacent point $(x'+lh, y'+mh, z'+nh)$, where (l, m, n) are the direction cosines of

† This was first pointed out by Stokes in 1849. See Rayleigh, *Theory of Sound*, **2** (1896), 101.

‡ The function $f(t)$ is not necessarily positive. If $f(t)$ is negative at any instant, it merely means that air is then being abstracted at P'. A most important case arises when $f(t) = \cos nt$, which varies periodically through positive and negative values.

the line joining the two sources. The velocity potential (3.41) of the source at P' is of the form

$$u = \phi(x-x', y-y', z-z', t).$$

Hence the velocity potential due to the two sources is

$$\begin{aligned} u &= \phi(x-x', y-y', z-z', t) - \phi(x-x'-lh, y-y'-mh, z-z'-nh, t) \\ &= \left\{l\frac{\partial}{\partial x} + m\frac{\partial}{\partial y} + n\frac{\partial}{\partial z}\right\} h\phi(x-x', y-y', z-z', t) + \dots \\ &= \left\{l\frac{\partial}{\partial x} + m\frac{\partial}{\partial y} + n\frac{\partial}{\partial z}\right\} \frac{h}{r} f\left(t-\frac{r}{c}\right) + \dots, \end{aligned}$$

where the terms omitted are of the order of h^2. If we write

$$hf(t) = F(t)$$

and keep F fixed whilst we make h tend to zero, we obtain the velocity potential

$$u = \left\{l\frac{\partial}{\partial x} + m\frac{\partial}{\partial y} + n\frac{\partial}{\partial z}\right\} \frac{1}{r} F\left(t-\frac{r}{c}\right), \tag{3.42}$$

which we may roughly describe as being due to two very large sources very close together. We call (3.42) the velocity potential of a *double source* or, more briefly, a doublet of moment $F(t)$; the line through P' with direction cosines (l, m, n) is called the axis of the doublet.

For example, the velocity potential of a doublet at the origin, whose moment is $F(t)$ and whose axis is the axis of x, is

$$u = -\left\{\frac{1}{r^2} F\left(t-\frac{r}{c}\right) + \frac{1}{cr} F'\left(t-\frac{r}{c}\right)\right\} \frac{x}{r}.$$

From this we see at once that a double source does not emit isotropic spherical waves.

§ 3.5. Poisson's solution of the equation of wave-motions

The simple solutions of the preceding section can be generalized in another way, namely, by considering volume distributions of simple sources. If the volume element $dx'dy'dz'$ at (x', y', z') is a simple source of strength $f(x', y', z')F(t)\,dx'dy'dz'$, the disturbance at (x, y, z) at the instant t is specified by the wave-function

$$\frac{f(x', y', z')}{r} F\left(t-\frac{r}{c}\right) dx'dy'dz',$$

where $$r^2 = (x-x')^2 + (y-y')^2 + (z-z')^2.$$

Integrating over the whole volume V containing the sources, we obtain the more general wave-function

$$u(x,y,z,t) = \iiint\limits_V \frac{f(x',y',z')}{r} F\left(t-\frac{r}{c}\right) dx'dy'dz',$$

it being supposed that (x,y,z) is not a point of V. If we transform to spherical polar coordinates with (x,y,z) as pole, this becomes

$$u(x,y,z,t) = \iiint f(x+lr,y+mr,z+nr)F\left(t-\frac{r}{c}\right) r\sin\theta\, drd\theta d\phi, \tag{3.51}$$

where (l,m,n) are the direction cosines of the line from (x,y,z) to (x',y',z').

In particular, we may suppose that the volume distribution of simple sources is active only for a very short interval of time, so that $F(t)$ is zero except when $-\epsilon < t < 0$; moreover, by multiplying by a suitable constant, we can choose $F(t)$ so that

$$\int\limits_{-\epsilon}^{0} F(t)\, dt = -\frac{1}{4\pi c^2}.$$

If we now make $\epsilon \to 0$ in (3.51), we obtain the wave-function

$$u(x,y,z,t) = \frac{t}{4\pi} \int\limits_0^{2\pi} d\phi \int\limits_0^{\pi} d\theta\, f(x+lct,y+mct,z+nct)\sin\theta, \tag{3.52}$$

and this may be written in the more concise form

$$u(P;t) = tM_{P,ct}\{f\}, \tag{3.53}$$

where $M_{P,r}\{f\}$ denotes the mean value of f over the sphere of radius r and centre P.

From (3.53) we have

$$\frac{\partial u(P:t)}{\partial t} = M_{P,ct}\{f\}+t\frac{\partial}{\partial t}M_{P,ct}\{f\}.$$

Hence $$u(P;t) \to 0, \qquad \frac{\partial u(P;t)}{\partial t} \to f(x,y,z)$$

as $t \to 0$. Thus (3.53) is the wave-function which satisfies the initial conditions $u = 0$, $\partial u/\partial t = f$ when† $t = 0$.

† More precisely, this is true when f satisfies certain conditions of continuity. In general, the value of u when $t = 0$ and the limit as $t \to 0$ are not necessarily equal.

If u is a wave-function, so also is $\partial u/\partial t$ since the wave-equation is a linear equation with constant coefficients. Hence

$$v = \frac{\partial}{\partial t}(tM_{P,ct}\{g\}) \tag{3.54}$$

is a wave-function which satisfies the initial condition $v = g$. Moreover,

$$\frac{\partial v}{\partial t} = 2\frac{\partial}{\partial t}M_{P,ct}\{g\}+t\frac{\partial^2}{\partial t^2}M_{P,ct}\{g\}. \tag{3.55}$$

Now from (3.52) we have

$$\frac{\partial}{\partial t}M_{P,ct}\{g\} = \frac{c}{4\pi}\int\limits_0^{2\pi}\int\limits_0^{\pi}\left\{l\frac{\partial}{\partial x'}+m\frac{\partial}{\partial y'}+n\frac{\partial}{dz'}\right\}g(x',y',z')\sin\theta\,d\theta d\phi$$

$$= \frac{1}{4\pi ct^2}\iint\limits_{r=ct}\frac{\partial g}{\partial r}\,dS$$

$$= \frac{1}{4\pi ct^2}\iiint\limits_{r\leqslant ct}\nabla^2 g\,dx'dy'dz'$$

$$= \frac{c^2t}{3}\{\nabla^2 g(x,y,z)+...\}\to 0$$

as $t\to 0$. Since $\partial^2 M/\partial t^2$ is bounded as $t\to 0$, it follows from (3.55) that

$$v = \frac{\partial}{\partial t}(tM_{P,ct}\{g\})$$

satisfies the initial conditions $v = g$, $\partial v/\partial t = 0$ when $t = 0$.

Finally, by adding (3.53) and (3.54), we obtain the wave-function

$$u(P;t) = tM_{P,ct}\{f\}+\frac{\partial}{\partial t}(tM_{P,ct}\{g\}) \tag{3.56}$$

which satisfies the initial conditions

$$u = g(x,y,z), \qquad \frac{\partial u}{\partial t} = f(x,y,z)$$

when $t = 0$. This formula, which is due to Poisson,† expresses the value of a wave-function u in terms of the values of u and $\partial u/\partial t$ at some fixed *previous* instant.

† Poisson, *Mémoires de l'Acad. Roy. des Sci.* III (1819), 121. Other proofs have been given by Liouville, *Journal de Math.* **1** (1856), 1; Boussinesq, *Comptes rendus*, **94** (1882), 1465; Rayleigh, *Theory of Sound*, **2** (1896), 97 *et seq.* The proof given here is not rigorous: we justify it later as a special case of a more general theorem of Kirchhoff. Poisson's formula has been extended to the case of wave motions in a space of constant curvature by E. Hölder, *Leipziger Berichte*, **19** (1938), 55–66.

That the solution (3.56) holds only when t is positive is of importance: it arises from the fact that the solution was generated by sources acting only when $-\epsilon < t < 0$, and these can evidently produce no effect when t is negative.

§ 3.6. Velocity waves and condensation waves

In the case when u is the velocity potential of sound waves in air, the velocity vector $\mathbf{q}$ and the condensation s are given by

$$\mathbf{q} = -\operatorname{grad} u, \qquad c^2 s = \frac{\partial u}{\partial t}.$$

Hence the function f in Poisson's solution (3.56) determines the initial condensation, the function g the initial distribution of velocity.

If the air is initially at rest, g is identically zero, and Poisson's solution reduces to its first term

$$u = tM_{P,ct}\{f\}.$$

The corresponding sound wave is due to an initial condensation in a medium initially at rest, and so may be called a *condensation wave.* If the condensation is initially zero everywhere, the solution reduces to its second term

$$u = \frac{\partial}{\partial t}(tM_{P,ct}\{g\});$$

the corresponding sound wave may be called a *velocity wave.*

Example. u is a wave function, and U is defined by

$$U(r, ct) = M_{0,r}\{u(x, y, z, ct)\}.$$

Prove that

$$U(r, ct) = \{\phi(ct+r)+\psi(ct-r)\}/r,$$

where ϕ and ψ are arbitrary functions. Deduce Poisson's formula for u.

§ 3.7. The verification of Huygens' principle for expanding isotropic spherical waves

The velocity potential of isotropic spherical waves with centre O is of the form

$$u = \frac{1}{R}F_1(R-ct)+\frac{1}{R}F_2(R+ct), \tag{3.71}$$

where F_1 and F_2 are arbitrary functions and where R denotes the distance from O. We can determine F_1 and F_2 from the initial values of u and $\partial u/\partial t$, and, as the form of (3.71) shows, an initial disturbance splits up in general into two isotropic waves, one converging to O, the other diverging from O. An expanding spherical wave is obtained only when the initial values of u and $\partial u/\partial t$ satisfy the relation

$$\frac{R}{c}\frac{\partial u}{\partial t}+\frac{\partial(Ru)}{\partial R} = 0. \tag{3.72}$$

We shall consider here the case of an isotropic expanding spherical wave-motion in which the initial disturbance is specified by

$$u = \frac{1}{R}F(R) \quad (t = 0),$$

where $F(R)$ is non-zero only when $R_2 \leqslant R \leqslant R_1$; thus the initial disturbance is null except in the shell bounded by the two spheres S_1 $(R = R_1)$ and S_2 $(R = R_2)$. Then by (3.72) the initial value of $\partial u/\partial t$ is

$$\frac{\partial u}{\partial t} = -\frac{c}{R}F'(R)$$

in the shell and is zero elsewhere. We show that Huygens' description of the propagation of this disturbance is in agreement with the analytical solution given by Poisson, namely,

$$u(P;t) = tM\{f\}+\frac{\partial}{\partial t}(tM\{g\}),$$

where $$f = -\frac{c}{R}F'(R), \qquad g = \frac{1}{R}F(R),$$

and M denotes† a mean value taken over the sphere S with centre P and radius ct. (See Fig. 1.)

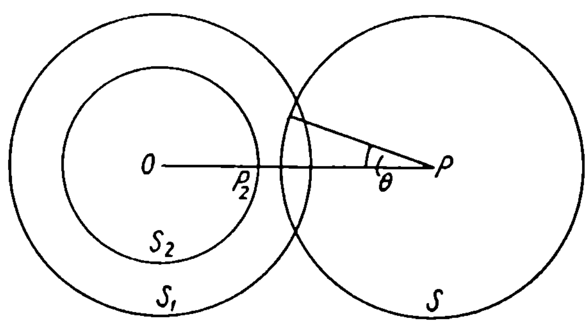


FIG. 1

Let R_0 be the distance of P from the origin. Then if P is outside S_1, the disturbance at P is null, except possibly during the interval

† We omit the double suffix P, ct indicating the centre and radius of the sphere over which the mean values are calculated. No confusion will be caused by this.

$(R_0-R_1)/c \leqslant t \leqslant (R_0+R_1)/c$; for at an instant outside this interval of time the sphere S of radius ct does not cut into the shell to which the initial disturbance is confined, and so the mean values of f and g are both zero. Similarly, if P is inside S_2, the disturbance is null except possibly during the interval $(R_2-R_0)/c \leqslant t \leqslant (R_0+R_1)/c$. It remains to consider Poisson's solution when S does cut into the shell bounded by S_1 and S_2.

Let θ be the angle between any radius vector r at P and the line PO; we then have

$$R^2 = R_0^2+r^2-2R_0 r\cos\theta, \qquad R_0 r\sin\theta\, d\theta = R\,dR.$$

Hence

$$M\{g\} = \frac{1}{4\pi}\iint g\sin\theta\, d\theta d\phi = \frac{1}{2}\int g\sin\theta\, d\theta$$

$$= \frac{1}{2}\int\limits_{R_2'}^{R_1'} \frac{F(R)}{R_0 r}\, dR,$$

where the limits of integration will be found later. This gives

$$\frac{\partial}{\partial t}(tM\{g\}) = \frac{1}{2cR_0}\left\{F(R_1')\frac{\partial R_1'}{\partial t} - F(R_2')\frac{\partial R_2'}{\partial t}\right\}.$$

Similarly,

$$tM\{f\} = \frac{r}{2c}\int f\sin\theta\, d\theta = -\frac{1}{2}\int\limits_{R_2'}^{R_1'} \frac{F'(R)}{R_0}\, dR$$

$$= \frac{1}{2R_0}\{F(R_2')-F(R_1')\}.$$

Substituting in Poisson's formula, we have

$$u(P;t) = \frac{F(R_2')}{2R_0}\left\{1-\frac{1}{c}\frac{\partial R_2'}{\partial t}\right\} - \frac{F(R_1')}{2R_0}\left\{1-\frac{1}{c}\frac{\partial R_1'}{\partial t}\right\}. \tag{3.73}$$

There are three cases to be considered, viz. (*a*) P outside S_1, (*b*) P inside S_2, (*c*) P between S_1 and S_2.

(*a*) *P outside* S_1; $(R_0 > R_1)$.

(i) $ct < R_0-R_1$. Then $u = 0$ since S does not cut into the shell.

(ii) $R_0-R_1 < ct < R_0-R_2$. In this case, we have

$$R_1' = R_1, \qquad R_2' = R_0-ct,$$

and so

$$u(P;t) = \frac{F(R_0-ct)}{R_0} - \frac{F(R_1)}{2R_0}.$$

This agrees with the known value

$$u(P;t) = \frac{F(R_0-ct)}{R_0}$$

only if $F(R_1) = 0$, i.e. if $F(R)$ is continuous at $R = R_1$. We shall, in what follows, assume that the initial data are continuous,† so that $F(R_1) = F(R_2) = 0$.

(iii) $R_0-R_2 < ct < R_0+R_2$. Then $R_1' = R_1$, $R_2' = R_2$, and so $u(P;t) = 0$.

(iv) $R_0+R_2 < ct < R_0+R_1$. Then $R_1' = R_1$, $R_2' = ct-R_0$; again $u(P;t) = 0$.

(v) $R_0+R_1 < ct$. In this case $u(P;t) = 0$ since S does not cut into the shell.

Thus Poisson's formula verifies Huygens' geometrical construction in the case when P lies outside the shell, provided that the initial data are continuous.

(*b*) *P inside* S_2; ($R_0 < R_2$).

(i) $ct < R_2-R_0$. Then $u = 0$, since S does not cut into the shell.

(ii) $R_2-R_0 < ct < R_1-R_0$. In this case we have

$$R_2' = R_2, \qquad R_1' = ct-R_0;$$

since $F(R_2) = 0$ by hypothesis, we have $u(P;t) = 0$.

(iii)‡ $R_1-R_0 < ct < R_2+R_0$ Then $R_1' = R_1$; $R_2' = R_2$, and $u(P;t) = 0$.

(iv) $R_2+R_0 < ct < R_1+R_0$. Then $R_1' = R_1$, $R_2' = ct-R_0$; hence $u(P;t) = 0$.

(v) $R_0+R_1 < ct$. In this case, the effect at P is null, since S does not cut into the shell.

Poisson's formula does not give rise to a returning wave, as the crude geometrical form of Huygens' principle would. Thus in the present analytical formulation there is no need of a special hypothesis discarding the inner sheet of the envelope of the secondary wave-fronts.

† For a discussion of wave-motions with discontinuities at the wave fronts, see Love, *Proc. London Math. Soc.* (2), **1** (1903), 37–62, 291–344; Rayleigh, ibid. **2** (1904), 266–9. The present example is taken from the first of these papers.

‡ We assume here that $R_0 > \frac{1}{2}(R_1-R_2)$. If not, the argument needs slight changes.

(c) *P lies between S_1 and S_2*; $(R_2 < R_0 < R_1)$.

A similar argument shows that

$$u(P;t) = \frac{1}{R_0}F(R_0-ct)$$

until the instant $t = (R_0-R_2)/c$, and that the disturbance is thereafter null at P.

The disturbance we have just considered is confined at the instant t to the shell bounded by the spheres $R = R_2+ct$, $R = R_1+ct$. In particular, if R_1-R_2 is very small, we have in effect a solitary spherical expanding wave of the type considered by Huygens. The initial disturbance is confined to a very thin layer on the surface of the sphere S_2. At the subsequent instant t the disturbance is confined to a layer of the same thickness on the surface of the sphere $R = R_2+ct$, which is the outer sheet of the envelope of spheres of radii ct, whose centres lie on S_2. Thus Huygens' geometrical construction, with its restriction that only one sheet of the envelope is to be considered, is justified by Poisson's analytical solution of the equation of wave-motions.

Moreover, Huygens' statement that a secondary wave is of effect only at the point where it touches the envelope also follows. For when the layer is very thin, Poisson's integrals for $M\{f\}$ and $M\{g\}$ are extended over a very small area on S near the point P_2 where OP cuts S_2; and if P is in the wave-front at the instant t, the secondary wave, with centre P_2 and radius ct touches the envelope at P.

In a similar way† we can justify Huygens' construction for an isolated spherical converging wave or an isolated plane wave; but only if the initial values of the velocity potential and condensation are suitably chosen. In other words, to justify Huygens' principle for isolated waves, we must have recourse to analysis and take into account the dynamics of the medium in which the wave-motion occurs.

Example. Apply Poisson's solution of the equation of wave-motions under the initial conditions

$$u = \frac{F(R)}{R} \quad (R_2 \leqslant R \leqslant R_1), \qquad = 0 \quad (R > R_1,\ R < R_2),$$

$$\frac{\partial u}{\partial t} = 0,$$

† See Croze, *Annales de Physique*, **5** (1926), 370–439 (383–98).

where $F(R)$ is continuous at R_1 and R_2. Discuss the nature of the solution so obtained. (Love.†)

§ 4. Huygens' principle for monochromatic phenomena

§ 4.1. Fresnel's extension of Huygens' principle

In his memoir on Diffraction, which won the Paris Academy's prize in 1818, Fresnel‡ made an important extension of Huygens' principle, in that he replaced Huygens' isolated spherical waves by purely periodic trains of spherical waves and made use of the principle of interference. On this theory, light ought to appear, not necessarily on the envelope of the secondary waves, but at every point where these secondary waves reinforce one another; on the other hand, there should be darkness wherever these secondary waves destroy one another. In this way, Fresnel was able to account, not only for the rectilinear propagation of light of very short wave-length and the laws of reflection and refraction, but also for certain diffraction phenomena.

Fresnel, at this stage, still regarded light as a disturbance in an aether analogous to sound in air. He did not yet realize that the phenomenon of polarization made such a theory untenable. (His elastic-solid theory of the luminiferous aether, in which light consisted of transverse vibrations, dates from 1821.) Accordingly, we shall discuss here the theory as applied to sound waves of small amplitude; the application to the scalar theory of light is merely a matter of changing the terminology.

Let us consider sound waves of a very general character, which are generated by making a certain part of the medium execute forced vibrations, not necessarily 'monochromatic'. Such a wave-motion is evidently much more complicated than that which arises when a portion of the fluid is initially disturbed from its equilibrium state and the disturbance is allowed to propagate itself freely; for a wave-motion of the former type is the result of superposing an infinite succession of wave-motions of the latter type. It seems likely that in a wave-motion produced by the forced vibrations of a portion V of the fluid, the effect at the instant t at the point P outside V would depend on the velocity and condensation at each point Q of V at

† Loc. cit. 44.

‡ *Mém. de l'Acad.* **5** (1826), 339; reprinted in Fresnel's *Œuvres complètes*, **1**, 247. For the history of this development, see E. T. Whittaker, *History of the Theories of Aether and Electricity* (1910), 113–15 and Croze's paper already cited.

the instant $t-PQ/c$. This idea, which is implicit in the work of Fresnel,† attained a rigorous analytical formulation in Kirchhoff's integral theorem.

In the case of 'monochromatic' forced vibrations the matter is somewhat simpler. Let us consider for definiteness the case considered by Fresnel, namely that of a 'monochromatic' source of expanding waves. If the source is at the origin, the wave-motion is characterized by the velocity potential

$$u = \frac{1}{r}e^{ik(r-ct)} \tag{4.11}$$

or, rather, by its real part.‡ In this motion the sphere S, whose equation is $r = r_0$, is a wave-surface; for all the particles of the fluid on S at the instant t are being affected by the disturbance which left the source at the instant $t-r_0/c$.

The disturbance which reaches S at the instant t continues to be propagated outwards and reaches the wave-surface S' of radius r_0' at the instant

$$t' = t+(r_0'-r_0)/c.$$

It seems natural to regard the vibrations of the particles on S' at the instant t' as being due to the vibrations of the particles on S at the instant t, instead of being due to the pulsations of the source at the instant $t'-r_0'/c$.

This is, in fact, the assumption which Fresnel made. He supposed that each element of the wave-surface S acts as a secondary source which is sustained by the displacement and velocity given to the particles of the surface element by the primary wave. The resultant effect is produced by the interference of these secondary waves.

If Fresnel's theory is to provide a valid extension of Huygens' principle, the secondary sources must produce not only the correct effect outside S but also a null effect inside S. Fresnel believed that this could be done only by taking account of the dynamical effect of the condensation and of the velocity in the primary wave at each point of S; that is, by supposing that each secondary source emits a 'condensation wave' and a 'velocity wave'. Thus an analytical formulation of Huygens' principle which involved only one of these

† See Croze, loc. cit. 398–405.

‡ As is usual in this sort of work, it is simpler to work with complex wave-functions, and then to take real parts at the end of the analysis.

two types of secondary wave would give rise to a non-null effect within S and would have no physical justification.

Unfortunately Fresnel was unable to carry out this programme completely. He considered only the case when the wave-surface S is of large radius compared with the wave-length λ. In this case (cf. § 3.3) the velocity q and the condensation s at points of S are connected approximately by the relation

$$q = cs.$$

He thought that it would then suffice to attribute to the elements of S the same velocity as is communicated to them by the primary wave in order to obtain the correct effect at points outside S and at a large distance from it. Actually considerations of this nature lead to incorrect results unless one makes, as Fresnel did, the following additional assumptions:

(i) the elements of S execute vibrations whose amplitude is to the amplitude in the primary wave as is $1 : \lambda$;

(ii) the elements of S are oscillating a quarter of a period ahead of the primary wave.

The necessity for these two additional assumptions† led many to regard Fresnel's theory merely as a convenient means of calculation which lacked any sound physical basis.

Various attempts have been made to overcome the phase difficulty. For example, Gouy‡ replaced each element dS by a small pulsating sphere (of surface area $2dS$), which emitted monochromatic isotropic waves. This does give the correct phase; but as it also gives rise to an effect inside S which is not null, it cannot be accepted as a satisfactory formulation of Huygens' principle.

Nevertheless, it is the case, as we shall see in § 4.4, that a careful analytical formulation of Fresnel's original idea, that each element of S gives rise to a condensation wave and a velocity wave which can be determined on dynamical principles, leads to correct results without any arbitrary assumptions regarding the phase and amplitude of the secondary sources. Moreover, the formula obtained in this way is in complete agreement with Helmholtz's analytical

† Fresnel also assumed that the effect at a point P outside S is due only to the secondary sources on the part of S visible from P, and not to the secondary sources all over S. This assumption makes little difference when the wave-length is small.

‡ *Comptes rendus*, **111** (1890), 910–12; *Annales de Physique* (6), **24** (1891), 145.

formulation of Huygens' principle for monochromatic phenomena, which we now discuss.

§ 4.2. Helmholtz's formula

The velocity potential of sound waves of small amplitude satisfies the equation of wave-motions

$$\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}.$$

If the disturbance is 'monochromatic', u will be of the form ve^{-ikct}, where v is independent of t and is a solution of

$$(\nabla^2+k^2)v = 0. \tag{4.21}$$

This equation reduces to Laplace's equation when $k = 0$; and the formula of Helmholtz,† which we are about to prove, then reduces to the well-known Green's equivalent layer formula in the theory of attractions.

We start with Green's identity

$$\iiint\limits_V (v\nabla^2 w - w\nabla^2 v)\, dxdydz = \iint\limits_S \left(v\frac{\partial w}{\partial \nu} - w\frac{\partial v}{\partial \nu}\right) dS, \tag{4.22}$$

where S is a closed surface bounding the volume V and $\partial/\partial\nu$ denotes differentiation along the outward normal to S. This identity is certainly valid‡ when v and w and their first- and second-order partial derivatives are continuous within and on S, or, as we shall say in future, when v and w are regular within and on S. If v and w satisfy (4.21) and the prescribed conditions of continuity, (4.22) becomes

$$\iint\limits_S \left(v\frac{\partial w}{\partial \nu} - w\frac{\partial v}{\partial \nu}\right) dS = 0.$$

In particular, if $w = e^{ikr}/r$, where r denotes distance from the fixed point P, we have

$$\iint\limits_S \left\{v\frac{\partial}{\partial \nu}\left(\frac{e^{ikr}}{r}\right) - \frac{e^{ikr}}{r}\frac{\partial v}{\partial \nu}\right\} dS = 0, \tag{4.23}$$

provided that P and the singularities‖ of v lie outside S.

† *Journal f. Math.* **57** (1859), 7.

‡ Here, and throughout this book, we take the simplest sufficient conditions for the validity of Green's identity. Lighter sufficient conditions can be obtained as in Kellogg's discussion of the Divergence Theorem (*Foundations of Potential Theory* (Berlin, 1929), Ch. IV).

‖ By a singularity of v we mean a point at which v or one of its first or second partial derivatives is discontinuous.

If P lies inside S but the singularities of v are still outside, equation (4.23) no longer holds, since w becomes infinite at P. To avoid this difficulty, we apply Green's identity (4.22) to the volume V_1 bounded externally by S and internally by the sphere σ, of centre P and small radius ϵ. This gives

$$\iint\limits_S \left\{v\frac{\partial}{\partial\nu}\left(\frac{e^{ikr}}{r}\right)-\frac{e^{ikr}}{r}\frac{\partial v}{\partial\nu}\right\}dS$$
$$= -\iint\limits_\sigma \left\{v\frac{\partial}{\partial\nu}\left(\frac{e^{ikr}}{r}\right)-\frac{e^{ikr}}{r}\frac{\partial v}{\partial\nu}\right\}dS$$
$$= \iint\limits_\sigma \left\{v\frac{\partial}{\partial r}\left(\frac{e^{ikr}}{r}\right)-\frac{e^{ikr}}{r}\frac{\partial v}{\partial r}\right\}dS$$
$$= \lim_{\epsilon\to 0}\iint\limits_\sigma \left\{v\frac{e^{ikr}}{r}\left(ik-\frac{1}{r}\right)-\frac{e^{ikr}}{r}\frac{\partial v}{\partial r}\right\}dS$$

since the integral over S is independent of ϵ. But since v and $\partial v/\partial r$ are continuous at P and the element of surface on σ is $\epsilon^2\sin\theta\,d\theta d\phi$, the value of this limit is $-4\pi v(P)$.

We have thus proved the following theorem of Helmholtz:

Let v be a solution of $(\nabla^2+k^2)v = 0$ whose first- and second-order partial derivatives are continuous within and on a closed surface S, and let

$$I(P) = \iint\limits_S \left\{\frac{e^{ikr}}{r}\frac{\partial v}{\partial\nu}-v\frac{\partial}{\partial\nu}\left(\frac{e^{ikr}}{r}\right)\right\}dS, \tag{4.24}$$

where r is distance from a fixed point P and $\partial/\partial\nu$ denotes differentiation along the outward normal to S. Then the value of $I(P)$ is $4\pi v(P)$ or zero according as P lies inside or outside S.

In this theorem the sources of the disturbance specified by the 'monochromatic' wave-function ve^{-ikct} lie outside a certain closed surface. In most of the applications all the sources lie at a finite distance and the effect at a distant point is to be expressed as an integral over a surface containing the sources. Accordingly we must modify the theorem to cover the case when all the sources lie inside a closed surface.

The function v is now a solution of $(\nabla^2+k^2)v = 0$ whose first- and second-order partial derivatives are continuous on and outside a closed surface S. We choose R so large that the sphere Σ, whose equation is $r = R$, encloses S; and we denote by V_2 the volume bounded externally by Σ and internally by S. Then if the normal ν is drawn into V_2, the value of

$$J(P) = \iint\limits_S + \iint\limits_\Sigma \left\{\frac{e^{ikr}}{r}\frac{\partial v}{\partial \nu} - v\frac{\partial}{\partial \nu}\left(\frac{e^{ikr}}{r}\right)\right\} dS$$

is $-4\pi v(P)$ or zero according as P is or is not a point of V_2.

The normal ν at points of S is the ordinary outward normal. Hence, in the notation of (4.24),

$$J(P) = I(P) - \iint\limits_\Sigma \frac{e^{ikr}}{r}\left\{\frac{\partial v}{\partial r} - ikv + \frac{v}{r}\right\} dS$$

$$= I(P) - \iint e^{ikR}\left\{r\left(\frac{\partial v}{\partial r} - ikv\right)\right\}_{r=R} d\omega - \iint e^{ikR}\{v\}_{r=R}\, d\omega,$$

where $d\omega$ is the solid angle subtended by an element of the sphere Σ at its centre P. Now make $R \to \infty$. The term

$$\iint e^{ikR}\{v\}_{r=R}\, d\omega$$

tends to zero if $v \to 0$ as $r \to \infty$, uniformly with respect to the polar angles θ and ϕ; in particular, it tends to zero if

$$|rv| < K \quad \text{as } r \to \infty. \tag{4.25}$$

The term $$\iint e^{ikR}\left\{r\left(\frac{\partial v}{\partial r} - ikv\right)\right\}_{r=R} d\omega$$

tends to zero if $$r\left(\frac{\partial v}{\partial r} - ikv\right) \to 0 \quad \text{as } r \to \infty \tag{4.26}$$

uniformly with respect to θ and ϕ. Condition (4.25) is called by Sommerfeld† the 'Endlichkeitsbedingung', condition (4.26) the 'Ausstrahlungsbedingung'. We shall return later to point out the importance of (4.26); at the moment, we merely remark that, when it is satisfied, the wave-function ve^{-ikct} is the velocity potential of a system of expanding waves.

When (4.25) and (4.26) are satisfied it follows that the value of $I(P)$ is $-4\pi v(P)$ or zero according as P lies outside or inside the closed surface S containing all the sources. Hence we have:

† *Jahresbericht der D.M.V.* **21** (1912), 309–53 (326–34). See also W. Magnus, ibid. **52** (1943), 177–88; F. Rellich, ibid. **53** (1943), 57–65, where it is shown that condition (4.25) is not necessary; F. V. Atkinson, *Phil. Mag.* (**40**), 1949, 645–51.

Let v be a solution of $(\nabla^2+k^2)v = 0$ whose first- and second-order partial derivatives are continuous outside and on a closed surface S, and let

$$|rv| < K,$$

$$r\left(\frac{\partial v}{\partial r}-ikv\right) \to 0$$

uniformly with respect to θ and ϕ as $r \to \infty$. Let

$$I(P) = \iint\limits_S \left\{\frac{e^{ikr}}{r}\frac{\partial v}{\partial \nu}-v\frac{\partial}{\partial \nu}\left(\frac{e^{ikr}}{r}\right)\right\} dS,$$

where r is the distance from a fixed point P and $\partial/\partial\nu$ denotes differentiation along the outward normal to S. Then the value of $I(P)$ is $-4\pi v(P)$ or zero according as P lies outside or inside S.

Example. ϕ and ψ are two solutions of $(\nabla^2+k^2)v = 0$ which satisfy the following conditions:

(i) ϕ and its first- and second-order partial derivatives are continuous within and on the closed surface S;

(ii) ψ and its first- and second-order partial derivatives are continuous outside and on S, and ψ satisfies Sommerfeld's conditions at infinity.

Prove that, if $\phi = \psi$ on S, the value of

$$\iint\limits_S \frac{e^{ikr}}{r}\left\{\frac{\partial\phi}{\partial\nu}-\frac{\partial\psi}{\partial\nu}\right\} dS$$

is $4\pi\phi(P)$ or $4\pi\psi(P)$ according as P is inside or outside S. Show also that if $\partial\phi/\partial\nu = \partial\psi/\partial\nu$ on S, the value of

$$\iint\limits_S (\phi-\psi)\frac{\partial}{\partial\nu}\left(\frac{e^{ikr}}{r}\right) dS$$

is $-4\pi\phi(P)$ or $-4\pi\psi(P)$ according as P is inside or outside S.

§ 4.3. Restrictions on the use of Helmholtz's formula

Helmholtz's formula

$$v(P) = \frac{1}{4\pi}\iint\limits_S \left\{\Psi\frac{e^{ikr}}{r}-\Phi\frac{\partial}{\partial\nu}\left(\frac{e^{ikr}}{r}\right)\right\} dS \tag{4.31}$$

expresses the value of the solution v of

$$(\nabla^2+k^2)v = 0, \tag{4.32}$$

regular within a closed surface S, at any point P inside S in terms of the values Φ and Ψ taken by v and $\partial v/\partial \nu$ on S. We can, however, express $v(P)$ in terms of Φ alone, by the equation†

$$v(P) = -\frac{1}{4\pi}\iint\limits_S \Phi\frac{\partial G}{\partial \nu}\,dS, \tag{4.33}$$

where G denotes the Green's function with singularity at P.

We see from (4.33) that a knowledge of the boundary values of v alone on S determines v inside S, and, moreover, determines v in general uniquely. In particular, we can find from (4.33) the boundary values of $\partial v/\partial \nu$ on S. Thus a knowledge of Φ alone determines Ψ, and in general does so uniquely. Conversely, it can be shown that a knowledge of Ψ alone determines Φ and, in general, does so uniquely. Hence the functions Φ and Ψ in (4.31) are related and cannot be assigned arbitrarily and independently of each other.

Of course, when Φ and Ψ are arbitrarily assigned, the expression on the right-hand side of (4.31) does satisfy the differential equation (4.32), since it is obtained by the addition of particular solutions of the form e^{ikr}/r and its normal derivative. But these arbitrary values of Φ and Ψ are not necessarily the boundary values of v and $\partial v/\partial \nu$.

There is, however, an exceptional case to which the theory of Green's function is not applicable, and in this case a knowledge of one of the functions Φ and Ψ does not determine the other. To see how this arises, let us suppose that $u = v\cos kct$ is the velocity potential of sound waves of small amplitude, so that v satisfies equation (4.32). Then if S is a rigid boundary, the air inside S possesses certain normal modes of vibration, corresponding to a sequence k_1, k_2,... of values of k, the *eigenvalues* of the problem. The velocity potential $u = v_r\cos k_r ct$ corresponding to one of these normal modes satisfies the boundary condition $\partial v_r/\partial \nu = 0$ on S. Hence the problem of finding the solution of

$$(\nabla^2+k_r^2)v = 0,$$

regular within S, given the boundary values of $\partial v/\partial \nu$ on S, is an indeterminate problem; for, if V is one solution,

$$v = V+Av_r$$

is also a solution for all values of the constant A. Similarly, there

† See Pockels, *Über die partielle Differentialgleichung* $(\Delta+k^2)u = 0$ (Leipzig, 1891), p. 280. The equation (4.33) does not provide an analytical formulation of Huygens' principle since, as we shall see later, it does not give a null effect outside S.

exists another sequence $k_1', k_2',...$ of values of k such that the problem of finding the solution of

$$(\nabla^2+k_r'^2)v = 0,$$

regular within S, given the boundary values of v on S, is also an indeterminate problem.

To sum up, except when the constant k is one of the eigenvalues k_r and k_r', a knowledge of one of the boundary functions Φ and Ψ in (4.31) determines the other uniquely, and together they determine v uniquely.

These exceptional cases do not occur in the exterior problem. It is true that there are solutions of the equation $(\nabla^2+k^2)v = 0$, for certain special values of k, which are regular outside S and satisfy the boundary condition $v = 0$ (or $\partial v/\partial\nu = 0$) on S and the condition of finiteness

$$|rv| < K \tag{4.34}$$

at infinity. But all these solutions (eigenfunctions) represent standing waves and do not satisfy Sommerfeld's *Ausstrahlungsbedingung*, namely that

$$r\left(\frac{\partial v}{\partial r}-ikv\right) \to 0 \tag{4.35}$$

uniformly with respect to the polar angles θ and ϕ as $r \to \infty$.

By imposing the additional condition (4.35), we can apply the Green's function argument to the exterior problem and assert that, for any value of k, a knowledge of one of the boundary values Φ or Ψ, taken on S by a solution v of (4.32), regular outside S, and its normal derivative, determines the other uniquely, and that together they determine v by means of the formula

$$v(P) = -\frac{1}{4\pi}\iint\limits_S \left\{\Psi\frac{e^{ikr}}{r}-\Phi\frac{\partial}{\partial\nu}\left(\frac{e^{ikr}}{r}\right)\right\}\, dS.$$

§ 4.4. The physical meaning of Helmholtz's formula

Let
$$u = ve^{-ikct}$$

be the velocity potential of sound waves in air due to 'monochromatic' sources all lying inside a certain closed surface S. Helmholtz's formula expresses the value of u at a point P outside S as due to a certain distribution of simple and double sources over S. Our proof of this was of an entirely analytical character; we now give an alternative proof by means of the dynamical arguments suggested by Fresnel.†

Let Q be a typical point of S, at a distance r_1 from P; let ν be the

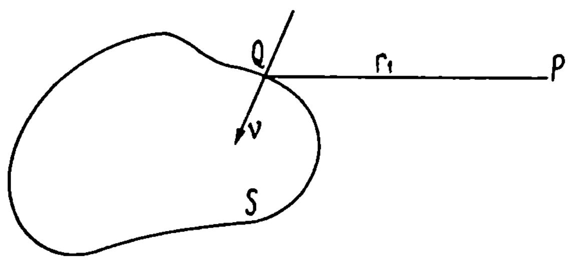


FIG. 2

unit vector drawn along the inward normal to S at Q. The particle velocity at Q has a component in the outward-normal direction

$$\frac{\partial v}{\partial \nu}e^{-ikct}.$$

Hence air flows across the element dS at Q at the rate

$$\frac{\partial v}{\partial \nu}e^{-ikct}\,dS.$$

In addition to producing a flux across dS, the sources within S also change the pressure at Q from the equilibrium pressure p to $p+\delta p$. If ρ is the equilibrium density of air, δp is given by

$$\delta p = \rho c^2 s,$$

where s is the condensation, and so

$$\delta p = -\rho ikcve^{-ikct}.$$

Hence there is a thrust on the area dS at Q of magnitude

$$-\rho ikcve^{-ikct}\,dS.$$

Now suppose that the sources and all the air inside S are destroyed. In order to get the effect outside S specified by the velocity potential ve^{-ikct}, we must introduce sources over S with the following properties:

(*a*) Air is created at each element dS at the rate

$$\frac{\partial v}{\partial \nu}e^{-ikct}\,dS.$$

† Croze, *Annales de Physique*, **5** (1926), 370 (408–11). See also Larmor, *Proc. London Math. Soc.* (2), **19** (1921), 169–80.

(*b*) A force $-\rho ikcve^{-ikct}\,dS$ perpendicular to dS acts on the air in contact with dS.

The creation of air at dS makes this element a source of strength

$$\frac{1}{4\pi}\frac{\partial v}{\partial \nu}e^{-ikct}\,dS;$$

the corresponding velocity potential at P is

$$\frac{1}{4\pi r_1}\frac{\partial v}{\partial \nu}e^{ik(r_1-ct)}\,dS. \tag{4.41}$$

It is well known† that a concentrated force Fe^{-ikct} acting at Q along the outward normal to S gives rise to a sound wave of velocity potential

$$-\frac{iF}{4\pi\rho kc}\frac{\partial}{\partial \nu}\left\{\frac{e^{ik(r_1-ct)}}{r_1}\right\}.$$

Hence when the force is given by (*b*) above, the resulting velocity potential at P is

$$-\frac{v}{4\pi}\frac{\partial}{\partial \nu}\left\{\frac{e^{ik(r_1-ct)}}{r_1}\right\}dS. \tag{4.42}$$

The total effect of the sources on dS is obtained by adding the velocity potentials (4.41) and (4.42). Of these, the first is determined by the normal velocities of the air particles on dS and so represents the velocity wave. The second depends on the condensation in the primary wave at Q and so represents the condensation wave (§ 3.6).

The effect of all the sources on S is obtained by adding (4.41) and (4.42) and integrating over S; this gives

$$u(P;t) = \frac{1}{4\pi}\iint\limits_S \left\{\frac{e^{ik(r_1-ct)}}{r_1}\frac{\partial v}{\partial \nu} - v\frac{\partial}{\partial \nu}\left(\frac{e^{ik(r_1-ct)}}{r_1}\right)\right\}dS,$$

which is Helmholtz's formula in the case when P lies outside a closed surface S containing all the sources. By a similar argument we can show that Helmholtz's formula also holds when P lies inside a closed surface S containing none of the sources.

Fresnel believed that, if all the sources were inside S, the secondary sources on each separate element dS would produce a null effect at points within S. This is evidently not the case. The effect inside S produced by the action of *all* the secondary sources on S is however null, as can be seen by the following simple argument.

Suppose that P is a point inside S. Introduce a diaphragm C

† Lamb, *Hydrodynamics* (Cambridge, 1916), 496.

dividing the space inside S into two regions, of which one contains P and the other contains all the sources. The rim of C divides S into two areas

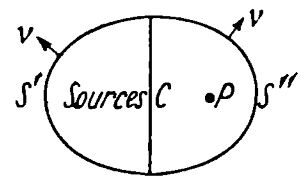


FIG. 3

S' and S'', as shown in the figure. We now apply Helmholtz's formula to the two surfaces $S'+C$ and $S''+C$. This gives

$$-u(P;t) = \frac{1}{4\pi}\iint\limits_{S'+C} \left\{\frac{e^{ik(r_1-ct)}}{r_1}\frac{\partial v}{\partial \nu} - v\frac{\partial}{\partial \nu}\left(\frac{e^{ik(r_1-ct)}}{r_1}\right)\right\} dS, \tag{4.43}$$

$$u(P;t) = \frac{1}{4\pi}\iint\limits_{S''+C} \left\{\frac{e^{ik(r_1-ct)}}{r_1}\frac{\partial v}{\partial \nu} - v\frac{\partial}{\partial \nu}\left(\frac{e^{ik(r_1-ct)}}{r_1}\right)\right\} dS. \tag{4.44}$$

Evidently on C the outward normals to the closed surfaces $S'+C$ and $S''+C$ are in opposite senses. Hence, when we add (4.43) and (4.44), we obtain

$$0 = \frac{1}{4\pi}\iint\limits_{S} \left\{\frac{e^{ik(r_1-ct)}}{r_1}\frac{\partial v}{\partial \nu} - v\frac{\partial}{\partial \nu}\left(\frac{e^{ik(r_1-ct)}}{r_1}\right)\right\} dS;$$

in other words, the effect inside S is null.

Apart from the fact that the secondary source on each separate element dS does not produce a null effect inside S, we now see that Helmholtz's formula presents an analytical form of Huygens' principle in the sense understood by Fresnel. It is not merely a convenient analytical formula, but has a sound dynamical basis; the secondary sources are real in the sense that their amplitude and phase are determined dynamically by the velocity and condensation in the air at points of the surface S.

If it were possible to find a distribution of sources on S, different from that of Helmholtz, which would give the correct effect at points outside S and a null effect inside S, such a distribution would have no physical meaning whatever, since it could not be obtained by considering the dynamics of sound waves in air. Actually no such alternative distribution exists. For if there were such a distribution, we could obtain by subtraction another distribution which would give a null effect everywhere. But a distribution of sources on S

giving a null effect everywhere would itself be null, contrary to our assumption. Hence Helmholtz's distribution of sources is the only one with the required property.

§ 4.5. Retarded values

It is often convenient to make use of the idea of the 'retarded value' of a function. If ϕ is a function of the coordinates (x,y,z) of a variable point Q and of the time t, say

$$\phi = \phi(x,y,z,t),$$

and if r is the distance of Q from a fixed point P, we write

$$[\phi] = \phi\left(x,y,z,t-\frac{r}{c}\right)$$

and call $[\phi]$ the retarded value of ϕ.

§ 4.6. Stokes's diffraction formula and the theory of Fresnel

We now derive from Helmholtz's integral an important approximate analytical formulation of Huygens' principle, due to Stokes; this approximation is valid when the wave-length is very small compared with the other distances involved.

Let us consider in the first instance the case of expanding monochromatic isotropic spherical waves with complex velocity potential

$$u = \frac{e^{ik(r-ct)}}{r}.$$

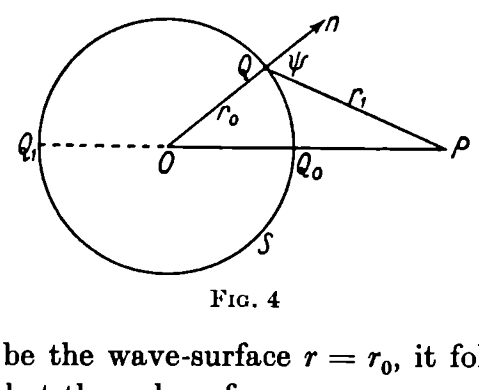


FIG. 4

If we take S to be the wave-surface $r = r_0$, it follows from Helmholtz's theorem that the value of

$$I = -\frac{1}{4\pi}\iint\limits_S \left\{\frac{e^{ik(r_1-ct)}}{r_1}\frac{\partial}{\partial r_0}\left(\frac{e^{ikr_0}}{r_0}\right)-\frac{e^{ikr_0}}{r_0}\frac{\partial}{\partial r_0}\left(\frac{e^{ik(r_1-ct)}}{r_1}\right)\right\} dS$$

is $u(P;t)$ or zero according as P lies outside or inside S. This integral transforms into

$$I = -\frac{1}{4\pi}\iint\limits_S \frac{e^{ik(r_0+r_1-ct)}}{r_0 r_1}\left\{ik(1+\cos\psi)-\frac{1}{r_0}+\frac{1}{r_1}\cos\psi\right\} dS,$$

where $$\cos\psi = \cos\angle PQn = -\frac{\partial r_1}{\partial r_0}.$$

Now, by hypothesis, the wave-length $2\pi/k$ is small compared with r_0 and r_1, and so $1/r_0$ and $1/r_1$ are negligible compared with k. This gives the approximate formula†

$$I = -\frac{1}{4\pi}\iint\limits_S e^{ik(r_0+r_1-ct)}\frac{ik}{r_0 r_1}(1+\cos\psi)\, dS, \qquad (4.61)$$

which is substantially *Fresnel's diffraction formula*.

But since $$u = \frac{e^{ik(r_0-ct)}}{r_0}$$
on S, we have there

$$\frac{\partial u}{\partial n} = \frac{e^{ik(r_0-ct)}}{r_0}\left(ik-\frac{1}{r_0}\right) = ik\frac{e^{ik(r_0-ct)}}{r_0},$$

since we are neglecting $1/r_0$ in comparison with k. Hence

$$\left[\frac{\partial u}{\partial n}\right] = ik\frac{e^{ik(r_0+r_1-ct)}}{r_0}$$

Substituting in equation (4.61), we have

$$u(P;t) = -\frac{1}{4\pi}\iint\limits_S \left[\frac{\partial u}{\partial n}\right]\frac{1+\cos\psi}{r_1}\, dS, \qquad (4.62)$$

when P is a point of the region into which the wave-surface S is progressing.

The case of a monochromatic progressive plane wave-motion is rather more difficult. We start with a complex velocity potential

$$U = e^{ik(x-ct)},$$

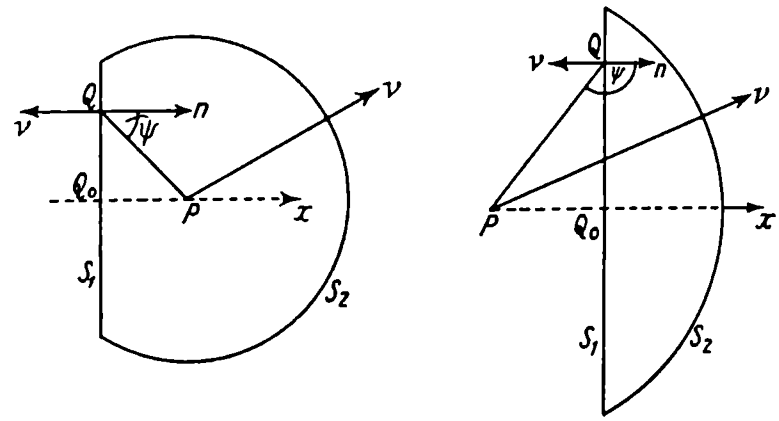
but have to take real parts before proceeding to a certain limit. We shall denote the real part of U by u.

† Note that, if λ is the wave-length, the contribution of the element dS is of the form

$$\frac{e^{ikr_0-\frac{1}{2}\pi i}}{2\lambda r_0}(1+\cos\psi)\frac{e^{ik(r_1-ct)}}{r_1}\, dS.$$

At first sight this seems to imply, as Fresnel believed, that the secondary sources oscillate a quarter of a period ahead of the primary wave.

Without loss of generality we may take the wave-surface S to be the plane $x = 0$; if P has coordinates (x_1, y_1, z_1), there are two cases to be considered according as x_1 is positive or negative. We apply Helmholtz's theorem, taking as the surface of integration S the boundary of the volume in which $r_1 \leqslant R$ and $x \geqslant 0$: S consists of



x_1 positive FIG. 5 *x_1 negative*

a plane area S_1 and a portion of a sphere S_2, as shown in the two figures. In either case Helmholtz's integral is

$$I = \frac{1}{4\pi} \iint\limits_{S_1+S_2} \left\{ \frac{e^{ikr_1}}{r_1} \frac{\partial}{\partial \nu}(e^{ik(x-ct)}) - e^{ik(x-ct)} \frac{\partial}{\partial \nu}\left(\frac{e^{ikr_1}}{r_1}\right) \right\} dS,$$

and this has the value $U(P;t)$ or zero according as x_1 is positive or negative.

The integral over S_1 is

$$I_1 = -\frac{1}{4\pi} \iint\limits_{S_1} \left\{ \frac{e^{ikr_1}}{r_1} \frac{\partial}{\partial x}(e^{ik(x-ct)}) + e^{ik(x-ct)} \frac{\partial}{\partial r_1}\left(\frac{e^{ikr_1}}{r_1}\right) \cos\psi \right\} dS.$$

If we neglect $1/r_1$ in comparison with k, we obtain the approximation

$$I_1 = -\frac{1}{4\pi} \iint\limits_{S_1} \frac{ik}{r_1} e^{ik(x+r_1-ct)}(1+\cos\psi)\, dS = -\frac{1}{4\pi} \iint\limits_{S_1} \frac{1+\cos\psi}{r_1} \left[\frac{\partial U}{\partial n}\right] dS,$$

where square brackets again indicate a retarded value.

The integral over S_2 is

$$I_2 = \frac{1}{4\pi} \iint\limits_{S_2} \left\{ \frac{e^{ikr_1}}{r_1} \frac{\partial}{\partial x}(e^{ik(x-ct)}) \frac{\partial x}{\partial r_1} - e^{ik(x-ct)} \frac{\partial}{\partial r_1}\left(\frac{e^{ikr_1}}{r_1}\right) \right\} dS$$

$$= \frac{1}{4\pi} \iint\limits_{S_2} \frac{e^{ik(r_1+x-ct)}}{r_1} \left\{ \frac{ik(x-x_1)}{r_1} - \left(ik - \frac{1}{r_1}\right) \right\} dS.$$

Now the integrand has a constant value on all the circles in which S_2 is cut by planes parallel to S_1, and so we can carry out the integration by dividing S_2 into zones bounded by these circles. The area on S_2 bounded by the two circles $x = x_1+\xi$, $x = x_1+\xi+\delta\xi$ is $2\pi R\,\delta\xi$, and $r_1 = R$ on S_2; hence

$$I_2 = \frac{e^{ik(R+x_1-ct)}}{2R}\int_{-x_1}^{R} e^{ik\xi}\{ik\xi-ikR+1\}\,d\xi$$

$$= \frac{e^{ik(R+x_1-ct)}}{2R}\left[e^{ik\xi}(\xi-R)\right]_{-x_1}^{R} = e^{ik(R-ct)}\frac{R+x_1}{2R}.$$

Combining these formulae for I_1 and I_2, we have the approximate formula

$$I = -\frac{1}{4\pi}\iint\limits_{S_1}\frac{1+\cos\psi}{r_1}\left[\frac{\partial U}{\partial n}\right]dS + e^{ik(R-ct)}\frac{R+x_1}{2R}. \tag{4.63}$$

We cannot make R tend to infinity in (4.63), since the second term on the right-hand side does not tend to a limit. This difficulty can be overcome by taking real parts and choosing special values for R.

If we write $u = \cos k(x-ct)$, it follows from (4.63) that

$$u(P;t) = -\frac{1}{4\pi}\iint\limits_{S_1}\frac{1+\cos\psi}{r_1}\left[\frac{\partial u}{\partial n}\right]dS + \frac{R+x_1}{2R}\cos k(R-ct), \tag{4.64}$$

if P is a point of the region into which the wave-surface S is advancing. If we give R the special value $ct+(m+\frac{1}{2})\pi/k$, where m is a positive integer, (4.64) becomes

$$u(P;t) = -\frac{1}{4\pi}\iint\limits_{S_1}\frac{1+\cos\psi}{r_1}\left[\frac{\partial u}{\partial n}\right]dS.$$

Hence, if we make m tend to infinity,

$$u(P;t) = -\frac{1}{4\pi}\iint\limits_{S}\frac{1+\cos\psi}{r_1}\left[\frac{\partial u}{\partial n}\right]dS, \tag{4.65}$$

where the integration is, in a sense, over the whole wave-surface S but has to be calculated by the special limiting process indicated.

We have thus proved that, for monochromatic plane or spherical waves with velocity potential u,

$$u(P;t) = -\frac{1}{4\pi}\iint\limits_{S}\frac{1+\cos\psi}{r_1}\left[\frac{\partial u}{\partial n}\right]dS,$$

where P is a point of the region into which the wave-surface S is progressing. This formula is usually known as *Stokes's diffraction formula*.† For the application of the formula to the approximate solution of diffraction problems, we refer the reader to the standard text-books.

§ 5. Wave-motions in three dimensions

§ 5.1. Kirchhoff's formula

In the earlier part of this chapter we have given the two analytical formulations of Huygens' principle for sound waves associated with the names of Poisson and Helmholtz. Poisson's formula justifies the principle in the case of an isolated wave due to an initial disturbance; Helmholtz's formula holds in the case of a 'monochromatic' disturbance. We now show that these formulae are particular cases of a general theorem due to Kirchhoff, concerning sound waves of any structure and origin.

We have seen that, if
$$u = ve^{-ikct}$$
is a 'monochromatic' wave-function with no singularities within or on the closed surface S, then the value of u at a point $P(x_1, y_1, z_1)$ within S at the instant t is given by
$$u(P;t) = \frac{1}{4\pi}\iint\limits_S \left\{v\frac{\partial}{\partial n}\left(\frac{e^{ik(r-ct)}}{r}\right) - \frac{e^{ik(r-ct)}}{r}\frac{\partial v}{\partial n}\right\} dS,$$
where r is the distance from P to a typical point (x, y, z) on S and $\partial/\partial n$ denotes differentiation along the inward-drawn normal to S. The value of the integral is zero when P is outside S.

Introducing the 'retarded' values‡ of the various functions involved, we find that Helmholtz's formula can be written as
$$u(P;t) = \frac{1}{4\pi}\iint\limits_S \left\{[u]e^{-ikr}\frac{\partial}{\partial n}\left(\frac{e^{ikr}}{r}\right) - \frac{1}{r}\left[\frac{\partial u}{\partial n}\right]\right\} dS$$
$$= \frac{1}{4\pi}\iint\limits_S \left\{[u]\frac{\partial r}{\partial n}\left(\frac{ik}{r} + \frac{d}{dr}\left(\frac{1}{r}\right)\right) - \frac{1}{r}\left[\frac{\partial u}{\partial n}\right]\right\} dS$$
or, finally,
$$u(P;t) = \frac{1}{4\pi}\iint\limits_S \left\{[u]\frac{\partial r}{\partial n}\frac{d}{dr}\left(\frac{1}{r}\right) - \frac{1}{cr}\frac{\partial r}{\partial n}\left[\frac{\partial u}{\partial t}\right] - \frac{1}{r}\left[\frac{\partial u}{\partial n}\right]\right\} dS,$$
since $\dfrac{\partial u}{\partial t} = -ikcu$ and therefore $\left[\dfrac{\partial u}{\partial t}\right] = -ikc[u]$.

† It was given by Stokes in his memoir on the 'Dynamical Theory of Diffraction' (*Trans. Camb. Phil. Soc.* **9** (1849), 1; *Math. and Phys. Papers*, **2**, 243).

‡ See § 4.5.

The formula we have just obtained is due to Kirchhoff. It expresses $u(P;t)$ as a surface integral in which the period $2\pi/(kc)$ does not occur explicitly, and so it is true for any period. Now an arbitrary function of t can be expressed as a sum of periodic constituents by means of a Fourier series or a Fourier integral. It follows that, since Kirchhoff's formula is linear in u, it holds for any solution of the equation of wave-motions, not merely for solutions corresponding to monochromatic disturbances.†

We have thus obtained the following theorem:‡

Let $u(x,y,z,t)$ be a solution of the equation

$$\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}$$

whose partial derivatives of the first and second orders are continuous within and on a closed surface S, and let (x_1,y_1,z_1) be a point within S. Then

$$u(x_1,y_1,z_1,t) = \frac{1}{4\pi}\iint\limits_S \left\{[u]\frac{\partial}{\partial n}\left(\frac{1}{r}\right) - \frac{1}{cr}\frac{\partial r}{\partial n}\left[\frac{\partial u}{\partial t}\right] - \frac{1}{r}\left[\frac{\partial u}{\partial n}\right]\right\} dS,$$

where r is the distance from (x_1,y_1,z_1) to a typical point of S, $\partial/\partial n$ denotes differentiation along the inward normal to S, and square brackets indicate retarded values. If, however, the point (x_1,y_1,z_1) lies outside S, the value of the integral is zero.

This theorem can also be applied when u has no singularities in the volume V bounded internally by a closed surface S and externally by a sphere Σ whose equation is $x^2+y^2+z^2 = R^2$; if we denote by $\partial/\partial n$ differentiation along the normal drawn into V, the value of the sum of the integrals over S and Σ is $u(x_1,y_1,z_1,t)$ or zero according as (x_1,y_1,z_1) is or is not a point of V.||

Now make $R \to \infty$; we then obtain a theorem valid when u has no singularities outside a closed surface S for all values of t from $-\infty$ up to the instant under consideration, provided only that the

† We do not go into a detailed application of the theory of Fourier series or integrals, as it is possible to give a direct proof of Kirchhoff's formula without using Helmholtz's formula. See § 5.2.

‡ See Kirchhoff, *Berliner Sitzungsber.* (1882), 641; *Annalen der Phys.* **18** (1883), 663; *Vorlesungen ü. math. Phys.* **2** (*Optik*), 23. The theorem has been generalized by W. R. Morgans in *Phil. Mag.* **9** (1930), 141, to cover the case when S is a moving surface.

|| It must not be a point on S or Σ.

integral over Σ tends to zero. This last condition is certainly satisfied if u behaves like $f(ct-R)/R$ for large values of R, where $f(ct)$ and $f'(ct)$ are bounded near $t = -\infty$. The result so obtained is:

Let $u(x,y,z,t)$ be a wave-function which has no singularities outside a closed surface S for all values of t from $-\infty$ up to the instant under consideration, and which behaves like $f(ct-R)/R$ at a large distance R from the origin, where $f(ct)$ and $f'(ct)$ are bounded near $t = -\infty$. Then if (x_1, y_1, z_1) is a point outside S, Kirchhoff's formula holds, provided that $\partial/\partial n$ means differentiation along the outward normal to S. If (x_1, y_1, z_1) is inside S, the value of the integral is zero.

These two theorems constitute Kirchhoff's integral formulation of Huygens' principle. As in the case of Helmholtz's formula, the boundary values of u and $\partial u/\partial n$ on S cannot be arbitrarily assigned independently of each other. (See § 4.3.)

§ 5.2. A direct proof† of Kirchhoff's formula

Let S be a closed surface within and on which the function $v(x,y,z)$ and its partial derivatives of orders one and two are continuous, and let r denote the distance from a fixed point P (x_1, y_1, z_1) outside S. Then, by Green's transformation, we have

$$\iint\limits_S \left\{v\frac{\partial}{\partial\nu}\left(\frac{1}{r}\right)-\frac{1}{r}\frac{\partial v}{\partial\nu}\right\}dS+\iiint\limits_V \frac{1}{r}\nabla^2 v\,dV = 0, \tag{5.21}$$

where V is the volume bounded by S and ν the normal to S directed out of V.

Let $u(x,y,z,t)$ be a solution of

$$\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}$$

which has no singularities inside or on S, and let us take

$$v = u\left(x,y,z,t-\frac{r}{c}\right) = [u]$$

in the formula (5.21). Differentiating with respect to x, we have

$$\frac{\partial v}{\partial x} = \left(\frac{\partial}{\partial x}-\frac{1}{c}\frac{\partial r}{\partial x}\frac{\partial}{\partial t}\right)u\left(x,y,z,t-\frac{r}{c}\right)$$

$$= \left[\frac{\partial u}{\partial x}\right]-\frac{1}{c}\frac{\partial r}{\partial x}\left[\frac{\partial u}{\partial t}\right], \tag{5.22}$$

† This proof is somewhat similar to one published by Gutzmer, *Journal f. Math.* **114** (1895), 333.

and therefore
$$\frac{\partial v}{\partial \nu} = \left[\frac{\partial u}{\partial \nu}\right] - \frac{1}{c}\frac{\partial r}{\partial \nu}\left[\frac{\partial u}{\partial t}\right]. \tag{5.23}$$

From (5.22), we have

$$\frac{\partial}{\partial x}\left[\frac{\partial u}{\partial x}\right] = \left[\frac{\partial^2 u}{\partial x^2}\right] - \frac{1}{c}\frac{\partial r}{\partial x}\left[\frac{\partial^2 u}{\partial x \partial t}\right],$$

$$\frac{\partial}{\partial x}\left[\frac{\partial u}{\partial t}\right] = \left[\frac{\partial^2 u}{\partial x \partial t}\right] - \frac{1}{c}\frac{\partial r}{\partial x}\left[\frac{\partial^2 u}{\partial t^2}\right],$$

and so

$$\frac{\partial^2 v}{\partial x^2} = \left[\frac{\partial^2 u}{\partial x^2}\right] - \frac{2}{c}\frac{\partial r}{\partial x}\left[\frac{\partial^2 u}{\partial x \partial t}\right] + \frac{1}{c^2}\left(\frac{\partial r}{\partial x}\right)^2\left[\frac{\partial^2 u}{\partial t^2}\right] - \frac{1}{c}\frac{\partial^2 r}{\partial x^2}\left[\frac{\partial u}{\partial t}\right].$$

Adding the three equations of this type, we obtain

$$\nabla^2 v = [\nabla^2 u] - \frac{2}{c}\sum \frac{\partial r}{\partial x}\left[\frac{\partial^2 u}{\partial x \partial t}\right] + \frac{1}{c^2}\left[\frac{\partial^2 u}{\partial t^2}\right]\sum\left(\frac{\partial r}{\partial x}\right)^2 - \frac{1}{c}\left[\frac{\partial u}{\partial t}\right]\sum \frac{\partial^2 r}{\partial x^2}$$

$$= \frac{2}{c^2}\left[\frac{\partial^2 u}{\partial t^2}\right] - \frac{2}{cr}\left[\frac{\partial u}{\partial t}\right] - \frac{2}{c}\sum \frac{\partial r}{\partial x}\left[\frac{\partial^2 u}{\partial x \partial t}\right],$$

where $\sum$ denotes summation over the three variables x, y, z.

The integrand of the volume integral in equation (5.21) is therefore

$$\frac{1}{r}\nabla^2 v = \frac{2}{c^2 r}\left[\frac{\partial^2 u}{\partial t^2}\right] - \frac{2}{cr^2}\left[\frac{\partial u}{\partial t}\right] - \frac{2}{c}\sum \frac{x-x_1}{r^2}\left[\frac{\partial^2 u}{\partial x \partial t}\right]. \tag{5.24}$$

If we could transform the expression on the right-hand side into a divergence, we could turn the volume integral in equation (5.21) into a surface integral by Green's theorem. To do so, we proceed as follows:

$$\frac{\partial}{\partial x}\left\{\frac{x-x_1}{r^2}\left[\frac{\partial u}{\partial t}\right]\right\} = \frac{\partial}{\partial x}\left(\frac{x-x_1}{r^2}\right)\left[\frac{\partial u}{\partial t}\right] + \frac{x-x_1}{r^2}\frac{\partial}{\partial x}\left[\frac{\partial u}{\partial t}\right]$$

$$= \frac{1}{r^2}\left[\frac{\partial u}{\partial t}\right] - \frac{2(x-x_1)^2}{r^4}\left[\frac{\partial u}{\partial t}\right] + \frac{x-x_1}{r^2}\left[\frac{\partial^2 u}{\partial x \partial t}\right] - \frac{(x-x_1)^2}{cr^3}\left[\frac{\partial^2 u}{\partial t^2}\right].$$

Adding the three equations of this type, we obtain

$$\sum \frac{\partial}{\partial x}\left\{\frac{x-x_1}{r^2}\left[\frac{\partial u}{\partial t}\right]\right\} = \frac{1}{r^2}\left[\frac{\partial u}{\partial t}\right] - \frac{1}{cr}\left[\frac{\partial^2 u}{\partial t^2}\right] + \sum \frac{x-x_1}{r^2}\left[\frac{\partial^2 u}{\partial x \partial t}\right].$$

Thus (5.24) becomes

$$\frac{1}{r}\nabla^2 v = -\frac{2}{c}\sum \frac{\partial}{\partial x}\left\{\frac{x-x_1}{r^2}\left[\frac{\partial u}{\partial t}\right]\right\},$$

and therefore

$$\iiint\limits_V \frac{1}{r}\nabla^2 v\,dV = -\frac{2}{c}\iiint\limits_V \sum \frac{\partial}{\partial x}\left\{\frac{1}{r}\frac{\partial r}{\partial x}\left[\frac{\partial u}{\partial t}\right]\right\} dV$$

$$= -\frac{2}{c}\iint\limits_S \frac{1}{r}\frac{\partial r}{\partial \nu}\left[\frac{\partial u}{\partial t}\right] dS. \tag{5.25}$$

Substituting in (5.21) from (5.23) and (5.25), we obtain

$$\iint\limits_S \left\{[u]\frac{\partial}{\partial \nu}\left(\frac{1}{r}\right) - \frac{1}{r}\left[\frac{\partial u}{\partial \nu}\right] - \frac{1}{cr}\frac{\partial r}{\partial \nu}\left[\frac{\partial u}{\partial t}\right]\right\} dS = 0,$$

which is Kirchhoff's formula in the case when P and all the singularities of u lie outside S.

If, however, P lies inside S and all the singularities of u lie outside, equation (5.21) has to be replaced by

$$\iint\limits_S + \iint\limits_\sigma \left\{v\frac{\partial}{\partial \nu}\left(\frac{1}{r}\right) - \frac{1}{r}\frac{\partial v}{\partial \nu}\right\} dS + \iiint\limits_{V_1} \frac{1}{r}\nabla^2 v\,dV = 0,$$

where V_1 is the volume bounded externally by S and internally by the small sphere σ with centre P and radius ϵ. A repetition of the previous argument gives

$$\iint\limits_S + \iint\limits_\sigma \left\{[u]\frac{\partial}{\partial \nu}\left(\frac{1}{r}\right) - \frac{1}{r}\left[\frac{\partial u}{\partial \nu}\right] - \frac{1}{cr}\frac{\partial r}{\partial \nu}\left[\frac{\partial u}{\partial t}\right]\right\} dS = 0.$$

If we now make ϵ tend to zero, remembering that $\partial/\partial\nu = -\partial/\partial r$ on σ, we obtain

$$u(x_1, y_1, z_1, t) = -\frac{1}{4\pi}\iint\limits_S \left\{[u]\frac{\partial}{\partial \nu}\left(\frac{1}{r}\right) - \frac{1}{r}\left[\frac{\partial u}{\partial \nu}\right] - \frac{1}{cr}\frac{\partial r}{\partial \nu}\left[\frac{\partial u}{\partial t}\right]\right\} dS,$$

which is Kirchhoff's formula in the case when P lies inside S and all the singularities lie outside.†

The proof of Kirchhoff's formulae in the case when all the singularities lie inside the closed surface S now proceeds as in § 5.1.

§ 5.3. The theorem of determinacy of the solution‡

The formulae of Kirchhoff enable us to express the value of u at a point P on one side of a closed surface S and at the instant t, in terms of the surface values of u and its first partial derivatives at

† The minus sign is due to the difference of definition of the normals ν and n.

‡ Love, *Proc. London Math. Soc.* (2), 1 (1903), 37–62 (42–3).

previous instants. Actually the data are redundant, inasmuch as a knowledge of the surface values of u alone for all values of t is sufficient to determine those of $\partial u/\partial t$, and is, in fact, sufficient, with a knowledge of the initial values of u and $\partial u/\partial t$, to determine u throughout the whole region of space in which it satisfies the equation of wave-motions and the prescribed conditions of continuity. This statement constitutes a part of the theorem of determinacy of the solution of the equation of wave-motions; it follows from it that, if u and $\partial u/\partial t$ are given everywhere initially and u is given for all values of t on the surface S, $\partial u/\partial n$ can have only one definite value at any point of S at any given instant. The surface values of $\partial u/\partial n$ would also suffice in the same way for the determination of u; this constitutes the other part of the theorem of determinacy.

To prove this theorem, let

$$W = \tfrac{1}{2}\iiint\limits_V \left\{\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2+\left(\frac{\partial u}{\partial z}\right)^2+\frac{1}{c^2}\left(\frac{\partial u}{\partial t}\right)^2\right\}\,dxdydz,$$

where integration is over the volume V, bounded by S, in which the wave-function u satisfies the conditions of continuity. Then

$$\begin{aligned}\frac{dW}{dt} &= \iiint\limits_V \left\{\frac{1}{c^2}\frac{\partial u}{\partial t}\frac{\partial^2 u}{\partial t^2}+\frac{\partial u}{\partial x}\frac{\partial^2 u}{\partial x\partial t}+\ldots\right\}\,dxdydz \\ &= \iiint\limits_V \left\{\frac{\partial u}{\partial t}\left(\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}-\nabla^2 u\right)+\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\frac{\partial u}{\partial t}\right)+\ldots\right\}\,dxdydz \\ &= -\iint\limits_S \frac{\partial u}{\partial n}\frac{\partial u}{\partial t}\,dS\end{aligned}$$

by Green's theorem. Integrating with respect to t from the initial instant $t = 0$, we have

$$W-W_0 = -\int\limits_0^t dt \iint\limits_S \frac{\partial u}{\partial n}\frac{\partial u}{\partial t}\,dS.$$

Now if there were two wave-functions satisfying the same initial and boundary conditions as well as the prescribed conditions of continuity, their difference would be a wave-function u for which W_0 vanishes, and, moreover, either $\partial u/\partial n$ or $\partial u/\partial t$ would vanish on S for all values of t, according as the surface values of $\partial u/\partial n$ or u are given. It follows that, for such a function, W is zero for positive values of t. But since W is the integral of a function which is never negative, the integrand

$$\frac{1}{2}\left\{\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2+\left(\frac{\partial u}{\partial z}\right)^2+\frac{1}{c^2}\left(\frac{\partial u}{\partial t}\right)^2\right\}$$

is zero everywhere in V. From this it follows that u is constant in V. Being initially zero, u must then be zero for all positive values of t. Hence there are not two wave-functions which satisfy the same initial and boundary conditions and the same conditions of continuity.

To extend the theorem to space outside a closed surface, it is necessary to prescribe the asymptotic behaviour of the wave-functions under consideration.

§ 5.4. The physical interpretation of Kirchhoff's formula

The velocity potential due to a simple source of strength $f(t)$ is

$$u=\frac{1}{r}f\left(t-\frac{r}{c}\right)=\frac{1}{r}[f(t)].$$

Hence the velocity potential at a point P outside a closed surface S due to a distribution over S of simple sources of strength $f(t)$ per unit area is

$$u=\iint\limits_S \frac{1}{r}[f]\,dS, \tag{5.41}$$

where r is the distance from P to a typical point of S and f may depend on the position of the element dS as well as on t.

Again, the velocity potential due to a doublet of strength $F(t)$ whose axis is directed along the unit vector $\mathbf{n}$ is

$$u=\frac{\partial}{\partial n}\left\{\frac{1}{r}F\left(t-\frac{r}{c}\right)\right\}=-\frac{1}{cr}\frac{\partial r}{\partial n}\left[\frac{\partial F}{\partial t}\right]+[F]\frac{\partial}{\partial n}\left(\frac{1}{r}\right).$$

Thus the disturbance at P outside S due to a distribution of doublets, directed normally to S and of strength $F(t)$ per unit area, is specified by

$$u=\iint\limits_S \left\{-\frac{1}{cr}\frac{\partial r}{\partial n}\left[\frac{\partial F}{\partial t}\right]+\frac{\partial}{\partial n}\left(\frac{1}{r}\right)[F]\right\}dS, \tag{5.42}$$

where $\partial/\partial n$ means differentiation along the outward normal to S; F may also depend on the position of the element dS.

It is readily seen† that across a sheet of doublets directed normally to S of strength F per unit area, there is a discontinuity

† Cf. Larmor, *Proc. London Math. Soc.* (2), 1 (1903), 1–13 (6–7).

in u of amount $4\pi F$ and no discontinuity in $\partial u/\partial n$; whereas in crossing a sheet of simple sources of strength f per unit area, there is a discontinuity $-4\pi f$ in $\partial u/\partial n$ and no discontinuity in u. Thus if we wish to have a distribution of sources and doublets on S such that the disturbance shall change suddenly from a null one just inside S to that specified by u and $\partial u/\partial n$ just outside S, then the distribution must consist of sources of strength $-\dfrac{1}{4\pi}\dfrac{\partial u}{\partial n}$ per unit area and normally directed doublets of strength $u/4\pi$ per unit area: and, by (5.41) and (5.42), these produce outside S a disturbance specified by

$$\frac{1}{4\pi}\iint\limits_S \left\{-\frac{1}{r}\left[\frac{\partial u}{\partial n}\right]-\frac{1}{cr}\frac{\partial r}{\partial n}\left[\frac{\partial u}{\partial t}\right]+\frac{\partial}{\partial n}\left(\frac{1}{r}\right)[u]\right\} dS.$$

This expression must represent u outside S, which is precisely Kirchhoff's result. Hence Kirchhoff's integral formula asserts that *the disturbance outside a closed surface S due to real sources inside S is the same as would be produced by a fictitious distribution of sources over S of strength* $-\dfrac{1}{4\pi}\dfrac{\partial u}{\partial n}$ *per unit area, together with normally directed doublets of strength $u/4\pi$ per unit area.*

The proof of Kirchhoff's formula given in § 5.2 and the argument of Larmor which we have just sketched both deal with the analytical theory of solutions of the equation of wave-motions. In the case when the wave-function u is the velocity potential of sound waves of small amplitude, the dynamical argument of § 4.4 is applicable with slight changes and leads at once to Kirchhoff's formula. Thus Kirchhoff's formula must be regarded as the analytical formulation of Huygens' principle for sound waves in air.

Kirchhoff's distribution is not the only distribution of secondary sources on the closed surface S which gives the same effect at a point P outside S as is given by the primary sources within S. For we can superpose on it any distribution which gives a null effect outside S, such as the Kirchhoff distribution corresponding to any set of primary sources outside S. But the distribution of secondary sources obtained in this way does not give a null effect inside S and so, if this distribution had any physical meaning, an expanding wave-front would appear to be propagated not only forwards but also backwards, which is impossible.

Kirchhoff's distribution is, in fact, the only distribution of secondary sources which gives the correct effect at points outside S and a null effect inside. For if it were not, it would be possible to construct, by subtraction, a distribution of secondary sources which would have a null effect everywhere.

The analysis of this chapter still leaves open the question of the validity of Huygens' principle in Optics. For, as we have already remarked, a theory which regards the propagation of light as analogous to that of sound is quite inadequate, since a luminous disturbance is characterized by a vector, not a scalar. A very close representation of the physical propagation of light was attained by Stokes in his memoir on the 'Dynamical Theory of Diffraction', in which the aether was regarded as an elastic solid whose state is specified everywhere and at every instant of time by the strain or displacement and the velocity of each element of volume: each element of volume thus disturbed is regarded as a source from which a secondary disturbance is propagated, and the law of the disturbance in a secondary wave is found on purely dynamical principles. The resultant disturbance at any point at a subsequent instant is then expressed as a volume integral. Whilst it is true that this volume integral can be reduced mathematically to a surface integral, similar to that which occurs in Kirchhoff's theory, the exact physical representation of the propagation of light according to the elastic solid theory is lost in the reduction.

We shall not refer further to this theory as it has been replaced by the electromagnetic theory developed in Chapter III.

Ex. 1. u is a wave-function whose singularities all lie on one side of a plane S. Prove that the value of u at a point P on the other side of S can be expressed in the forms

(i) $$u(P;t) = -\frac{1}{2\pi}\iint\limits_S \frac{1}{r}\left[\frac{\partial u}{\partial n}\right] dS,$$

involving only simple sources, and

(ii) $$u(P;t) = \frac{1}{2\pi}\iint\limits_S \left\{[u]\frac{\partial}{\partial n}\left(\frac{1}{r}\right) - \frac{1}{cr}\frac{\partial r}{\partial n}\left[\frac{\partial u}{\partial t}\right]\right\} dS,$$

involving only doublets. Show also that neither of these distributions give a null effect at a point on the same side of S as the singularities of u.

[Consider Kirchhoff's integral for P', the image of P in S.]

Ex. 2. Prove that, if $\quad [u] = u\left(x, y, z, t-\frac{r}{c}\right),$

where r is the distance from (x_1, y_1, z_1), then

$$\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial x_1}\right)[u] = \left[\frac{\partial u}{\partial x}\right].$$

Hence prove that,† if u is a wave-function, the value for $\partial u/\partial x$ at (x_1, y_1, z_1) is obtained either by differentiating the Kirchhoff integral for u with respect to x_1 or by forming the Kirchhoff integral for $\partial u/\partial x$.

Ex. 3. By means of Kirchhoff's formula prove that, if u is a solution of the equation of wave-motions,

$$u(P;t) = \frac{1}{4\pi}\iint\limits_S \left\{\frac{1}{r^2}(u)_0+\frac{1}{cr}\left(\frac{\partial u}{\partial t}\right)_0+\frac{1}{r}\left(\frac{\partial u}{\partial r}\right)_0\right\} dS,$$

where S is the sphere of centre P and radius ct and $(\quad)_0$ indicates that the expression inside the brackets is to be evaluated at the instant $t = 0$. (Poincaré.)

Ex. 4. A spherical expanding wave is defined by the initial values

$$u = \frac{F(R)}{R}, \qquad \frac{\partial u}{\partial t} = -\frac{cF'(R)}{R} \quad (t = 0).$$

By means of Poincaré's formula (Ex. 3), show that at a point P at a distance R_0 from O and at the instant t,

$$u = \left[\frac{(R-ct)^2-R_0^2}{2RR_0\,ct}F(R)\right]_{R_2'}^{R_1'}.$$

Apply this to the case when $F(R) = 0$ except when $R_2 \leqslant R \leqslant R_1$, and deduce the results of § 3.7. (Love.)

Ex. 5. Deduce Poisson's formula

$$u(P;t) = tM_{P,ct}\{f\}+\frac{\partial}{\partial t}(tM_{P,ct}\{g\})$$

for a wave-function satisfying the initial conditions

$$u = g, \qquad \frac{\partial u}{\partial t} = f \quad (t = 0)$$

from Poincaré's formula (Ex. 3).

§ 6. Wave motions in two dimensions

§ 6.1. The equation of cylindrical waves

A wave-function u which does not involve one of the Cartesian coordinates, z say, represents a disturbance which is the same in all planes perpendicular to Oz. Such a disturbance is usually called

† This theorem shows that Kirchhoff's formula is consistent. It is of importance in electromagnetic theory, where we may apply Kirchhoff's formula either to the components of the electric and magnetic vectors or to the scalar and vector potentials and then use the formulae

$$\mathbf{d} = -\frac{1}{c}\dot{\mathbf{a}}-\operatorname{grad}\phi, \qquad \mathbf{h} = \operatorname{curl}\mathbf{a};$$

by the theorem, both methods give the same results.

a cylindrical wave, since the wave-fronts are cylinders with generators parallel to Oz; the corresponding wave-function satisfies the equation of cylindrical waves

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}. \tag{6.11}$$

This differential equation is also called the equation of two-dimensional wave-motions, since it occurs in problems, such as the problem of the normal vibrations of a membrane, which are actually two-dimensional.

Let us consider the solution of the wave-equation

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} \tag{6.12}$$

in the case where the initial conditions

$$u = g(x,y), \qquad \frac{\partial u}{\partial t} = f(x,y) \quad (t = 0) \tag{6.13}$$

do not involve the variable z. If the solution were

$$u = \phi(x,y,z,t),$$

involving the variable z, then

$$u = \phi(x,y,z+h,t)$$

would also be a solution for any value of the constant h; and this is impossible since there is only one wave-function which satisfies the conditions (6.13). Hence a wave-function which satisfies initial conditions not involving the variable z must be a cylindrical wave-function.

From this it follows that a method of determining a wave-function in three dimensions under given conditions can also be used for finding cylindrical wave-functions. This method is called by Hadamard the *Method of Descent*, for it involves descending from three spatial dimensions to two.

There is, however, an important difference between wave-motions in two and three dimensions. To see how this difference arises, let us consider first the three-dimensional wave-motion under the initial conditions

$$u = g(x,y,z), \qquad \frac{\partial u}{\partial t} = f(x,y,z) \quad (t = 0),$$

where f and g vanish whenever

$$R^2 \equiv x^2+y^2+z^2 > a^2;$$

the initial disturbance is thus confined to the region $R \leqslant a$. By Poisson's formula

$$u(P;t) = tM\{f\}+\frac{\partial}{\partial t}(tM\{g\}),$$

we see that the disturbance at a point P at a distance $R > a$ from the origin is null until the instant $t = (R-a)/c$, and that it is again null when $t > (R+a)/c$. Thus the disturbance at P lasts for a time $2a/c$; in other words, we have a clean-cut disturbance.

Let us compare this result with the cylindrical wave-motion defined by

$$u = g(x,y), \qquad \frac{\partial u}{\partial t} = f(x,y) \quad (t = 0),$$

where f and g vanish whenever

$$r^2 \equiv x^2+y^2 > a^2;$$

the initial disturbance is thus confined to the cylinder $r \leqslant a$. An application of Poisson's formula shows, as before, that at a point P at a distance $r > a$ *from the axis of* z the effect is null until the instant $t = (r-a)/c$. But after this instant the disturbance at P is never null; for the sphere with centre P and radius ct always intersects the cylinder $r = a$, no matter how large t may be, and so the mean values which appear in Poisson's formula are not zero. Thus, if the disturbance in a two-dimensional wave-motion is initially confined to the region $x^2+y^2 \leqslant a^2$, the head of the wave is sharply defined; but instead of having a sharply defined rear, such a wave-motion possesses a 'tail'. This characteristic of a two-dimensional wave-motion is sometimes called *diffusion*.

In most cases this residual after-effect would be expected by the physicist. For in such problems the sources of disturbance are line sources extending to infinity in a three-dimensional space: even though such line sources may act only for a finite time, the disturbance from each element of a line source travels with a finite speed and so the whole disturbance can never pass any given point completely. There are, however, some problems, such as the problem of the vibrating membrane, in which physical intuition would not lead us to expect diffusion.

From the purely mathematical point of view, the phenomenon of diffusion is of great interest. As Hadamard† has shown, it also occurs

† See his Yale *Lectures on Cauchy's Problem* (1923), 236.

in more general partial differential equations and its occurrence is intimately associated with the properties of Hadamard's 'elementary solution'.

Ex. 1. Prove by means of Poisson's formula that the cylindrical wave-function which satisfies the initial conditions

$$u = g(x,y), \qquad \partial u/\partial t = f(x,y) \quad (t = 0)$$

is

$$u(x,y,t) = \frac{1}{2\pi c}\int\limits_0^{ct}\int\limits_0^{2\pi} f(x+r\cos\theta,\, y+r\sin\theta)\frac{r\,d\theta\,dr}{\sqrt{\{c^2t^2-r^2\}}}+$$
$$+\frac{1}{2\pi c}\frac{\partial}{\partial t}\int\limits_0^{ct}\int\limits_0^{2\pi} g(x+r\cos\theta,\, y+r\sin\theta)\frac{r\,d\theta\,dr}{\sqrt{\{c^2t^2-r^2\}}}.$$

Ex. 2.† Prove that a line source of strength $f(t)$ along Oz generates expanding cylindrical waves specified by

$$u = \frac{1}{2\pi}\int\limits_{-\infty}^{t-r/c} cf(\theta)\frac{d\theta}{\sqrt{\{c^2(t-\theta)^2-r^2\}}},$$

where $r^2 = x^2+y^2$, provided that $f'(t) = 0$ when $t < -\alpha$.

If the source acts only during the interval $0 < t < T$, verify that the head of the wave is propagated cleanly and that there is a 'tail'

$$u = \frac{c}{2\pi\sqrt{\{c^2t^2-r^2\}}}\int\limits_0^T f(\theta)\,d\theta$$

when $ct-r$ is large compared with cT.

Ex. 3. The function u satisfies the differential equation

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}$$

and the initial conditions $u = \phi(x)$, $\partial u/\partial t = \psi(x)$ when $t = 0$, where ϕ and ψ are non-zero only in the small interval $x' \leqslant x \leqslant x''$. Discuss the behaviour of u at the point $x = x_0$ as t varies, and show that there is a residual after-effect

$$u = \frac{1}{2c}\int\limits_{x'}^{x''}\psi(\tau)\,d\tau. \qquad \text{(Hadamard.)}$$

§ 6.2. Weber's solution for 'monochromatic' cylindrical waves

Although it is possible to determine cylindrical wave-functions by applying the method of descent, it is usually more convenient to discuss the equation of cylindrical waves,

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2},$$

† See Lamb, *Hydrodynamics* (Cambridge, 1916), 291.

on its own merits as a two-dimensional problem, without introducing a third Cartesian coordinate z. All the formulae concerning wave-motion in three dimensions have their two-dimensional analogues, but the proofs of the two-dimensional formulae are usually much more difficult.

We prove first a formula, due to Weber, which is the analogue of Helmholtz's formula for 'monochromatic' wave-functions. Helmholtz's theory depended on the existence of a monochromatic spherical wave-function $e^{ik(R-ct)}/R$. To extend his theory to two dimensions, we must find a monochromatic cylindrical wave-function $u = ve^{-ikct}$, where v is a function of $r = \sqrt{(x^2+y^2)}$ alone. Since the equation of cylindrical waves in plane polar coordinates is

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2},$$

it is evident that v is a solution of

$$\frac{d^2 v}{dr^2}+\frac{1}{r}\frac{dv}{dr}+k^2 v = 0, \tag{6.21}$$

and this is Bessel's equation of order zero and independent variable kr.

The solution $J_0(kr)$ of equation (6.21) has no singularity for any finite value of r. It follows that the cylindrical wave-function $J_0(kr)e^{-ikct}$ is not the velocity potential due to a source† at a finite distance, and so is unsuited to our purpose.

The other well-known solution $Y_0(kr)$ of (6.21) is also unsuitable for a different reason. The most convenient solutions to use in the theory of cylindrical waves are the Hankel functions, which may be defined by the equations‡

$$H_0^{(1)}(kr) = \frac{2}{\pi i}\int\limits_0^\infty e^{ikr\cosh\tau}\,d\tau, \qquad H_0^{(2)}(kr) = -\frac{2}{\pi i}\int\limits_0^\infty e^{-ikr\cosh\tau}\,d\tau, \tag{6.22}$$

when kr is positive. These functions are linearly independent; they behave near the origin like $\pm(2i/\pi)\log r$, but have no other singularity at a finite distance. The Hankel functions can be expressed in terms of Bessel functions by the equations

$$H_0^{(1)}(kr) = J_0(kr)+iY_0(kr), \qquad H_0^{(2)}(kr) = J_0(kr)-iY_0(kr). \tag{6.23}$$

† The reader who regards cylindrical waves as existing in a three-dimensional space will understand by the word 'source' a uniform distribution of centres of disturbance along a line parallel to Oz and extending to infinity in both directions.

‡ See, for example, Watson, *Theory of Bessel Functions* (Cambridge, 1922), 73, 180, 198.

When r is large, it can be shown that

$$H_0^{(1)}(kr) \sim \left(\frac{2}{\pi kr}\right)^{\frac{1}{2}} e^{i(kr-\frac{1}{4}\pi)}, \qquad H_0^{(2)}(kr) \sim \left(\frac{2}{\pi kr}\right)^{\frac{1}{2}} e^{-i(kr-\frac{1}{4}\pi)}. \quad (6.24)$$

Hence $H_0^{(1)}(kr)e^{-ikct}$ represents cylindrical waves diverging from the origin, whereas $H_0^{(2)}(kr)e^{-ikct}$ represents cylindrical waves converging to the origin. It follows from (6.23) that $Y_0(kr)e^{-ikct}$ represents standing waves, and so is unsuited to our purpose. The function which plays in the theory of cylindrical waves the same part as $e^{ik(R-ct)}/R$ plays in the theory of spherical waves is $H_0^{(1)}(kr)e^{-ikct}$.

The most general cylindrical wave-function of period $2\pi/(kc)$ is of the form

$$u = ve^{-ikct},$$

where v does not depend on t and is a solution of

$$(\nabla^2+k^2)v = 0. \quad (6.25)$$

Now let Γ be a closed contour bounding the region D in the (x,y) plane. If v and w are two functions whose first- and second-order partial derivatives are continuous within and on Γ, Green's transformation gives

$$\int_\Gamma \left(v\frac{\partial w}{\partial \nu}-w\frac{\partial v}{\partial \nu}\right) ds = \iint_D (v\nabla^2 w - w\nabla^2 v)\, dxdy, \quad (6.26)$$

where $\partial/\partial\nu$ means differentiation along the normal to Γ drawn out of D. If v and w are both solutions of equation (6.25), this formula becomes

$$\int_\Gamma \left(v\frac{\partial w}{\partial \nu}-w\frac{\partial v}{\partial \nu}\right) ds = 0.$$

In particular, by taking $w = H_0^{(1)}(kr_1)$, where r_1 denotes the distance from $P(x_1, y_1)$, we obtain

$$\int_\Gamma \left\{v\frac{\partial}{\partial \nu}H_0^{(1)}(kr_1)-H_0^{(1)}(kr_1)\frac{\partial v}{\partial \nu}\right\} ds = 0 \quad (6.27)$$

provided that P and the singularities of v lie outside Γ.

If, however, P lies inside Γ and the singularities of v lie outside, equation (6.27) no longer holds since $H_0^{(1)}(kr_1)$ has a logarithmic singularity at P. In this case we apply Green's transformation to the region D_1 bounded externally by Γ and internally by a circle σ of centre P and radius ϵ. Formula (6.26) becomes

$$\int\limits_{\Gamma} \left(v\frac{\partial w}{\partial \nu}-w\frac{\partial v}{\partial \nu}\right) ds + \int\limits_{\sigma} \left(v\frac{\partial w}{\partial \nu}-w\frac{\partial v}{\partial \nu}\right) ds = \iint\limits_{D_1} (v\nabla^2 w - w\nabla^2 v)\, dxdy.$$

Proceeding as before, we obtain

$$\int\limits_{\Gamma} \left\{v\frac{\partial}{\partial \nu} H_0^{(1)}(kr_1) - H_0^{(1)}(kr_1)\frac{\partial v}{\partial \nu}\right\} ds$$

$$= \int\limits_{\sigma} \left\{v\frac{\partial}{\partial r_1} H_0^{(1)}(kr_1) - H_0^{(1)}(kr_1)\frac{\partial v}{\partial r_1}\right\} ds$$

$$= \lim_{\epsilon\to 0} \int\limits_{\sigma} \left\{v\frac{\partial}{\partial r_1} H_0^{(1)}(kr_1) - H_0^{(1)}(kr_1)\frac{\partial v}{\partial r_1}\right\} ds$$

since the integral round Γ does not depend on ϵ. Now on σ the dominant parts of $H_0^{(1)}(kr_1)$ and $\frac{\partial}{\partial r_1} H_0^{(1)}(kr_1)$ are respectively $(2i/\pi)\log\epsilon$ and $2i/(\pi\epsilon)$; moreover v and its first derivatives are continuous at P. It follows at once that

$$\lim_{\epsilon\to 0} \int\limits_{\sigma} = 4iv(P).$$

We have thus proved the following theorem:

Let v be a solution of the equation

$$\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}+k^2 v = 0$$

whose partial derivatives of the first and second orders are continuous within and on a closed curve Γ; and let

$$I = \int\limits_{\Gamma} \left\{v\frac{\partial}{\partial \nu} H_0^{(1)}(kr_1) - H_0^{(1)}(kr_1)\frac{\partial v}{\partial \nu}\right\} ds, \tag{6.28}$$

where r_1 is the distance from a fixed point P and $\partial/\partial\nu$ means differentiation along the outward normal to Γ. Then $I = 0$ or $4iv(P)$ according as P lies outside or inside Γ.

This is Weber's† analogue of Helmholtz's theorem. There is a corresponding theorem in the case when v has continuous partial derivatives of the first and second orders everywhere outside a closed curve Γ. This is most easily obtained by applying the previous argument first to the case when v has no singularities in the annulus

† *Math. Ann.* **1** (1869), 1–36.

bounded internally by Γ and externally by a circle Σ, whose equation is $r = R$. The value of

$$\int_{\Gamma} + \int_{\Sigma} \left\{ v \frac{\partial}{\partial \nu} H_0^{(1)}(kr_1) - H_0^{(1)}(kr_1) \frac{\partial v}{\partial \nu} \right\} ds,$$

where $\partial/\partial\nu$ means differentiation along the normal out of the annulus, then turns out to be $4iv(P)$ or zero according as P is or is not a point of the annulus. The required result follows by making $R \to \infty$, provided that the integral over Σ tends to zero; and a sufficient condition for this is that v should behave like $H_0^{(1)}(kr)$ for large values of r. The details, which involve a use of the asymptotic property (6.24) of $H_0^{(1)}(kr)$, are left to the reader.

In the formula (6.28) we may evidently replace $H_0^{(1)}(kr)$ by any other Bessel function which behaves like $(2i/\pi)\log r$ near the origin, and, in particular, by the function $iY_0(kr)$. This leads to the form of the result actually proved by Weber.

§ 6.3. Volterra's analogue of Kirchhoff's formula

We proved in § 6.2 that, if $u = ve^{-ikct}$ is a cylindrical wave-function in which v does not depend on t, then

$$v(P) = \frac{1}{4i} \int_{\Gamma} \left\{ v \frac{\partial}{\partial \nu} H_0^{(1)}(kr_1) - H_0^{(1)}(kr_1) \frac{\partial v}{\partial \nu} \right\} ds$$

when P lies inside the closed contour Γ But, by (6.22), we have

$$H_0^{(1)}(kr) = \frac{2}{\pi i} \int_0^\infty e^{ikr\cosh\tau}\, d\tau = \frac{2}{\pi i} \int_1^\infty e^{ikr\theta} \frac{d\theta}{\sqrt{(\theta^2-1)}}.$$

It follows that

$$u(P;t) = \frac{1}{2\pi} \int_{\Gamma} \left\{ \frac{\partial v}{\partial \nu} \int_1^\infty e^{ik(r_1\theta - ct)} \frac{d\theta}{\sqrt{(\theta^2-1)}} - v \frac{\partial}{\partial \nu} \int_1^\infty e^{ik(r_1\theta - ct)} \frac{d\theta}{\sqrt{(\theta^2-1)}} \right\} ds,$$

where P lies inside the closed contour Γ which contains no singularities of u; but when P is outside Γ, the value of the integral is zero.

If we put $r_1\theta = \psi$, we obtain

$$u(P;t) = \frac{1}{2\pi} \int_{\Gamma} \left\{ \frac{\partial v}{\partial \nu} \int_{r_1}^\infty e^{ik(\psi - ct)} \frac{d\psi}{\sqrt{(\psi^2-r_1^2)}} - v \frac{\partial}{\partial \nu} \int_{r_1}^\infty e^{ik(\psi - ct)} \frac{d\psi}{\sqrt{(\psi^2-r_1^2)}} \right\} ds$$

$$= \frac{1}{2\pi} \int_{\Gamma} \left\{ \frac{\partial}{\partial \nu} \int_{r_1}^\infty ve^{-ik(ct-\psi)} \frac{d\psi}{\sqrt{(\psi^2-r_1^2)}} - \frac{\delta}{\delta \nu} \int_{r_1}^\infty ve^{-ik(ct-\psi)} \frac{d\psi}{\sqrt{(\psi^2-r_1^2)}} \right\} ds$$

where, if $f = f(x, y, r_1)$, we write

$$\frac{\partial f}{\partial \nu} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \nu}+\frac{\partial f}{\partial y}\frac{\partial y}{\partial \nu}, \qquad \frac{\delta f}{\delta \nu} = \frac{\partial f}{\partial r_1}\frac{\partial r_1}{\partial \nu},$$

the variables x, y, and r_1 being regarded as independent. By a slight change of notation, this becomes

$$u(x_1, y_1, t) = \frac{1}{2\pi}\int\limits_{\Gamma}\left\{\left(\frac{\partial}{\partial \nu}-\frac{\delta}{\delta \nu}\right)\int\limits_{r_1}^{\infty} u\left(x, y, t-\frac{\psi}{c}\right)\frac{d\psi}{\sqrt{(\psi^2-r_1^2)}}\right\} ds. \quad (6.31)$$

The equation (6.31) expresses the 'monochromatic' cylindrical wave-function $u = ve^{-ikct}$ as an integral in which there is no explicit mention of the period $2\pi/(kc)$, and so it holds for any period whatever. Moreover, as it is linear in u, (6.31) also holds for any cylindrical wave-function obtained by adding 'monochromatic' solutions of different frequencies. But as an arbitrary function of t can be expressed as a sum of periodic constituents by means of a Fourier series or integral, the equation (6.31) holds for all cylindrical wave-functions. This result, which is due to Volterra,† is the analogue of Kirchhoff's theorem for two-dimensional wave-motions.

By a similar argument, we can show that Volterra's formula holds when (x_1, y_1) lies outside a closed curve Γ containing all the singularities of u, provided that ν denotes the normal drawn inwards and provided that u behaves suitably at infinity.

The phenomenon of 'diffusion', to which attention was drawn in § 6.1, is evident from Volterra's formula (6.31). For whereas in Kirchhoff's formula (§ 5.1) the value of a three-dimensional wave-function at P at the instant t depended on the disturbance at points Q of a surface S at the instant $t-PQ/c$, the value of a cylindrical wave-function is expressed in terms of the disturbance at points Q of a curve Γ at the instant $t-PQ/c$ and all previous instants.

§ 6.4. Note on the proof of Volterra's formula

Whilst it is possible under certain conditions to justify the appeal made in § 6.3 to the theory of Fourier series and integrals, it is most desirable to give a direct proof of the analogue of Kirchhoff's formula. Volterra‡ has proved directly a more general theorem which includes the result of § 6.3 as a particular case.

† *Acta Math.* **18** (1894), 161. Volterra's formula has been used by Sommerfeld (*Zeits. f. Math. u. Phys.* **46** (1901), 11–97 (§ 9)) in his discussion of the diffraction of X-rays.

‡ *Rend. dei Lincei* (5), I_2 (1892), 161, 265; *Stockholm Lectures* (1912).

In this more general theorem, which can be regarded as the analogue of the Riemann-Green theory of one-dimensional wave-propagation†, (x, y, t) are interpreted as Cartesian coordinates in a three-dimensional space. On a certain surface σ we are given the values taken by a cylindrical wave-function u and its first partial derivatives.‡ Then

$$u(x_1, y_1, t_1) = \frac{1}{2\pi}\iint\limits_{\sigma_1} \frac{c}{\sqrt{\{c^2(t-t_1)^2-r^2\}}}\left\{\frac{\partial u}{\partial x}\frac{\partial x}{\partial \nu}+\frac{\partial u}{\partial y}\frac{\partial y}{\partial \nu}-\frac{1}{c^2}\frac{\partial u}{\partial t}\frac{\partial t}{\partial \nu}\right\} dS - $$
$$-\frac{1}{2\pi}\frac{\partial}{\partial t_1}\iint\limits_{\sigma_1} \frac{cu}{\sqrt{\{c^2(t-t_1)^2-r^2\}}}\left\{\frac{t-t_1}{r}\frac{\partial r}{\partial \nu}+\frac{1}{c^2}\frac{\partial t}{\partial \nu}\right\} dS,$$

where $$r^2 = (x-x_1)^2+(y-y_1)^2;$$

in this formula σ_1 is the area cut out of σ by the cone $r = c(t-t_1)$ and $\partial/\partial\nu$ denotes differentiation along the normal to σ.

The surface σ in this theorem has to satisfy certain conditions, which we shall not attempt to discuss. We may, however, remark that, in deducing the formula of § 6.3 from the general theorem, we take σ to be a cylinder whose generators are parallel to Ot and which has the curve Γ as cross-section. Volterra's proof of the more general theorem just enunciated and a later proof due to Hadamard are both too difficult to give here.

Quite recently Professor Marcel Riesz has discovered an elegant and simple method for dealing with problems of this type. We conclude this chapter by showing where the difficulties in the work of Volterra and Hadamard lie and how Riesz has been able to avoid these difficulties by making use of the theory of analytical continuation of a function of a complex variable. For the sake of simplicity we restrict ourselves to the initial value problem, when Volterra's surface σ reduces to a plane perpendicular to the axis of t.

§ 7. Marcel Riesz's solution of the equation of cylindrical waves

§ 7.1. An analogy with potential theory

Let u be a solution of Laplace's equation

$$\nabla^2 u \equiv \frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2} = 0 \qquad (7.11)$$

† For an account of this theory see, for example, Goursat, *Cours d'Analyse*, **3** (1923), 137–55.

‡ We recall that a knowledge of u and one non-tangential partial derivative determines the other two partial derivatives of the first order.

which has no singularities within or on a closed surface S, and let

$$v = \frac{1}{\sqrt{\{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2\}}}, \tag{7.12}$$

where $P(x_0, y_0, z_0)$ is a fixed point inside S. Then if V is the volume bounded externally by S and internally by a small sphere S_0 with centre P, we have

$$\iiint\limits_V \{u\nabla^2 v - v\nabla^2 u\}\, dxdydz = 0;$$

from this it follows by Green's transformation that

$$\iint\limits_{S+S_0} \left\{u\frac{\partial v}{\partial \nu} - v\frac{\partial u}{\partial \nu}\right\} dS = 0,$$

where $\partial/\partial\nu$ denotes differentiation along the normal out of V. If we now make the radius of S_0 tend to zero, we obtain

$$u(P) = \frac{1}{4\pi}\iint\limits_S \left\{v\frac{\partial u}{\partial \nu} - u\frac{\partial v}{\partial \nu}\right\} dS. \tag{7.13}$$

When this equation, which is a special case of Helmholtz's formula of § 4.2, is applied to gravitation, it states that the potential at a point P inside S due to matter outside S is the same as the potential due to 'Green's Equivalent Layer' of matter and normal doublets on S. Alternatively we may regard (7.13) as providing the solution of the boundary value problem for Laplace's equation in which the values† of u and $\partial u/\partial\nu$ are given on S. It is the latter interpretation of the Green's equivalent layer theorem which interests us here.

If we replace x and y by ix and iy respectively, u becomes a solution of the equation of cylindrical waves

$$L(u) \equiv \frac{\partial^2 u}{\partial z^2} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = 0, \tag{7.14}$$

and v becomes the particular solution of this equation

$$v = \frac{1}{\sqrt{\{(z-z_0)^2-(x-x_0)^2-(y-y_0)^2\}}}. \tag{7.15}$$

By analogy with potential theory, we should expect that an application of Green's transformation to the identity

$$\iiint\limits_V \{uL(v) - vL(u)\}\, dxdydz = 0 \tag{7.16}$$

† The boundary values of u and $\partial u/\partial\nu$ are not independent.

would enable us to find u in terms of the boundary values of u and $\partial u/\partial \nu$ on a surface S. But if we attempt to carry out this process, we immediately meet difficulties which do not appear in potential theory.

In the first place, the function v defined by (7.15) is real only when
$$(x-x_0)^2+(y-y_0)^2 \leqslant (z-z_0)^2.$$
Accordingly we take the volume† V to be bounded by part of the cone
$$(x-x_0)^2+(y-y_0)^2 = (z-z_0)^2 \tag{7.17}$$
and by the area which this cone cuts out of S; then v is real everywhere in V.

The more serious difficulty is that, when we choose V in this way and apply Green's transformation to (7.16), we obtain an integral of the required form over a part of S together with an integral over the cone (7.17), and the latter integral is infinite since v and its derivatives are infinite on the cone.

There are two classical ways of avoiding this difficulty. Volterra‡ replaced v by its integral with respect to z_0; in fact, he wrote
$$v = \int \frac{dz_0}{\sqrt{\{(z-z_0)^2-(x-x_0)^2-(y-y_0)^2\}}} = \cosh^{-1}\frac{|z-z_0|}{\sqrt{\{(x-x_0)^2+(y-y_0)^2\}}}$$
in the identity (7.16). This function has no singularity on the cone, but has a line of singularities on the axis of the cone. By cutting out the singularities on the axis by means of a small coaxial cylinder and then applying Green's transformation, Volterra obtained a formula for
$$\int u(x_0,y_0,z_0)\,dz_0$$
in terms of the boundary values of u and $\partial u/\partial \nu$, and readily deduced the required expression for $u(x_0,y_0,z_0)$. The values of u and $\partial u/\partial \nu$ on the cone do not appear in the solution on account of the properties of Volterra's function v.

Hadamard,|| on the other hand, used the identity (7.16) with
$$v = \frac{1}{\sqrt{\{(z-z_0)^2-(x-x_0)^2-(y-y_0)^2\}}},$$

† This should be compared with the corresponding step in the Riemann-Green theory of hyperbolic equations in two independent variables.

‡ *Rend. dei Lincei* (5), I_2 (1892), 161, 265. See also Volterra's *Stockholm Lectures* (1912).

|| See, for example, his Yale *Lectures on Cauchy's Problem* (1923). An excellent account of his method is given in Courant and Hilbert, *Methoden der math. Physik*, **2** (1938), 438–42.

and applied Green's transformation. He did not try to avoid the occurrence of divergent integrals, but developed a new method of picking out the 'finite part' of a divergent integral.

Professor Marcel Riesz† has shown how all the difficulties of Hadamard's method disappear if we introduce an additional complex parameter α. The real part of α can be chosen so large that Green's transformation is immediately applicable; the determination of the finite part of Hadamard's integral is then replaced by the process of continuing analytically an analytic function of the complex variable α. We give below an account of Riesz's method as it applies to a simple yet typical problem in the theory of cylindrical waves.

§ 7.2. Integrals of fractional order

The repeated integral of order n

$$\int_a^x \int_a^{x_1} \dots \int_a^{x_{n-1}} f(x_n)\, dx_n\, dx_{n-1} \dots dx_1$$

can be easily transformed into the simple integral

$$\frac{1}{\Gamma(n)} \int_a^x (x-t)^{n-1} f(t)\, dt. \tag{7.21}$$

The latter expression has a meaning for non-integer values of n provided that $\mathrm{Rl}\, n > 0$, and so provides what is usually called the Riemann-Liouville definition of the integral of $f(x)$ of fractional order n. Professor M. Riesz's method of solving the equation of cylindrical waves depends on an extension of this idea to functions of more than one variable.

For simplicity, we shall consider here the following problem:

To find the value at the point (x_0, y_0) and at the instant $t_0\,(> 0)$ of the solution of the equation

$$L(u) \equiv \frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = f(x,y,t), \tag{7.22}$$

given the values of u and $\partial u/\partial t$ everywhere at the instant $t = 0$.

† *Comptes rendus du congrès int. des math.* **2** (Oslo, 1936), 44–5; *Acta Math.* **81** (1949), 1–223. See also E. T. Copson, *Proc. R.S. Edin.* (A) **61** (1943), 260–72, *Proc. Edin. Math. Soc.* (2) **8** (1947), 25–36; N. E. Fremberg, *Kungl. Fysiogr. Sällsk. i Lund Forhandl.* **15** (1945), No. 27, *Meddelanden från Lunds Universitets Matematiska Seminarium*, **7** (1946); H. Malmheden, ibid. **8** (1947); L. Gårding, *Annals of Math.* **48** (1947), 785–826.

Actually it is more convenient to regard (x, y, t) as rectangular Cartesian coordinates: our problem is then to find the value of the solution of

$$L(u) = f(x, y, t)$$

at a point $P(x_0, y_0, t_0)$, given the values of u and $\partial u/\partial t$ on the plane $t = 0$. The definition of the fractional integral *appropriate to this problem*† is then as follows:

The αth integral of u is

$$I^\alpha u(x_0, y_0, t_0) = \frac{1}{2\pi\Gamma(\alpha-1)} \iiint_D u(x, y, t)\Gamma^{(\alpha-3)/2}\, dxdydt, \quad (7.23)$$

where $$\Gamma = (t-t_0)^2-(x-x_0)^2-(y-y_0)^2$$

and D is the volume bounded by the plane $t = 0$ and the cone $\Gamma = 0$.

If we assume merely the integrability of u, we can easily show that the function $I^\alpha u$ so defined is an analytic function of the complex variable α, regular when $\mathrm{Rl}\,\alpha > 1$. In the application we have in view, we shall make further restrictions on the nature of u and shall be able to continue $I^\alpha u$ analytically into the wider doman $\mathrm{Rl}\,\alpha > -1$.

The integral on the right-hand side of (7.23) is, in form, similar to the Riemann-Liouville integral (7.21), but the reason why $I^\alpha u$ is called the αth integral of u lies deeper than mere similarity of form. It can be shown that, for quite a wide class of functions u,

$$\left(\frac{\partial^2}{\partial t_0^2}-\frac{\partial^2}{\partial x_0^2}-\frac{\partial^2}{\partial y_0^2}\right)I^2u(x_0, y_0, t_0) = u(x_0, y_0, t_0),$$

and so the operator I^2 is the inverse of the differential operator‡ L_0. Moreover, we write I^α rather than I_α, since the parameter α obeys the law of indices

$$I^\alpha I^\beta u(x_0, y_0, t_0) = I^{\alpha+\beta}u(x_0, y_0, t_0).$$

Because of these two properties we can regard $I^\alpha u$ as being a generalization of the Riemann-Liouville integral of fractional order, though this aspect of the function does not concern us here.

Ex. 1. The function $F(X, Y, Z)$, where $X^2+Y^2 < Z^2$, is defined by the equation

$$F(X, Y, Z) = \iiint_V (z^2-x^2-y^2)^p\{(Z-z)^2-(X-x)^2-(Y-y)^2\}^q\, dxdydz,$$

† See § 7.6.

‡ The suffix in L_0 indicates that x, y, t in L are replaced by x_0, y_0, t_0 respectively.

where V is the volume specified by the inequalities

$$x^2+y^2 \leqslant z^2, \qquad (X-x)^2+(Y-y)^2 \leqslant (Z-z)^2, \qquad 0 \leqslant z \leqslant Z.$$

Prove that

$$F(X,Y,Z) = F(0,0,a),$$

where

$$a = +\sqrt{(Z^2-X^2-Y^2)}.$$

Hence show that

$$F(X,Y,Z) = 2\pi(Z^2-X^2-Y^2)^{p+q+3/2}\frac{\Gamma(2p+2)\Gamma(2q+2)}{\Gamma(2p+2q+5)}.$$

Ex. 2. Use Ex. 1 to prove that

$$I^{\alpha}I^{\beta}u(x_0,y_0,t_0) = I^{\alpha+\beta}u(x_0,y_0,t_0).$$

§ 7.3. A transformation of $I^{\alpha}u$

We have already remarked that, if u is integrable, the function $I^{\alpha}u$ is an analytic function of α, regular when $\mathrm{Rl}\,\alpha > 1$. But if we assume that u has continuous first partial derivatives, we can continue $I^{\alpha}u$ analytically into an analytic function, regular when $\mathrm{Rl}\,\alpha > -1$. As a first step in this analytical continuation, we need the following

Lemma. *If* $\mathrm{Rl}\,\alpha > 1$ *and u is continuous,*

$$I^{\alpha}u(x_0,y_0,t_0) = \frac{1}{2\pi\Gamma(\alpha-1)}\iiint\limits_V u(x,y,t_0-r)\frac{z^{\alpha-2}}{r}\,dxdydz, \tag{7.31}$$

where V is the hemispherical volume defined by

$$r^2 \equiv (x-x_0)^2+(y-y_0)^2+z^2 \leqslant t_0^2, \quad z \geqslant 0.$$

Under the conditions of the lemma, we can write the triple integral (7.23) defining $I^{\alpha}u$ as a repeated integral, namely

$$I^{\alpha}u(x_0,y_0,t_0) = \frac{1}{2\pi\Gamma(\alpha-1)}\int\limits_0^{t_0} dt \iint\limits_{\Sigma} u(x,y,t)\Gamma^{(\alpha-3)/2}\,dxdy,$$

where Σ is the area in which

$$(x-x_0)^2+(y-y_0)^2 \leqslant (t-t_0)^2.$$

If we make the substitution $t = t_0-r$, we obtain

$$I^{\alpha}u(x_0,y_0,t_0)$$

$$= \frac{1}{2\pi\Gamma(\alpha-1)}\int\limits_0^{t_0} dr \iint\limits_{\Sigma'} u(x,y,t_0-r)\{r^2-(x-x_0)^2-(y-y_0)^2\}^{(\alpha-3)/2}\,dxdy,$$

where Σ' is the area

$$(x-x_0)^2+(y-y_0)^2 \leqslant r^2.$$

This equation can be written in the form

$$I^\alpha u(x_0, y_0, t_0) = \frac{1}{2\pi\Gamma(\alpha-1)} \int_0^{t_0} dr \iint_{\Sigma'} u(x, y, t_0-r) z^{\alpha-3}\, dxdy,$$

where $$(x-x_0)^2+(y-y_0)^2+z^2 = r^2.$$

We now regard (x, y, z) as Cartesian coordinates in a new space. The area Σ' is then the projection of the hemisphere S specified by

$$(x-x_0)^2+(y-y_0)^2+z^2 = r^2, \quad z \geqslant 0,$$

on the plane $z = 0$, and the element of area on this hemisphere is

$$dS = \frac{r}{z} dxdy.$$

Hence

$$I^\alpha u(x_0, y_0, t_0) = \frac{1}{2\pi\Gamma(\alpha-1)} \int_0^{t_0} dr \iint_S u(x, y, t_0-r) \frac{z^{\alpha-2}}{r}\, dS$$

$$= \frac{1}{2\pi\Gamma(\alpha-1)} \iiint_V u(x, y, t_0-r) \frac{z^{\alpha-2}}{r}\, dxdydz,$$

which proves the lemma.

§ 7.4. The analytical continuation of $I^\alpha u$

We now transform the triple integral (7.31) to polar coordinates defined by

$$x = x_0+r\sin\theta\cos\phi, \qquad y = y_0+r\sin\theta\sin\phi, \qquad z = r\cos\theta$$

and, for brevity, we write

$$\bar{u} = u(x_0+r\sin\theta\cos\phi,\ y_0+r\sin\theta\sin\phi,\ t_0-r).$$

Then we have

$$I^\alpha u(x_0, y_0, t_0) = \frac{1}{2\pi\Gamma(\alpha-1)} \int_0^{t_0} dr \int_0^{2\pi} d\phi \int_0^{\frac{1}{2}\pi} \bar{u} r^{\alpha-1} \cos^{\alpha-2}\theta \sin\theta\, d\theta, \tag{7.41}$$

the representation as a repeated integral being valid since u is continuous. We also make the additional assumptions that the three partial derivatives u_x, u_y, and u_t are continuous.

If $\mathrm{Rl}\,\alpha > 1$, we may integrate by parts, to obtain

$$\int\limits_0^{\frac{1}{2}\pi} \bar{u}\cos^{\alpha-2}\theta\sin\theta\,d\theta = \frac{1}{\alpha-1}\left\{-\left[\bar{u}\cos^{\alpha-1}\theta\right]_0^{\frac{1}{2}\pi} + \int\limits_0^{\frac{1}{2}\pi}\bar{u}_\theta\cos^{\alpha-1}\theta\,d\theta\right\}$$

$$= \frac{1}{\alpha-1}u(x_0,y_0,t_0-r) + \frac{1}{\alpha-1}\int\limits_0^{\frac{1}{2}\pi}(\bar{u}_{x_0}\cos\phi+\bar{u}_{y_0}\sin\phi)r\cos^{\alpha}\theta\,d\theta.$$

Hence we have

$$I^\alpha u(x_0,y_0,t_0) = \frac{1}{\Gamma(\alpha)}\int\limits_0^{t_0} u(x_0,y_0,t_0-r)r^{\alpha-1}\,dr + $$
$$+\frac{1}{2\pi\Gamma(\alpha)}\iiint(\bar{u}_{x_0}\cos\phi+\bar{u}_{y_0}\sin\phi)r^\alpha\cos^\alpha\theta\,drd\theta d\phi.$$

A further integration by parts gives

$$I^\alpha u(x_0,y_0,t_0)$$
$$= \frac{1}{\Gamma(\alpha+1)}\left[u(x_0,y_0,t_0-r)r^\alpha\right]_0^{t_0} + \frac{1}{\Gamma(\alpha+1)}\int\limits_0^{t_0} u_{t_0}(x_0,y_0,t_0-r)r^\alpha\,dr + $$
$$+\frac{1}{2\pi\Gamma(\alpha)}\iiint(\bar{u}_{x_0}\cos\phi+\bar{u}_{y_0}\sin\phi)r^\alpha\cos^\alpha\theta\,drd\theta d\phi.$$

We have thus proved that, *if u and its first partial derivatives are continuous and if* $\mathrm{Rl}\,\alpha > 1$, *Riesz's fractional integral of order α can be written in the form*

$$I^\alpha u(x_0,y_0,t_0)$$
$$= \frac{1}{\Gamma(\alpha+1)}u(x_0,y_0,0)t_0^\alpha + \frac{1}{\Gamma(\alpha+1)}\int\limits_0^{t_0} u_{t_0}(x_0,y_0,t_0-r)r^\alpha\,dr + $$
$$+\frac{1}{2\pi\Gamma(\alpha)}\int\limits_0^{t_0}\int\limits_0^{2\pi}\int\limits_0^{\frac{1}{2}\pi}(\bar{u}_{x_0}\cos\phi+\bar{u}_{y_0}\sin\phi)r^\alpha\cos^\alpha\theta\,drd\theta d\phi. \quad (7.42)$$

The function $I^\alpha u$, as originally defined, is an analytic function of the complex variable α, regular when $\mathrm{Rl}\,\alpha > 1$. The expression on the right-hand side of (7.42) is, however, an analytic function regular when $\mathrm{Rl}\,\alpha > -1$, and so this equation provides the analytical continuation of $I^\alpha u$ across the line $\mathrm{Rl}\,\alpha = 1$. When u and its first partial derivatives are continuous, we define $I^\alpha u$ by equation (7.42) in the domain $\mathrm{Rl}\,\alpha > -1$.

In particular, we have

$$I^0 u(x_0,y_0,t_0) = u(x_0,y_0,0)+\int\limits_0^{t_0} u_{t_0}(x_0,y_0,t_0-r)\,dr$$
$$= u(x_0,y_0,t_0). \qquad (7.43)$$

The transformation I^0 is then the identical transformation.

Example. Prove that

$$\left(\frac{\partial^2}{\partial t_0^2}-\frac{\partial^2}{\partial x_0^2}-\frac{\partial^2}{\partial y_0^2}\right)I^{\alpha+2}u = I^\alpha u.$$

§ 7.5. Riesz's solution of the initial value problem

The problem is to find the value of the solution of

$$L(u) \equiv \frac{\partial^2 u}{\partial t^2}-\frac{\partial^2 u}{\partial x^2}-\frac{\partial^2 u}{\partial y^2} = f(x,y,t) \qquad (7.51)$$

at the 'point-event' (x_0,y_0,t_0), given the values of u and $\partial u/\partial t$ when $t = 0$. We suppose that $t_0 > 0$, so that the problem is an initial value problem; the case when $t_0 < 0$ can be similarly treated but is of less importance.

As in § 7.2, we write

$$\Gamma = (t-t_0)^2-(x-x_0)^2-(y-y_0)^2,$$

and we denote by D the volume in (x,y,t) space bounded by the cone $\Gamma = 0$ and the plane $t = 0$. If (l,m,n) are the direction cosines of the outward normal to S, the boundary of D, we have, by Green's transformation,

$$\iiint\limits_D \{uL(v)-vL(u)\}\,dxdydt$$
$$= \iint\limits_S \left\{n\left(u\frac{\partial v}{\partial t}-v\frac{\partial u}{\partial t}\right)-l\left(u\frac{\partial v}{\partial x}-v\frac{\partial u}{\partial x}\right)-m\left(u\frac{\partial v}{\partial y}-v\frac{\partial u}{\partial y}\right)\right\}dS, \qquad (7.52)$$

a result which is certainly valid when u and v possess continuous partial derivatives of the first and second orders. In this identity we take u to be the required solution of (7.51) and

$$v = \Gamma^{(\alpha-1)/2},$$

where $\mathrm{Rl}\,\alpha > 5$ to ensure the validity† of (7.52). Since

$$L(v) = \alpha(\alpha-1)\Gamma^{(\alpha-3)/2},$$

† The validity of (7.52) could be proved for smaller values of $\mathrm{Rl}\,\alpha$. There is no point in doing so as we shall ultimately use analytical continuation.

this gives

$$\iiint_D \{\alpha(\alpha-1)u\Gamma^{(\alpha-3)/2}-f\Gamma^{(\alpha-1)/2}\}\,dxdydt$$
$$= \iint_S \Big[(\alpha-1)u\Gamma^{(\alpha-3)/2}\{l(x-x_0)+m(y-y_0)+n(t-t_0)\}+$$
$$+\Gamma^{(\alpha-1)/2}\Big\{l\frac{\partial u}{\partial x}+m\frac{\partial u}{\partial y}-n\frac{\partial u}{\partial t}\Big\}\Big]\,dS.$$

Now S consists of two parts, a portion of the cone $\Gamma = 0$ and the area Σ specified by

$$(x-x_0)^2+(y-y_0)^2 \leqslant t_0^2, \quad t = 0;$$

and the integrand of the surface integral vanishes on the cone. Hence

$$\iiint_D \{\alpha(\alpha-1)u\Gamma^{(\alpha-3)/2}-f\Gamma^{(\alpha-1)/2}\}\,dxdydt$$
$$= \iint_\Sigma \Big\{\Gamma^{(\alpha-1)/2}\frac{\partial u}{\partial t}-(\alpha-1)(t-t_0)\Gamma^{(\alpha-3)/2}u\Big\}_{t=0}\,dxdy,$$

or, in Riesz's fractional-integral notation,

$$I^\alpha u-I^{\alpha+2}f = \frac{1}{2\pi\Gamma(\alpha+1)}\iint_\Sigma \Big\{\Gamma^{(\alpha-1)/2}\frac{\partial u}{\partial t}+(\alpha-1)t_0\Gamma^{(\alpha-3)/2}u\Big\}_{t=0}\,dxdy. \tag{7.53}$$

Position in Σ can be specified by polar coordinates (ρ,ϕ), where

$$x = x_0+\rho\cos\phi, \qquad y = y_0+\rho\sin\phi \qquad (0 \leqslant \rho \leqslant t_0;\ 0 \leqslant \phi \leqslant 2\pi).$$

The given initial values of u and $\partial u/\partial t$ are then functions of ρ and ϕ,

$$u = U(\rho,\phi), \qquad \frac{\partial u}{\partial t} = V(\rho,\phi) \quad (t = 0)$$

say. With this notation, (7.53) becomes

$$I^\alpha u-I^{\alpha+2}f$$
$$= \frac{1}{2\pi\Gamma(\alpha+1)}\iint_\Sigma \{(t_0^2-\rho^2)^{(\alpha-1)/2}V+(\alpha-1)t_0(t_0^2-\rho^2)^{(\alpha-3)/2}U\}\rho\,d\rho d\phi$$

or

$$I^\alpha u = I^{\alpha+2}f+\frac{1}{2\pi\Gamma(\alpha+1)}\iint_\Sigma (t_0^2-\rho^2)^{(\alpha-1)/2}V\rho\,d\rho d\phi+$$
$$+\frac{1}{2\pi\Gamma(\alpha+1)}\frac{\partial}{\partial t_0}\iint_\Sigma (t_0^2-\rho^2)^{(\alpha-1)/2}U\rho\,d\rho d\phi. \tag{7.54}$$

By § 7.4, $I^\alpha u$ is an analytic function of α, regular when $\mathrm{Rl}\,\alpha > -1$, since, by hypothesis, u and its first partial derivatives are continuous. The three terms on the right-hand side of (7.54) are evidently also regular when $\mathrm{Rl}\,\alpha > -1$. Hence, by analytical continuation, the equation (7.54) which was proved under the assumption $\mathrm{Rl}\,\alpha > 5$ is valid when $\mathrm{Rl}\,\alpha > -1$. In particular, when $\alpha = 0$, we have

$$I^0 u = I^2 f + \frac{1}{2\pi}\left[\iint\limits_\Sigma (t_0^2-\rho^2)^{-\frac{1}{2}} V\,dxdy + \frac{\partial}{\partial t_0}\iint\limits_\Sigma (t_0^2-\rho^2)^{-\frac{1}{2}} U\,dxdy\right],$$

and so, by (7.43),

$$u(x_0,y_0,t_0) = \frac{1}{2\pi}\iiint\limits_D f(x,y,t)\{(t-t_0)^2-(x-x_0)^2-(y-y_0)^2\}^{-\frac{1}{2}}\,dxdydt +$$
$$+\frac{1}{2\pi}\iint\limits_\Sigma \left(\frac{\partial u}{\partial t}\right)_{t=0}\{t_0^2-(x-x_0)^2-(y-y_0)^2\}^{-\frac{1}{2}}\,dxdy +$$
$$+\frac{1}{2\pi}\frac{\partial}{\partial t_0}\iint\limits_\Sigma (u)_{t=0}\{t_0^2-(x-x_0)^2-(y-y_0)^2\}^{-\frac{1}{2}}\,dxdy. \quad (7.55)$$

This is *Volterra's formula* expressing the value of a solution of $L(u) = f$ in terms of the values of u and $\partial u/\partial t$ when $t = 0$. When $f \equiv 0$, (7.55) reduces to a special case of the general theorem enunciated in § 6.4, viz. the case when σ is a plane.

We cannot simplify the last term on the right-hand side of (7.55) by differentiating under the sign of integration, since that would give rise to a divergent integral; but by a slightly more elaborate procedure we can express (7.55) in a form which does not involve the differential coefficient of a double integral. Let us write

$$\bar{U}(\phi) = U(t_0,\phi),$$

so that $\bar{U}$ is the value of u on the boundary of Σ. Then equation (7.53) becomes

$$I^\alpha u = I^{\alpha+2} f + \frac{1}{2\pi\Gamma(\alpha+1)}\left[\iint\limits_\Sigma (t_0^2-\rho^2)^{(\alpha-1)/2} V\rho\,d\rho d\phi +\right.$$
$$\left.+\iint\limits_\Sigma (\alpha-1)t_0(t_0^2-\rho^2)^{(\alpha-3)/2}(U-\bar{U})\rho\,d\rho d\phi + t_0^\alpha\int\limits_0^{2\pi}\bar{U}\,d\phi\right]. \quad (7.56)$$

Now suppose that U possesses a continuous partial derivative with respect to ρ; then

$$U-\bar{U} = -(t_0-\rho)U_\tau(\tau,\phi),$$

where $\rho \leqslant \tau \leqslant t_0$, and so the term involving $U-\overline{U}$ in (7.56) is an analytic function of α, regular when $\mathrm{Rl}\,\alpha > -1$. Putting $\alpha = 0$, we obtain

$$u(x_0,y_0,t_0) = \frac{1}{2\pi}\iiint\limits_D f(x,y,t)\{(t-t_0)^2-(x-x_0)^2-(y-y_0)^2\}^{-\frac{1}{2}}\,dxdydt +$$

$$+\frac{1}{2\pi}\iint\limits_\Sigma \left(\frac{\partial u}{\partial t}\right)_{t=0}\{t_0^2-(x-x_0)^2-(y-y_0)^2\}^{-\frac{1}{2}}\,dxdy -$$

$$-\frac{1}{2\pi}\iint\limits_\Sigma \{(u)_{t=0}-\overline{U}\}t_0\{t_0^2-(x-x_0)^2-(y-y_0)^2\}^{-\frac{3}{2}}\,dxdy +$$

$$+\frac{1}{2\pi}\int\limits_0^{2\pi} \overline{U}\,d\phi.$$

This transformation of Volterra's formula is a particular case of a general formula of Hadamard.†

§ 7.6. The advantages of Riesz's method

The first and most obvious advantage of Riesz's method is that it avoids the awkward limiting processes which appear in the theories of Volterra and Hadamard, by introducing an arbitrary complex parameter α and using analytical continuation.

The second advantage lies much deeper. We remarked in § 6.1 that there is a striking difference between the solution of the equation of wave-motions in two and in three dimensions; for in three dimensions wave-propagation is clean-cut, whereas in two dimensions it is diffused. More generally, diffusion always occurs in space of an even number of dimensions, but may or may not occur in space of an odd number of dimensions. This difference depends ultimately, as Hadamard has shown, on the different characters of the elementary solution of the equation of wave-motions in spaces with an even or odd number of dimensions. Hadamard's theory of the finite part of a divergent integral can be applied in any space with an even number of dimensions. But when the number of spatial dimensions is odd, Hadamard's theory does not apply; we have to use either Hadamard's method of descent or else the idea of the logarithmic part‡ of a divergent integral.

† See Hadamard's Yale *Lectures on Cauchy's Problem* (1923), 208, equation (60).

‡ See, for example, Courant-Hilbert, *Methoden der math. Physik*, **2** (1938), 443–8, for a discussion of the three-dimensional case.

Professor Marcel Riesz's method, on the other hand, does not involve any distinction between spaces of an even and odd number of dimensions. The solution is, in fact, independent of the number of spatial dimensions.

Let us suppose we wish to find the solution of

$$L(u) \equiv \frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x_1^2} - \frac{\partial^2 u}{\partial x_2^2} - \dots - \frac{\partial^2 u}{\partial x_{m-1}^2} = f(x_1, x_2, \dots, x_{m-1}, t)$$

at $(x_1^0, x_2^0, \dots, x_{m-1}^0, t^0)$, given the values of u and its 'conormal' derivative on a certain hypersurface S in $(x_1, x_2, \dots, x_{m-1}, t)$ space. The function $I^\alpha u$ is now defined to be

$$I^\alpha u(x_1^0, x_2^0, \dots, t^0) = \frac{1}{H_m(\alpha)} \iiint\limits_D u(x_1, x_2, \dots, t) \Gamma^{(\alpha-m)/2} \, dx_1 \, dx_2 \dots dt,$$

where $$\Gamma = (t-t^0)^2 - \sum_{r=1}^{m-1} (x_r - x_r^0)^2$$

and D is the volume bounded by S and the hypercone $\Gamma = 0$. The constant $H_m(\alpha)$ is equal to $\pi^{(m-2)/2} 2^{\alpha-1} \Gamma(\frac{1}{2}\alpha) \Gamma(1+\frac{1}{2}\alpha-\frac{1}{2}m)$.

The function $I^\alpha u$ so defined possesses all the properties of the simpler function (7.23). It is regular when $\mathrm{Rl}\,\alpha > m-2$, and can be continued analytically into $\mathrm{Rl}\,\alpha > -1$ when u is sufficiently well behaved. Riesz then proves the identity

$$I^\alpha u = I^{\alpha+2} f - \frac{1}{H_m(\alpha+2)} \int\limits_\Sigma \left(\frac{\partial u}{\partial \nu} \Gamma^{(\alpha+2-m)/2} - u \frac{\partial \Gamma^{(\alpha+2-m)/2}}{\partial \nu} \right) dS,$$

where Σ is the part cut out of S by the hypercone and ν denotes the outward 'conormal'. By analytical continuation, this identity holds when $\alpha = 0$ and provides Riesz's 'solution invariantive' of the generalized equation of wave-motions.

Finally, we wish to draw attention to the work of Dr. M. Mathisson, who discovered a new method of solving the initial value problem for equations of normal hyperbolic type,† and gave‡ a criterion for the existence or non-existence of diffusion.

Ex. 1. When $\mathrm{Rl}\,\alpha > 1$, the function $J^\alpha u$ is defined by

$$J^\alpha u(x_0, y_0, t_0) = \frac{1}{2\pi\Gamma(\alpha-1)} \iiint\limits_D u(x,y,t) \cosh(k\sqrt{\Gamma}) \Gamma^{(\alpha-3)/2} \, dx dy dt$$

in the notation of § 7.2. Show that, if u and its first partial derivatives are

† *Math. Annalen*, **107** (1933), 400–19; *Comptes rendus*, **208** (1939), 1776–8.

‡ *Acta Math.* **71** (1939), 249–82. See also Hadamard, *Annals of Math.* **43** (1942), 510–22.

continuous, $J^\alpha u$ can be continued into an analytic function regular when $\mathrm{Rl}\,\alpha > -1$, and that, when this continuation is carried out,

$$J^0 u(x_0, y_0, t_0) = u(x_0, y_0, t_0).$$

[Express $J^\alpha u$ as a power series in k.]

Ex. 2. Prove that
$$u = \frac{\cosh k\sqrt{\Gamma}}{\sqrt{\Gamma}}$$
is a solution of
$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + k^2 u = \frac{\partial^2 u}{\partial t^2}.$$

Ex. 3. Prove that the solution of the initial value problem for the equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} - k^2 u = f(x, y, t)$$

is

$$u(x_0, y_0, t_0) = \frac{1}{2\pi} \iiint\limits_D f(x, y, t) \frac{\cosh k\sqrt{\Gamma}}{\sqrt{\Gamma}}\, dx dy dt + $$
$$+ \frac{1}{2\pi} \iint\limits_\Sigma \left(\frac{\partial u}{\partial t}\right)_{t=0} \frac{\cosh k\sqrt{\Gamma_0}}{\sqrt{\Gamma_0}}\, dx dy +$$
$$+ \frac{1}{2\pi} \frac{\partial}{\partial t_0} \iint\limits_\Sigma (u)_{t=0} \frac{\cosh k\sqrt{\Gamma_0}}{\sqrt{\Gamma_0}}\, dx dy,$$

where Γ_0 is the value of Γ when $t = 0$.

II

HUYGENS' PRINCIPLE AND THE DIFFRACTION OF LIGHT

§ 1. Kirchhoff's theory of diffraction

§ 1.1. Kirchhoff's application of Helmholtz's formula

ACCORDING to geometrical optics, rays of light are straight lines. With this assumption, light from a point-source incident on a non-reflecting opaque screen gives a sharply defined shadow, which it is convenient to call the geometrical shadow. Actually it is observed experimentally that the light is propagated up to the screen as if the screen were absent, but that, beyond the screen, light enters the geometrical shadow. This phenomenon, which violates the laws of geometrical optics, is known as the *diffraction of light*.

It was Fresnel who first discovered the real cause of diffraction, namely the mutual interference of the secondary waves emitted by those parts of the original wave-front which were not obstructed by the diffraction screen. But, in order to obtain satisfactory results, Fresnel had to make somewhat arbitrary assumptions on the nature of the secondary waves.† Most of the difficulties of Fresnel's theory were overcome by Kirchhoff, who used Helmholtz's formulation of Huygens' principle for monochromatic phenomena.‡ We give here a brief critical account of Kirchhoff's work,|| and refer the reader to the standard works on physical optics for a detailed discussion of special diffraction problems.

Before we can apply Helmholtz's formula to monochromatic diffraction problems, we have to overcome two rather serious difficulties: we have to express mathematically what is meant by a point-source of light, and we have to give mathematical conditions for a screen to be opaque and non-reflecting.

In the first place, the notion of a point-source of monochromatic light is a somewhat idealized one. On the electromagnetic theory of light it would be reasonable to regard it as a Hertzian electric oscillator or as a Fitzgerald magnetic oscillator.†† Kirchhoff was, however, working with the older theory which regarded the aether

† Ch. I, § 4.1. ‡ Ch. I, § 4.2.

|| For Kirchhoff's own account of his theory, see his *Vorlesungen über math. Physik*, **2** (*Optik*), (Leipzig, 1891).

†† See Whittaker, *History of the Theories of Aether and Electricity*, 345–7, 360–2.

as an elastic solid, and he assumed that the displacement vector in the aether is of the form†

$$\frac{\mathbf{A}}{r} e^{ik(r-ct)}$$

when monochromatic light is emitted by a point-source at $r = 0$, $\mathbf{A}$ being a constant complex vector. The intensity of the light is then measured by $\mathbf{A} \cdot \bar{\mathbf{A}}/r^2$, where $\bar{\mathbf{A}}$ is the complex vector conjugate to $\mathbf{A}$. He then applied Helmholtz's formula to each component of the displacement vector separately.

Following Kirchhoff, we shall assume in this chapter that monochromatic light from a point-source L can be represented by one or more wave-functions of the form

$$u_0 = v_0 e^{-ikct} = \frac{e^{ik(r-ct)}}{r}$$

when there is no diffracting screen, r being the distance from L. We have to find the corresponding wave-function $u = ve^{-ikct}$ when diffraction occurs.

Let us suppose, then, that monochromatic light from a point-source L defined in this way is diffracted by an opaque non-reflecting body whose boundary is the closed surface S. Let S_0 be a small sphere with centre L; then if P is a point exterior to S and S_0, Helmholtz's formula gives

$$4\pi u(P) = \iint_S + \iint_{S_0} \left\{ \frac{e^{ik(r_1-ct)}}{r_1} \frac{\partial v}{\partial \nu} - v \frac{\partial}{\partial \nu}\left(\frac{e^{ik(r_1-ct)}}{r_1} \right) \right\} dS,$$

where r_1 is the distance from P to a typical point of a surface of integration and $\partial/\partial\nu$ denotes differentiation along the inward normal. This formula would give the value of u everywhere if we knew the boundary values of v and $\partial v/\partial\nu$ on S and S_0, and it is in the determination of these boundary values that our second difficulty arises.

Actually the surface S_0 causes no trouble; for it is evident that, when the radius of S_0 is small, the screen has no effect at points of S_0 and so $v = v_0$, $\partial v/\partial\nu = \partial v_0/\partial\nu$ there. There is, however, a serious difficulty in that we know very little about what happens on S; roughly speaking, all we can say is that the part of S invisible from

† Actually Kirchhoff used the real part of this complex vector. The use of the complex vector simplifies the analysis.

L is very feebly illuminated. Kirchhoff made the following assumptions, which, though reasonable, are nevertheless quite arbitrary:

(i) $v = v_0$, $\partial v/\partial \nu = \partial v_0/\partial \nu$ on S_1, the part of S visible from L;

(ii) $v = 0$, $\partial v/\partial \nu = 0$ on S_2, the part of S invisible from L.

In words, he assumed that there is no change in the light on S_1 and that S_2 is quite dark. In this way, Kirchhoff obtained the formula

$$4\pi u(P) = \iint_{S_0} + \iint_{S_1} \left\{ \frac{e^{ik(r_1-ct)}}{r_1} \frac{\partial v_0}{\partial \nu} - v_0 \frac{\partial}{\partial \nu}\left(\frac{e^{ik(r_1-ct)}}{r_1} \right) \right\} dS. \quad (1.11)$$

(See Fig. 6.)

Before discussing the validity of Kirchhoff's assumptions we introduce a modification of his formula (1.11) which is often more

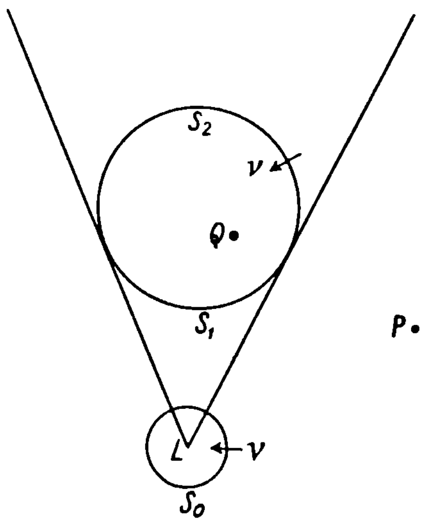


FIG. 6

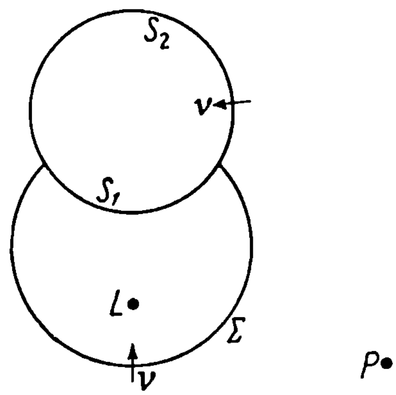


FIG. 7

convenient to apply. Let us suppose for simplicity that the tangent cone from L to S has contact along a single connected curve Γ, which divides S_1 from S_2. Γ may be either a simple closed curve or an arc which extends to infinity at both ends; for definiteness we consider only the former case, and leave the modifications when Γ extends to infinity to the reader.

We now construct an unclosed surface Σ with Γ as rim with the following properties:

(i) Σ and S_1 form a closed surface Σ';

(ii) L lies inside Σ', P outside.

(See Fig. 7.) The surface S_0 can be continuously deformed into Σ' without crossing L or P, and this deformation does not alter the value of the integral over S_0. Thus in formula (1.11) we can replace S_0 by Σ'. Taking into account the opposite directions of the normal ν to S and Σ' on their common part, we find that

$$4\pi u(P) = \iint\limits_{\Sigma} \left\{\frac{e^{ik(r_1-ct)}}{r_1}\frac{\partial v_0}{\partial \nu} - v_0\frac{\partial}{\partial \nu}\left(\frac{e^{ik(r_1-ct)}}{r_1}\right)\right\} dS, \qquad (1.12)$$

where ν denotes the normal from the 'dark' to the 'illuminated' side† of Σ.

In the simplest type of diffraction problem, light passes through a hole in an opaque non-reflecting plane screen. The surface Σ can then be taken to be a plane area bridging the gap in the screen. The extension to more complicated plane screens can be left to the reader.

Example. Prove that Kirchhoff's formula (1.11) can also be written in the form

$$u(P) = u_0(P) + \frac{1}{4\pi}\iint\limits_{S_1} \left\{\frac{e^{ik(r_1-ct)}}{r_1}\frac{\partial v_0}{\partial \nu} - v_0\frac{\partial}{\partial \nu}\left(\frac{e^{ik(r_1-ct)}}{r_1}\right)\right\} dS,$$

where ν is the normal from the illuminated to the dark side of S_1.

§ 1.2. Criticisms of Kirchhoff's theory

All we know from experiment is that the part of the screen invisible from the source is very slightly illuminated; Kirchhoff assumes it to be perfectly dark. This assumption makes the boundary values of v and $\partial v/\partial \nu$ on the screen discontinuous across the curve Γ, which is certainly not the case physically. It might further be objected that Helmholtz's formula was proved on the basis of continuous boundary values, so that it is not evident without further investigation that the function $u(P)$ does satisfy Kirchhoff's boundary conditions. In point of fact, it does not. For Poincaré‡ has shown that *Kirchhoff's boundary values of v are incompatible with his boundary values of*|| $\partial v/\partial \nu$.

To prove this, let Q be a point within S, and let r_2 denote distance measured from Q. Then, since v_0 and e^{ikr_2}/r_2 have no singularities outside the closed surface $\Sigma + S_2$,

$$\iint\limits_{\Sigma} + \iint\limits_{S_2} \left\{\frac{e^{ikr_2}}{r_2}\frac{\partial v_0}{\partial \nu} - v_0\frac{\partial}{\partial \nu}\left(\frac{e^{ikr_2}}{r_2}\right)\right\} dS = 0. \qquad (1.21)$$

† This is the most convenient way of describing the sense of ν. Of course, as Σ is not a screen, both sides of Σ are, in fact, illuminated.

‡ *Théorie mathématique de la lumière*, **2** (Paris, 1892), 187–8.

|| We recall that a knowledge of the boundary values of v suffices to determine those of $\partial v/\partial \nu$. (Ch. I, § 4.3.)

Similarly, if ve^{-ikct} is the wave-function satisfying Kirchhoff's boundary conditions,

$$\iint_{\Sigma} + \iint_{S_1} \left\{\frac{e^{ikr_2}}{r_2}\frac{\partial v}{\partial \nu} - v\frac{\partial}{\partial \nu}\left(\frac{e^{ikr_2}}{r_2}\right)\right\} dS = 0.$$

But by the boundary conditions, the latter equation reduces to

$$\iint_{\Sigma} \left\{\frac{e^{ikr_2}}{r_2}\frac{\partial v_0}{\partial \nu} - v_0\frac{\partial}{\partial \nu}\left(\frac{e^{ikr_2}}{r_2}\right)\right\} dS = 0.$$

Hence (1.21) becomes

$$\iint_{S_1} \left\{\frac{e^{ikr_2}}{r_2}\frac{\partial v_0}{\partial \nu} - v_0\frac{\partial}{\partial \nu}\left(\frac{e^{ikr_2}}{r_2}\right)\right\} dS = 0. \tag{1.22}$$

This is, however, impossible. For we can construct a diffraction problem in which S_2 plays the part of Σ in the analysis of the preceding section and, by (1.12), the right-hand side of (1.22) should be, not zero, but the non-zero value of ue^{ikct} at Q in this new problem.

In spite of this difficulty, Kirchhoff's formula does give results which agree very well with experiment. Some authors explain this by saying that the formula gives the first step in an accurate solution by successive approximations. The next step would be to obtain a second approximation by substituting in Helmholtz's formula the boundary values given by Kirchhoff's first approximation; and so on indefinitely. If this iterative process were a rapidly convergent one, the success of the method would be explained; but the iteration has never been carried beyond the first stage and the convergence never discussed.

Other authors, notably Kottler, regard Kirchhoff's formula as an accurate solution, not of a boundary value problem, but of a 'saltus problem'. A brief account of this work is given later.

§ 1.3. The Fresnel-Kirchhoff approximate diffraction formula

It turns out to be difficult to apply the accurate Kirchhoff formula (1.12) to special diffraction problems. However, in most problems, the wave-length λ is very small compared with the other distances involved, and we can then use a simple approximation which we now obtain.

If L is a point-source of light characterized by the wave-function $e^{ik(r-ct)}/r$, Kirchhoff's diffraction formula can be written in the form

$$4\pi u(P) = \iint\limits_{\Sigma} \left\{ \frac{e^{ikr_0}}{r_0} \frac{\partial}{\partial n}\left(\frac{e^{ik(r_1-ct)}}{r_1}\right) - \frac{e^{ik(r_1-ct)}}{r_1} \frac{\partial}{\partial n}\left(\frac{e^{ikr_0}}{r_0}\right) \right\} dS, \quad (1.31)$$

where r_0 is the distance from L to a typical point Q of Σ and $\partial/\partial n$ denotes differentiation along the normal $\mathbf{n}$ from the illuminated to the dark side of Σ. (See Fig. 8.) Since

$$\frac{\partial r_0}{\partial n} = \cos\theta_0, \qquad \frac{\partial r_1}{\partial n} = \cos\theta_1,$$

where θ_0 and θ_1 are the angles the normal $\mathbf{n}$ makes with $\overrightarrow{LQ}$ and $\overrightarrow{PQ}$ respectively, the equation (1.31) may be written as

$$4\pi u(P) = \iint\limits_{\Sigma} e^{ik(r_0+r_1-ct)} \left\{ \frac{ik}{r_0 r_1}(\cos\theta_1 - \cos\theta_0) - \left(\frac{\cos\theta_1}{r_0 r_1^2} - \frac{\cos\theta_0}{r_0^2 r_1}\right) \right\} dS. \quad (1.32)$$

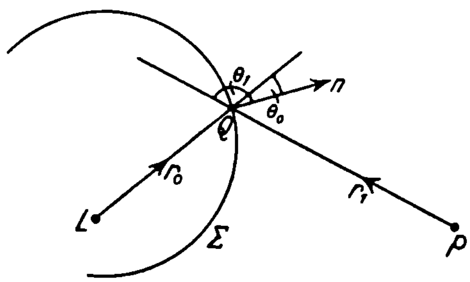


FIG. 8

Now, by hypothesis, the wave-length $2\pi/k$ is small compared with r_0 and r_1, and so we may neglect $1/r_0$ and $1/r_1$ in comparison with k in (1.32). This gives

$$u(P) = \frac{1}{4\pi} \iint\limits_{\Sigma} \frac{e^{ik(r_0+r_1-ct)}}{r_0 r_1} ik(\cos\theta_1 - \cos\theta_0)\, dS, \quad (1.33)$$

a result more usually written in the form

$$u(P) = \frac{1}{2\lambda} \iint\limits_{\Sigma} e^{ik(r_0+r_1-ct)-\frac{1}{2}\pi i}(\cos\theta_0 - \cos\theta_1)\frac{dS}{r_0 r_1}. \quad (1.34)$$

This approximation is usually called the Fresnel-Kirchhoff diffraction formula. (Cf. Ch. I, § 4.6.)

As Larmor† has remarked, the formula (1.34) 'puts in evidence the factor of attenuation $(r_0 r_1)^{-1}$ and the phase depending on the path r_0+r_1'. Moreover, it is a consequence of Kelvin's Principle of

† *Proc. London Math. Soc.* (2), **19** (1919), 169–80 (174).

Stationary Phase that, when k is large, the important part of the integral (1.34) arises from the part of the range of integration near which the phase of $e^{ik(r_0+r_1)}$ is stationary, and so the Fresnel-Kirchhoff formula 'shows that it is only the elements δS, that lie near the line joining the source to any point, which produce the disturbance at that point'.

§ 2. Maggi's transformation

§ 2.1. Helmholtz's integral over an unclosed surface

We have seen that when Kirchhoff applied Helmholtz's formula to diffraction problems, he was led to an equation (1.12), which expresses a component of the light-vector as an integral of Helmholtz's type, extended over a certain unclosed surface Σ. We now show that such an integral can be expressed as a line integral over Γ, the rim of Σ, and carry out the transformation in detail in the two most important cases which arise. The resulting expression as a line integral provides the most suitable way of applying Kirchhoff's ideas accurately.

We recall that the value taken at a point P on one side of a closed surface S by a monochromatic wave-function u of period $2\pi/(kc)$ due to sources on the other side of S is given by

$$4\pi u(P) = \iint\limits_S \left\{\frac{e^{ikr_1}}{r_1}\frac{\partial u}{\partial n} - u\frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right)\right\} dS,$$

where $\partial/\partial n$ is differentiation along the normal out of the region containing P. But if P and the sources are all on the same side of S, the value of the integral is zero. It follows that, if we deform S continuously, the value of the integral is unaltered, provided that, in the deformation, S does not cross P or any of the sources.

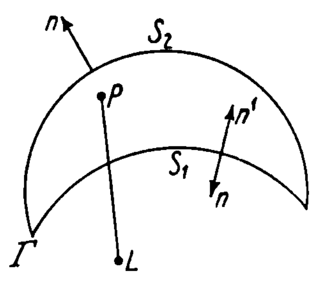


FIG. 9

Now consider the case when there is only one source at the point L. Take any simple closed curve Γ in space and through it draw two unclosed surfaces S_1 and S_2, each having Γ as rim; we suppose that S_1 intersects the segment LP but that S_2 does not. Then, by the formula of Helmholtz,

$$4\pi u(P) = \iint\limits_{S_1+S_2} \left\{\frac{e^{ikr_1}}{r_1}\frac{\partial u}{\partial n} - u\frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right)\right\} dS.$$

Now keep S_1 fixed and continuously deform S_2 in such a way that it never crosses L or P. Then since the integral over S_1+S_2 is unaltered, the value of the integral over S_2 does not change and so it depends only on the form of the rim Γ and not on the actual shape of S_2. This means that it must be possible to express the integral over S_2 as a line integral round Γ, say

$$\iint\limits_{S_2} \left\{\frac{e^{ikr_1}}{r_1}\frac{\partial u}{\partial n} - u\frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right)\right\} dS = \int\limits_{\Gamma} (a_x\,dx + a_y\,dy + a_z\,dz).$$

If we denote by $\mathbf{n}'$ the unit vector normal to S_1 drawn in the opposite sense to $\mathbf{n}$, we have, by Helmholtz's formula,

$$\iint\limits_{S_1} \left\{\frac{e^{ikr_1}}{r_1}\frac{\partial u}{\partial n'} - u\frac{\partial}{\partial n'}\left(\frac{e^{ikr_1}}{r_1}\right)\right\} dS$$
$$= -\iint\limits_{S_1+S_2} + \iint\limits_{S_2} \left\{\frac{e^{ikr_1}}{r_1}\frac{\partial u}{\partial n} - u\frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right)\right\} dS$$
$$= -4\pi u(P) + \int\limits_{\Gamma} (a_x\,dx + a_y\,dy + a_z\,dz).$$

To sum up: *if u is a monochromatic wave-function of period $2\pi/(kc)$ due to a source at L, and if S is an unclosed surface with Γ as rim, then*

$$\iint\limits_{S} \left\{\frac{e^{ikr_1}}{r_1}\frac{\partial u}{\partial n} - u\frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right)\right\} dS = -4\pi\epsilon u(P) + \int\limits_{\Gamma} (a_x\,dx + a_y\,dy + a_z\,dz), \tag{2.11}$$

where $\epsilon = 0$ if the segment LP does not cut S, $\epsilon = 1$ if it cuts S once. The normal $\mathbf{n}$ is supposed to be drawn from the 'illuminated' to the 'dark' side of S.

In the diffraction problem which will concern us most, we have $u = e^{ik(r_0-ct)}/r_0$ where r_0 is distance from the source L.

We shall find an explicit formula, due to Maggi,† for the vector $\mathbf{a}$ in this case. We may evidently suppose that the segment LP does not cut S; then, dropping the time factor e^{-ikct}, the formula (2.11) can be written in vector notation as

$$\iint\limits_{S} \mathbf{B}\cdot\mathbf{n}\,dS = \int\limits_{\Gamma} \mathbf{a}\cdot\mathbf{t}\,ds, \tag{2.12}$$

† *Annali di Mat.* (2) **16** (1888), 21–48. See also Kottler, *Annalen der Phys.* **70** (1923), 405–56; Rubinowicz, ibid. **53** (1917), 257–78.

where **t** is the unit vector along the tangent to Γ and

$$\mathbf{B} = \frac{e^{ikr_1}}{r_1}\operatorname{grad}\left(\frac{e^{ikr_0}}{r_0}\right) - \frac{e^{ikr_0}}{r_0}\operatorname{grad}\left(\frac{e^{ikr_1}}{r_1}\right), \tag{2.13}$$

$$\mathbf{a} = a_x\mathbf{i}+a_y\mathbf{j}+a_z\mathbf{k}. \tag{2.14}$$

If we adopt the usual convention that the directions of **t** and **n** are connected by the right-hand screw law, an application of Stokes's theorem gives

$$\iint\limits_S \mathbf{B}\cdot\mathbf{n}\,dS = \iint\limits_S \operatorname{curl}\mathbf{a}\cdot\mathbf{n}\,dS,$$

and so **a** is a solution of

$$\operatorname{curl}\mathbf{a} = \mathbf{B}. \tag{2.15}$$

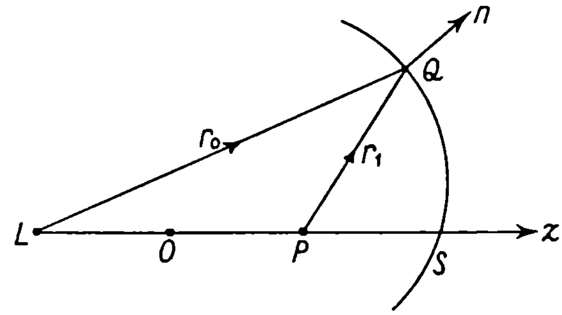


FIG. 10

Suppose that the distance between L and P is $2f$. Then we can choose our axes so that L is $(0, 0, -f)$, P $(0, 0, f)$. As our problem is a 'bipolar' problem, it is most convenient to introduce spheroidal coordinates (λ, μ, ν) defined by the equations

$$x = f\sinh\lambda\sin\mu\cos\nu, \qquad y = f\sinh\lambda\sin\mu\sin\nu, \qquad z = f\cosh\lambda\cos\mu.$$

The surfaces $\lambda =$ constant are confocal ellipsoids of revolution with L and P as foci, $\mu =$ constant confocal hyperboloids of revolution with the same foci, and the surfaces $\nu =$ constant are planes through the line LP. We make the orthogonal curvilinear coordinates (λ, μ, ν) quite definite by the conventions

$$\lambda \geqslant 0, \qquad 0 \leqslant \mu \leqslant \pi, \qquad 0 \leqslant \nu < 2\pi.$$

In this system of coordinates we have

$$r_0 = LQ = f(\cosh\lambda+\cos\mu), \qquad r_1 = PQ = f(\cosh\lambda-\cos\mu).$$

If we denote by **r** the position vector $\overrightarrow{OQ}$, $\partial\mathbf{r}/\partial\lambda$ is a tangent vector to the curve $\mu =$ constant, $\nu =$ constant. Denoting its magnitude by h_1, we have

$$h_1 = f\sqrt{(\cosh^2\lambda-\cos^2\mu)};$$

and so we may write $$\frac{\partial \mathbf{r}}{\partial \lambda} = h_1 \mathbf{i}_1,$$

where $\mathbf{i}_1$ is a unit vector tangent to μ = constant, ν = constant.

Similarly we define unit vectors $\mathbf{i}_2$ and $\mathbf{i}_3$ by

$$\frac{\partial \mathbf{r}}{\partial \mu} = h_2 \mathbf{i}_2, \qquad \frac{\partial \mathbf{r}}{\partial \nu} = h_3 \mathbf{i}_3,$$

where $h_2 = f\sqrt{(\cosh^2\lambda - \cos^2\mu)}, \qquad h_3 = f \sinh\lambda \sin\mu.$

The three vectors $\mathbf{i}_1$, $\mathbf{i}_2$, $\mathbf{i}_3$ are mutually orthogonal vectors, though they change their directions in space as the point at which they are drawn varies. Any vector $\mathbf{A}$ at a point Q can be resolved into components A_1, A_2, A_3 in the directions of the vectors $\mathbf{i}_1$, $\mathbf{i}_2$, and $\mathbf{i}_3$ at Q; these components are defined by

$$\mathbf{A} = A_1 \mathbf{i}_1 + A_2 \mathbf{i}_2 + A_3 \mathbf{i}_3.$$

The differential of $\mathbf{r}$ is given by

$$d\mathbf{r} = \frac{\partial \mathbf{r}}{\partial \lambda} d\lambda + \frac{\partial \mathbf{r}}{\partial \mu} d\mu + \frac{\partial \mathbf{r}}{\partial \nu} d\nu = h_1 \mathbf{i}_1 d\lambda + h_2 \mathbf{i}_2 d\mu + h_3 \mathbf{i}_3 d\nu.$$

Since the magnitude of $d\mathbf{r}$ is ds, the element of length in space, it follows that $$ds^2 = h_1^2 d\lambda^2 + h_2^2 d\mu^2 + h_3^2 d\nu^2.$$

By well-known formulae, we now have

$$B_1 = \frac{1}{h_1}\left\{\frac{e^{ikr_1}}{r_1}\frac{\partial}{\partial \lambda}\left(\frac{e^{ikr_0}}{r_0}\right) - \frac{e^{ikr_0}}{r_0}\frac{\partial}{\partial \lambda}\left(\frac{e^{ikr_1}}{r_1}\right)\right\},$$

$$(\operatorname{curl} \mathbf{a})_1 = \frac{1}{h_2 h_3}\left\{\frac{\partial}{\partial \mu}(a_3 h_3) - \frac{\partial}{\partial \nu}(a_2 h_2)\right\},$$

and similarly for the other components. The equations (2.15) to determine a_1, a_2, a_3 are then

$$\frac{\partial}{\partial \mu}(a_3 h_3) - \frac{\partial}{\partial \nu}(a_2 h_2) = \frac{h_2 h_3}{h_1}\left\{\frac{e^{ikr_1}}{r_1}\frac{\partial}{\partial \lambda}\left(\frac{e^{ikr_0}}{r_0}\right) - \frac{e^{ikr_0}}{r_0}\frac{\partial}{\partial \lambda}\left(\frac{e^{ikr_1}}{r_1}\right)\right\}, \tag{2.16}$$

$$\frac{\partial}{\partial \nu}(a_1 h_1) - \frac{\partial}{\partial \lambda}(a_3 h_3) = \frac{h_3 h_1}{h_2}\left\{\frac{e^{ikr_1}}{r_1}\frac{\partial}{\partial \mu}\left(\frac{e^{ikr_0}}{r_0}\right) - \frac{e^{ikr_0}}{r_0}\frac{\partial}{\partial \mu}\left(\frac{e^{ikr_1}}{r_1}\right)\right\}, \tag{2.17}$$

$$\frac{\partial}{\partial \lambda}(a_2 h_2) - \frac{\partial}{\partial \mu}(a_1 h_1) = 0, \tag{2.18}$$

since r_0 and r_1 are independent of ν. (2.18) gives

$$a_1 h_1 = \frac{\partial \phi}{\partial \lambda}, \qquad a_2 h_2 = \frac{\partial \phi}{\partial \mu},$$

where ϕ is an arbitrary function of λ, μ, ν. We can, however, take ϕ to be identically zero; for if not, **a** would involve an added term $\operatorname{grad}\phi$, which would disappear on integration round Γ. Hence $a_1 = a_2 = 0$.

To simplify the remaining equations, we observe that, if

$$\psi(r) = e^{ikr}/r,$$

then

$$\sin\mu\frac{\partial\psi(r_0)}{\partial\lambda} = -\sinh\lambda\frac{\partial\psi(r_0)}{\partial\mu}, \qquad \sin\mu\frac{\partial\psi(r_1)}{\partial\lambda} = \sinh\lambda\frac{\partial\psi(r_1)}{\partial\mu};$$

moreover, $\qquad (\sinh^2\lambda+\sin^2\mu)\psi(r_0)\psi(r_1) = e^{2ikf\cosh\lambda}.$

The equations (2.16), (2.17) then reduce to

$$\frac{\partial}{\partial\lambda}(a_3 h_3) = f\sin^2\mu\frac{\partial}{\partial\lambda}\{\psi(r_0)\psi(r_1)\} = \frac{\partial}{\partial\lambda}\{f\sin^2\mu\,\psi(r_0)\psi(r_1)\},$$

$$\frac{\partial}{\partial\mu}(a_3 h_3) = -f\sinh^2\lambda\frac{\partial}{\partial\mu}\{\psi(r_0)\psi(r_1)\} = -\frac{\partial}{\partial\mu}\{f\sinh^2\lambda\,\psi(r_0)\psi(r_1)\}$$

$$= \frac{\partial}{\partial\mu}\{f\sin^2\mu\,\psi(r_0)\psi(r_1)\}.$$

Integrating, we obtain

$$a_3 h_3 = f\sin^2\mu\,\psi(r_0)\psi(r_1) = f\sin^2\mu\frac{e^{ik(r_0+r_1)}}{r_0 r_1},$$

and so

$$\mathbf{a} = a_3\mathbf{i}_3 = \frac{\sin\mu}{\sinh\lambda}\frac{e^{ik(r_0+r_1)}}{r_0 r_1}\mathbf{i}_3. \tag{2.19}$$

It remains to put this in a form independent of our special choice of coordinates.

We have

$$\mathbf{i}_3 = \frac{1}{h_3}\left(\mathbf{i}\frac{\partial x}{\partial\nu}+\mathbf{j}\frac{\partial y}{\partial\nu}+\mathbf{k}\frac{\partial z}{\partial\nu}\right) = -\mathbf{i}\sin\nu+\mathbf{j}\cos\nu.$$

Thus $\mathbf{i}_3$ is perpendicular to the plane LPQ, and so is a constant multiple of $\mathbf{r}_0\times\mathbf{r}_1$. But since

$$\mathbf{r}_0 = \mathbf{r}+f\mathbf{k}, \qquad \mathbf{r}_1 = \mathbf{r}-f\mathbf{k},$$

we have

$$\mathbf{r}_0\times\mathbf{r}_1 = -2f\mathbf{r}\times\mathbf{k} = -2f^2\sinh\lambda\sin\mu(\mathbf{i}\sin\nu-\mathbf{j}\cos\nu),$$

and so

$$\mathbf{i}_3 = \frac{\mathbf{r}_0\times\mathbf{r}_1}{2f^2\sinh\lambda\sin\mu}.$$

Substituting in the formula for $\mathbf{a}$, we obtain

$$\mathbf{a} = \frac{e^{ik(r_0+r_1)}}{r_0 r_1} \frac{\mathbf{r}_0 \times \mathbf{r}_1}{2f^2 \sinh^2\lambda}.$$

Again, it is easily seen that

$$2f^2 \sinh^2\lambda = r_0 r_1 + \mathbf{r}_0 \cdot \mathbf{r}_1,$$

and so, finally,

$$\mathbf{a} = \frac{e^{ik(r_0+r_1)}}{r_0 r_1} \frac{\mathbf{r}_0 \times \mathbf{r}_1}{r_0 r_1 + \mathbf{r}_0 \cdot \mathbf{r}_1}. \tag{2.110}$$

We have thus proved that, if S is an unclosed surface with rim Γ,

$$\iint\limits_S \left\{ \frac{e^{ikr_1}}{r_1} \frac{\partial}{\partial n}\left(\frac{e^{ikr_0}}{r_0}\right) - \frac{e^{ikr_0}}{r_0} \frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right) \right\} dS = \int\limits_\Gamma \frac{e^{ik(r_0+r_1)}}{r_0 r_1} \frac{(\mathbf{r}_0 \times \mathbf{r}_1)\cdot \mathbf{t}}{r_0 r_1 + \mathbf{r}_0 \cdot \mathbf{r}_1}\, ds, \tag{2.111}$$

provided that the segment LP does not intersect S. The modification when LP does intersect S is evident.

Example. $\mathbf{p}$ is a fixed unit vector and $u = \exp(ik\mathbf{r}\cdot\mathbf{p})$. Prove that

$$u = \lim_{f\to\infty} \frac{fe^{ikr_0}}{r_0 e^{ikf}},$$

where

$$\mathbf{r}_0 = \mathbf{r} + f\mathbf{p}.$$

Hence show that

$$\iint\limits_S \left\{ \frac{e^{ikr_1}}{r_1} \frac{\partial}{\partial n} \exp(ik\mathbf{r}\cdot\mathbf{p}) - \exp(ik\mathbf{r}\cdot\mathbf{p}) \frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right) \right\} dS$$

$$= \int\limits_\Gamma \frac{\exp\{ik(\mathbf{r}\cdot\mathbf{p}+r_1)\}}{r_1} \frac{(\mathbf{p}\times\mathbf{r}_1)\cdot\mathbf{t}}{r_1 + \mathbf{p}\cdot\mathbf{r}_1}\, ds,$$

provided that the vector through P in the direction $-\mathbf{p}$ does not intersect S.

§ 2.2. Geometrical optics as a limiting form of physical optics

There is a general theorem due to Kirchhoff† which states that geometrical optics is a limiting form of physical optics. More precisely, the diffuse boundary of the shadow in diffraction phenomena becomes the sharp shadow of geometrical optics as the wave-length of the light tends to zero. Kirchhoff's proof was based on his theory of diffraction as outlined in § 1.1, and depended on a transformation of Helmholtz's integral and the use of what would now be called the Riemann-Lebesgue lemma.‡ But, owing to the generality of his work,

† *Vorlesungen ü. math. Phys.* **2** (*Optik*), 35. See also *Encyc. der math. Wissen.* V_3, 437; Sommerfeld and Runge, *Ann. d. Phys.* (4) **35** (1911), 277–98; Friedlander, *Proc. Camb. Phil. Soc.* **43** (1947), 284–6.

‡ We refer here to the 'Hilfsatz' on p. 33 of Kirchhoff's *Vorlesungen*. Actually the 'Hilfsatz' is not correctly stated.

it is not easy to present Kirchhoff's argument in a satisfactory form.† In the present section we show how Maggi's transformation of Helmholtz's integral enables us to prove Kirchhoff's result under conditions which are sufficiently general for most purposes.

Let us suppose that L is a point-source‡ of light which can, in some way, be specified by the wave-function

$$u_0 = \frac{e^{ik(r-ct)}}{r}.$$

This light is diffracted by a non-reflecting opaque body with bounding surface S. By (1.12) the effect at a point P is specified by the wave-function

$$u = \frac{1}{4\pi}\iint\limits_{\Sigma} \left\{\frac{e^{ikr_0}}{r_0}\frac{\partial}{\partial n}\left(\frac{e^{ik(r_1-ct)}}{r_1}\right) - \frac{e^{ik(r_1-ct)}}{r_1}\frac{\partial}{\partial n}\left(\frac{e^{ikr_0}}{r_0}\right)\right\} dS,$$

where Σ is an unclosed surface whose rim is the curve Γ along which the tangent cone from L to S touches S, and $\partial/\partial n$ denotes differentiation along the normal|| drawn from the 'illuminated' to the 'dark' side of Σ.

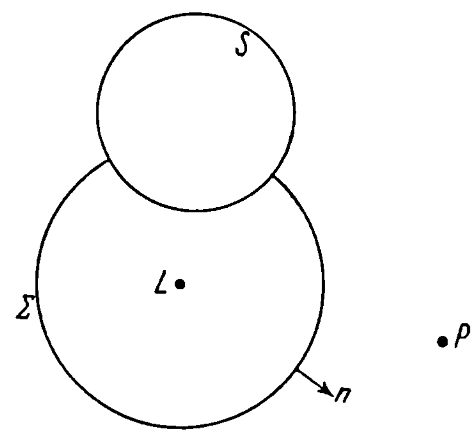


FIG. 11

Applying Maggi's transformation, we obtain

$$u = \epsilon u_0 - \frac{1}{4\pi}\int\limits_{\Gamma} \mathbf{a}\cdot\mathbf{t}\,ds, \tag{2.21}$$

where $\epsilon = 0$ or 1 according as P is or is not in the geometrical shadow.

† The difficulty is to show that the 'Hilfsatz' or the Riemann-Lebesgue lemma is applicable.

‡ The proof when the light consists of plane waves (L at infinity) follows similar lines and is omitted here.

|| The sense of the normal in formula (1.12) has been reversed.

To prove Kirchhoff's theorem, we have to show that the line integral in (2.21) tends to zero as the wave-length $2\pi/(kc)$ tends to zero, i.e. as $k \to \infty$.

Using the spheroidal coordinates of § 2.1, we have

$$\mathbf{t} = \frac{d\mathbf{r}}{ds} = h_1 \mathbf{i}_1 \frac{d\lambda}{ds} + h_2 \mathbf{i}_2 \frac{d\mu}{ds} + h_3 \mathbf{i}_3 \frac{d\nu}{ds},$$

and so $$\mathbf{a}\cdot\mathbf{t} = a_3 h_3 \frac{d\nu}{ds} = \frac{\sin^2\mu}{f(\cosh^2\lambda - \cos^2\mu)} e^{ik(2f\cosh\lambda - ct)} \frac{d\nu}{ds}.$$

Hence the integral to be discussed is

$$\int\limits_{\Gamma} \mathbf{a}\cdot\mathbf{t}\, ds = \int\limits_{\Gamma} e^{2ikf\cosh\lambda} \frac{\sin^2\mu}{f(\cosh^2\lambda - \cos^2\mu)}\, d\nu, \tag{2.22}$$

the time factor e^{-ikct} being omitted.

We now make the following assumptions:

(i) P is not on the boundary of the geometrical shadow. This implies that $\lambda > 0$ on Γ, and so the integrand in (2.22) is a continuous function of λ and μ which can be differentiated as often as we please.

(ii) On no arc of Γ is λ constant. This means that no arc of Γ lies on a spheroid $r_0 + r_1 =$ constant.

(iii) Γ is a simple closed curve† on which λ, μ, ν are continuous functions of the arc s, which can be differentiated as often as we please.

(iv) Γ can be divided into arcs on which λ is strictly monotonic. By (iii), $d\lambda/ds$ is of one sign on such an arc and vanishes only at its ends.

Of these assumptions, (i) and (ii) are necessary for the truth of the theorem, whilst (iii) and (iv) are conveniently simple sufficient conditions.

Consider an arc γ on which $\cosh\lambda$ increases steadily from a to b; by (i), $a > 1$. The integral along γ is

$$I \equiv \int\limits_{\gamma} \mathbf{a}\cdot\mathbf{t}\, ds = \int\limits_{\gamma} e^{2ikf\cosh\lambda} \frac{\sin^2\mu}{f(\cosh^2\lambda - \cos^2\mu)} \frac{d\nu}{ds}\, ds$$

$$= \int\limits_{a}^{b} e^{inu} \frac{\sin^2\mu}{f(u^2 - \cos^2\mu)} \frac{d\nu}{ds} \frac{ds}{du}\, du,$$

† The case when Γ is a simple curve extending to infinity is ruled out here for simplicity. The reader will see that the result still holds in this case provided that the integral along Γ is absolutely convergent.

where $u = \cosh\lambda$, $n = 2kf$. We write this in the form

$$I = \int_a^b e^{inu}\phi(u)\frac{ds}{du}\,du.$$

We must next discuss how s depends on u.

Take any two numbers a' and b' such that $a < a' < b' < b$. Then, by the inverse function theorem, s is a steadily increasing function of u in the interval $a' \leqslant u \leqslant b'$, and can be differentiated as often as we please. Hence, by an integration by parts beginning with

$$\frac{1}{ni}\int_{a'}^{b'} \phi(u)\frac{ds}{du}\,d(e^{niu}),$$

we see that the contribution to I of the interval (a', b') tends to zero as $n \to \infty$.

To deal with the interval (a, a') we measure s from $u = a$ and observe that, by (iv), an expansion

$$u = a + a_m s^m + a_{m+1} s^{m+1} + \dots,$$

where m is a positive integer greater than unity, holds in an interval $0 \leqslant s \leqslant s_0$; we can choose a' so that $a' < u(s_0)$. By reversion of series we deduce an expansion

$$s = \sum_{r=1}^{\infty} b_r (u-a)^{r/m},$$

valid when $a \leqslant u \leqslant a'$, the real positive mth root being understood here. It then follows that

$$\phi(u)\frac{ds}{du} = \frac{1}{u-a}\sum_{r=1}^{\infty} c_r (u-a)^{r/m}.$$

We have now to consider the behaviour of

$$\int_a^{a'} e^{niu} \sum_{r=1}^{\infty} c_r (u-a)^{r/m} \frac{du}{u-a}$$

as $n \to \infty$.

We can divide the infinite series into two parts. To the sum of those terms for which $r \geqslant m$, our previous argument depending on integration by parts can be applied, and it turns out that they have no effect when $n \to \infty$. If $r < m$, we have

$$\int_a^{a'} e^{inu}(u-a)^{r/m-1}\,du = \int_0^{n(a'-a)} e^{ina+iv}\frac{v^{r/m-1}}{n^{r/m}}\,dv$$

$$\sim \frac{e^{ina}}{n^{r/m}}\int_0^{\infty} e^{iv}v^{r/m-1}\,dv = \frac{e^{ina}}{n^{r/m}}e^{\frac{1}{2}\pi ir/m}\Gamma\left(\frac{r}{m}\right),$$

on using a well-known integral valid when $0 < r < m$. Hence

$$\int_a^{a'} e^{inu}\phi(u)\frac{ds}{du}\,du \to 0$$

as $n \to \infty$. A similar argument shows that the integral over (b', b) also tends to zero. Thus, finally,

$$\lim_{k\to\infty}\int_\gamma \mathbf{a}\cdot\mathbf{t}\,ds = 0$$

for each finite arc γ of Γ having the property assumed under condition (iv).

We have thus proved that, under quite general conditions,

$$u = \epsilon u_0 - \frac{1}{4\pi}\int_\Gamma \mathbf{a}\cdot\mathbf{t}\,ds$$

$$\to \epsilon u_0$$

as the wave-length $2\pi/(kc)$ tends to zero. In other words, the limit of the wave-function u has a sharp discontinuity on the edge of the geometrical shadow; in the shadow there is absolute darkness.

By our assumption (ii) we ruled out of consideration the case when there is an arc of Γ on which λ has a constant value. This excluded case is of some experimental interest. Let us suppose that γ is an arc on which λ has the constant value α. Then

$$\int_\gamma \mathbf{a}\cdot\mathbf{t}\,ds = e^{2ikf\cosh\alpha}\int_\gamma \frac{\sin^2\mu}{f(\cosh^2\alpha-\cos^2\mu)}\,d\nu,$$

and this certainly does not tend to zero as $k \to \infty$. Thus, when an arc of Γ lies on one of the spheroids $r_0+r_1 =$ constant, there is always illumination at the point P, even if it is in the geometrical shadow, no matter how small the wave-length is.

This result agrees with experiment. For Fresnel observed that, when there is a source of light on the axis of a circular screen, there is always brightness on the axis beyond the screen, no matter what the wave-length may be.

§ 3. Diffraction by a black half-plane

§ 3.1. The diffraction of plane waves of monochromatic light

The simplest type of diffraction problem is that in which plane monochromatic light is incident on a plane black screen, i.e. on a very thin non-reflecting opaque plane surface. In spite of its simplicity a problem of this type is of considerable experimental interest: for it can be realized by means of a spectrometer in which the collimator and telescope are both focused on infinity and are separated by the diffraction screen. We consider here the simplest problem of this type, namely that in which the black screen is of infinite extent and is bounded by a straight edge.

Let us suppose that the incident monochromatic light is specified by the wave-function
$$u = e^{-ik(x+ct)},$$
or rather by its real part. The wave-surfaces are all perpendicular to Ox and are travelling with velocity $-c$ parallel to Ox. If P is a point at which $x\cos\alpha+y\sin\alpha < 0$, where $-\frac{1}{2}\pi < \alpha < \frac{1}{2}\pi$, Helmholtz's formula gives
$$u(P) = \frac{1}{4\pi}\iint\limits_{S} \left\{u\frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right)-\frac{e^{ikr_1}}{r_1}\frac{\partial u}{\partial n}\right\} dS, \tag{3.11}$$
where integration is over the whole plane S whose equation is
$$x\cos\alpha+y\sin\alpha = 0$$
and $\mathbf{n}$ is the unit vector $-\mathbf{i}\cos\alpha-\mathbf{j}\sin\alpha$ normal to S.

If, however, a black screen covers up the part of S on which y is negative, the wave-function U specifying the effect beyond S is obtained, according to Kirchhoff's assumption, by putting
$$u = \partial u/\partial n = 0$$
on the screen. Hence (3.11) then becomes
$$U(P) = \frac{1}{4\pi}\iint\limits_{S_1} \left\{u\frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right)-\frac{e^{ikr_1}}{r_1}\frac{\partial u}{\partial n}\right\} dS, \tag{3.12}$$
where S_1 is the part of S on which $y \geqslant 0$. (See Fig. 12.)

We now apply to (3.12) Maggi's transformation in the form appropriate to plane waves. As the source L is the point at infinity on the x-axis, the formula of § 2.1, Ex., gives

$$U(P) = -\frac{1}{4\pi}\int\limits_{\Gamma} \mathbf{a}\cdot\mathbf{t}\,ds \tag{3.13}$$

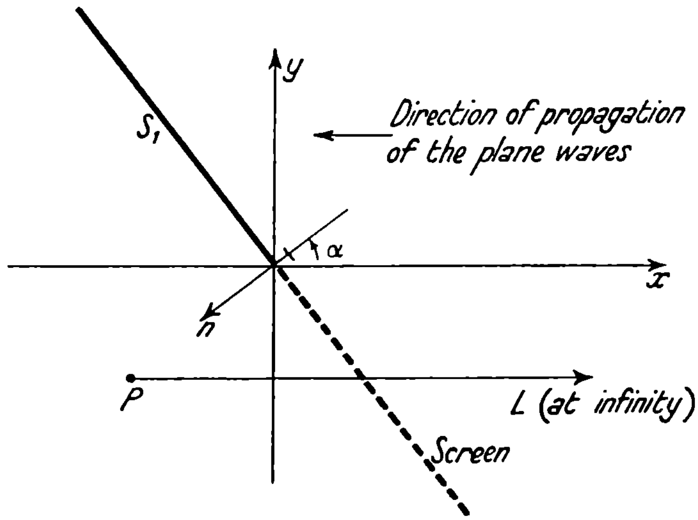


FIG. 12

when P is in the geometrical shadow; in this formula Γ is the straight edge of the screen (the axis of z), $\mathbf{t}$ the unit vector tangent to Γ, and

$$\mathbf{a} = -\frac{e^{ik(r_1-ct)}}{r_1(r_1-\mathbf{i}\cdot\mathbf{r}_1)}\mathbf{i}\times\mathbf{r}_1.$$

If P has coordinates (x_1, y_1, z_1), we have

$$\mathbf{r}_1 = \overrightarrow{PQ} = -\mathbf{i}x_1-\mathbf{j}y_1+\mathbf{k}(z-z_1)$$

when Q is a point on Γ; hence

$$\mathbf{i}\cdot\mathbf{r}_1 = -x_1, \qquad \mathbf{i}\times\mathbf{r}_1 = -\mathbf{k}y_1-\mathbf{j}(z-z_1).$$

But since $\mathbf{n} = -\mathbf{i}\cos\alpha-\mathbf{j}\sin\alpha$, the right-hand screw rule gives $\mathbf{t} = \mathbf{k}$, and so

$$(\mathbf{i}\times\mathbf{r}_1)\cdot\mathbf{t} = -y_1.$$

Hence (3.13) becomes

$$U(P) = -\frac{1}{4\pi}\int\limits_{-\infty}^{\infty} \frac{e^{ik(r_1-ct)}}{r_1(r_1+x_1)}y_1\,dz. \tag{3.14}$$

It is more convenient to use, not the cartesian coordinates of P, but its cylindrical coordinates (ρ, ϕ, z_1) with Oz as axis. This gives

$$x_1 = \rho\cos\phi, \quad y_1 = \rho\sin\phi, \quad r_1^2 = \rho^2+(z-z_1)^2.$$

If we substitute these values in (3.14) and then change the variable of integration to τ, where

$$z = z_1+\rho\sinh\tau,$$

we obtain

$$U(P) = -\frac{1}{4\pi}\int\limits_{-\infty}^{\infty} e^{ik\rho\cosh\tau-ikct}\frac{\sin\phi}{\cosh\tau+\cos\phi}\,d\tau. \tag{3.15}$$

The formula (3.15) holds only when P is in the geometrical shadow. If the incident light falls directly on P we must add to the right-hand side the term $u(P)$. Hence

$$U(P) = \epsilon u(P)-\frac{1}{4\pi}\int\limits_{-\infty}^{\infty} e^{ik\rho\cosh\tau-ikct}\frac{\sin\phi}{\cosh\tau+\cos\phi}\,d\tau, \tag{3.16}$$

where $\epsilon = 0$ or 1 according as P is or is not in the geometrical shadow. We shall write equation (3.16) in the form

$$U = u^*+u^B, \tag{3.17}$$

where u^* is written for ϵu, the wave-function according to the laws of geometrical optics. The term u^B then represents the effect of diffraction.

§ 3.2. Some analytical transformations and approximations

As we have just seen, when plane monochromatic light is incident on a black half-plane, the wave-function u^B, where

$$u^B = -\frac{1}{2\pi}\int\limits_{0}^{\infty} e^{ik(\rho\cosh\tau-ct)}\frac{\sin\phi}{\cosh\tau+\cos\phi}\,d\tau, \tag{3.21}$$

gives the correction which has to be added to the wave-function u^* of geometrical optics in order to account for the diffraction of light into the geometrical shadow. The formula (3.21) is rather difficult to apply as it stands. In the present section we obtain two transformations of the formula which are much easier to handle, and we deduce approximate formulae valid when ρ, the distance from the edge of the screen, is either very small or very large compared with the wave-length $2\pi/k$.

The function u^B is evidently a periodic function of ϕ of period 2π. We consider then only the range $-\pi < \phi < \pi$; the part of this range in which $-\pi < \phi < \alpha - \frac{1}{2}\pi$ forms the geometrical shadow. The two formulae we shall prove are

$$u^B = -\frac{1}{2\pi} \int\limits_0^{\infty+\frac{1}{2}\pi i} e^{ik(\rho\cosh\zeta - ct)} \frac{\sin\phi}{\cosh\zeta+\cos\phi}\, d\zeta \tag{3.22}$$

and

$$u^B = -\frac{1}{2\pi} e^{-ik(\rho\cos\phi+ct)} \left[\phi - \tfrac{1}{2}\pi\sin\phi \int\limits_0^{k\rho} e^{i\xi\cos\phi} H_0^{(1)}(\xi)\, d\xi\right], \tag{3.23}$$

where $H_0^{(1)}(\xi)$ denotes the Hankel function $J_0(\xi)+iY_0(\xi)$.

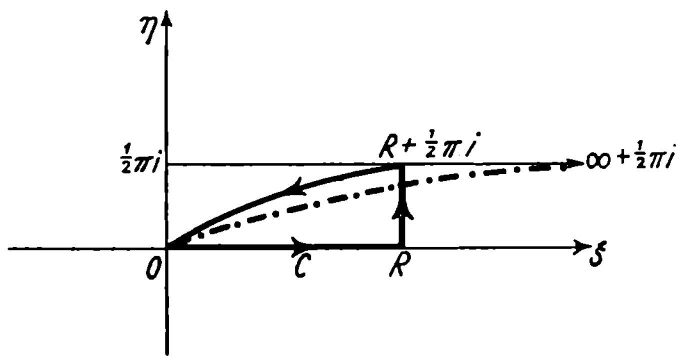


FIG. 13

To prove (3.22) we consider the contour integral

$$\int\limits_C e^{ik\rho\cosh\zeta} \frac{\sin\phi}{\cosh\zeta+\cos\phi}\, d\zeta$$

taken round the closed contour C of the figure. Since the integrand, regarded as a function of the complex variable ζ, has all its singularities on the imaginary axis, the value of the integral is zero. Hence

$$\int\limits_0^R + \int\limits_R^{R+\frac{1}{2}\pi i} e^{ik\rho\cosh\zeta} \frac{\sin\phi}{\cosh\zeta+\cos\phi}\, d\zeta = \int\limits_0^{R+\frac{1}{2}\pi i} e^{ik\rho\cosh\zeta} \frac{\sin\phi}{\cosh\zeta+\cos\phi}\, d\zeta. \tag{3.24}$$

But on the side of C parallel to the imaginary axis, we have $\zeta = R+i\eta$, and so

$$\left|\int\limits_R^{R+\frac{1}{2}\pi i} e^{ik\rho\cosh\zeta} \frac{\sin\phi}{\cosh\zeta+\cos\phi}\, d\zeta\right|$$

$$\leqslant \int\limits_0^{\frac{1}{2}\pi} e^{-k\rho\sinh R\sin\eta} \frac{|\sin\phi|}{|\cosh(R+i\eta)|-|\cos\phi|}\, d\eta \leqslant \int\limits_0^{\frac{1}{2}\pi} \frac{|\sin\phi|}{\sinh R-|\cos\phi|}\, d\eta.$$

Hence $$\int_{R}^{R+\frac{1}{2}\pi i} \to 0 \quad \text{as} \quad R \to \infty.$$

Making R tend to infinity in (3.24), we obtain

$$\int_{0}^{\infty} e^{ik\rho\cosh\zeta}\frac{\sin\phi}{\cosh\zeta+\cos\phi}\,d\zeta = \int_{0}^{\infty+\frac{1}{2}\pi i} e^{ik\rho\cosh\zeta}\frac{\sin\phi}{\cosh\zeta+\cos\phi}\,d\zeta,$$

the path of integration in the latter integral being the dotted curve in the figure which is asymptotic to the line $\eta = \frac{1}{2}\pi$. If we substitute in (3.21), we find that

$$u^B = -\frac{1}{2\pi}\int_{0}^{\infty+\frac{1}{2}\pi i} e^{ik(\rho\cosh\zeta-ct)}\frac{\sin\phi}{\cosh\zeta+\cos\phi}\,d\zeta,$$

which is equation (3.22).

The simplest way of proving (3.23) is to show that u^B satisfies the differential equation

$$\frac{\partial u^B}{\partial(k\rho)}+i\cos\phi\, u^B = \tfrac{1}{4}\sin\phi\, H_0^{(1)}(k\rho)e^{-ikct}.$$

It is more convenient to derive $\partial u^B/\partial(k\rho)$ from equation (3.22) since we then obtain an absolutely convergent integral instead of a conditionally convergent one. We then have

$$\frac{\partial u^B}{\partial(k\rho)} = -\frac{1}{2\pi}\int_{0}^{\infty+\frac{1}{2}\pi i} e^{ik(\rho\cosh\zeta-ct)}\frac{i\sin\phi\cosh\zeta}{\cosh\zeta+\cos\phi}\,d\zeta$$

provided that this integral converges uniformly with respect to ρ; this is certainly the case since the integrand behaves like

$$e^{-k\rho\sinh\xi},$$

where $\zeta = \xi+\frac{1}{2}\pi i$ and ξ is large and positive. Hence we have

$$\frac{\partial u^B}{\partial(k\rho)}+i\cos\phi\, u^B = -\frac{i}{2\pi}\int_{0}^{\infty+\frac{1}{2}\pi i} e^{ik(\rho\cosh\zeta-ct)}\sin\phi\,d\zeta.$$

Using the formula†

$$H_0^{(1)}(x) = \frac{2}{\pi i}\int_{0}^{\infty+\frac{1}{2}\pi i} e^{ix\cosh\zeta}\,d\zeta,$$

† Watson, *Bessel Functions*, 180 (8).

we obtain the required differential equation

$$\frac{\partial u^B}{\partial(k\rho)}+i\cos\phi\, u^B = \tfrac{1}{4}\sin\phi\, H_0^{(1)}(k\rho)e^{-ikct}. \tag{3.25}$$

From this it follows that

$$u^B = u^B(0)e^{-ik\rho\cos\phi}+\tfrac{1}{4}\sin\phi\, e^{-ik(\rho\cos\phi+ct)}\int\limits_0^{k\rho} e^{i\xi\cos\phi}H_0^{(1)}(\xi)\,d\xi.$$

But, by (3.21),

$$u^B(0) = -\frac{1}{2\pi}\int\limits_0^\infty e^{-ikct}\frac{\sin\phi}{\cosh\tau+\cos\phi}\,d\tau = -\frac{\phi}{2\pi}e^{-ikct},$$

provided that $-\pi<\phi<\pi$. Hence we have

$$u^B = -\frac{1}{2\pi}e^{-ik(\rho\cos\phi+ct)}\left[\phi-\tfrac{1}{2}\pi\sin\phi\int\limits_0^{k\rho} e^{i\xi\cos\phi}H_0^{(1)}(\xi)\,d\xi\right] \tag{3.23}$$

if $-\pi<\phi<\pi$, a formula due to Dr. Erdélyi.

The equation (3.23) enables us to find very easily an approximate formula for u^B valid when ρ is small compared with the wave-length, i.e. when $k\rho$ is small. For

$$H_0^{(1)}(\xi) = J_0(\xi)+iY_0(\xi) = \frac{2i}{\pi}\log\xi+O(1)$$

as $\xi\to 0$, and so

$$\int\limits_0^{k\rho} e^{i\xi\cos\phi}H_0^{(1)}(\xi)\,d\xi = \frac{2i}{\pi}k\rho\log(k\rho)+O(k\rho).$$

Hence
$$u^B = -\frac{1}{2\pi}e^{-ikct}[\phi-i\sin\phi\, k\rho\log(k\rho)+O(k\rho)], \tag{3.26}$$

when $k\rho$ is small.

To find an asymptotic formula for u^B valid when $k\rho$ is large, we go back to (3.22). Making the substitution

$$\cosh\zeta = 1+iv,$$

we have

$$u^B = -\frac{e^{-ikct}}{2\pi}\int\limits_0^{\infty+\frac{1}{2}\pi i} e^{ik\rho\cosh\zeta}\frac{\sin\phi}{\cosh\zeta+\cos\phi}\,d\zeta$$

$$= -\frac{1}{2\pi}e^{ik(\rho-ct)+\frac{1}{4}\pi i}\int\limits_0^{-\infty i} e^{-k\rho v}\frac{\sin\phi}{(1+\cos\phi+iv)\sqrt{(2v+iv^2)}}\,dv$$

$$= -\frac{1}{2\pi}e^{ik(\rho-ct)+\frac{1}{4}\pi i}\int\limits_0^{\infty} e^{-k\rho v}\frac{\sin\phi}{(1+\cos\phi+iv)\sqrt{(2v+iv^2)}}\,dv,$$

the rotation of the path of integration through a right angle being justified by Cauchy's theorem.† It follows by Watson's lemma‡ that

$$u^B \sim -\frac{1}{2\pi} e^{ik(\rho-ct)+\frac{1}{4}\pi i} \int\limits_0^\infty e^{-k\rho v} \frac{\sin\phi}{(1+\cos\phi)\sqrt{(2v)}}\, dv$$

as $k\rho \to \infty$. Hence

$$u^B \sim -\frac{1}{2\sqrt{(2\pi k\rho)}} \tan\tfrac{1}{2}\phi\, e^{ik\rho+\frac{1}{4}\pi i-ikct} \tag{3.27}$$

when ρ is large compared with the wave-length, provided that $-\pi < \phi < \pi$.

In conclusion we observe that u^B can be expanded as a Fourier series of the form||

$$u^B = \mp \eta e^{-ik\rho\cos\phi-ikct} - \frac{1}{2i}\sum_{n=1}^\infty e^{-\frac{1}{4}n\pi i} h_n^{(1)}(k\rho) e^{-ikct}\sin n\phi, \tag{3.28}$$

where the upper or lower sign is taken according as ϕ is positive or negative and $\eta = 0$, $\frac{1}{4}$, or $\frac{1}{2}$ according as $0 \leqslant |\phi| < \frac{1}{2}\pi$, $|\phi| = \frac{1}{2}\pi$, or $\frac{1}{2}\pi < |\phi| < \pi$. The function $h_n^{(1)}(\xi)$ is a 'cut' Hankel function, that is, it is the function

$$H_n^{(1)}(\xi) = J_n(\xi) + iY_n(\xi)$$

deprived of terms involving negative powers of ξ.

Although this expansion leads at once to the approximation (3.26), its real interest lies in the occurrence of the 'cut' Hankel functions in a physical problem. The function u^B is a solution of the equation of cylindrical waves, and so we should expect the Fourier series for u^B to be in terms of the normal functions for cylindrical waves. These normal functions are, as is well known,

$$H_n^{(m)}(k\rho) \begin{matrix}\sin\\ \cos\end{matrix} n\phi e^{-ikct}, \tag{3.29}$$

where $$H_n^{(m)}(k\rho) = J_n(k\rho) \pm iY_n(k\rho),$$

the upper or lower sign being taken according as $m = 1$ or 2: moreover, when $k\rho$ is large, these functions behave like

$$\left(\frac{2}{\pi k\rho}\right)^{\frac{1}{2}} \exp\left\{\pm i\left(k\rho - \frac{2n+1}{4}\pi\right) - ikct\right\} \begin{matrix}\sin\\ \cos\end{matrix} n\phi,$$

† The argument consists essentially in considering the integral round the complete boundary of the fourth quadrant of the circle $|v| = R$, and then making R tend to infinity.

‡ Watson, *Bessel Functions*, 236.

|| Copson and Ferrar, *Proc. Edin. Math. Soc.* (2), **5** (1938), 159–68. See also Watson, ibid. **5** (1938), 174–81; Erdélyi, ibid. **6** (1939), 11.

so that (3.29) represents an expanding or contracting cylindrical wave-motion according as $m = 1$ or 2.

The expansion (3.28) consists, however, of a term representing a plane wave-motion together with terms of the form

$$h_n^{(1)}(k\rho)\sin n\phi\, e^{-ikct}$$

which cannot be formed from the normal functions (3.29) and do not satisfy the equation of cylindrical waves. A phenomenon of this type was first noticed by Whipple† in his work on diffraction by a wedge.

§ 3.3. The diffracted waves

So far, we have shown that the diffraction of plane monochromatic light by a black half-plane is characterized by the wave-function

$$U = u^* + u^B,$$

where u^* represents the effect according to the laws of geometrical optics and u^B is the correction term which takes account of diffraction. We now consider the function U in more detail.

Although u^* is discontinuous across the geometrical shadow, U is continuous, since u^B has a discontinuity which compensates the discontinuity of u^*. For we have

$$\lim_{\phi\to\pi-0} u^* = e^{-ik(\rho+ct)}, \qquad \lim_{\phi\to-\pi+0} u^* = 0;$$

also, by (3.23),

$$\lim_{\to\pi-0} u^B = -\tfrac{1}{2}e^{-ik(\rho+ct)}, \qquad \lim_{\phi\to-\pi+0} u^B = \tfrac{1}{2}e^{-ik(\rho+ct)}.$$

Hence
$$\lim_{\phi\to\pi-0} U = \lim_{\phi\to-\pi+0} U = \tfrac{1}{2}e^{-ik(\rho+ct)},$$

which proves that U is continuous across the boundary of the geometrical shadow. There is thus no sudden change from light to darkness, but a gradual change.

Moreover, in the geometrical shadow, $U = u^B$ and so the intensity of illumination is measured by

$$|U|^2 = |u^B|^2 \sim \frac{1}{8\pi k\rho}\tan^2\tfrac{1}{2}\phi,$$

when ρ is large compared with the wave-length. Since the intensity in the incident light has measure unity, the illumination of the geometrical shadow is very feeble, but there is nowhere absolute darkness

† *Proc. London Math. Soc.* (2), **16** (1917), 94–111 (104).

since $\tan^2 \frac{1}{2}\phi$ is never zero in the shadow. The intensity gradually decreases as we move farther into the shadow.

When ρ is large compared with the wave-length

$$u^B \sim -\frac{1}{2\sqrt{(2\pi k\rho)}} \tan \tfrac{1}{2}\phi \, e^{ik\rho+\frac{1}{4}\pi i-ikct},$$

so that the diffracted wave is a non-isotropic cylindrical wave propagated outwards from the edge of the screen. This means that the edge of the screen should appear to be luminous when observed from points in the geometrical shadow, a result which can be verified experimentally. Outside the shadow the effect is masked by the incident light. Actually diffraction by a black half-plane is purely an edge effect; for neither u^* nor u^B depends on the angle of incidence α and so U is the same for all black screens having the same edge and the same geometrical shadow.

Lastly, we observe that outside the geometrical shadow the two wave-motions specified by u^* and u^B interfere and produce interference fringes—the diffraction pattern. But when ρ is large compared with the wave-length, the amplitude of u^B is very small except near $\phi = \pi$. Thus the diffraction pattern is observable only near the boundary of the geometrical shadow.

The actual diffraction pattern can be calculated by a careful approximation to u^B near the boundary of the shadow. In spite of the defects of the theory, the results agree quite well with the experiments.

§ 3.4. The diffraction of spherical waves by a black half-plane

The problem we have just discussed is the limiting case of the more general problem of the diffraction of monochromatic spherical waves by a black half-plane, which we shall now consider. The incident light is generated by a point-source L and is characterized by the wave-function

$$u = \frac{1}{r_0} e^{ik(r_0-ct)},$$

where r_0 is distance measured from the source.

We choose axes of coordinates so that L lies on the axis of x and the straight edge of the screen lies along the z-axis. Then the screen covers the half-plane $x\cos\alpha+y\sin\alpha = 0$, $y < 0$, where the angle α lies between $\pm\frac{1}{2}\pi$. On Kirchhoff's theory, the effect at a point P

beyond the plane of the screen is characterized by the wave-function

$$U(P) = \frac{1}{4\pi}\iint\limits_{S_1} \left\{u\frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right) - \frac{e^{ikr_1}}{r_1}\frac{\partial u}{\partial n}\right\} dS,$$

where S_1 is the half-plane $x\cos\alpha + y\sin\alpha = 0$, $y \geqslant 0$; in this formula r_1 is the distance measured from P and $\mathbf{n}$ is the unit vector

$$-\mathbf{i}\cos\alpha - \mathbf{j}\sin\alpha$$

normal to S_1.

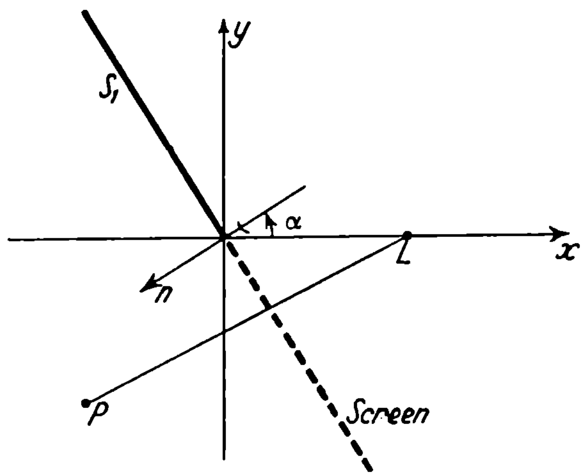


FIG. 14

If we apply Maggi's transformation (§ 2.1) to Kirchhoff's integral, we obtain

$$U(P) = \epsilon u(P) - \frac{1}{4\pi}\int\limits_{\Gamma} \mathbf{a}\cdot\mathbf{t}\, ds, \tag{3.41}$$

where Γ is the straight edge of the screen and $\epsilon = 0$ or 1 according as P is or is not in the geometrical shadow. In this equation

$$\mathbf{t} = \mathbf{k}, \qquad \mathbf{a} = \frac{e^{ik(r_0+r_1-ct)}}{r_0 r_1}\frac{\mathbf{r}_0\times\mathbf{r}_1}{r_0 r_1 + \mathbf{r}_0\cdot\mathbf{r}_1}. \tag{3.42}$$

Let L and P be at distances ρ_0 and ρ_1 respectively from the edge of the black screen, so that their coordinates are $(\rho_0, 0, 0)$ and $(\rho_1\cos\phi, \rho_1\sin\phi, z_1)$. Then

$$\mathbf{r}_0 = -\mathbf{i}\rho_0 + \mathbf{k}z, \qquad \mathbf{r}_1 = -\mathbf{i}\rho_1\cos\phi - \mathbf{j}\rho_1\sin\phi + \mathbf{k}(z-z_1),$$

since $x = y = 0$ on Γ. Hence

$$\mathbf{r}_0\cdot\mathbf{r}_1 = \rho_0\rho_1\cos\phi + z(z-z_1),$$

$$\mathbf{r}_0\times\mathbf{r}_1 = \mathbf{i}z\rho_1\sin\phi + \mathbf{j}\rho_0(z-z_1) - \mathbf{j}z\rho_1\cos\phi + \mathbf{k}\rho_0\rho_1\sin\phi,$$

$$(\mathbf{r}_0\times\mathbf{r}_1)\cdot\mathbf{t} = \rho_0\rho_1\sin\phi,$$

and so (3.41) becomes

$$U(P) = \epsilon u(P) - \frac{1}{4\pi} \int\limits_{-\infty}^{\infty} \frac{e^{ik(r_0+r_1-ct)}}{r_0 r_1} \frac{\rho_0 \rho_1 \sin\phi}{r_0 r_1 + \rho_0 \rho_1 \cos\phi + z(z-z_1)}\, dz. \tag{3.43}$$

We now make the substitutions

$$z = \rho_0 \sinh\tau_0, \qquad z - z_1 = \rho_1 \sinh\tau_1,$$

so that $\qquad r_0 = \rho_0 \cosh\tau_0, \qquad r_1 = \rho_1 \cosh\tau_1.$

It follows from (3.43) that

$$U(P) = \epsilon u(P) - \frac{1}{4\pi} \int\limits_{-\infty}^{\infty} \frac{e^{ik(R-ct)}}{\rho_0 \rho_1 \cosh\tau_0 \cosh\tau_1} \cdot \frac{\sin\phi}{\cosh(\tau_0+\tau_1)+\cos\phi}\, dz, \tag{3.44}$$

where $R = r_0 + r_1$. Lastly, we take a new variable of integration

$$\tau = \tau_0 + \tau_1 = \sinh^{-1}\frac{z}{\rho_0} + \sinh^{-1}\frac{z-z_1}{\rho_1}.$$

It is easily shown that

$$\frac{d\tau}{dz} = \frac{R}{\rho_0 \rho_1 \cosh\tau_0 \cosh\tau_1}$$

and that $\qquad R^2 = \rho_0^2 + \rho_1^2 + z_1^2 + 2\rho_0 \rho_1 \cosh\tau.$

Thus, finally,

$$U(P) = \epsilon u(P) - \frac{1}{4\pi} \int\limits_{-\infty}^{\infty} \frac{e^{ik(R-ct)}}{R} \frac{\sin\phi}{\cosh\tau + \cos\phi}\, d\tau. \tag{3.45}$$

This formula, which should be compared with the corresponding formula (3.16) for plane waves is due to Rubinowicz.†

Equation (3.45) is usually written in the form

$$U = u^* + u^B,$$

where u^* is the wave-function according to the laws of geometrical optics and u^B is the correction term which has to be added to account for the diffraction of light into the geometrical shadow. As in the case of plane waves, neither u^* nor u^B depends on the angle α, and so U is the same for all black screens having the same straight edge and the same geometrical shadow: more briefly, diffraction by a black half-plane is an edge effect.

† *Annalen der Phys.* **53** (1917), 257. The proof given here is due to Kottler (ibid. **70** (1923), 405).

§ 3.5. The diffracted wave and Fresnel's integrals

When monochromatic spherical waves are diffracted by a straight edge, the diffraction wave is, as we have just proved, characterized by the wave-function

$$u^B = -\frac{1}{2\pi}\int_0^\infty \frac{e^{ik(R-ct)}}{R}\frac{\sin\phi}{\cosh\tau+\cos\phi}\,d\tau, \tag{3.51}$$

where $$R^2 = \rho_0^2+\rho_1^2+z_1^2+2\rho_0\rho_1\cosh\tau.$$

We now consider† the asymptotic behaviour of u^B where ρ_0 and ρ_1 are large compared with the wave-length $2\pi/k$.

Since k is very large compared with $1/\rho_0$ and $1/\rho_1$, the important part of the integral (3.51) is due to that part of the range of integration near which the phase of e^{ikR} is stationary, and so is due to the part of the range of integration near $\tau = 0$. Now when τ is very small, we have

$$R = R_1+\frac{\rho_0\rho_1}{2R_1}\tau^2$$

correct to the second order in τ, where

$$R_1 = \sqrt{\{(\rho_0+\rho_1)^2+z_1^2\}}.$$

It follows that

$$u^B \doteqdot -\frac{1}{2\pi}\frac{e^{ik(R_1-ct)}}{R_1}\psi^2\tan\tfrac{1}{2}\phi\int_0^\infty e^{iq^2\tau^2}\frac{d\tau}{\tau^2+\psi^2}, \tag{3.52}$$

where $$q = \sqrt{\left(\frac{k\rho_0\rho_1}{2R_1}\right)}, \qquad \psi = 2\cos\tfrac{1}{2}\phi.$$

We have now to discuss the behaviour of

$$F(q) = \int_0^\infty e^{iq^2\tau^2}\frac{d\tau}{\tau^2+\psi^2} \tag{3.53}$$

for large positive values of q.

By rotating the path of integration in (3.53) through 45° by means of Cauchy's theorem, we obtain

$$F(q) = \int_0^\infty e^{-q^2\tau^2}\frac{e^{\frac{1}{4}\pi i}\,d\tau}{i\tau^2+\psi^2}.$$

† The subsequent analysis is a modification of a similar investigation made by Whipple (*Proc. London Math. Soc.* (2), **16** (1917), 103). The main difference is the rotation of the path of integration in (3.53) which simplifies the justification of what follows.

Hence we have

$$F(q) = e^{-\frac{1}{4}\pi i}\int\limits_0^\infty e^{-q^2\tau^2}\frac{d\tau}{\tau^2-i\psi^2}$$

$$= e^{-i(q^2\psi^2+\frac{1}{4}\pi)}\int\limits_0^\infty e^{-q^2(\tau^2-i\psi^2)}\frac{d\tau}{\tau^2-i\psi^2}$$

$$= e^{-i(q^2\psi^2+\frac{1}{4}\pi)}\int\limits_0^\infty d\tau\int\limits_q^\infty d\eta\, 2\eta e^{-\eta^2(\tau^2-i\psi^2)}.$$

As this repeated integral is absolutely convergent, we may invert the order of integration, to obtain

$$F(q) = e^{-i(q^2\psi^2+\frac{1}{4}\pi)}\int\limits_q^\infty d\eta\, 2\eta e^{i\eta^2\psi^2}\int\limits_0^\infty e^{-\eta^2\tau^2}\,d\tau$$

$$= \frac{\sqrt{\pi}}{\psi}e^{-i(q^2\psi^2+\frac{1}{4}\pi)}\int\limits_{q\psi}^\infty e^{i\xi^2}\,d\xi$$

$$= \frac{\sqrt{\pi}}{\psi}e^{-i(q^2\psi^2+\frac{1}{4}\pi)}f(q\psi), \tag{3.54}$$

where $f(\xi)$ denotes the complex Fresnel integral

$$f(\xi) = \int\limits_\xi^\infty e^{i\xi^2}\,d\xi.$$

Finally, if we substitute for $F(q)$ from (3.54) in (3.52), we find that

$$u^B \doteqdot -\frac{1}{2\sqrt{\pi}}\frac{e^{ik(R_1-ct)}}{R_1}\psi\tan\tfrac{1}{2}\phi\, e^{-i(q^2\psi^2+\frac{1}{4}\pi)}f(q\psi),$$

when the wave-length $2\pi/k$ is small compared with the distances ρ_0 and ρ_1 of the source and the point of observation from the edge of the screen.

The intensity of the diffracted wave is measured by

$$|u^B|^2 \doteqdot \frac{\psi^2\tan^2\frac{1}{2}\phi}{4\pi R_1^2}|f(q\psi)|^2.$$

A qualitative discussion of the variation of intensity can then be carried out geometrically by means of the curve (known as Cornu's Spiral), whose parametric equations are

$$X = \int\limits_\xi^\infty \cos(\xi^2)\,d\xi, \qquad Y = \int\limits_\xi^\infty \sin(\xi^2)\,d\xi;$$

for the distance from the origin to the point of parameter $q\psi$ on this curve is $|f(q\psi)|$.

Ex. 1. Show that the equation (3.45) can be written in the form

$$U(P) = \epsilon u(P) - \frac{2}{\pi}\int\limits_{R_1}^{\infty} \frac{e^{ik(R-ct)}}{R^2-r^2} \frac{\rho_0\rho_1\sin\phi}{\sqrt{\{(R^2-R_1^2)(R^2-R_2^2)\}}}\,dR,$$

where $R_1^2 = (\rho_0+\rho_1)^2+z_1^2$, $R_2^2 = (\rho_0-\rho_1)^2+z_1^2$, $r = PL$.

Ex. 2. u is the wave-function of a line of simple sources through the point L parallel to Oz. Prove that

$$u = \pi i H_0^{(1)}(kr_0)e^{-ikct},$$

where $$r_0^2 = \rho_0^2+\rho_1^2-2\rho_0\rho_1\cos\phi.$$

Hence show that, if the light from this line-source is diffracted by the black half-plane $x\cos\alpha+y\sin\alpha = 0$, $y < 0$, then

$$U(P) = \epsilon u(P) - \tfrac{1}{2}i\int\limits_{-\infty}^{\infty} H_0^{(1)}(kR)\frac{\sin\phi}{\cosh\tau+\cos\phi}\,d\tau,$$

where $R^2 = \rho_0^2+\rho_1^2+2\rho_0\rho_1\cosh\tau$.

§ 3.6. The asymptotic behaviour of the diffracted wave

An alternative method of obtaining the asymptotic behaviour of the diffracted wave when ρ_0 and ρ_1 are large compared with the wave-length depends on a transformation of the result of Ex. 1 above. This example shows that

$$u^B = -\frac{2}{\pi}\int\limits_{R_1}^{\infty} \frac{e^{ik(R-ct)}}{R^2-r^2} \frac{\rho_0\rho_1\sin\phi}{\sqrt{\{(R^2-R_1^2)(R^2-R_2^2)\}}}\,dR, \tag{3.61}$$

where
$$R_1^2 = (\rho_0+\rho_1)^2+z_1^2,$$
$$R_2^2 = (\rho_0-\rho_1)^2+z_1^2,$$
$$r = PL = \sqrt{\{\rho_0^2+\rho_1^2-2\rho_0\rho_1\cos\phi+z^2\}}.$$

In equation (3.61) we make the substitution $R = R_1(1+\tau)$ and then apply Cauchy's theorem to rotate the path of integration through a right angle. This gives

$$u^B = -\frac{2}{\pi}\rho_0\rho_1\sin\phi\, e^{ik(R_1-ct)}\int\limits_0^{\infty} \frac{e^{ikR_1\tau}\,d\tau}{\{R_1^2(1+\tau)^2-r^2\}\sqrt{\{2\tau+\tau^2\}}\sqrt{\{R_1^2(1+\tau)^2-R_2^2\}}}$$
$$= -\frac{2}{\pi}\rho_0\rho_1\sin\phi\, e^{ik(R_1-ct)+\frac{1}{4}\pi i}\times$$
$$\times\int\limits_0^{\infty} \frac{e^{-kR_1u}\,du}{\{R_1^2(1+iu)^2-r^2\}\sqrt{\{2u+iu^2\}}\sqrt{\{R_1^2(1+iu)^2-R_2^2\}}}.$$

When the wave-length is small compared with ρ_0 and ρ_1, kR_1 is large; the asymptotic expansion of u^B can then be obtained by applying Watson's lemma.† The dominant term is

$$u^B \doteqdot -\frac{2\rho_0\rho_1 \sin\phi}{\pi(R_1^2-r^2)\sqrt{(R_1^2-R_2^2)}} e^{ik(R_1-ct)+\frac{1}{4}\pi i} \int\limits_0^\infty e^{-kR_1 u} \frac{du}{\sqrt{(2u)}}$$

$$= -\frac{2\rho_0\rho_1 \sin\phi}{\pi(R_1^2-r^2)\sqrt{(R_1^2-R_2^2)}} e^{ik(R_1-ct)+\frac{1}{4}\pi i} \sqrt{\left(\frac{\pi}{2kR_1}\right)}.$$

Hence follows the approximation

$$u^B \doteqdot -\frac{\tan\frac{1}{2}\phi}{\sqrt{(8\pi kR_1\rho_0\rho_1)}} e^{ik(R_1-ct)+\frac{1}{4}\pi i}. \tag{3.62}$$

This result does not hold when P is near the boundary of the geometrical shadow, on which $\phi = \pm\pi$.

§ 4. Kottler's theory of diffraction by a black screen

Instead of regarding Kirchhoff's solution of the problem of diffraction by a black screen as a fairly accurate first approximation to a boundary value problem, Kottler‡ has shown that it is the rigorous solution of a somewhat different problem; in fact, it is the solution, not of a boundary value problem, but of what we may term a 'saltus problem'.

Let us consider the case of diffraction by a thin black screen in the form of a surface S bounded by a rim Γ; for simplicity, we take Γ to be a simple closed curve and leave the reader to make the appropriate modifications when Γ either extends to infinity or consists of several closed curves. We suppose that one face, S_+ say, of this screen is illuminated, and that the other, S_-, is dark. We shall regard S_+ and S_- as constituting a single degenerate closed surface S'.

If the undisturbed light is specified by the wave-function u_0 then, according to Kirchhoff, the effect when the screen is present is specified by a wave-function u satisfying the boundary conditions

$$u = u_0, \quad \frac{\partial u}{\partial n} = \frac{\partial u_0}{\partial n} \quad \text{on } S_+,$$

$$u = 0, \quad \frac{\partial u}{\partial n} = 0 \quad \text{on } S_-.$$

Kottler uses a different definition of blackness; he says that S is a

† Watson, *Bessel Functions*, 236. ‡ *Annalen der Phys.* **70** (1923), 405–56.

black screen if u and $\partial u/\partial n$ are discontinuous across S, the discontinuity being defined by

$$[u]_-^+ = u_0, \qquad \left[\frac{\partial u}{\partial n}\right]_-^+ = \frac{\partial u_0}{\partial n};$$

the actual values of u and $\partial u/\partial n$ on the two sides of S' are unknown.

As an example of this idea, let us suppose that the incident light is monochromatic light due to a simple source at the point L, the undisturbed wave-function being

$$u_0 = v_0 e^{-ikct} = e^{ik(r_0-ct)}/r_0.$$

The wave-function u will be of the form ve^{-ikct}, where v is independent of t and satisfies the following conditions:

(i) v is a one-valued function of the coordinates of the point of observation P, and satisfies the equation

$$\nabla^2 v + k^2 v = 0,$$

except when P is at L or on S'.

(ii) v becomes infinite at L, its principal part near L being $1/r_0$.

(iii) v has a discontinuity across the degenerate closed surface S' specified by

$$[v]_-^+ = U, \qquad \left[\frac{\partial v}{\partial n}\right]_-^+ = V,$$

where U and V are known functions, and $\mathbf{n}$ is the normal from S_+ to S_-.

(iv) v behaves like e^{ikr}/r on a sphere of large radius r.

Let D be the volume bounded externally by a sphere Σ of large radius R, and internally by the degenerate closed surface S', the small sphere σ_0 with centre L and radius ϵ_0, and the small sphere σ_1 with centre P and radius ϵ_1. Then by the usual application of Green's transformation, we have

$$\iint_{\Sigma} + \iint_{S'} + \iint_{\sigma_0} + \iint_{\sigma_1} \left\{v\frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right) - \frac{e^{ikr_1}}{r_1}\frac{\partial v}{\partial n}\right\} dS = 0, \quad (4.01)$$

where $\mathbf{n}$ denotes the normal drawn into D.

We consider separately the four terms on the left-hand side of equation (4.01). By hypothesis (iv), the integral over Σ tends to zero as $R \to \infty$. The integral over σ_1 tends to $-4\pi v(P)$ as $\epsilon_1 \to 0$, where

$v(P)$ denotes the value of v at P. By hypothesis (ii), the integral over σ_0 has the limit, as $\epsilon_0 \to 0$,

$$4\pi \frac{e^{ikLP}}{LP} = 4\pi v_0(P).$$

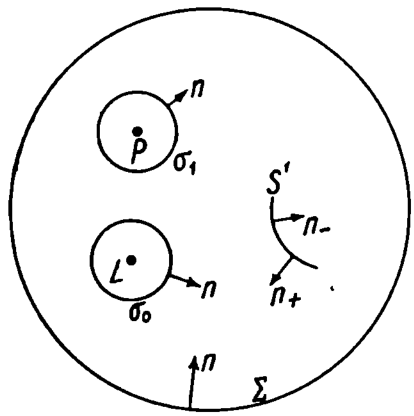


FIG. 15

Hence

$$4\pi v(P) = 4\pi v_0(P) + \iint\limits_{S'} \left\{ v\frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right) - \frac{e^{ikr_1}}{r_1}\frac{\partial v}{\partial n} \right\} dS. \tag{4.02}$$

It remains to consider the integral over the two faces of S'.

Now

$$\iint\limits_{S'} = \iint\limits_{S_+} + \iint\limits_{S_-}$$

$$= \iint\limits_{S} \left\{ v^+ \frac{\partial}{\partial n_+}\left(\frac{e^{ikr_1}}{r_1}\right) + v^- \frac{\partial}{\partial n_-}\left(\frac{e^{ikr_1}}{r_1}\right) - \frac{e^{ikr_1}}{r_1}\frac{\partial v^+}{\partial n_+} - \frac{e^{ikr_1}}{r_1}\frac{\partial v^-}{\partial n_-} \right\} dS$$

$$= -\iint\limits_{S} \left\{ [v]_-^+ \frac{\partial}{\partial n_-}\left(\frac{e^{ikr_1}}{r_1}\right) - \frac{e^{ikr_1}}{r_1}\left[\frac{\partial v}{\partial n_-}\right]_-^+ \right\} dS$$

$$= -\iint\limits_{S} \left\{ U\frac{\partial}{\partial n_-}\left(\frac{e^{ikr_1}}{r_1}\right) - \frac{e^{ikr_1}}{r_1} V \right\} dS$$

by condition (iii). It follows from (4.02) that

$$v(P) = v_0(P) - \frac{1}{4\pi}\iint\limits_{S} \left\{ U\frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right) - \frac{e^{ikr_1}}{r_1} V \right\} dS, \tag{4.03}$$

where $\mathbf{n}$ is the normal to S, drawn from the illuminated to the dark side.

In the problem of diffraction by a thin black screen having the form S, Kottler's hypothesis is that

$$U = v_0, \qquad V = \partial v_0/\partial n.$$

Hence, by (4.03), his solution of the problem is

$$v(P) = v_0(P) - \frac{1}{4\pi} \iint\limits_{S} \left\{ v_0 \frac{\partial}{\partial n}\left(\frac{e^{ikr_1}}{r_1}\right) - \frac{e^{ikr_1}}{r_1} \frac{\partial v_0}{\partial n} \right\} dS. \qquad (4.04)$$

Equation (4.04) is the formula Kirchhoff gave for diffraction by a black screen (§ 1.1, Ex.). But whereas Kirchhoff's formula purported to give a solution of a boundary value problem in which the boundary data turned out to be incompatible, Kottler derived the same formula rigorously from a different definition of 'blackness'. It is impossible to give a satisfactory *physical* definition of a thin black screen; Kottler's work shows us what *analytical* definition of 'blackness' gives rise to Kirchhoff's formula.

It should be noticed that it is not possible to apply Kottler's definition of 'blackness' to a thick screen. A further assumption is needed, namely that the illuminated part of a thick screen behaves like a thin screen and that the shape of the dark part is of no importance.

III

HUYGENS' PRINCIPLE FOR ELECTROMAGNETIC WAVES

§ 1. Huygens' principle and Maxwell's equations

§ 1.1. The formulae of Larmor and Tedone†

KIRCHHOFF's formula gives an analytic representation of Huygens' principle for a field specified by a single scalar potential such as the velocity potential of sound waves in air. But the original and still the most important applications of the principle relate to fields of radiation which, owing to the property of polarization, cannot be represented by a single scalar potential. To deal with radiation, it is in fact necessary to have recourse to the electromagnetic theory of light. Whilst it is true that Kirchhoff's formula can be applied to each of the components of the electric and magnetic vectors, this does not constitute a valid formulation of Huygens' principle as it possesses no physical interpretation. We shall therefore consider the analytical formulation of Huygens' principle for an electromagnetic field in free aether, regarded as a single entity.

Let the electric and magnetic forces be given vectorially by

$$\mathbf{d} = \mathbf{i}d_x+\mathbf{j}d_y+\mathbf{k}d_z,$$
$$\mathbf{h} = \mathbf{i}h_x+\mathbf{j}h_y+\mathbf{k}h_z,$$

it being supposed that $\mathbf{d}$ and $\mathbf{h}$ are measured in electrostatic and electromagnetic units respectively. Then Maxwell's equations of the electromagnetic field in free aether are

$$\left.\begin{aligned} c\,\mathrm{curl}\,\mathbf{h} &= \dot{\mathbf{d}}, & \mathrm{div}\,\mathbf{h} &= 0,\\ c\,\mathrm{curl}\,\mathbf{d} &= -\dot{\mathbf{h}}, & \mathrm{div}\,\mathbf{d} &= 0, \end{aligned}\right\} \tag{1.11}$$

where c is the velocity of light. It follows from these equations that each component of $\mathbf{d}$ and $\mathbf{h}$ satisfies the wave equation

$$\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}.$$

As we have already remarked, each of these six components can be represented by a Kirchhoff integral, depending on the boundary

† We are greatly indebted to Professor E. T. Whittaker for drawing our attention to the importance of the work of Larmor and Tedone, and for allowing us to make free use of his lecture notes in § 1 of this chapter.

values of the component under consideration. But it turns out, as we shall see, that the secondary disturbances obtained in this way are not solutions of Maxwell's equations, and so such a representation of the electromagnetic field is not satisfactory from the physical point of view; each secondary source in Huygens' principle ought to give rise to an electromagnetic wave. We must then consider, not the equation of wave-motions satisfied by each component of **d** and **h**, but Maxwell's equations as a whole.

Consider an electromagnetic field specified by the vectors **d** and **h** which have no singularities in the region D bounded by the closed surface S. We wish to find a surface distribution of electric and magnetic charges, etc., on S which will give rise to the actual electromagnetic field in D and a null field on the other side of S. This way of approaching the problem was first suggested by Larmor.†

At a point P on S we can resolve **d** into a tangential component $\mathbf{d}_t$ and a component $\mathbf{d}_n$ in the direction of the inward normal unit vector **n**; similarly for **h**. Our proposed distribution must be such that the corresponding components of **d** and **h** are discontinuous across S. Now in order to produce a zero field outside and a tangential component $\mathbf{h}_t$ just inside, there must be an electric current flowing at P of strength $h_t/(4\pi)$; vectorially the current in the element dS at P is

$$\mathbf{I} = \frac{1}{4\pi}\mathbf{n}\times\mathbf{h}_t = \frac{1}{4\pi}\mathbf{n}\times\mathbf{h}. \tag{1.12}$$

This current gives rise to a surface distribution of electric charge which proves to be just sufficient to produce the required discontinuity in the normal component of **d**.

The discontinuity in $\mathbf{d}_n$ in crossing S at P will be produced only by the part of S very near to P, and this neighbourhood of P can be regarded as a plane in the usual way. We take Oz to be in the direction of **n**, Oy to be the direction of $\mathbf{h}_t$. Then

$$\mathbf{h} = \mathbf{j}h_y+\mathbf{k}h_z, \quad \mathbf{n} = \mathbf{k}$$

so that

$$\mathbf{I} = -\mathbf{i}h_y/(4\pi).$$

† *Proc. London Math. Soc.* (2), **1** (1903), 1–13. For an interesting but entirely different method of attacking this problem, see Love, *Phil. Trans.* (A), **197** (1901), 1–45. Stratton and Chu (*Phys. Rev.* **56** (1939), 99–107) have proved a vector analogue of Green's Theorem, from which they deduce an analogue of Helmholtz's formula without using scalar and vector potentials or the Hertzian vector. See also Stratton, *Electromagnetic Theory* (New York, 1941), 464.

The increase of charge per unit time on the surface element $dxdy$ is then

$$\dot{\sigma}\, dxdy = \frac{1}{4\pi}\frac{\partial h_y}{\partial x}dxdy = \frac{1}{4\pi c}\dot{d}_z\, dxdy$$

by Maxwell's equations. Hence the charge per unit area is

$$\sigma = \frac{1}{4\pi c}d_z = \frac{1}{4\pi c}\mathbf{d}\cdot\mathbf{n}. \tag{1.13}$$

Thus the effect of the current is to produce a charge $\mathbf{d}\cdot\mathbf{n}/(4\pi c)$ per unit area distributed over S, the charge being measured in electromagnetic units. But a surface charge of density $\mathbf{d}\cdot\mathbf{n}/(4\pi)$ in electrostatic units produces a discontinuity $\mathbf{d}\cdot\mathbf{n}$ in the normal electric force, i.e. it produces the required discontinuity in $\mathbf{d}_n$.

Since Maxwell's equations are invariant under the transformation

$$\mathbf{d}\rightarrow\mathbf{h},\quad \mathbf{h}\rightarrow -\mathbf{d},$$

we require a magnetic current in S of density

$$\mathbf{K} = -\frac{1}{4\pi}\mathbf{n}\times\mathbf{d} \tag{1.14}$$

and a magnetic charge $$\tau = \frac{1}{4\pi c}\mathbf{h}\cdot\mathbf{n} \tag{1.15}$$

per unit area to produce the required discontinuities in the normal component of $\mathbf{h}$ and the tangential component of $\mathbf{d}$.

The field at a point $P_0\,(x_0, y_0, z_0)$ within S, due to the electric current $\mathbf{I}$ and electric charge σ, is given by the equations†

$$\mathbf{d}_1 = -\mathrm{grad}_0\,\Phi - \frac{1}{c}\dot{\mathbf{A}},$$

$$\mathbf{h}_1 = \mathrm{curl}_0\,\mathbf{A},$$

where Φ and $\mathbf{A}$ are the scalar and vector potentials

$$\Phi = \iint\limits_S [c\sigma]\frac{dS}{r} = \frac{1}{4\pi}\iint\limits_S [\mathbf{d}\cdot\mathbf{n}]\frac{dS}{r}, \tag{1.161}$$

$$\mathbf{A} = \iint\limits_S [\mathbf{I}]\frac{dS}{r} = \frac{1}{4\pi}\iint\limits_S [\mathbf{n}\times\mathbf{h}]\frac{dS}{r}, \tag{1.162}$$

† The operators curl_0 and grad_0 refer to variation of the position of the point $P_0\,(x_0, y_0, z_0)$.

where r is the distance from P_0 to a typical point $P(x,y,z)$ of S, and square brackets denote retarded values. (See Ch. I, § 4.5.)

Similarly the field at P_0, due to the magnetic current $\mathbf{K}$ and magnetic charge τ, is given by

$$\mathbf{h}_2 = -\text{grad}_0\Psi - \frac{1}{c}\dot{\mathbf{B}},$$

$$\mathbf{d}_2 = -\text{curl}_0\mathbf{B},$$

where†

$$\Psi = \frac{1}{4\pi}\iint\limits_S [\mathbf{h}\cdot\mathbf{n}]\frac{dS}{r}, \tag{1.171}$$

$$\mathbf{B} = -\frac{1}{4\pi}\iint\limits_S [\mathbf{n}\times\mathbf{d}]\frac{dS}{r}. \tag{1.172}$$

The field due to the electric and magnetic currents and charges distributed over S is

$$\mathbf{d} = \mathbf{d}_1+\mathbf{d}_2, \qquad \mathbf{h} = \mathbf{h}_1+\mathbf{h}_2$$

and so is given by

$$\mathbf{d} = -\text{grad}_0\Phi - \frac{1}{c}\dot{\mathbf{A}} - \text{curl}_0\mathbf{B}, \tag{1.18}$$

$$\mathbf{h} = -\text{grad}_0\Psi - \frac{1}{c}\dot{\mathbf{B}} + \text{curl}_0\mathbf{A}. \tag{1.19}$$

Equations (1.18) and (1.19), where Φ, Ψ, $\mathbf{A}$, and $\mathbf{B}$ are defined by (1.161), (1.162), (1.171), (1.172) as integrals over the surface S, provide the required analytical expressions for Huygens' principle in an electromagnetic field. We shall call these formulae the *Larmor-Tedone formulae.*

It might be expected that, from their mode of derivation, the secondary waves emitted by the surface elements according to the Larmor-Tedone formulae would be electromagnetic waves satisfying Maxwell's equations, but this is however not the case. The Larmor-Tedone formulae suffer from precisely the same defect as Kirchhoff's formula did. We return to this point in § 2.2.

§ 1.2. The complex form of the Larmor-Tedone formulae

If we introduce the complex vector‡

$$\mathbf{q} = \mathbf{d}+i\mathbf{h},$$

† Professor Max Born calls Ψ and $\mathbf{B}$ the scalar and vector antipotentials.

‡ Bateman (*Electrical and Optical Wave Motions* (Cambridge, 1915), p. 4) points out that if we use a complex time factor e^{-ikct} with a complex electromagnetic vector, it is necessary to take $\mathbf{q} = \mathbf{d}\pm i\mathbf{h}$ in order to avoid confusion in equating real and imaginary parts. A simpler way would be to keep $\mathbf{q} = \mathbf{d}+i\mathbf{h}$ and to follow engineering practice in writing the time factor as e^{-jkct}.

Maxwell's equations take the simple form

$$c\,\mathrm{curl}\,\mathbf{q} = i\dot{\mathbf{q}}, \qquad \mathrm{div}\,\mathbf{q} = 0. \tag{1.21}$$

The use of this vector enables us to write the Larmor-Tedone formulae very compactly; for, if we write

$$\Omega = \Phi + i\Psi, \qquad \mathbf{C} = \mathbf{A} + i\mathbf{B},$$

we have the following theorem:

Let S be a closed surface within and on which the complex electromagnetic vector $\mathbf{q}(x, y, z, t)$ and its first partial derivatives are continuous. Let r be the distance from $P_0(x_0, y_0, z_0)$ to a typical point P of S and let $\mathbf{n}$ be the unit normal vector at P, drawn inwards. Then, if

$$\mathbf{Q} = -\mathrm{grad}_0\,\Omega - \frac{1}{c}\dot{\mathbf{C}} + i\,\mathrm{curl}_0\,\mathbf{C}, \tag{1.22}$$

where

$$4\pi\Omega = \iint\limits_S [\mathbf{q}\cdot\mathbf{n}]\frac{dS}{r}, \tag{1.23}$$

$$4\pi i\mathbf{C} = \iint\limits_S [\mathbf{n}\times\mathbf{q}]\frac{dS}{r}, \tag{1.24}$$

the value of $\mathbf{Q}$ is $\mathbf{q}(x_0, y_0, z_0, t)$ or zero according as P_0 lies inside or outside S.

§ 1.3. Tedone's proof of the Larmor-Tedone formulae

The way in which the analytical formulation of Huygens' principle for an electromagnetic field was obtained in § 1.1, though valuable as affording insight into the physical meaning of the expressions which occur, can hardly be regarded as sufficient from the logical point of view. The following analytical proof is based on one due to Tedone.†

Let the direction cosines of $\mathbf{n}$ be (λ, μ, ν). We shall denote by $\partial/\partial x$ a partial differentiation with respect to x as it occurs explicitly, ignoring the fact that r depends on x, and by d/dx a partial differentiation with respect to x taking into account the dependence of r on x.

We shall suppose that $\mathbf{d}$ and $\mathbf{h}$ and their first partial derivatives

† *Rendiconti dei Lincei*, **26** (1917), 286–9.

are continuous in the volume D bounded by S, and, in the first instance, that P_0 lies outside S. Then, by Green's transformation, we have

$$\iint_S \nu[q_y]\frac{dS}{r} = -\iiint_D \frac{d}{dz}\left\{\frac{[q_y]}{r}\right\} dxdydz$$

$$= -\iiint_D \frac{1}{r}\frac{\partial[q_y]}{\partial z} dxdydz - \iiint_D \left\{[q_y]\frac{\partial}{\partial z}\left(\frac{1}{r}\right) + \frac{1}{r}\frac{\partial r}{\partial z}\frac{\partial}{\partial r}[q_y]\right\} dxdydz$$

$$= -\iiint_D \frac{1}{r}\frac{\partial[q_y]}{\partial z} dxdydz + \frac{\partial}{\partial z_0}\iiint_D \frac{[q_y]}{r} dxdydz,$$

since $$r^2 = (x-x_0)^2+(y-y_0)^2+(z-z_0)^2.$$

The x component of the vector $\mathbf{C}$ defined by (1.24) is given by

$$4\pi i C_x = \iint_S \{-\nu[q_y]+\mu[q_z]\}\frac{dS}{r}$$

$$= -\frac{\partial}{\partial z_0}\iiint_D [q_y]\frac{dxdydz}{r} + \frac{\partial}{\partial y_0}\iiint_D [q_z]\frac{dxdydz}{r} - \iiint_D \left[\frac{\partial q_z}{\partial y} - \frac{\partial q_y}{\partial z}\right]\frac{dxdydz}{r}$$

$$= -\frac{\partial}{\partial z_0}\iiint_D [q_y]\frac{dxdydz}{r} + \frac{\partial}{\partial y_0}\iiint_D [q_z]\frac{dxdydz}{r} - \frac{i}{c}\iiint_D [\dot{q}_x]\frac{dxdydz}{r},$$

after using the complex form of Maxwell's equations. Hence

$$4\pi i \mathbf{C} = \mathrm{curl}_0\,\mathbf{F} - \frac{i}{c}\dot{\mathbf{F}}, \tag{1.31}$$

where $$\mathbf{F} = \iiint_D [\mathbf{q}]\frac{dxdydz}{r}. \tag{1.32}$$

By a similar argument

$$4\pi\Omega = \iint_S \{\lambda[q_x]+\mu[q_y]+\nu[q_z]\}\frac{dS}{r}$$

$$= \mathrm{div}_0 \iiint_D [\mathbf{q}]\frac{dxdydz}{r} - \iiint_D [\mathrm{div}\,\mathbf{q}]\frac{dxdydz}{r}.$$

The second term on the right-hand side vanishes by Maxwell's equations, and so

$$4\pi\Omega = \operatorname{div}_0 \mathbf{F}. \tag{1.33}$$

The equations (1.31), (1.32), (1.33) also hold when P_0 is a point of D, but the proof just given has to be modified since $1/r$ becomes infinite at P_0. In this case we apply Green's transformation to the volume D' bounded externally by S and internally by a small sphere Σ with centre P_0 and radius ϵ. We then obtain

$$4\pi i(\mathbf{C}+\mathbf{C}') = \operatorname{curl}_0 \mathbf{F}' - \frac{i}{c}\dot{\mathbf{F}}',$$

$$4\pi(\Omega+\Omega') = \operatorname{div}_0 \mathbf{F}',$$

where

$$4\pi i\mathbf{C}' = \iint\limits_{\Sigma} [\mathbf{n}\times\mathbf{q}]\frac{dS}{r},$$

$$4\pi\Omega' = \iint\limits_{\Sigma} [\mathbf{q}\cdot\mathbf{n}]\frac{dS}{r},$$

$$\mathbf{F}' = \iiint\limits_{D'} [\mathbf{q}]\frac{dxdydz}{r}.$$

The equations (1.31), (1.32), (1.33) follow at once, since

$$\mathbf{C}' \to 0, \quad \Omega' \to 0, \quad \mathbf{F}' \to \mathbf{F}$$

as $\epsilon \to 0$.

We now consider the vector

$$\mathbf{Q} = -\operatorname{grad}_0\Omega - \frac{1}{c}\dot{\mathbf{C}} + i\operatorname{curl}_0\mathbf{C} \tag{1.34}$$

which appears in the Larmor-Tedone formula. By (1.31) and (1.33) we have

$$4\pi\mathbf{Q} = -\operatorname{grad}_0\operatorname{div}_0\mathbf{F} + \frac{i}{c}\operatorname{curl}_0\dot{\mathbf{F}} + \frac{1}{c^2}\ddot{\mathbf{F}} + \operatorname{curl}_0\operatorname{curl}_0\mathbf{F} - \frac{i}{c}\operatorname{curl}_0\dot{\mathbf{F}}$$

$$= -\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathbf{F},$$

since

$$\operatorname{curl}\operatorname{curl}\mathbf{X} = \operatorname{grad}\operatorname{div}\mathbf{X} - \nabla^2\mathbf{X}.$$

Hence

$$\mathbf{Q} = -\frac{1}{4\pi}\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\iiint\limits_{D}[\mathbf{q}]\frac{dxdydz}{r}. \tag{1.35}$$

But by the well-known properties of retarded potentials the expression on the right-hand side of (1.35) has the value $\mathbf{q}(x_0, y_0, z_0, t)$ or zero according as P_0 is or is not a point of D. This completes the analytical proof of the Larmor-Tedone formula.

§ 1.4. A modification of the Larmor-Tedone formulae

The case when all the singularities of the electromagnetic field lie inside the closed surface S may be discussed by applying to the result of § 1.2 an argument which will by now be familiar (cf. Chapter I, § 5.1). The result is as follows:

Let the complex electromagnetic vector $\mathbf{q}$ *and its first partial derivatives be continuous on and outside the closed surface* S, *and let* $r\mathbf{q} \to 0$ *as* $r \to \infty$. *Let* $\mathbf{n}$ *be the unit vector, normal to* S, *drawn inwards. Then if*

$$\mathbf{Q} = -\mathrm{grad}_0 \Omega - \frac{1}{c}\dot{\mathbf{C}} + i\,\mathrm{curl}_0 \mathbf{C},$$

where $$4\pi\Omega = \iint\limits_S [\mathbf{q}\cdot\mathbf{n}]\frac{dS}{r}, \qquad 4\pi i\mathbf{C} = \iint\limits_S [\mathbf{n}\times\mathbf{q}]\frac{dS}{r},$$

the value of $\mathbf{Q}$ *is* $-\mathbf{q}(x_0, y_0, z_0, t)$ *or zero according as* (x_0, y_0, z_0) *lies outside or inside* S.

§ 2. The failure of the Kirchhoff and the Larmor-Tedone formulae

§ 2.1. The connexion between the Larmor-Tedone formulae and the Kirchhoff formula

The Larmor-Tedone formula for Q_x is

$$4\pi Q_x = -\frac{\partial}{\partial x_0}\iint\limits_S \{\lambda[q_x]+\mu[q_y]+\nu[q_z]\}\frac{dS}{r} + \frac{i}{c}\frac{\partial}{\partial t}\iint\limits_S \{\mu[q_z]-\nu[q_y]\}\frac{dS}{r} + $$
$$+\frac{\partial}{\partial y_0}\iint\limits_S \{\lambda[q_y]-\mu[q_x]\}\frac{dS}{r} - \frac{\partial}{\partial z_0}\iint\limits_S \{\nu[q_x]-\lambda[q_z]\}\frac{dS}{r}.$$

For definiteness we consider here the case when $\mathbf{q}$ is regular inside and on S, so that (λ, μ, ν) are the direction cosines of the inward normal. We shall write this formula as

$$4\pi Q_x = I_1 + I_2 + I_3,$$

where I_1, I_2, I_3 denote the parts of $4\pi Q_x$ involving λ, μ, or ν respectively. Then

$$I_1 = -\frac{\partial}{\partial x_0}\iint\limits_S \lambda[q_x]\frac{dS}{r} + \frac{\partial}{\partial y_0}\iint\limits_S \lambda[q_y]\frac{dS}{r} + \frac{\partial}{\partial z_0}\iint\limits_S \lambda[q_z]\frac{dS}{r}. \tag{2.11}$$

Let $\partial/\partial x$ denote partial differentiation with respect to x ignoring the fact that r depends on x, and let d/dx denote partial differentiation when we take account of the dependence of r on x. Evidently

$$\frac{d}{dx}[q_x] = \left[\frac{\partial q_x}{\partial x}\right] - \frac{1}{c}\frac{dr}{dx}[\dot{q}_x],$$

and so
$$\frac{d}{dx}\left[\frac{q_x}{r}\right] = \frac{1}{r}\left[\frac{\partial q_x}{\partial x}\right] - \frac{1}{cr}\frac{dr}{dx}[\dot{q}_x] + [q_x]\frac{d}{dx}\left(\frac{1}{r}\right). \tag{2.12}$$

Also
$$\frac{\partial}{\partial x_0}[q_x] = -\frac{1}{c}\frac{\partial r}{\partial x_0}[\dot{q}_x] = \frac{1}{c}\frac{dr}{dx}[\dot{q}_x],$$

and so
$$\frac{\partial}{\partial x_0}\left[\frac{q_x}{r}\right] = \frac{1}{cr}\frac{dr}{dx}[\dot{q}_x] - [q_x]\frac{d}{dx}\left(\frac{1}{r}\right). \tag{2.13}$$

From (2.12) and (2.13) we have

$$\left(\frac{d}{dx} + \frac{\partial}{\partial x_0}\right)\left[\frac{q_x}{r}\right] = \frac{1}{r}\left[\frac{\partial q_x}{\partial x}\right]. \tag{2.14}$$

Similar formulae hold for differentiation with respect to the other variables.

Applying (2.13) to the first term in I_1 and the analogues of (2.14) to the second and third terms, we obtain

$$I_1 = \iint\limits_S \lambda\left\{-\frac{1}{cr}\frac{dr}{dx}[\dot{q}_x] + [q_x]\frac{d}{dx}\left(\frac{1}{r}\right) + \frac{1}{r}\left[\frac{\partial q_y}{\partial y}\right] - \frac{d}{dy}\left[\frac{q_y}{r}\right] + \right.$$
$$\left. + \frac{1}{r}\left[\frac{\partial q_z}{\partial z}\right] - \frac{d}{dz}\left[\frac{q_z}{r}\right]\right\}\, dS.$$

But since $\dfrac{\partial q_y}{\partial y} + \dfrac{\partial q_z}{\partial z} = \operatorname{div}\mathbf{q} - \dfrac{\partial q_x}{\partial x}$, and $\operatorname{div}\mathbf{q} = 0$,

this reduces to

$$I_1 = \iint\limits_S \lambda\left\{-\frac{1}{cr}\frac{dr}{dx}[\dot{q}_x] + [q_x]\frac{d}{dx}\left(\frac{1}{r}\right) - \frac{1}{r}\left[\frac{\partial q_x}{\partial x}\right] - \frac{d}{dy}\left[\frac{q_y}{r}\right] - \frac{d}{dz}\left[\frac{q_z}{r}\right]\right\}\, dS. \tag{2.15}$$

Applying the same argument to

$$I_2 = -\frac{\partial}{\partial x_0}\iint\limits_S \mu[q_y]\frac{dS}{r} + \frac{i}{c}\frac{\partial}{\partial t}\iint\limits_S \mu[q_z]\frac{dS}{r} - \frac{\partial}{\partial y_0}\iint\limits_S \mu[q_x]\frac{dS}{r},$$

we obtain

$$I_2 = \iint\limits_S \mu\left\{\frac{d}{dx}\left[\frac{q_y}{r}\right]-\frac{1}{r}\left[\frac{\partial q_y}{\partial x}\right]+\frac{i}{c}[\dot{q}_z]\frac{1}{r}-\frac{1}{cr}\frac{dr}{dy}[\dot{q}_x]+[q_x]\frac{d}{dy}\left(\frac{1}{r}\right)\right\}\,dS.$$

But since $$i\dot{q}_z = c\left\{\frac{\partial q_y}{\partial x}-\frac{\partial q_x}{\partial y}\right\},$$
this becomes

$$I_2 = \iint\limits_S \mu\left\{\frac{d}{dx}\left[\frac{q_y}{r}\right]-\frac{1}{r}\left[\frac{\partial q_x}{\partial y}\right]-\frac{1}{cr}\frac{dr}{dy}[\dot{q}_x]+[q_x]\frac{d}{dy}\left(\frac{1}{r}\right)\right\}\,dS. \quad (2.16)$$

In the same way, we may prove that

$$I_3 = \iint\limits_S \nu\left\{\frac{d}{dx}\left[\frac{q_z}{r}\right]-\frac{1}{r}\left[\frac{\partial q_x}{\partial z}\right]-\frac{1}{cr}\frac{dr}{dz}[\dot{q}_x]+[q_x]\frac{d}{dz}\left(\frac{1}{r}\right)\right\}\,dS. \quad (2.17)$$

If we add equations (2.15), (2.16), (2.17) and denote, as usual, differentiation along $\mathbf{n}$ by $\partial/\partial n$, we find that

$$4\pi Q_x = \iint\limits_S \left\{-\frac{1}{r}\left[\frac{\partial q_x}{\partial n}\right]-\frac{1}{cr}\frac{\partial r}{\partial n}[\dot{q}_x]+[q_x]\frac{\partial}{\partial n}\left(\frac{1}{r}\right)\right\}\,dS+$$
$$+\iint\limits_S \left\{\lambda\left(-\frac{d}{dy}\left[\frac{q_y}{r}\right]-\frac{d}{dz}\left[\frac{q_z}{r}\right]\right)+\mu\frac{d}{dx}\left[\frac{q_y}{r}\right]+\nu\frac{d}{dx}\left[\frac{q_z}{r}\right]\right\}\,dS.$$

The second term on the right-hand side vanishes by Green's theorem, and so we have reduced the Larmor-Tedone formula for Q_x to†

$$4\pi Q_x = \iint\limits_S \left\{-\frac{1}{r}\left[\frac{\partial q_x}{\partial n}\right]-\frac{1}{cr}\frac{\partial r}{\partial n}[\dot{q}_x]+[q_x]\frac{\partial}{\partial n}\left(\frac{1}{r}\right)\right\}\,dS.$$

But this is precisely Kirchhoff's integral for the wave-function q_x; it represents $4\pi q_x(x_0, y_0, z_0, t)$ or zero according as P lies inside or outside S. Vectorially, we may write this transformation in the form

$$4\pi\mathbf{Q} = \iint\limits_S \left\{-\frac{1}{r}\left[\frac{\partial \mathbf{q}}{\partial n}\right]-\frac{1}{cr}\frac{\partial r}{\partial n}[\dot{\mathbf{q}}]+[\mathbf{q}]\frac{\partial}{\partial n}\left(\frac{1}{r}\right)\right\}\,dS+$$
$$+\iint\limits_S (\mathbf{n}\times\nabla)\times\left[\frac{\mathbf{q}}{r}\right]dS.$$

† A shorter 'proof' could be given by considering first the monochromatic case and then using Fourier's Theorem. Cf. Ch. I, § 5.1 on p. 36.

The first term is Kirchhoff's integral representing the vector $4\pi\mathbf{q}$ or zero, whilst the second term vanishes since S is a closed surface. This result is due to E. T. Whittaker.

§ 2.2. The inadequacy of the Kirchhoff and Larmor-Tedone formulae for dealing with diffraction problems

We have just seen that the formulae of Kirchhoff and of Larmor and Tedone are equivalent, in that both express the components of the electric and magnetic vectors as integrals over a closed surface; the two formulae differ by a quantity which vanishes in virtue of Green's theorem. In the application to diffraction problems, integration is not over a closed surface but over a cap bridging the gap in the diffracting body; an application of Stokes's theorem shows that in this case the two formulae give different results. The question arises, 'Which is the correct formula to apply to diffraction problems?' It turns out that neither is suitable; for the vectors **d** and **h** given by either formula when S is not a closed surface do not satisfy Maxwell's equations.

We demonstrate this by considering the special case when S is the half-plane $x = 0$, $y \geqslant 0$; and the given field is that of a plane wave, viz.

$$d_x = 0, \qquad d_y = 0, \qquad d_z = e^{-ik(x+ct)},$$
$$h_x = 0, \qquad h_y = e^{-ik(x+ct)}, \qquad h_z = 0.$$

This is the problem of the diffraction of plane monochromatic light incident normally on the 'black' half-plane $x = 0$, $y < 0$.

In the first instance we use Helmholtz's formula, the form of Kirchhoff's integral suitable for dealing with monochromatic phenomena. The resulting field turns out to be

$$d_x = 0, \qquad d_y = 0, \qquad d_z = U,$$
$$h_x = 0, \qquad h_y = U, \qquad h_z = 0,$$

where
$$U = u^* + u^B$$

in the notation of Chapter II, equation (3.17). Now U involves t only in the presence of the time factor e^{-ikct}; hence, by Maxwell's equations, U should be of the form $Ae^{-ik(x+ct)}$, where A is a constant. The term u^* is of this form, but u^B is not. Hence the field given by Kirchhoff's integral does not satisfy Maxwell's equations.

In applying the Larmor-Tedone formulae, we observe that, by § 2.1, the field obtained is of the form

$$D_x = D_x^K + \frac{1}{4\pi}\iint\limits_S \left\{\lambda\left(-\frac{d}{dy}\left[\frac{d_y}{r}\right]-\frac{d}{dz}\left[\frac{d_z}{r}\right]\right)+\mu\frac{d}{dx}\left[\frac{d_y}{r}\right]+\nu\frac{d}{dx}\left[\frac{d_z}{r}\right]\right\}\,dS,$$

etc., where the superscript K indicates the field given by Kirchhoff's formula. Using Stokes's theorem, we have

$$D_x = D_x^K + \frac{1}{4\pi}\int\limits_\Gamma \left\{\left[\frac{d_z}{r}\right]dy - \left[\frac{d_y}{r}\right]dz\right\},$$

etc., where Γ is the rim of S, i.e. the axis of z described in the positive sense. In our case the formulae reduce to

$$D_x = D_x^K, \qquad H_x = H_x^K - \frac{1}{4\pi}\int\limits_\Gamma \left[\frac{h_y}{r}\right]dz,$$
$$D_y = D_y^K, \qquad H_y = H_y^K,$$
$$D_z = D_z^K, \qquad H_z = H_z^K.$$

The field given by Kirchhoff's formula we have already found; it follows that the Larmor-Tedone formulae give

$$D_x = 0, \qquad D_y = 0, \qquad D_z = u^* + u^B,$$
$$H_x = v, \qquad H_y = u^* + u^B, \qquad H_z = 0,$$

where $$v = -\frac{1}{4\pi}\int\limits_{-\infty}^{\infty} e^{ik(r-ct)}\frac{dz}{r} = -\frac{1}{4\pi}\int\limits_{-\infty}^{\infty} e^{ik\rho\cosh\tau - ikct}\,d\tau$$

by the transformation $z = z_1 + \rho\sinh\tau$ of Chapter II, § 3.1. Using the result given by Watson in his *Bessel Functions*, p. 180 (10), we get

$$v = -\frac{i}{4}H_0^{(1)}(k\rho)e^{-ikct}.$$

This field can be obtained by adding to the field of geometrical optics, namely

$$\mathbf{d} = \mathbf{k}u^*, \qquad \mathbf{h} = \mathbf{j}u^*$$

which satisfies Maxwell's equations, the field specified by

$$\mathbf{d} = \mathbf{k}u^B, \qquad \mathbf{h} = \mathbf{i}v + \mathbf{j}u^B.$$

To prove that the latter field does not satisfy Maxwell's equations, it suffices to consider the two equations

$$\frac{\partial d_x}{\partial z} - \frac{\partial d_z}{\partial x} = -\frac{1}{c}\frac{\partial h_y}{\partial t}, \qquad \frac{\partial h_y}{\partial x} - \frac{\partial h_x}{\partial y} = \frac{1}{c}\frac{\partial d_z}{\partial t}.$$

Since u^B and v involve t only in the factor e^{-ikct}, these equations reduce to

$$-\frac{\partial u^B}{\partial x} = iku^B, \qquad \frac{\partial u^B}{\partial x} - \frac{\partial v}{\partial y} = -iku^B;$$

but these imply that $\partial v/\partial y = 0$, which is certainly not the case.

Thus neither the Kirchhoff formula nor the Larmor-Tedone formulae, applied to the components of an electromagnetic field with an unclosed surface of integration, give solutions of Maxwell's equations.

§ 3. Kottler's formulation of Huygens' principle for electromagnetic waves

It is not surprising that Kirchhoff's formula, applied to each of the components of **d** and **h**, should not give vectors satisfying Maxwell's equations when S is an unclosed surface. For the secondary waves emitted by each element of S are not electromagnetic waves; an electromagnetic wave is obtained only by the superposition of the secondary waves due to all the elements of a closed surface.

The reason why the Larmor-Tedone formulae also fail is not quite so obvious; the example of § 2.2 shows that the secondary waves emitted by each surface element are not electromagnetic waves, though one would expect them to be in virtue of the physical argument by which the formulae were derived.

We recall that, according to Larmor, each element of S carries

(i) an electric current-sheet of density

$$\mathbf{I} = \frac{1}{4\pi}\mathbf{n}\times\mathbf{h},$$

(ii) an electric surface charge of density

$$\sigma = \frac{1}{4\pi c}\mathbf{d}\cdot\mathbf{n},$$

(iii) a magnetic current-sheet of density

$$\mathbf{K} = -\frac{1}{4\pi}\mathbf{n}\times\mathbf{d},$$

(iv) a magnetic surface charge of density

$$\tau = \frac{1}{4\pi c}\mathbf{h}\cdot\mathbf{n},$$

all in electromagnetic units.

Kottler† has pointed out that, *if we wish each element of a surface S to emit electromagnetic waves, we must consider, not only the electric and magnetic charges and currents on dS, but also, possibly, electric and magnetic line charges on the boundary of dS.* The effects of these line charges cancel out when S is a closed surface; but if S is not closed, additional terms due to the line charges on the rim of S are added to the Larmor-Tedone formula.

Let us consider then an unclosed surface S with rim Γ. Unless the vector $\mathbf{I}$ at every point of Γ is tangential to Γ, each element ds of Γ must carry a charge which varies with the time in such a way as to provide the charge carried away by the component of the current normal to ds.

At any point O of Γ, $\mathbf{n}$, the unit vector normal to S, and $\mathbf{ds}$, the elementary tangent vector to Γ, are at right angles. Choose axes so that $\mathbf{ds}$ is along Ox, $\mathbf{n}$ along Oz. The current-density at O has components

$$I_x = -\frac{1}{4\pi}h_y, \qquad I_y = \frac{1}{4\pi}h_x;$$

the former is tangential to Γ at O, the latter normal to Γ. The charge carried away from dx in time t is then

$$\frac{1}{4\pi}dx\int^t h_x\,dt.$$

Hence if the line-density of electric charge on dx is Σ in electromagnetic units, we have

$$\Sigma\,dx = -\frac{1}{4\pi}dx\int^t h_x\,dt,$$

or, vectorially,

$$\Sigma\,ds = -\frac{1}{4\pi}\mathbf{ds}\cdot\int^t \mathbf{h}\,dt.$$

But as O was any point of Γ, this formula gives the line-density of electric charge at every point of Γ. Similarly we can show that the line-density of magnetic charge is Θ in electromagnetic units, where

$$\Theta\,ds = \frac{1}{4\pi}\mathbf{ds}\cdot\int^t \mathbf{d}\,dt.$$

† *Annalen der Phys.* **71** (1923), 457–508. This paper contains an interesting critical discussion of the various attempts to provide a rigorous electromagnetic theory of diffraction by a black screen of arbitrary form. See also Stratton and Chu, loc. cit.

Naturally these formulae hold no matter how small the surface S may be. In particular, if we wish each element of a surface to emit electromagnetic waves satisfying Maxwell's equations, we must suppose it to carry, not only electric and magnetic current-sheets and surface charges, but also line charges of electricity and magnetism on its boundary. The effects of these line charges cancel when we consider the resultant of the electromagnetic waves due to all the elements of a closed surface, but they must be retained if we wish the secondary sources to behave like real sources of electromagnetic disturbances. Moreover, in the theory of diffraction by a black screen, we are concerned with integration over an unclosed surface, and the line charges on the boundaries of the surface elements are then of importance, since they give rise to a non-vanishing integral along the rim of the surface.

The introduction of line charges on the boundary of each surface element necessitates a modification of the formulae of § 1.1. The field due to the electric current-sheet and the electric charges is now

$$\mathbf{d}_1 = -\text{grad}_0\,\Phi - \frac{1}{c}\dot{\mathbf{A}}, \qquad \mathbf{h}_1 = \text{curl}_0\,\mathbf{A},$$

where†

$$\Phi = \frac{1}{4\pi}\iint\limits_S [\mathbf{d}\cdot\mathbf{n}]\frac{dS}{r} - \frac{c}{4\pi}\int\limits_\Gamma \int^{t-r/c} \mathbf{h}\cdot d\mathbf{s}\,\frac{dt}{r}, \tag{3.01}$$

$$\mathbf{A} = \frac{1}{4\pi}\iint\limits_S [\mathbf{n}\times\mathbf{h}]\frac{dS}{r}. \tag{3.02}$$

Similarly, the field due to the magnetic current-sheet and magnetic charges is

$$\mathbf{d}_2 = -\text{curl}_0\,\mathbf{B}, \qquad \mathbf{h}_2 = -\text{grad}_0\,\Psi - \frac{1}{c}\dot{\mathbf{B}},$$

where

$$\Psi = \frac{1}{4\pi}\iint\limits_S [\mathbf{h}\cdot\mathbf{n}]\frac{dS}{r} + \frac{c}{4\pi}\int\limits_\Gamma \int^{t-r/c} \mathbf{d}\cdot d\mathbf{s}\,\frac{dt}{r}, \tag{3.03}$$

$$\mathbf{B} = -\frac{1}{4\pi}\iint\limits_S [\mathbf{n}\times\mathbf{d}]\frac{dS}{r}. \tag{3.04}$$

† The formula for Φ can also be deduced from the relation $\dot{\Phi} = -c\,\text{div}_0\,\mathbf{A}$. Similarly for Ψ below.

The total field is then given by

$$\mathbf{d} = -\text{grad}_0\Phi - \frac{1}{c}\dot{\mathbf{A}} - \text{curl}_0\mathbf{B}, \tag{3.05}$$

$$\mathbf{h} = -\text{grad}_0\Psi - \frac{1}{c}\dot{\mathbf{B}} + \text{curl}_0\mathbf{A}. \tag{3.06}$$

We shall refer to the equations (3.01)–(3.06) as *Kottler's formulae.*

We can most simply describe the field specified by Kottler's formulae as being the field obtained by adding to that of Larmor and Tedone the terms due to the effect of the line charges on Γ, viz.

$$\mathbf{d} = \frac{c}{4\pi}\text{grad}_0 \int_\Gamma \int^{t-r/c} \mathbf{h}\cdot d\mathbf{s}\frac{dt}{r},$$

$$\mathbf{h} = -\frac{c}{4\pi}\text{grad}_0 \int_\Gamma \int^{t-r/c} \mathbf{d}\cdot d\mathbf{s}\frac{dt}{r}.$$

These extra terms vanish when S is a closed surface.

Example. Prove that the complex form of Kottler's formula is

$$\mathbf{q} = -\text{grad}_0\Omega - \frac{1}{c}\dot{\mathbf{C}} + i\,\text{curl}_0\mathbf{C},$$

where

$$4\pi\Omega = \iint_S [\mathbf{q}\cdot\mathbf{n}]\frac{dS}{r} + ic \int_\Gamma \int^{t-r/c} \mathbf{q}\cdot d\mathbf{s}\frac{dt}{r},$$

$$4\pi i\mathbf{C} = \iint_S [\mathbf{n}\times\mathbf{q}]\frac{dS}{r}.$$

§ 4. The diffraction of electromagnetic waves by a black screen

§ 4.1. Kottler's definition of a black screen

We have seen that the problem of the diffraction by a black screen of waves, characterized by a single scalar potential, can be regarded either as a boundary value problem or as a saltus problem. Kirchhoff's formula solves the saltus problem accurately in the case of a thin screen, but gives only a first approximation to the solution of the boundary-value problem.

The same difficulty arises in the case of electromagnetic waves, and in a somewhat more acute form, since it is very difficult to formulate the electromagnetic properties of a black screen. One way out of the difficulty is to regard the screen as perfectly reflecting, and then neglect the effect of the reflected wave (cf. Chapter IV,

§ 3.1). Alternatively, we can define 'blackness' in terms of the discontinuity across the screen of the vectors **d** and **h**. This idea has been developed by Kottler (loc. cit.).

Let us suppose that we have a point-source of light at the point L, and that this light is diffracted by a thin black screen in the form of a surface S with rim Γ. We suppose that one face, S_+ say, of the screen is illuminated, and that the other, S_-, is dark. We denote by $\mathbf{d}_0$ and $\mathbf{h}_0$ the electric and magnetic forces when the screen is absent. We have to find solutions **d** and **h** of Maxwell's equations, with the following properties:

(i) The vectors $\mathbf{d}-\mathbf{d}_0$ and $\mathbf{h}-\mathbf{h}_0$ are regular at L.

(ii) On S, $[\mathbf{d}]_-^+ = \mathbf{d}_0$, $[\mathbf{h}]_-^+ = \mathbf{h}_0$.

(iii) At infinity, **d** and **h** vanish to an appropriate order.

We regard S_+ and S_- as constituting a single degenerate closed surface S', and apply the complex form of Kottler's modification of the Larmor-Tedone formulae to the volume bounded externally by a sphere Σ of large radius R and internally by S' and a small sphere Σ_0 with centre L. Remembering that each surface element carries line charges on its boundary, we obtain the result

$$\mathbf{q}(P) = -\operatorname{grad}\Omega - \frac{1}{c}\dot{\mathbf{C}} + i\operatorname{curl}\mathbf{C}, \tag{4.11}$$

where

$$4\pi\Omega = \iint\limits_{\Sigma} [\mathbf{q}\cdot\mathbf{n}]\frac{dS}{r} + \iint\limits_{\Sigma_0} [\mathbf{q}\cdot\mathbf{n}]\frac{dS}{r} + \iint\limits_{S} [\mathbf{q}_0\cdot\mathbf{n}]\frac{dS}{r} + ic\int\limits_{\Gamma}\int^{t-r/c} \mathbf{q}_0\cdot d\mathbf{s}\,\frac{dt}{r},$$

$$4\pi i C = \iint\limits_{\Sigma} [\mathbf{n}\times\mathbf{q}]\frac{dS}{r} + \iint\limits_{\Sigma_0} [\mathbf{n}\times\mathbf{q}]\frac{dS}{r} + \iint\limits_{S} [\mathbf{n}\times\mathbf{q}_0]\frac{dS}{r},$$

where **n** on S means $\mathbf{n}_+$. (See Fig. 16.) Moreover, we can omit the terms depending on Σ, since they tend to zero as $R \to \infty$, by hypothesis (iii). An application of the Larmor-Tedone formula to the closed surface Σ_0 gives

$$\mathbf{q}_0(P) = -\operatorname{grad}\Omega' - \frac{1}{c}\dot{\mathbf{C}}' + i\operatorname{curl}\mathbf{C}', \tag{4.12}$$

where

$$4\pi\Omega' = \iint\limits_{\Sigma_0} [\mathbf{q}_0\cdot\mathbf{n}]\frac{dS}{r}, \qquad 4\pi i\mathbf{C}' = \iint\limits_{\Sigma_0} [\mathbf{n}\times\mathbf{q}_0]\frac{dS}{r}.$$

Hence, by (4.11) and (4.12), we have

$$\mathbf{q}(P) = \mathbf{q}_0(P) - \operatorname{grad}(\Omega - \Omega') - \frac{1}{c}(\dot{\mathbf{C}} - \dot{\mathbf{C}}') + i \operatorname{curl}(\mathbf{C} - \mathbf{C}'). \quad (4.13)$$

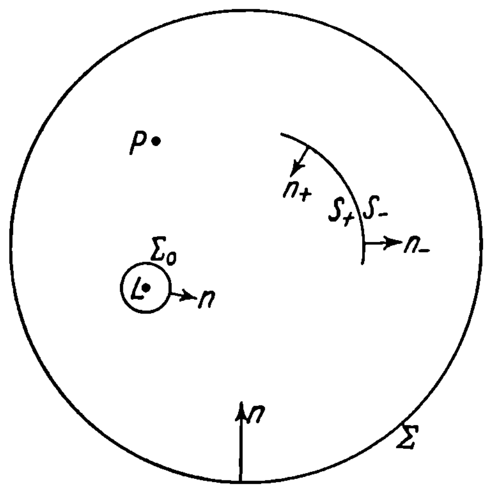


FIG. 16

Now, by hypothesis (i), $\mathbf{q} - \mathbf{q}_0$ is regular at L. Hence if we make the radius of Σ_0 tend to zero, we obtain from (4.13) the equation

$$\mathbf{q}(P) = \mathbf{q}_0(P) - \operatorname{grad}\Omega_0 - \frac{1}{c}\dot{\mathbf{C}}_0 + i \operatorname{curl}\mathbf{C}_0, \quad (4.14)$$

where

$$4\pi\Omega_0 = \iint_S [\mathbf{q}_0 \cdot \mathbf{n}] \frac{dS}{r} + ic \int_\Gamma \int^{t-r/c} \mathbf{q}_0 \cdot d\mathbf{s} \frac{dt}{r}, \quad (4.15)$$

$$4\pi i \mathbf{C}_0 = \iint_S [\mathbf{n} \times \mathbf{q}_0] \frac{dS}{r}; \quad (4.16)$$

the normal $\mathbf{n}$ is drawn from the dark to the illuminated side of S and the direction of description of Γ is related to $\mathbf{n}$ by the right-hand screw law. Equations (4.14), (4.15), (4.16) solve the problem of diffraction by a 'black' screen.

It is often more convenient to use integrals, not over the diffraction screen, but over surfaces bridging the gaps in the screen. We discuss here the simplest case, when Γ is a simple closed curve. We construct a surface S_1 with Γ as rim, such that $S+S_1$ is a closed surface which encloses L but not P. The direction of the normal vector $\mathbf{n}$ on S_1 is chosen so that $\mathbf{n}$ and the direction of description of

Γ are connected by the right-hand screw law; this means that $\mathbf{n}$ is drawn from the 'illuminated' to the 'dark' side of S_1. (See Fig. 17.)

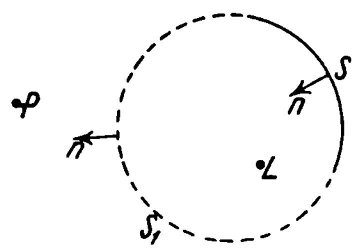


FIG. 17

If we apply the Larmor-Tedone formula to the closed surface $S+S_1$, bearing in mind the difference in the directions of $\mathbf{n}$ on S and S_1, we obtain

$$\mathbf{q}_0(P) = \operatorname{grad}\Omega_1+\frac{1}{c}\dot{\mathbf{C}}_1-i\operatorname{curl}\mathbf{C}_1,$$

where

$$4\pi\Omega_1 = \iint\limits_S [\mathbf{q}_0\cdot\mathbf{n}]\frac{dS}{r}-\iint\limits_{S_1} [\mathbf{q}_0\cdot\mathbf{n}]\frac{dS}{r},$$

$$4\pi i\mathbf{C}_1 = \iint\limits_S [\mathbf{n}\times\mathbf{q}_0]\frac{dS}{r}-\iint\limits_{S_1} [\mathbf{n}\times\mathbf{q}_0]\frac{dS}{r}.$$

Substituting for $\mathbf{q}_0(P)$ in (4.14), we obtain

$$\mathbf{q}(P) = -\operatorname{grad}\Omega-\frac{1}{c}\dot{\mathbf{C}}+i\operatorname{curl}\mathbf{C}, \tag{4.17}$$

where

$$4\pi\Omega = \iint\limits_{S_1} [\mathbf{q}_0\cdot\mathbf{n}]\frac{dS}{r}+ic\int\limits_\Gamma\int^{t-r/c} \mathbf{q}_0\cdot d\mathbf{s}\frac{dt}{r}, \tag{4.18}$$

$$4\pi i\mathbf{C} = \iint\limits_{S_1} [\mathbf{n}\times\mathbf{q}_0]\frac{dS}{r}, \tag{4.19}$$

which provide the required modification of formula (4.14).

§ 4.2. The diffraction of plane waves by a black half-plane

As an example† we consider the diffraction of plane waves of light by a black half-plane. We suppose that the half-plane is $x = 0, y \leqslant 0$, and that that the light is incident normally‡ on the screen.

† For an account of the application of the theory to the diffraction of electromagnetic waves from a Hertzian oscillator, we refer the reader to Kottler's paper.

‡ There is no loss of generality in making the assumption, since, on the present theory, diffraction by a black screen is an edge effect.

If the incident light is plane-polarized perpendicular to the edge of the screen, the electric and magnetic forces are given by

$$\mathbf{d}_0 = \mathbf{k}e^{-ik(x+ct)}, \qquad \mathbf{h}_0 = \mathbf{j}e^{-ik(x+ct)}$$

before diffraction—more precisely by the real parts of these expressions. After diffraction, the field is specified by

$$\mathbf{d} = \mathbf{d}_1+\mathbf{d}_2, \qquad \mathbf{h} = \mathbf{h}_1+\mathbf{h}_2,$$

where $\mathbf{d}_1$ and $\mathbf{h}_1$ are the vectors given by the Larmor-Tedone formulae applied to the surface $x = 0$, $y > 0$, and $\mathbf{d}_2$ and $\mathbf{h}_2$ are the terms due to Kottler's line charges on the rim Γ of the screen.

We have already proved (§ 2.2) that

$$\mathbf{d}_1 = \mathbf{k}(\epsilon e^{-ik(x+ct)}+u^B),$$
$$\mathbf{h}_1 = \mathbf{i}v+\mathbf{j}(\epsilon e^{-ik(x+ct)}+u^B),$$

where $\epsilon = 0$ or 1 according as P is or is not in the geometrical shadow,

$$u^B = -\frac{1}{4\pi}\int\limits_{-\infty}^{\infty} e^{ik(\rho\cosh\tau-ct)}\frac{\sin\phi}{\cosh\tau+\cos\phi}\,d\tau,$$

and
$$v = -\frac{i}{4}H_0^{(1)}(k\rho)e^{-ikct}.$$

We write these expressions in the form

$$\mathbf{d}_1 = \mathbf{d}^*+\mathbf{k}u^B, \qquad \mathbf{h}_1 = \mathbf{h}^*+\mathbf{i}v+\mathbf{j}u^B,$$

where $\mathbf{d}^*$ and $\mathbf{h}^*$ form the field according to the laws of geometrical optics.

The terms due to the line charges on Γ turn out to be

$$\mathbf{d}_2 = 0,$$

$$\mathbf{h}_2 = -\frac{1}{4k}\operatorname{grad} H_0^{(1)}(k\rho)e^{-ikct} = \tfrac{1}{4}H_1^{(1)}(k\rho)e^{-ikct}(\mathbf{i}\cos\phi+\mathbf{j}\sin\phi),$$

or
$$\mathbf{h}_2 = w(\mathbf{i}\cos\phi+\mathbf{j}\sin\phi)$$

say. The total effect after diffraction can therefore be written in the form

$$\mathbf{d} = \mathbf{d}^*+\mathbf{d}^B, \qquad \mathbf{h} = \mathbf{h}^*+\mathbf{h}^B,$$

where
$$\mathbf{d}^B = \mathbf{k}u^B,$$
$$\mathbf{h}^B = \mathbf{i}(v+w\cos\phi)+\mathbf{j}(u^B+w\sin\phi).$$

To obtain the effect when the incident light is plane-polarized parallel to the edge of the screen, so that

$$\mathbf{d}_0 = -\mathbf{j}e^{-ik(x+ct)}, \qquad \mathbf{h}_0 = \mathbf{k}e^{-ik(x+ct)},$$

we recall that Maxwell's equations are invariant under the transformation $\mathbf{d} \to \mathbf{h}$, $\mathbf{h} \to -\mathbf{d}$. The total effect is in this case therefore

$$\mathbf{d} = \mathbf{d}^* + \mathbf{d}^B, \qquad \mathbf{h} = \mathbf{h}^* + \mathbf{h}^B,$$

where $\quad \mathbf{d}^B = -\mathbf{i}(v + w\cos\phi) - \mathbf{j}(u^B + w\sin\phi), \qquad \mathbf{h}^B = \mathbf{k}u^B.$

So far we have been working with complex wave-functions. To deal with a real problem, the final step is to take real parts of these complex functions: but the change from light polarized perpendicular to the edge of the screen to light polarized parallel to the edge of the screen is still obtained by the transformation $\mathbf{d} \to \mathbf{h}$, $\mathbf{h} \to -\mathbf{d}$. When we are using real wave-functions the intensity of light is measured by $d^2 + h^2$. It follows that when plane-polarized light is diffracted by a black half-plane, there is no difference in intensity of the diffracted light corresponding to a difference in the planes of polarization. In this respect diffraction by a black half-plane differs from diffraction by a perfectly reflecting screen, a point to which we return in Chapter IV.

§ 4.3. The behaviour of the diffracted wave at large distances

To discuss the behaviour of the diffracted light when ρ, the distance from the edge of the screen, is large compared with the wave-length $2\pi/k$, we use the asymptotic formulae

$$u^B \sim -\frac{1}{2\sqrt{(2\pi k\rho)}} \tan\tfrac{1}{2}\phi\, e^{i\zeta},$$

$$v = -\tfrac{1}{4}iH_0^{(1)}(k\rho)e^{-ikct} \sim -\frac{1}{2\sqrt{(2\pi k\rho)}} e^{i\zeta},$$

$$w = \tfrac{1}{4}H_1^{(1)}(k\rho)e^{-ikct} \sim \frac{1}{2\sqrt{(2\pi k\rho)}} e^{i\zeta},$$

where $$\zeta = k(\rho - ct) + \tfrac{1}{4}\pi.$$

The first of these formulae was proved in Chapter II, § 3.2, the other two are immediate consequences of the asymptotic expansions of the Bessel functions.

It follows that, when the incident light is plane-polarized perpendicular to the edge of the screen,

$$\mathbf{d}^B \sim -\mathbf{k}\frac{\tan\frac{1}{2}\phi}{2\sqrt{(2\pi k\rho)}} e^{i\zeta},$$

$$\mathbf{h}^B \sim (-\mathbf{i}\sin\phi + \mathbf{j}\cos\phi)\frac{\tan\frac{1}{2}\phi}{2\sqrt{(2\pi k\rho)}} e^{i\zeta}.$$

Thus, at large distances, the electric force in the diffracted light is parallel to the edge of the screen, and the magnetic force is in the transverse direction perpendicular to the edge of the screen.

The corresponding result when the incident light is plane-polarized parallel to the edge of the screen is obtained by the transformation $\mathbf{d} \to \mathbf{h}$, $\mathbf{h} \to -\mathbf{d}$.

In both cases the edge of the screen appears luminous, and the intensity of the diffracted light is proportional to

$$\frac{1}{8\pi k\rho}\tan^2\tfrac{1}{2}\phi,$$

which agrees with the result obtained on the scalar theory in Chapter II.

The Poynting flux-of-energy vector in the diffracted light is

$$\frac{1}{8\pi}\mathbf{d}^B \times \bar{\mathbf{h}}^B \sim (\mathbf{i}\cos\phi+\mathbf{j}\sin\phi)\frac{\tan^2\frac{1}{2}\phi}{16\pi k\rho}.$$

Hence, at large distances, the flux of energy is directed radially outwards from the edge of the screen. In particular, at points of the illuminated face of the screen, the flux of energy in the diffracted light is along the screen. Kottler's definition of 'blackness' is thus satisfactory in so far as no energy is reflected from the screen.

IV

SOMMERFELD'S THEORY OF DIFFRACTION

§ 1. Sommerfeld's many-valued wave-functions

§ 1.1. Introduction

In his work on diffraction by a black screen Kirchhoff was attempting to solve a boundary-value problem, but he met the difficulty that it is impossible to give a correct physical statement of the conditions of the problem. He therefore made what seemed to be a reasonable assumption, namely that the field on the illuminated part of the screen was the same as in the incident light and that the field on the dark part was null. Unfortunately it turned out that his solution did not satisfy his boundary values, which were actually incompatible. The results of his theory agreed quite well with experiment, so that most authors regarded Kirchhoff's solution as a fairly accurate first approximation.

When we deal with a perfectly reflecting screen, a quite different state of affairs arises, since the boundary conditions at the surface of the screen are well known from the electromagnetic theory of light. At the surface of a perfect reflector the electric force is normal to the surface and the magnetic force tangential: and it is quite easy to solve such problems as that of the reflection of light from a perfectly reflecting plane mirror by constructing solutions of Maxwell's equations which satisfy these boundary conditions.

Theoretically it should be possible to solve diffraction problems relating to perfectly reflecting screens in the same way, but the analytical difficulties are considerable. It is only in the case of two-dimensional problems that much progress has been made. The fundamental idea, due to Sommerfeld, of using many-valued solutions of Maxwell's equations has been applied successfully to the problems of diffraction by a perfectly reflecting half-plane or wedge. In the present chapter we give an account of Sommerfeld's method, and work out as an example the problem of the diffraction of plane monochromatic light by a half-plane.

§ 1.2. The case of an electrostatic field

As a preliminary to a discussion of the diffraction of electromagnetic waves by a perfectly reflecting screen, the consideration of a similar problem in electrostatics is likely to prove instructive, since Maxwell's equations include the equations of an electrostatic field as the particular case when **h** is zero and **d** is independent of the time t.

The potential V in a two-dimensional electrostatic problem is a solution of Laplace's equation

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}=0 \tag{1.21}$$

which is constant on the surface of a conductor. Certain problems in electrostatics can be solved by the method of images. Let us

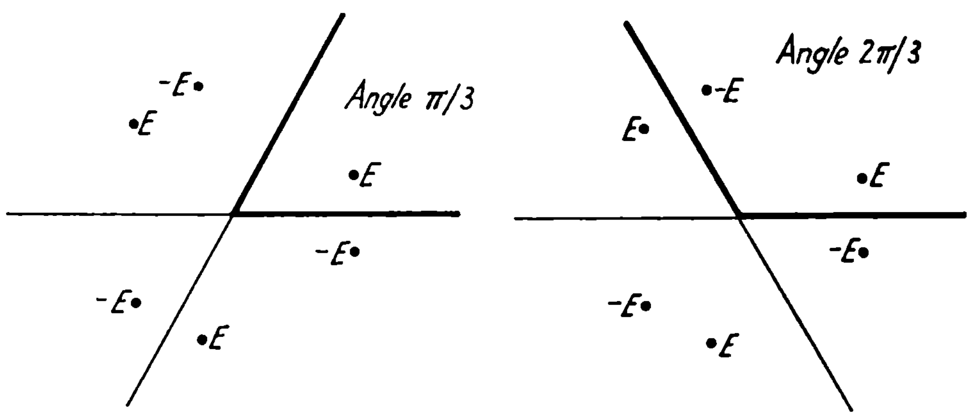


Fig. 18

consider, for example, the two-dimensional problem in which the field† is produced by a charge E at (x_1, y_1), where $y_1 > 0$ in the presence of an earthed conductor along the axis of x. This field is null in the half-plane $y < 0$: but in the region $y > 0$ the field is the same as if the conductor were removed and an additional charge $-E$ placed at the point $(x_1, -y_1)$, which is the image of (x_1, y_1) with respect to the conductor. A similar method will give the field due to a charge placed between two earthed conductors whose equations are $\theta = 0$ and $\theta = \pi/n$ in polar coordinates, n being a positive integer.

This method of images fails completely when we try to find the two-dimensional potential due to a charge in the presence of an earthed conductor along the radius $\theta = 0$ or, more generally, that due to a charge placed in the angle between two earthed conductors whose equations are $\theta = 0$ and $\theta = m\pi/n$, where $m\,(> 1)$ and n are

† Strictly we should describe this field as being due to a uniform line-charge parallel to an earthed plane conductor.

integers without common factor. For the image process introduces additional charges into the part of the plane considered. The diagrams illustrate the two cases of a charge placed in angles $\pi/3$ and $2\pi/3$ respectively.

A similar difficulty arises when we try to discuss the reflection of light by a perfectly reflecting half-plane or wedge: the image process introduces additional sources of light into the part of space outside the wedge.

In electrostatics the difficulty may be overcome by means of the theory of conformal representation. We recall that, if V is a solution of Laplace's equation (1.21), there exists a conjugate function U such that

$$U+iV = f(z), \tag{1.22}$$

where $f(z)$ is an analytic function of the complex variable $z = x+iy$. If the complex variables z and $Z = X+iY$ are connected by a relation

$$z = \phi(Z), \tag{1.23}$$

where $\phi(Z)$ is an analytic function of Z, areas in the z plane are mapped conformally† on the corresponding areas in the Z plane. If we apply this transformation to the complex potential $U+iV$ we obtain

$$U+iV = f\{\phi(Z)\},$$

so that V is also an electrostatic potential in the Z plane. Equipotential curves are transformed into equipotentials and, in particular, earthed conductors into earthed conductors. If V is the potential due to a charge E in the presence of a system of earthed conductors in the z plane, it can be shown that the transformation (1.23) turns V into the potential due to a charge E', not always the same as E, at the corresponding point of the Z plane in the presence of the transformed system of earthed conductors.‡

As an application of these ideas, we find the potential due to a charge E at the point z_1 in the plane of the complex variable

$$z = x+iy = re^{\theta i},$$

there being an earthed conductor along the positive half of the real axis. The variable θ, which is usually called the argument of z, is undetermined to an additive multiple of 2π; we fix it by requiring

† A mapping is 'conformal' if it leaves unaltered the angle between any two intersecting curves.

‡ See Jeans, *The Mathematical Theory of Electricity and Magnetism* (Cambridge, 1915), 264–83.

that θ shall lie between 0 and 2π. The two faces of the conductor can then be described as lying along the radii $\theta = 0$ and $\theta = 2\pi$. Now apply the transformation

$$Z = z^{\frac{1}{2}} = r^{\frac{1}{2}}e^{i\theta/2}, \tag{1.24}$$

where $r^{\frac{1}{2}}$ has its positive value. The restriction on θ implies that the argument of Z lies between 0 and π. Hence this transformation provides a (1, 1) conformal mapping of the whole z plane, apart from the radius $\theta = 0$ (or 2π), on the part of the Z plane in which $Y > 0$, and the radii $\theta = 0$ and $\theta = 2\pi$ are turned respectively into the positive and negative parts of the real axis. We have then to find the potential due to a charge E' at a point Z_1 in the region $Y > 0$, there being an earthed conductor along the line $Y = 0$.

If the conductor in the Z plane were absent the potential would be given by

$$U+iV = -2iE'\log(Z-Z_1);$$

but when the conductor is present we have

$$V = 0 \quad (Y < 0),$$

$$U+iV = -2iE'\log\frac{Z-Z_1}{Z-\bar{Z}_1} \quad (Y \geqslant 0),$$

the extra term being due to the charge $-E'$ at the image point $\bar{Z}_1$.

To obtain the potential due to the charge E at z_1 in the presence of an earthed conductor along $\theta = 0$, we apply the transformation (1.24). This gives†

$$U+iV = -2iE\log\frac{\sqrt{z}-\sqrt{z_1}}{\sqrt{z}-\sqrt{z_2}}, \tag{1.25}$$

where the arguments (or phases) of the complex numbers $\sqrt{z}$ and $\sqrt{z_1}$ lie between 0 and π, and $\sqrt{z_2}$ denotes the complex number conjugate to $\sqrt{z_1}$.

Now consider the function $U+iV$ defined in (1.25) as a function of the complex variable z, $\sqrt{z_1}$ and $\sqrt{z_2}$ being kept fixed. If we relax the restriction that the argument of z shall lie between 0 and 2π, $\sqrt{z}$ has two values whose arguments differ by π, and so the function $U+iV$ is a two-valued function of z.

The analytical difficulties of handling a two-valued function may be overcome by introducing the idea of a *Riemann surface*. Take two superimposed copies of the z plane, and cut each of them along

† That $E' = E$ follows from a consideration of the behaviour of $U+iV$ near the point z_1.

the positive part of the real axis. Then bind the lower edge of the cut in the upper sheet to the upper edge of the cut in the lower sheet, and finally bind together the remaining edges. In this way we have constructed what is called a two-sheeted Riemann surface. If we start at any point on the upper sheet and describe a circle about the origin, we reach the congruent point in the lower sheet; a second circuit about the origin brings us back to our starting-point.

We distinguish between the two sheets of the Riemann surface by writing $z = re^{\theta i}$ and requiring that θ shall lie between 0 and 2π in the upper sheet, between 2π and 4π in the lower sheet. With this convention we define $\sqrt{z}$ as being equal to the function

$$f(r,\theta) = r^{\frac{1}{2}}e^{i\theta/2},$$

where $r^{\frac{1}{2}}$ has its positive value. Then evidently

$$f(r,\theta+2\pi) = -f(r,\theta),$$

so that what we previously regarded as the two values of $\sqrt{z}$ are the values taken by $f(r,\theta)$ at congruent points in the two sheets. The function $\sqrt{z}$ defined in this way is a continuous one-valued function of position on the Riemann surface.

We now return to the function

$$U+iV = -2iE\log\frac{\sqrt{z}-\sqrt{z_1}}{\sqrt{z}-\sqrt{z_2}},$$

where $\sqrt{z_2}$ is the complex number conjugate to $\sqrt{z_1}$. With our new definition of $\sqrt{z}$, $U+iV$ is a one-valued function of position on the Riemann surface. In our physical problem θ lies between 0 and 2π, and so z lies in the upper sheet: accordingly we describe the upper sheet as physical space, the lower sheet as non-physical space.

The function $U+iV$ has two singular points on the Riemann surface, namely z_1 and z_2. Of these, z_1, the point at which the charge E is placed, lies in physical space, but z_2 does not. For if

$$\sqrt{z_1} = ae^{\alpha i} \quad (0 < \alpha < \pi),$$

we have
$$\sqrt{z_2} = ae^{-\alpha i}$$

and so
$$z_1 = a^2e^{2\alpha i}, \qquad z_2 = a^2e^{-2\alpha i}.$$

Hence z_2 is the point of the lower sheet congruent to

$$\bar{z}_1 = a^2e^{2(\pi-\alpha)i}$$

in the upper sheet.

To sum up: in order to obtain by the method of images the potential due to a charge E in the presence of an earthed conductor along the line $\theta = 0$ we have to consider a two-sheeted Riemann surface of which only the upper sheet corresponds to physical space. The image charge $-E$ lies in the lower sheet.

In a similar manner we can apply the method of images to find the potential due to a charge placed in the angle between two earthed conductors along the lines $\theta = 0$ and $\theta = m\pi/n$, where $m\,(>1)$ and n are integers without common factor. Since this angle can be turned into an angle π/n by the transformation

$$Z = z^{1/m},$$

we shall have to use an m-sheeted Riemann surface on which the function $z^{1/m}$ is one-valued.

§ 1.3. The introduction of a 'Riemann space' into diffraction problems

From the analogy with two-dimensional problems in electrostatics Sommerfeld saw that, in order to treat a diffraction problem as a boundary-value problem, it was necessary to consider many-valued solutions of Maxwell's equations which are single-valued functions of position in a 'Riemann space', this space being an imaginary space which bears the same relation to ordinary three-dimensional space as a Riemann surface bears to a plane. Such 'Riemann spaces' must not be confused with the Riemannian spaces considered in modern differential geometry.

For suppose we have a screen with rim Γ. We construct a 'Riemann space' out of a number of superimposed ordinary three-dimensional spaces. The sheets of this 'Riemann space' are joined together along the boundary of the geometrical shadow, so that we pass from one sheet to another by crossing the shadow. The rim Γ plays the same part as the point O did in our simpler Riemann surface in § 1.2. Physical space near the screen does not all belong to the same sheet of the 'Riemann space', since we can get from one side of the screen to the other only by crossing the boundary of the shadow. It follows that light incident on the screen leaves physical space and enters non-physical space. On the other hand, the reflected light originates in non-physical space and can be observed only when it crosses the screen into physical space.

On account of the analytical difficulties nearly all the problems which have been solved in this way are two-dimensional problems.

§ 1.4. Two-dimensional solutions of Maxwell's equations

An electromagnetic field is said to be two-dimensional if the effect is the same in all planes perpendicular to a given straight line. If we take this line to be the axis of z, the electric and magnetic vectors **d** and **h** are independent of z and so Maxwell's equations reduce to

$$\begin{aligned} &\frac{1}{c}\dot{d}_x = \frac{\partial h_z}{\partial y}, && -\frac{1}{c}\dot{h}_x = \frac{\partial d_z}{\partial y}, \\ &\frac{1}{c}\dot{d}_y = -\frac{\partial h_z}{\partial x}, && -\frac{1}{c}\dot{h}_y = -\frac{\partial d_z}{\partial x}, \\ &\frac{1}{c}\dot{d}_z = \frac{\partial h_y}{\partial x}-\frac{\partial h_x}{\partial y}, && -\frac{1}{c}\dot{h}_z = \frac{\partial d_y}{\partial x}-\frac{\partial d_x}{\partial y}, \\ &\frac{\partial d_x}{\partial x}+\frac{\partial d_y}{\partial y} = 0, && \frac{\partial h_x}{\partial x}+\frac{\partial h_y}{\partial y} = 0. \end{aligned} \tag{1.41}$$

These equations fall into two groups. In the first group, namely

$$\begin{aligned} \frac{\partial d_z}{\partial x} &= \frac{1}{c}\dot{h}_y, \\ \frac{\partial d_z}{\partial y} &= -\frac{1}{c}\dot{h}_x, \\ \frac{1}{c}\dot{d}_z &= \frac{\partial h_y}{\partial x}-\frac{\partial h_x}{\partial y}, \\ 0 &= \frac{\partial h_x}{\partial x}+\frac{\partial h_y}{\partial y}, \end{aligned} \tag{1.42}$$

only h_x, h_y, and d_z occur; whereas in the second group,

$$\begin{aligned} \frac{\partial h_z}{\partial x} &= -\frac{1}{c}\dot{d}_y, \\ \frac{\partial h_z}{\partial y} &= \frac{1}{c}\dot{d}_x, \\ \frac{1}{c}\dot{h}_z &= -\frac{\partial d_y}{\partial x}+\frac{\partial d_x}{\partial y}, \\ 0 &= \frac{\partial d_x}{\partial x}+\frac{\partial d_y}{\partial y}, \end{aligned} \tag{1.43}$$

only d_x, d_y, and h_z occur.

If h_x, h_y, and d_z satisfy (1.42), it is evident that the vectors

$$\mathbf{d} = d_z\,\mathbf{k}, \qquad \mathbf{h} = h_x\,\mathbf{i}+h_y\,\mathbf{j}$$

satisfy the equations (1.41). Hence the group of equations (1.42) specify an electromagnetic disturbance polarized perpendicular to Oz. Similarly, if d_x, d_y, and h_z satisfy (1.43), the vectors

$$\mathbf{d} = d_x \mathbf{i} + d_y \mathbf{j}, \qquad \mathbf{h} = h_z \mathbf{k},$$

specify an electromagnetic disturbance polarized parallel to Oz. The general solution of (1.41) is obtained by adding solutions of (1.42) and (1.43).

By eliminating h_x and h_y from (1.42) we find that the third equation becomes

$$\frac{\partial^2 d_z}{\partial x^2} + \frac{\partial^2 d_z}{\partial y^2} = \frac{1}{c^2}\frac{\partial^2 d_z}{\partial t^2},$$

and that the fourth equation is satisfied identically. Thus d_z satisfies the equation of cylindrical waves, and, if we can determine d_z, the components h_x and h_y are given by the first two equations of the group. Similarly, if we eliminate d_x and d_y from (1.43), the third equation of the group becomes

$$\frac{\partial^2 h_z}{\partial x^2} + \frac{\partial^2 h_z}{\partial y^2} = \frac{1}{c^2}\frac{\partial^2 h_z}{\partial t^2},$$

and the fourth equation is satisfied identically. Thus h_z also satisfies the equation of cylindrical waves, and a knowledge of h_z determines d_x and d_y.

The problem of finding a two-dimensional electromagnetic field reduces in this way to that of solving the equation of cylindrical waves, but the boundary conditions which have to be satisfied at the surface of a perfect reflector depend on whether the field is polarized perpendicular to or parallel to the axis of z.

At the surface of a perfectly reflecting screen the electric force is in a direction normal to the surface, the magnetic force in a tangential direction. In a two-dimensional problem a perfectly reflecting screen is a cylinder with generators parallel to Oz. It follows that, if the field is polarized perpendicular to Oz so that

$$d_x = d_y = h_z = 0,$$

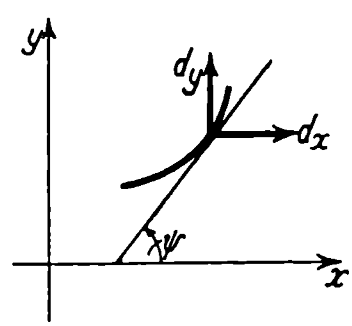


FIG. 19

we have to find a solution d_z of the equation of cylindrical waves which vanishes on the surface of the screen. But if the field is polarized parallel to Oz, so that $h_x = h_y = d_z = 0$, the boundary condition is

$$d_x \cos\psi + d_y \sin\psi = 0$$

in the notation of the figure. By equations (1.43) this reduces to

$$\sin\psi\frac{\partial h_z}{\partial x}-\cos\psi\frac{\partial h_z}{\partial y}=0,$$

that is,

$$\frac{\partial h_z}{\partial n}=0$$

where $\partial/\partial n$ denotes differentiation along the normal to the screen. Thus our problem is to find a solution h_z of the equation of cylindrical waves whose normal derivative vanishes on the screen.

It will be observed that, when the field is polarized parallel to Oz, h_z satisfies the same conditions as the velocity potential of sound waves in air in the presence of a rigid boundary.

§ 1.5. Sommerfeld's many-valued solutions† of the wave-equation

We now introduce the many-valued wave-functions which Sommerfeld used to solve two-dimensional problems of diffraction by a wedge or half-plane. If we take the straight edge of the wedge or half-plane as axis of z, we have to solve

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}$$

under the appropriate boundary conditions. Actually it is more convenient to use polar coordinates (ρ,ϕ), in terms of which the equation becomes

$$\frac{\partial^2 u}{\partial \rho^2}+\frac{1}{\rho}\frac{\partial u}{\partial \rho}+\frac{1}{\rho^2}\frac{\partial^2 u}{\partial \phi^2}=\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}. \tag{1.51}$$

As the problem is two-dimensional we need to consider, not a 'Riemann space', but only the Riemann surface constructed in the following way. We take p superimposed planes each slit along the radius $\phi=\phi'+\pi$, where ϕ' is any given angle between $\pm\pi$; we then bind the edge $\phi=\phi'+\pi-0$ of each cut to the edge $\phi=\phi'+\pi+0$ of the cut in the plane below, and finally bind the remaining edges of the top and bottom sheets. We thus obtain a Riemann surface of p sheets with the origin as branch-point; we can pass from one sheet to the next by the cross-bridge at $\phi=\phi'+\pi$. The diagram shows the connexion of the sheets in the case $p=4$, as seen looking along the cross-bridges towards the origin.

† The relevant papers are *Math. Ann.* **45** (1894), 263; **47** (1896), 317. *Gött. Nach.* (1894), 338; (1895), 267. *Proc. London Math. Soc.* (1), **28** (1897), 395. *Zeitschrift f. Math. u. Phys.* **46** (1901), 11. See also Carslaw, *Proc. London Math. Soc.* (1), **30** (1899), 121; Lamb, ibid. (2), **8** (1910), 422; Hanson, *Phil. Trans.* (A) **237** (1938), 48.

If we start from any point on this Riemann surface and make p complete circuits about the origin, we return to our starting-point. In doing so, the polar angle ϕ is increased by $2p\pi$. We specify the particular sheet in which the point (ρ, ϕ) lies by saying that it lies in the nth sheet if ϕ lies between $\phi' \pm \pi + 2(n-1)\pi$. The coordinates (ρ, ϕ) and $(\rho, \phi+2p\pi)$ specify the same point.

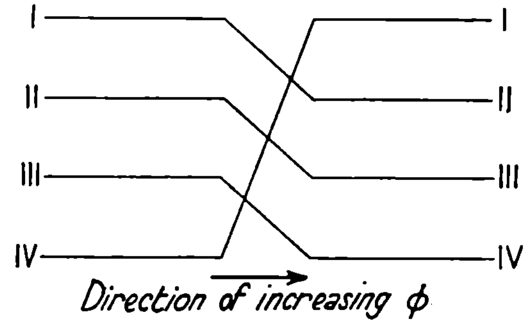


FIG. 20

Our first task is to find solutions of the wave-equation which are one-valued on this Riemann surface.† We start by considering the simplest case $p = 1$, so that the wave-functions in question are periodic in ϕ of period 2π. The simplest such solution is that which represents 'plane waves', viz.

$$e^{ik\rho\cos(\phi-\alpha)+ikct},$$

where α is a constant, real or complex. A more general solution is obtained by multiplying by a function of α and then integrating with respect to α. In particular,

$$u = \frac{e^{ikct}}{2\pi} \int e^{ik\rho\cos(\phi-\alpha)} \frac{e^{i\alpha}}{e^{i\alpha}-e^{i\phi'}}\, d\alpha$$

integrated round any closed contour in the plane of the complex variable α satisfies the wave-equation. If we write $\alpha = \phi+\zeta$ this becomes

$$u = \frac{e^{ikct}}{2\pi} \int\limits_L e^{ik\rho\cos\zeta} \frac{e^{i\zeta}}{e^{i\zeta}-e^{i(\phi'-\phi)}}\, d\zeta. \tag{1.52}$$

The integrand in (1.52) has simple poles at the points

$$\zeta = \phi'-\phi+2n\pi.$$

When the contour L is a simple closed contour about the pole $\zeta = \phi'-\phi$, the solution (1.52) reduces to

$$u_0 = e^{ik\rho\cos(\phi-\phi')+ikct}, \tag{1.53}$$

† The present discussion is based on that of Carslaw, loc. cit.

which represents monochromatic plane-waves. Moreover u remains equal to u_0 when L is continuously deformed, so long as L remains closed and does not cross any of the singularities. In particular, we may take L to be the contour in Fig. 21. (In the figure it is assumed that $\phi'-\phi$ lies between $\pm\pi$; if it does not, there is, in any case, one pole $\phi'-\phi+2n\pi$ lying between $\pm\pi$, and the same conclusion follows.)

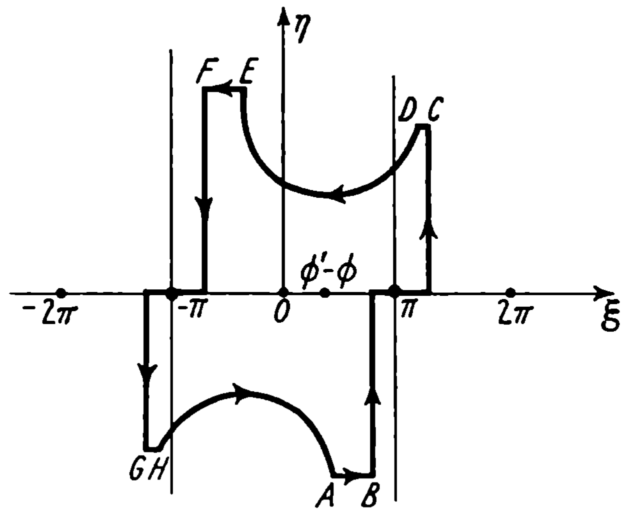


FIG. 21

On the line AB we have $\zeta = \xi - iR$, where ξ varies from α to β and $0 < \alpha < \beta < \pi$. Since

$$\cos\zeta = \cos\xi\cosh R + i\sin\xi\sinh R,$$

we see that

$$\left|\int\limits_{AB}\right| \leqslant \frac{e^R}{2\pi(e^R-1)}\int\limits_{\alpha}^{\beta} e^{-kp\sin\xi\sinh R}\,d\xi$$

$$< \frac{1}{\pi}\int\limits_{\alpha}^{\beta} e^{-kpR\sin\xi}\,d\xi$$

$$\to 0$$

as $R \to \infty$. Again, on the line CD, we have $\zeta = \xi + iS$, where ξ varies from $\pi+\gamma$ to $\pi+\delta$ and $0 < \delta < \gamma < \pi$. Hence

$$\left|\int\limits_{CD}\right| \leqslant \frac{e^{-S}}{2\pi(1-e^{-S})}\int\limits_{\pi+\delta}^{\pi+\gamma} e^{kp\sin\xi\sinh S}\,d\xi$$

$$< \frac{1}{\pi}\int\limits_{\delta}^{\gamma} e^{-kpS\sin\theta}\,d\theta$$

$$\to 0$$

as $S \to \infty$. Similarly, the integrals along EF and GH tend to zero as these segments recede to infinity, remaining parallel to the real axis.

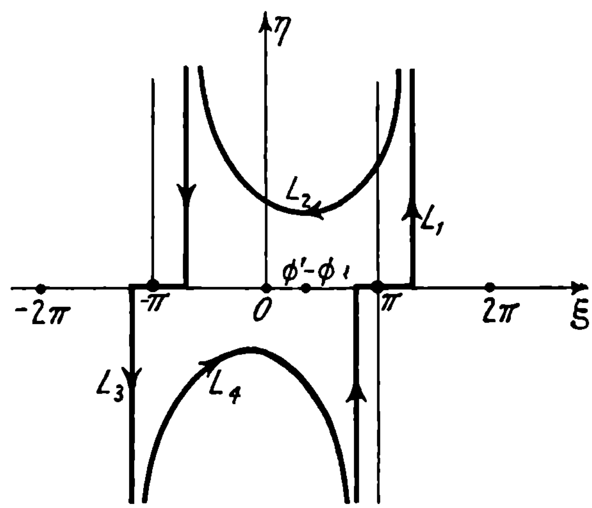


FIG. 22

It follows from this argument that we can take the contour L to be that of Fig. 22. But it is evident that the portion L_1 is obtained from L_3 by translation through a distance 2π and, since the integrand has period 2π, the integrals along L_1 and L_3 cancel. Hence

$$u_0 = \frac{e^{ikct}}{2\pi} \int_{L_2+L_4} e^{ik\rho\cos\zeta} \frac{e^{i\zeta}}{e^{i\zeta} - e^{i(\phi'-\phi)}} d\zeta.$$

For brevity, we shall call the two curves L_2 and L_4, taken together, the path A.

We next consider the case $p = 2$. The expression

$$\frac{e^{ikct}}{4\pi} \int e^{ik\rho\cos(\phi-\alpha)} \frac{e^{i\alpha/2}}{e^{i\alpha/2} - e^{i\phi'/2}} d\alpha, \tag{1.54}$$

integrated along any fixed complex path, is a solution of the equation of wave-motions, provided that we can perform the necessary differentiations under the sign of integration. It follows that, if A is the path just defined, the function

$$u(\rho, \phi) = \frac{e^{ikct}}{4\pi} \int_A e^{ik\rho\cos\zeta} \frac{e^{i\zeta/2}}{e^{i\zeta/2} - e^{i(\phi'-\phi)/2}} d\zeta \tag{1.55}$$

satisfies the wave-equation; for it can be expressed as a sum of two expressions of the form (1.54).

The wave-function (1.55), regarded as a function of ϕ, does not have period 2π. For

$$u(\rho,\phi+2\pi) = \frac{e^{ikct}}{4\pi}\int_{A'} e^{ik\rho\cos\zeta}\frac{e^{i\zeta/2}}{e^{i\zeta/2}-e^{i(\phi'-\phi)/2}}\,d\zeta, \tag{1.56}$$

where A' is obtained from A by a translation to the left through a distance 2π (Fig. 23): and as the integrand in (1.55), regarded as a function of ζ, does not have the period 2π, $u(\rho,\phi)$ and $u(\rho,\phi+2\pi)$ are in general unequal. By a similar argument, we can show that $u(\rho,\phi)$ is, nevertheless, periodic in ϕ, but that its period is 4π. Thus the wave-function defined by (1.55) is not a one-valued function in the ordinary plane, but is a one-valued function on the two-sheeted Riemann surface. Moreover, it is finite and continuous for all real values of ρ.

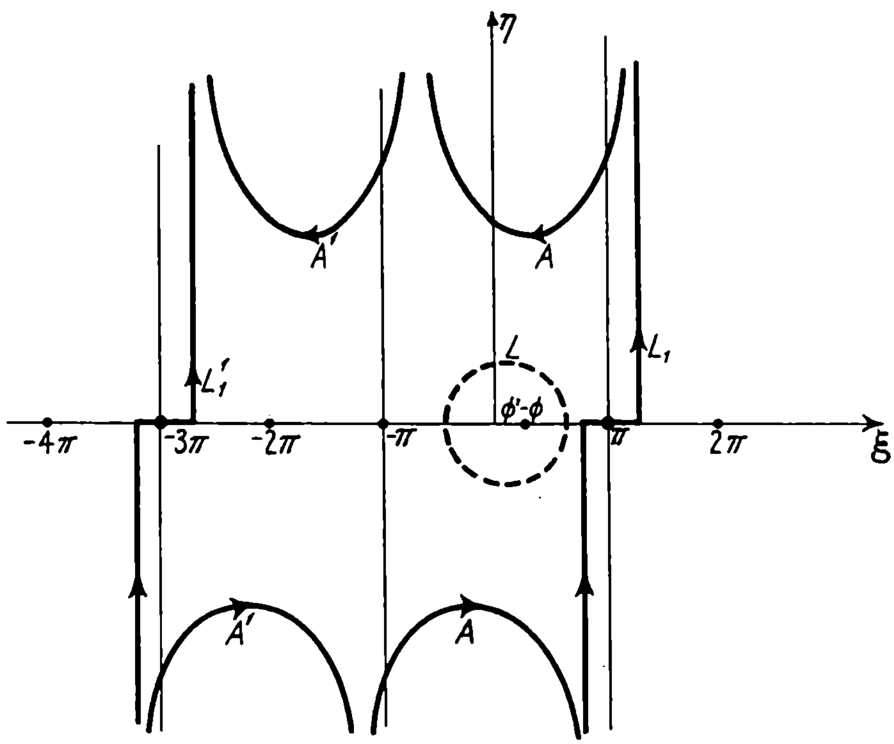


FIG. 23

Keeping the point (ρ,ϕ) in the first sheet of our Riemann surface we have

$$u(\rho,\phi)+u(\rho,\phi+2\pi) = \frac{e^{ikct}}{4\pi}\int_{A+A'} e^{ik\rho\cos\zeta}\frac{e^{i\zeta/2}}{e^{i\zeta/2}-e^{i(\phi'-\phi)/2}}\,d\zeta.$$

But if the polygonal path L_1' is obtained from L_1 by a translation to the left through a distance 4π, this equation gives

$$u(\rho,\phi)+u(\rho,\phi+2\pi) = \frac{e^{ikct}}{4\pi}\int\limits_{A+A'+L_1-L_1'} e^{ik\rho\cos\zeta}\frac{e^{i\zeta/2}}{e^{i\zeta/2}-e^{i(\phi'-\phi)/2}}\,d\zeta,$$

since the integrals along L_1 and $-L_1'$ cancel. Now the only singularities of the integrand are simple poles at the points

$$\zeta = (\phi'-\phi)+4n\pi.$$

Hence if we deform the contour $A+A'+L_1-L_1'$ into a simple closed curve L surrounding the point $\zeta = \phi'-\phi$, we obtain

$$u(\rho,\phi)+u(\rho,\phi+2\pi) = e^{ik\rho\cos(\phi-\phi')+ikct},$$

by the calculus of residues, and so

$$u(\rho,\phi)+u(\rho,\phi+2\pi) = u_0(\rho,\phi). \tag{1.57}$$

This relation connects the values of the function u at corresponding points of the two sheets of the Riemann surface.

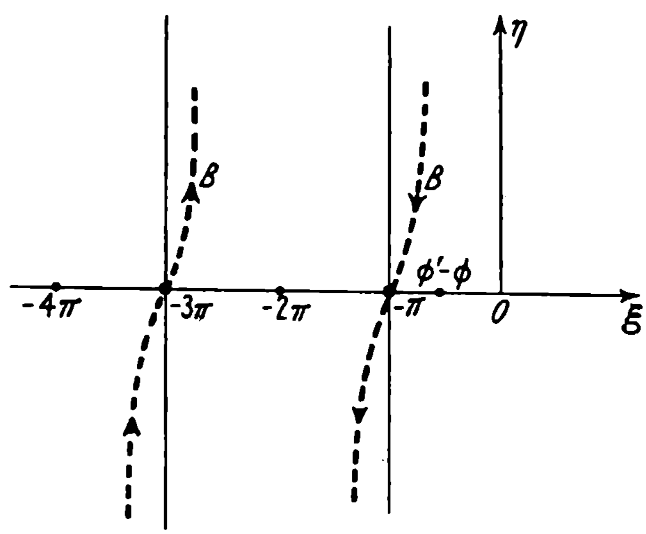


FIG. 24

Still keeping the point (ρ,ϕ) in the first sheet, the path A' can be deformed into the two-branched path B (Fig. 24) without altering the value of the integral (1.56). Now

$$|e^{ik\rho\cos\zeta}| = e^{k\rho\sin\xi\sinh\eta}$$

and $\sin\xi \sinh\eta$ is negative on B. Hence

$$u(\rho,\phi+2\pi) \to 0$$

as $\rho \to \infty$. From (1.57) it follows that, as $\rho \to \infty$,

$$u(\rho,\phi)-u_0(\rho,\phi) \to 0.$$

The same type of argument can be applied to the function

$$u = \frac{e^{ikct}}{2\pi p}\int\limits_A e^{ik\rho\cos\zeta}\frac{e^{i\zeta/p}}{e^{i\zeta/p}-e^{i(\phi'-\phi)/p}}\,d\zeta, \tag{1.58}$$

where p is any positive integer. This function has the following properties:

(i) *it is a solution of the equation of wave-motions which is p-valued in the ordinary plane*;

(ii) *on the p-sheeted Riemann surface it is one-valued, finite, and continuous*;

(iii) *as $\rho \to \infty$, $u \to 0$ in all sheets except the first*;

(iv) *on the first sheet, $u-u_0 \to 0$ as $\rho \to \infty$*;

(v) *the sum of the p values taken by u at any point in the first sheet and the corresponding points in the other sheets is u_0*.

In the first sheet of the Riemann surface the wave-function (1.58) behaves like the wave-function u_0 of plane waves, at any rate at large distances from the origin.

The corresponding solution for cylindrical waves expanding from a line-source at (ρ',ϕ') is†

$$u = \frac{e^{ikct}}{2\pi p}\int\limits_A H_0^{(2)}\{k\sqrt{(\rho^2+\rho'^2-2\rho\rho'\cos\zeta)}\}\frac{e^{i\zeta/p}}{e^{i\zeta/p}-e^{i(\phi'-\phi)/p}}\,d\zeta,$$

where $H_0^{(2)}$ denotes the Bessel function of the third kind J_0-iY_0.

§ 1.6. A transformation of the formulae for the case $p = 2$

If (ρ,ϕ) is a point in the first sheet, we have

$$u(\rho,\phi+2\pi) = \frac{e^{ikct}}{4\pi}\int\limits_B e^{ik\rho\cos\zeta}\frac{e^{i\zeta/2}}{e^{i\zeta/2}-e^{i(\phi'-\phi)/2}}\,d\zeta, \tag{1.61}$$

where B is the two-branched path of Fig. 24. This path can be deformed into the two lines $\mathrm{Rl}\,\zeta = -\pi$, $\mathrm{Rl}\,\zeta = -3\pi$ without alter-

† Cf. Carslaw, loc. cit. 147 et seq. Carslaw also discusses the corresponding three-dimensional problem with a point-source.

ing the value of the integral. Hence if we write $\zeta = -\pi+i\eta$, $\zeta = -3\pi+i\eta$ on these lines, equation (1.61) becomes

$$u(\rho,\phi+2\pi) = \frac{e^{ikct}}{4\pi}\int_{-\infty}^{\infty} e^{-ik\rho\cosh\eta}\left\{\frac{ie^{-\eta/2}}{ie^{-\eta/2}-e^{i(\phi'-\phi)/2}}-\frac{ie^{-\eta/2}}{ie^{-\eta/2}+e^{i(\phi'-\phi)/2}}\right\}i\,d\eta.$$

This reduces to

$$u(\rho,\phi+2\pi) = \frac{e^{ikct}}{4\pi}\int_{-\infty}^{\infty} e^{-ik\rho\cosh\eta}\frac{d\eta}{\cos\frac{1}{2}(\phi'-\phi-i\eta)}$$

$$= \frac{e^{ikct}}{4\pi}\int_{0}^{\infty} e^{-ik\rho\cosh\eta}\left\{\frac{1}{\cos\frac{1}{2}(\phi'-\phi-i\eta)}+\right.$$

$$\left.+\frac{1}{\cos\frac{1}{2}(\phi'-\phi+i\eta)}\right\}d\eta$$

$$= \frac{e^{ikct}}{\pi}\int_{0}^{\infty} e^{-ik\rho\cosh\eta}\frac{\cosh\frac{1}{2}\eta\cos\frac{1}{2}(\phi'-\phi)}{\cosh\eta+\cos(\phi'-\phi)}d\eta.$$

Hence we have
$$u(\rho,\phi+2\pi) = Xu_0(\rho,\phi), \tag{1.62}$$
where

$$X = \frac{1}{\pi}\cos\tfrac{1}{2}(\phi'-\phi)\int_{0}^{\infty} e^{-ik\rho\{\cosh\eta+\cos(\phi'-\phi)\}}\frac{\cosh\frac{1}{2}\eta}{\cosh\eta+\cos(\phi'-\phi)}d\eta.$$

If we now make the substitution $\tau = \sinh\frac{1}{2}\eta$, the formula foı X becomes

$$X = \frac{1}{\pi}\cos\tfrac{1}{2}(\phi'-\phi)\int_{0}^{\infty} e^{-2ik\rho\{\tau^2+\cos^2(\phi'-\phi)/2\}}\frac{d\tau}{\tau^2+\cos^2\frac{1}{2}(\phi'-\phi)}.$$

It follows that

$$\frac{\partial X}{\partial\rho} = -\frac{2ik}{\pi}\cos\tfrac{1}{2}(\phi'-\phi)e^{-2ik\rho\cos^2(\phi'-\phi)/2}\int_{0}^{\infty} e^{-2ik\rho\tau^2}\,d\tau,$$

the differentiation under the sign of integration being justified by the uniform convergence of the resulting integral. Using a well-known result, we have

$$\frac{\partial X}{\partial\rho} = -e^{\pi i/4}\sqrt{\left(\frac{k}{2\pi\rho}\right)}\cos\tfrac{1}{2}(\phi'-\phi)e^{-2ik\rho\cos^2(\phi'-\phi)/2}$$

$$= \frac{e^{\pi i/4}}{\sqrt{\pi}}\frac{\partial}{\partial\rho}\int_{\alpha}^{-T} e^{-i\lambda^2}\,d\lambda,$$

where $$T = \cos\tfrac{1}{2}(\phi'-\phi)\sqrt{(2k\rho)}$$

and α is independent of ρ. Hence

$$X \doteq \frac{e^{\pi i/4}}{\sqrt{\pi}} \int\limits_{\alpha}^{-T} e^{-i\lambda^2}\, d\lambda,$$

and so, by (1.62),

$$u(\rho,\phi+2\pi) = \frac{e^{\pi i/4}}{\sqrt{\pi}} u_0(\rho,\phi) \int\limits_{\alpha}^{-T} e^{-i\lambda^2}\, d\lambda.$$

But since $u(\rho,\phi+2\pi) \to 0$ as $\rho \to \infty$, the lower limit α is $-\infty$; therefore

$$u(\rho,\phi+2\pi) = \frac{e^{\pi i/4}}{\sqrt{\pi}} u_0(\rho,\phi) \int\limits_{-\infty}^{-T} e^{-i\lambda^2}\, d\lambda. \tag{1.63}$$

The point $(\rho,\phi+2\pi)$ is in the lower sheet of the Riemann surface. To obtain the value of u at the corresponding point (ρ,ϕ) of the upper sheet, we use the relation

$$u(\rho,\phi)+u(\rho,\phi+2\pi) = u_0(\rho,\phi) = \frac{e^{\pi i/4}}{\sqrt{\pi}} u_0(\rho,\phi) \int\limits_{-\infty}^{\infty} e^{-i\lambda^2}\, d\lambda.$$

It follows that $$u(\rho,\phi) = \frac{e^{\pi i/4}}{\sqrt{\pi}} u_0(\rho,\phi) \int\limits_{-T}^{\infty} e^{-i\lambda^2}\, d\lambda,$$

or, by a change of variable,

$$u(\rho,\phi) = \frac{e^{\pi i/4}}{\sqrt{\pi}} u_0(\rho,\phi) \int\limits_{-\infty}^{T} e^{-i\lambda^2}\, d\lambda. \tag{1.64}$$

Finally, we observe that

$$-T = \cos\tfrac{1}{2}(\phi'-\phi-2\pi)\sqrt{(2k\rho)}.$$

Hence (1.63) and (1.64) can be comprised in the single formula

$$u(\rho,\phi) = \frac{e^{\pi i/4}}{\sqrt{\pi}} u_0(\rho,\phi) \int\limits_{-\infty}^{T} e^{-i\lambda^2}\, d\lambda, \tag{1.65}$$

where $$T = \cos\tfrac{1}{2}(\phi'-\phi)\sqrt{(2k\rho)}$$

no matter in which sheet of the Riemann surface the point (ρ,ϕ) lies.

§ 1.7. Some properties of Fresnel's integral

In § 1.6 we expressed Sommerfeld's two-valued wave-function in a form involving the function

$$Z(t) = \int_{-\infty}^{t} e^{-it^2}\,dt.$$

This function has real and imaginary parts

$$X(t) = \int_{-\infty}^{t} \cos t^2\,dt, \qquad Y(t) = -\int_{-\infty}^{t} \sin t^2\,dt,$$

which are known as Fresnel's integrals. In order to discuss qualitatively the behaviour of $Z(t)$ it is simplest to make use of the curve, known as Cornu's spiral, which has parametric equations

$$x = X(t), \qquad y = Y(t).$$

For our purposes we need to know rather more about $Z(t)$ when t is large than can be found by consideration of a graph. To get an approximate formula for $Z(t)$ when t is large and negative we integrate by parts twice; this gives

$$Z(t) = -\int_{-\infty}^{t} \frac{1}{2it}\,de^{-it^2} = -\frac{e^{-it^2}}{2it} - \int_{-\infty}^{t} \frac{1}{4t^3}\,de^{-it^2}$$

$$= -\frac{e^{-it^2}}{2it} - \frac{e^{-it^2}}{4t^3} - \int_{-\infty}^{t} \frac{3}{4t^4} e^{-it^2}\,dt = -\frac{e^{-it^2}}{2it} + \eta_0(t), \quad \text{say.}$$

Here
$$|\eta_0(t)| \leqslant -\frac{1}{4t^3} + \frac{3}{4}\int_{-\infty}^{t} \frac{dt}{t^4} = -\frac{1}{2t^3}.$$

Hence, when t is negative,

$$Z(t) = -\frac{e^{-it^2}}{2it} + \eta_0(t), \tag{1.71}$$

where
$$|\eta_0(t)| \leqslant -\frac{1}{2t^3}.$$

When t is positive we have

$$Z(t) = \int_{-t}^{\infty} e^{-it^2}\,dt = \int_{-\infty}^{\infty} e^{-it^2}\,dt - \int_{-\infty}^{-t} e^{-it^2}\,dt = \surd\pi\, e^{-\pi i/4} - Z(-t).$$

By (1.71) we find that

$$Z(t) = \sqrt{\pi}\, e^{-\pi i/4} - \frac{e^{-it^2}}{2it} + \eta_1(t), \tag{1.72}$$

where

$$|\eta_1(t)| \leqslant \frac{1}{2t^3}.$$

§ 2. The diffraction of plane-polarized light by a reflecting half-plane

§ 2.1. Sommerfeld's solution

Let us suppose in the first instance that the plane of polarization is perpendicular to the edge of the screen. If we choose the axes so that the screen is the half-plane $x = 0$, $y < 0$, the incident light is specified by

$$d_z = e^{ik\rho\cos(\phi-\phi')+ikct} \equiv u_0(\rho, \phi; \phi'), \tag{2.11}$$

where ϕ', the angle of incidence, may be supposed, without loss of generality, to lie between $\pm\frac{1}{2}\pi$. We have to find a solution d_z of the equation of wave-motions which behaves like u_0 when ρ is large compared with the wave-length $2\pi/k$, and which vanishes on the screen.

At first sight it would appear that the appropriate solution ought to be $d_z = u$, where

$$u = u_0(\rho, \phi; \phi') - u_0(\rho, \phi; \pi-\phi'). \tag{2.12}$$

But this solution vanishes all over the plane $x = 0$ and so solves, not the problem of diffraction by a half-plane, but the problem of reflection by a perfectly conducting plane. If, however, we replace u_0 by Sommerfeld's two-valued wave-function

$$u(\rho, \phi; \phi') = \frac{e^{\pi i/4}}{\sqrt{\pi}}\, u_0(\rho, \phi; \phi') \int\limits_{-\infty}^{T} e^{-i\lambda^2}\, d\lambda, \tag{2.13}$$

where

$$T = \cos\tfrac{1}{2}(\phi-\phi')\sqrt{(2k\rho)},$$

and consider the wave-function

$$U(\rho, \phi; \phi') = u(\rho, \phi; \phi') - u(\rho, \phi; 3\pi-\phi'), \tag{2.14}$$

then $d_z = U$ provides the solution of our problem, it being supposed that in physical space ϕ lies between $-\frac{1}{2}\pi$ and $\frac{3}{2}\pi$. This solution is obtained by applying the method of images to Sommerfeld's two-valued wave-function, just as the simpler solution (2.12) was obtained from u_0. There is a slight difference, in that the angle of incidence of the reflected wave must be $3\pi-\phi'$, not $\pi-\phi'$, since this wave originates in non-physical space.

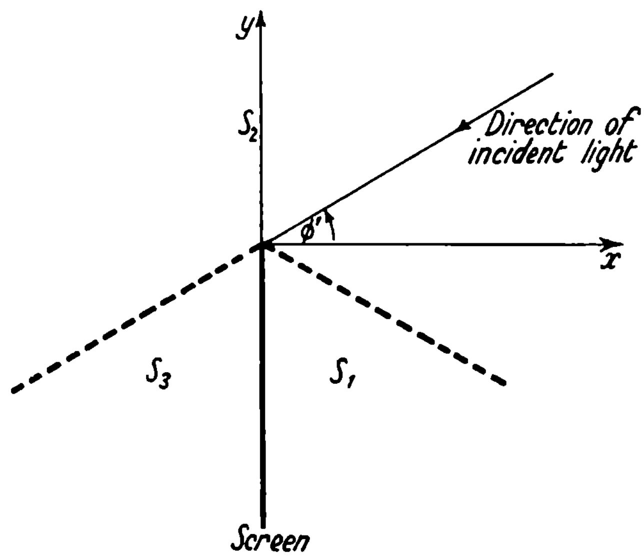


FIG. 25

Each of the wave-functions $u(\rho,\phi;\,\phi')$, $u(\rho,\phi;\,3\pi-\phi')$ has its own associated Riemann surface, the cross-bridges in physical space being $\phi = \pi+\phi'$ and $\phi = -\phi'$ respectively. These cross-bridges divide physical space into three sectors S_1, S_2, and S_3, as shown in the figure, Of these S_1 and S_2 belong to the upper sheet, S_3 to the lower sheet of the Riemann surface associated with $u(\rho,\phi;\,\phi')$; but S_1 belongs to the upper sheet, and S_2 and S_3 to the lower sheet of the Riemann surface associated with $u(\rho,\phi;\,3\pi-\phi')$. In both cases S_1 and S_3 belong to different sheets. This means that the incident light, which is specified by $u(\rho,\phi;\,\phi')$, at any rate at large distances from the screen, remains in the upper sheet on crossing the screen and so disappears from physical space; on the other hand, the reflected light specified by $-u(\rho,\phi;\,3\pi-\phi')$ originates in non-physical space and enters physical space by crossing the screen into the sector S_1. There is, however, no discontinuity of $U(\rho,\phi;\,\phi')$ in physical space except across the screen.

We have, however, somewhat anticipated the fact that $U(\rho,\phi;\,\phi')$ does provide the solution of our diffraction problem. In the first place, U is a solution of the equation of wave-motions and evidently vanishes on the two sides of the screen, viz. the radii $\phi = -\frac{1}{2}\pi$ and $\phi = \frac{3}{2}\pi$.

Secondly, when ρ is large compared with the wave-length $2\pi/k$ and (ρ, ϕ) lies in the sector S_2, it follows from (1.71) and (1.72) that

$$U(\rho, \phi; \phi') = u_0(\rho, \phi; \phi') + O(k\rho)^{-\frac{1}{2}}.$$

Thus $U(\rho, \phi; \phi')$ gives a field which is practically unaffected by the presence of the screen at points in the sector S_2 at large distances from the origin.

Next, in the geometrical shadow S_3 we have

$$U(\rho, \phi; \phi') = O(k\rho)^{-\frac{1}{2}}$$

when ρ is large compared with the wave-length. Thus the geometrical shadow is practically dark.

Finally, in the sector S_1

$$U(\rho, \phi; \phi') = u_0(\rho, \phi; \phi') - u_0(\rho, \phi; \pi - \phi') + O(k\rho)^{-\frac{1}{2}},$$

where ρ is large compared with $2\pi/k$. Hence in the sector S_1 we have, in effect, the ordinary field of equation (2.12) above due to the incident and reflected light.

The function $U(\rho, \phi; \phi')$ does satisfy all the conditions of the diffraction problem. A closer examination of the order-terms $O(k\rho)^{-\frac{1}{2}}$ will give the diffracted light.

If, however, the plane of polarization of the incident light is parallel to the edge of the screen, the original field is

$$h_z = u_0(\rho, \phi; \phi'),$$

where ϕ' is the angle of incidence, supposed to lie between $\pm\frac{1}{2}\pi$. The diffraction problem is solved by taking $h_z = V$, where V is a wave-function which behaves like u_0 when ρ is large compared with the wave-length and which satisfies the boundary condition $\partial V/\partial\phi = 0$ on the faces $\phi = -\frac{1}{2}\pi$ and $\phi = \frac{3}{2}\pi$ of the screen. An argument similar to that used above shows that

$$V(\rho, \phi; \phi') = u(\rho, \phi; \phi') + u(\rho, \phi; 3\pi - \phi') \tag{2.15}$$

satisfies these conditions. When $k\rho$ is large, we have

(i) in S_2,

$$V(\rho, \phi; \phi') = u_0(\rho, \phi; \phi') + O(k\rho)^{-\frac{1}{2}};$$

(ii) in S_3,

$$V(\rho, \phi; \phi') = O(k\rho)^{-\frac{1}{2}};$$

(iii) in S_1,

$$V(\rho, \phi; \phi') = u_0(\rho, \phi; \phi') + u_0(\rho, \phi; \pi - \phi') + O(k\rho)^{-\frac{1}{2}}.$$

These equations can be interpreted in a manner similar to that given above, and a closer examination of the order terms will give the diffraction effect.

§ 2.2. The diffracted light

The function $T = \cos\frac{1}{2}(\phi-\phi')\sqrt{(2k\rho)}$, which occurs in the wave-function $u(\rho,\phi;\phi')$, vanishes on the radius vector $\phi = \phi'+\pi$. To avoid difficulties which this might cause, we consider only the part of physical space which lies outside the parabola $\pi\epsilon^2T^2 = 1$; this involves no loss of generality when we consider the effect when $k\rho$ is large, since we are cutting out only a finite part of each radius vector except the critical one. In the parts of S_1 and S_2 which lie outside this parabola

$$T > \frac{1}{\epsilon\sqrt{\pi}};$$

but in the part of S_3 outside the parabola

$$T < -\frac{1}{\epsilon\sqrt{\pi}}.$$

Again, since the function $T' = \cos\frac{1}{2}(\phi+\phi'-3\pi)\sqrt{(2k\rho)}$, which occurs in the wave-function $u(\rho,\phi;3\pi-\phi')$, vanishes on the radius

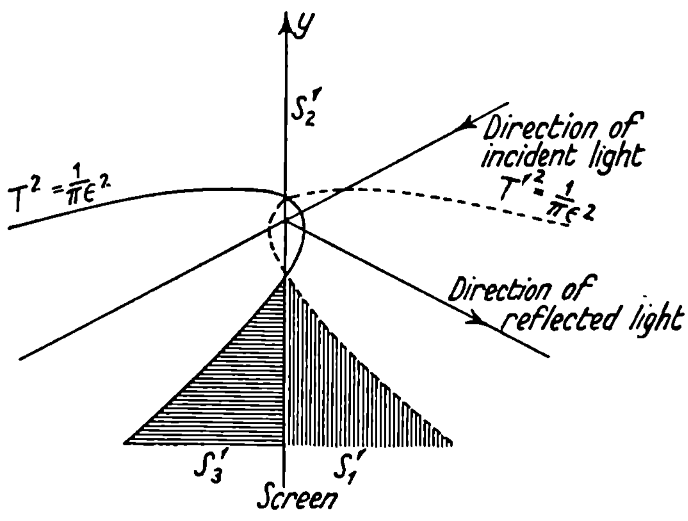


FIG. 26

vector $\phi = -\phi'$, we leave out of consideration the part of physical space within the parabola $\pi\epsilon^2T'^2 = 1$. In the part of S_1 which lies outside this parabola, we have

$$T' > \frac{1}{\epsilon\sqrt{\pi}},$$

but in the parts of S_2 and S_3 outside the parabola,

$$T' < -\frac{1}{\epsilon\sqrt{\pi}}.$$

We denote by S_1', S_2', S_3' the parts of S_1, S_2, and S_3 respectively which lie outside both parabolas.

By using the results of §§ 1.6 and 1.7 we readily obtain the following approximate formulae:

$$u(\rho,\phi;\,\phi') = u_0(\rho,\phi;\,\phi') - \frac{e^{-ik\rho+ikct-\frac{1}{4}\pi i}}{2\cos\frac{1}{2}(\phi-\phi')\sqrt{(2\pi k\rho)}} \quad \text{(in } S_1' \text{ and } S_2'\text{)},$$

$$u(\rho,\phi;\,\phi') = -\frac{e^{-ik\rho+ikct-\frac{1}{4}\pi i}}{2\cos\frac{1}{2}(\phi-\phi')\sqrt{(2\pi k\rho)}} \quad \text{(in } S_3'\text{)},$$

$$u(\rho,\phi;\,3\pi-\phi') = u_0(\rho,\phi;\,\pi-\phi') + \frac{e^{-ik\rho+ikct-\frac{1}{4}\pi i}}{2\sin\frac{1}{2}(\phi+\phi')\sqrt{(2\pi k\rho)}} \quad \text{(in } S_1'\text{)},$$

$$u(\rho,\phi;\,3\pi-\phi') = \frac{e^{-ik\rho+ikct-\frac{1}{4}\pi i}}{2\sin\frac{1}{2}(\phi+\phi')\sqrt{(2\pi k\rho)}} \quad \text{(in } S_2' \text{ and } S_3'\text{)};$$

the error in each case being less than $\frac{1}{2}\pi\epsilon^3$. We now apply these approximations to the diffraction problems considered in the preceding section.

When the incident light is polarized perpendicular to the edge of the screen, the total resultant field is given by

$$d_z = d_z^* + d_z^B,$$

where d_z^* represents the effect according to geometrical optics† and

$$d_z^B = -\frac{e^{-ik\rho+ikct-\frac{1}{4}\pi i}}{2\sqrt{(2\pi k\rho)}}\{\sec\tfrac{1}{2}(\phi-\phi')+\operatorname{cosec}\tfrac{1}{2}(\phi+\phi')\} \tag{2.21}$$

gives the diffracted light in the regions S_1', S_2', S_3' with an error less than $\frac{1}{2}\pi\epsilon^3$. To the same order, the magnetic vector in the diffracted light is found to be

$$\mathbf{h}^B = -(\mathbf{i}\sin\phi-\mathbf{j}\cos\phi)\frac{e^{-ik\rho+ikct-\frac{1}{4}\pi i}}{2\sqrt{(2\pi k\rho)}}\{\sec\tfrac{1}{2}(\phi-\phi')+\operatorname{cosec}\tfrac{1}{2}(\phi+\phi')\}.$$

When the incident light is polarized parallel to the edge of the screen the total resultant field is given by

$$h_z = h_z^* + h_z^B,$$

where h_z^* represents the effect according to geometrical optics and

$$h_z^B = -\frac{e^{-ik\rho+ikct-\frac{1}{4}\pi i}}{2\sqrt{(2\pi k\rho)}}\{\sec\tfrac{1}{2}(\phi-\phi')-\operatorname{cosec}\tfrac{1}{2}(\phi+\phi')\} \tag{2.22}$$

† By this, we mean that there is the ordinary incident and reflected light in S_1, only the incident light in S_2, and darkness in S_3.

gives the diffracted light in the regions S_1', S_2', S_3' with an error less than $\frac{1}{2}\pi\epsilon^3$. To the same order the electric vector in the diffracted light is found to be

$$\mathbf{d}^B = (\mathbf{i}\sin\phi - \mathbf{j}\cos\phi)\frac{e^{-ik\rho+ikct-\frac{1}{4}\pi i}}{2\sqrt{(2\pi k\rho)}}\{\sec\tfrac{1}{2}(\phi-\phi') - \operatorname{cosec}\tfrac{1}{2}(\phi+\phi')\}.$$

The diffracted light rays are all perpendicular to the edge of the screen since, in both cases, the diffracted light diverges from the edge like a non-isotropic cylindrical wave. Thus, when the eye is focused on the edge of the screen, the edge should appear luminous although it is not a true source of light. This phenomenon can actually be observed in the geometrical shadow where it is not masked by the much greater intensity of the incident light. In the region S_2 the incident and diffracted light interfere and produce the interference fringes observed experimentally near the boundary of the geometrical shadow. (Cf. Ch. II, § 3.3).

The amplitude of the magnetic vector in the diffracted light is given by the approximate formulae

$$A_\perp = \frac{1}{2\sqrt{(2\pi k\rho)}}\{\sec\tfrac{1}{2}(\phi-\phi') + \operatorname{cosec}\tfrac{1}{2}(\phi+\phi')\}$$

or

$$A_\parallel = \frac{1}{2\sqrt{(2\pi k\rho)}}\{\sec\tfrac{1}{2}(\phi-\phi') - \operatorname{cosec}\tfrac{1}{2}(\phi+\phi')\},$$

according as the incident light is polarized perpendicular or parallel to the edge of the screen. The ratio of these amplitudes is

$$\frac{A_\parallel}{A_\perp} = -\cot(\tfrac{1}{4}\pi + \tfrac{1}{2}\phi')\cot(\tfrac{1}{4}\pi + \tfrac{1}{2}\phi).$$

This should be compared with the equation $A_\parallel = A_\perp$ obtained in the theory of the black screen (Chapter III, § 4.2).

The angle between the diffracted ray at a point P and an incident ray produced through P is called the *angle of diffraction*, the plane of these rays the *plane of diffraction*. In the present case the angle of diffraction at (ρ, ϕ) is

$$\delta = \phi - \phi' - \pi,$$

the plane of diffraction being the plane of incidence. With this notation,

$$\frac{A_\parallel}{A_\perp} = \cot(\tfrac{1}{4}\pi + \tfrac{1}{2}\phi')\tan(\tfrac{1}{4}\pi + \tfrac{1}{2}\delta + \tfrac{1}{2}\phi'). \tag{2.23}$$

We notice at once one important difference between this theory and the theory of the black screen; for the diffraction effect at a point P due to a black half-plane depends only on the angle of diffraction and not on the angle of incidence, whereas with a reflecting screen it depends on both.

Although the diffracted light is not plane-polarized, it does behave like plane-polarized light at large distances, and so we may still call the plane which contains the diffracted ray and the magnetic vector $\mathbf{h}^B$ the plane of polarization.

The plane of polarization of the diffracted light is parallel to that of the incident light in the two cases we have considered so far. We now consider the more general case of plane-polarized incident light in which the plane of polarization makes an angle α with the plane of incidence $z = 0$. The incident field is then

$$\mathbf{d} = (\mathbf{i}\sin\phi'\sin\alpha - \mathbf{j}\cos\phi'\sin\alpha + \mathbf{k}\cos\alpha)e^{ik\rho\cos(\phi-\phi')+ikct},$$

$$\mathbf{h} = (-\mathbf{i}\sin\phi'\cos\alpha + \mathbf{j}\cos\phi'\cos\alpha + \mathbf{k}\sin\alpha)e^{ik\rho\cos(\phi-\phi')+ikct}.$$

The resultant field† turns out to be

$$\mathbf{d} = \mathbf{d}^* + \mathbf{d}^B, \qquad \mathbf{h} = \mathbf{h}^* + \mathbf{h}^B,$$

where $\mathbf{d}^*$, $\mathbf{h}^*$ is the field of geometrical optics, and

$$\mathbf{h}^B = -\{A_\perp \mathbf{i}\sin\phi\cos\alpha - A_\perp \mathbf{j}\cos\phi\cos\alpha + A_\parallel \mathbf{k}\sin\alpha\}e^{-ik\rho+ikct-\frac{1}{4}\pi i} \tag{2.24}$$

is the approximate magnetic force in the diffracted light.

The amplitude of $\mathbf{h}^B$ is

$$A = \sqrt{\{A_\perp^2\cos^2\alpha + A_\parallel^2\sin^2\alpha\}}$$

$$= A_\perp\sqrt{\left\{\cos^2\alpha + \frac{\tan^2(\frac{1}{4}\pi + \frac{1}{2}\delta + \frac{1}{2}\phi')}{\tan^2(\frac{1}{4}\pi + \frac{1}{2}\phi')}\sin^2\alpha\right\}}.$$

The intensity of the diffracted light is proportional to A^2 and so varies with α. Moreover, since $0 < \delta < \frac{1}{2}\pi - \phi'$ in the geometrical shadow, the intensity is a maximum when $\alpha = \frac{1}{2}\pi$, that is, when the incident light is polarized parallel to the edge of the screen. This effect does not occur with a black screen since $A_\parallel = A_\perp$ in that case.

The plane of polarization of the diffracted light (2.24) at the point (ρ, ϕ, z') is

$$-A_\parallel x\sin\phi\sin\alpha + A_\parallel y\cos\phi\sin\alpha + A_\perp(z - z')\cos\alpha = 0.$$

This plane makes an angle

$$\tan^{-1}\left(\frac{A_\parallel}{A_\perp}\tan\alpha\right)$$

† This is obtained by supposing that the incident light is the result of superposing a field with $\alpha = 0$ on a field with $\alpha = \frac{1}{2}\pi$.

with the plane of diffraction $z = z'$, so that the plane of polarization is rotated in diffraction by a perfectly reflecting screen.† Since

$$\frac{A_{\parallel}}{A_{\perp}} = \frac{\tan\frac{1}{2}(\frac{1}{2}\pi+\delta+\phi')}{\tan\frac{1}{2}(\frac{1}{2}\pi+\phi')},$$

the direction of rotation depends on the angle δ. In the geometrical shadow $A_{\parallel}/A_{\perp} > 1$ since $0 < \delta < \frac{1}{2}\pi-\phi'$, and so the plane of polarization is more nearly perpendicular to the plane of diffraction in the diffracted than in the incident light.

The phenomenon of the rotation of the plane of polarization in diffraction was first discovered theoretically by Stokes,‡ who found experimentally a rotation in the sense opposite to that which we have just found—a result of considerable interest from the point of view of the elastic-solid theory of light with which Stokes worked. The experiment was repeated by Holtzmann, L. Lorenz, and Quincke with very divergent results. All these experiments were conducted with a grating, and, as Poincaré pointed out,‖ the effect looked for may be masked by a rotation due to refraction. The experiments of Gouy with a very acute-angled reflecting wedge give a rotation in the sense which the present theory indicates. The experiments of Wien†† with very sharp steel blades require a more rapid increase of $A_{\parallel}: A_{\perp}$ with δ than is given by (2.23), which may be due to the finite thickness‡‡ of the blade or to its not being a perfect conductor.

§ 3. Diffraction of plane-polarized light by a black half-plane

§ 3.1. Voigt's theory

When plane-polarized light specified by

$$\mathbf{d} = \mathbf{k}e^{ik\rho\cos(\phi-\phi')+ikct}$$

is incident on a reflecting screen $x = 0$, $y < 0$, the effect is specified, as we have seen, by

$$\mathbf{d} = \mathbf{k}\{u(\rho,\phi;\,\phi')-u(\rho,\phi;\,3\pi-\phi')\}.$$

† There is no rotation with a black screen, as is confirmed by experiments of Gouy. *Ann. de chim. et de phys.* (6), **8** (1886), 145.

‡ In his paper on the 'Dynamical Theory of Diffraction', *Trans. Camb. Phil. Soc.* **9** (1849), 1, reprinted in Stokes's *Papers*, **2**, 243.

‖ *Théorie math. de la Lumière*, **2** (1892), 213–26.

†† *Wied. Ann.* **28** (1886), 117.

‡‡ The problem of diffraction of plane electromagnetic waves by a thick half-plane has been discussed by Hanson, *Phil. Trans.* (A) **229** (1930), 87–124. His method is to consider two parallel thin half-planes whose distance apart is small. See also A. E. Heins, *Quarterly App. Math.* **5** (1947), 157–66; **6** (1948), 215–20.

In this equation the first term represents the effect of the incident light, the second the effect of the reflected light. Voigt† suggested that, if the screen is black, the effect should be obtained by omitting the term due to the reflected light, so that

$$\mathbf{d} = \mathbf{k}u(\rho,\phi;\phi'). \tag{3.11}$$

Similarly, when the incident light is polarized parallel to the edge of the black screen,

$$\mathbf{h} = \mathbf{k}u(\rho,\phi;\phi').$$

There is no theoretical reason for making this assumption, since an opaque non-reflecting screen must absorb all the incident light, which is impossible if it is very thin. It is, however, interesting to work out the consequences of the assumption, especially as it gives results which differ from those obtained by using Kottler's modified Larmor-Tedone formula.

In the first place, it suffices to consider only the case when the incident light is polarized perpendicular to the edge of the screen, since the results when the plane of polarization is parallel to the edge can be obtained by the transformation

$$\mathbf{d} \to \mathbf{h}, \qquad \mathbf{h} \to -\mathbf{d}.$$

It follows immediately that there will be no rotation of the plane of polarization, just as in Kottler's theory.

The field specified by (3.11) can be written in the form

$$\mathbf{d} = \mathbf{d}^* + \mathbf{d}^B, \qquad \mathbf{h} = \mathbf{h}^* + \mathbf{h}^B,$$

where $(\mathbf{d}^*, \mathbf{h}^*)$ represents the field of geometrical optics and $(\mathbf{d}^B, \mathbf{h}^B)$ the field of the diffracted light. In the part of physical space $(-\frac{1}{2}\pi < \phi < \frac{3}{2}\pi)$ outside the parabola

$$2k\rho\cos^2\tfrac{1}{2}(\phi-\phi') = 1/(\pi\epsilon^2),$$

we have

$$\mathbf{d}^B = -\mathbf{k}\frac{e^{-ik\rho+ikct-\frac{1}{4}\pi i}}{2\cos\frac{1}{2}(\phi-\phi')\sqrt{(2\pi k\rho)}},$$

$$\mathbf{h}^B = -(\mathbf{i}\sin\phi-\mathbf{j}\cos\phi)\frac{e^{-ik\rho+ikct-\frac{1}{4}\pi i}}{2\cos\frac{1}{2}(\phi-\phi')\sqrt{(2\pi k\rho)}},$$

with an error less than $\frac{1}{4}\pi\epsilon^3$. The diffracted light, as before, diverges like a non-isotropic cylindrical wave from the edge of the screen; the apparent brightness of the edge again agrees with experiment.

There is, however, an important difference from the theory of the reflecting screen in that the diffracted light depends only on the angle

† *Gött. Nach.* (1899), 1.

of diffraction δ, so that diffraction is here purely an edge effect. The amplitude of the diffracted light is

$$A = \frac{1}{2\sin\frac{1}{2}\delta\sqrt{(2\pi k\rho)}}. \tag{3.12}$$

This should be compared with the amplitude

$$A = \frac{1}{2\tan\frac{1}{2}\delta\sqrt{(2\pi k\rho)}} \tag{3.13}$$

given by Kottler's formulae or by Kirchhoff's scalar theory in its exact form.

Actually, (3.13) agrees with experiment better than (3.12), so that Voigt's black screen is not black enough. It has been suggested† that matters could be improved by increasing the number of sheets p of Sommerfeld's Riemann surface; for then light would be less likely to enter physical space from non-physical space. In particular by making $p \to \infty$, we obtain Voigt's 'blackest' screen. The corresponding solution is then

$$\mathbf{d} = \mathbf{k}\frac{e^{ikct}}{2\pi i}\int\limits_{A} e^{ik\rho\cos\zeta}\frac{d\zeta}{\zeta-\phi'+\phi}, \tag{3.14}$$

where A is the two-branched path of Fig. 23 on p. 136.

By using Cauchy's theorem it can be shown that, in this field,

$$\mathbf{d} = \mathbf{d}^*+\mathbf{d}^B,$$

where $\mathbf{d}^*$ is the electric force according to the laws of geometrical optics and

$$\mathbf{d}^B = -2\mathbf{k}e^{ikct}\int\limits_0^\infty e^{-ik\rho\cosh\tau}\frac{\pi^2-\theta^2+\tau^2}{(\pi^2-\theta^2+\tau^2)^2+4\theta^2\tau^2}\,d\tau, \tag{3.15}$$

(where $\theta = \phi-\phi'$) is the electric force in the diffracted light.

The approximate formula

$$\mathbf{d}^B = -\mathbf{k}\frac{2\pi e^{-ik\rho+ikct-\frac{1}{4}\pi i}}{(\pi^2-\theta^2)\sqrt{(2\pi k\rho)}},$$

valid when ρ is large compared with the wave-length $2\pi/k$, can be deduced from (3.15) by the principle of stationary phase. Hence the amplitude of the diffracted light with Voigt's 'blackest' screen is

$$A = \frac{1}{\delta\left(1+\frac{\delta}{2\pi}\right)\sqrt{(2\pi k\rho)}}, \tag{3.16}$$

when ρ is large.

† See, for example, Sommerfeld, *Zeits. f. Math. u. Phys.* **46** (1901), 11–97, § 5.

Formulae (3.12), (3.13), and (3.16) all give the same approximate amplitude $1/\{\delta\sqrt{(2\pi k\rho)}\}$, when δ, the angle of diffraction, is small, but do not agree when the angle of diffraction is large.

§ 3.2. The unsatisfactory state of the theory of the black screen

Sommerfeld's theory of diffraction by a perfectly reflecting screen can be regarded as a satisfactory one. The boundary conditions satisfied by his solution are easy to formulate on theoretical grounds, and the structure of his 'Riemann space' is fixed by the conditions of the problem. On the other hand, although each of the theories of diffraction by a black screen gives results which agree fairly well with the available experimental evidence, not one of them rests on a sound theoretical basis.

Voigt's work suffers from two disadvantages: the electric and magnetic vectors satisfy no prescribed boundary conditions and the number of sheets in the 'Riemann space' is quite arbitrary. Kottler tried to overcome these difficulties by regarding the problem as a saltus problem instead of a boundary-value problem. After a critical study of Kottler's work, Ignatowsky† has concluded that there is no definition of a black screen applicable in all cases, and he asserts that each black-screen diffraction problem must be discussed independently on its own merits.

The real difficulty lies in the fact that the black screen is an idealization which cannot be attained experimentally and which has no precise definition in electromagnetic theory. What is needed is a rigorous theory which does not assume that the screen is perfectly reflecting but takes into account the properties of the material of the screen. An empirical way of finding the effect due to a screen which is neither perfectly black nor perfectly reflecting is to suppose that the reflected wave has its amplitude and phase reduced in some definite way, due to the imperfections of the reflecting power of the screen. This amounts to multiplying the term representing the reflected wave in Sommerfeld's solution by a complex constant. The effect of this modification of Sommerfeld's theory has been worked out in some detail by Raman and Krishnan.‡

† *Annalen d. Phys.* **77** (1925), 589–643.
‡ *Proc. R.S.* (A), **116** (1927), 254.

V

DIFFRACTION BY A PLANE SCREEN

§ 1. Introduction

§ 1.1. Recent results in the theory of diffraction

SINCE 1939 there have been many interesting developments in the theory of diffraction. Work in the theory of sound was concerned mainly with the diffraction of sound pulses. Friedlander† discussed the problem of diffraction by a perfectly reflecting half-plane; using the formulae of Sommerfeld and Lamb, he considered the problem in some detail and gave interesting numerical results for different shapes of pulse. E. N. Fox‡ considered the diffraction of sound pulses of arbitrary form by an infinitely long strip or slit; he used an integral equation method, and his solution by successive substitutions converges rapidly and is well suited for obtaining numerical results.

The first use of integral equations in diffraction theory seems to be due to Rayleigh,|| who deduced approximate solutions for the diffraction of plane waves normally incident on a perfectly reflecting plane having an aperture or slit whose dimensions were small compared with the wave-length. A little later, Schwarzschild†† formulated the problem of diffraction at a slit virtually by means of a pair of simultaneous integral equations. Actually Fox's integral equation can be obtained from Schwarzschild's (for normal incidence) by a Laplace transform; but in the case of monochromatic waves considered by Schwarzschild, the convergence of the process of successive substitutions is too slow to be of practical use.

More recently Magnus‡‡ reduced the problem of diffraction of plane waves by a half-plane to the solution of what is now called a Wiener-Hopf integral equation, which he solved by using an infinite series of Bessel Functions. In America during the war Schwinger and others showed that certain diffraction problems of importance in

† *Proc. Roy. Soc.* (A) **186** (1946), 322–67.

‡ *Phil. Trans.* (A), **241** (1948), 71–103; **242** (1949), 1–32.

|| *Phil. Mag.* **43** (1897), 259–72; *Proc. Roy. Soc.* (A), **89** (1913), 194–219.

†† *Math. Ann.* **55** (1902), 177–247.

‡‡ *Zeitschrift f. Phys.* **117** (1941), 168–79. See also Copson, *Quarterly J. Math.* (Oxford), **17** (1946), 19–34, where the integral equation is solved by complex Fourier transforms.

radar can be formulated as Wiener-Hopf integral equations, which they solved by means of complex Laplace transforms.†

In the present chapter an account is given of the application of integral equation methods to plane diffraction problems, mainly in the theory of sound. The extension to electromagnetic problems is more complicated.‡

§ 1.2. The radiation condition

In this chapter we shall be concerned with wave-functions of the form $u = ve^{i\omega t}$, where $\omega > 0$. The function v is independent of t, and satisfies the equation

$$(\nabla^2+k^2)v = 0. \tag{1.21}$$

The factor $e^{i\omega t}$ will usually be omitted.

If u is the velocity-potential of sound waves, k is real and positive, except in the case when the gas has slight viscosity; in this exceptional case, k is a complex number of the form $p-iq$ where p and q are positive and q is small. Similarly if u is a component of the electric or magnetic vector in monochromatic waves, k is real and positive, except when the medium has slight conductivity; in the exceptional case, k is again of the form $p-iq$.

The assumption that k is complex simplifies the analysis considerably. It will be recalled that, in Chapter I, it was found necessary to introduce two conditions, due to Sommerfeld,‖ in order to ensure that the exterior boundary-value problem for equation (1.21) had a unique solution and represented expanding waves. The situation is much simpler when $k = p-iq$; for then the velocity-potential of waves diverging from a point source is

$$\frac{1}{r}e^{i\omega t-ikr} = \frac{1}{r}e^{i(\omega t-pr)-qr}$$

whereas that for waves converging to a focus is

$$\frac{1}{r}e^{i\omega t+ikr} = \frac{1}{r}e^{i(\omega t+pr)+qr}.$$

† Amongst the papers which have appeared are J. F. Carlson and A. E. Heins, *Quarterly App. Math.* **4** (1947), 313–29, **5** (1947), 82–8; A. E. Heins, ibid. **5** (1947), 157–66, **6** (1948), 215–20; A. E. Heins and H. Feshbach, *Journal Math. Phys.* **26** (1947), 143–55; H. Levine and J. Schwinger, *Phys. Rev.* **73** (1948), 383–406; J. W. Miles, *Journal Acoustical Soc. America*, **20** (1948), 370–4. An account of Schwinger's work is promised in a forthcoming M.I.T. Radiation Laboratory publication, *Theory of Guided Waves.*

‡ See Copson, *Proc. Roy. Soc.* (A), **186** (1946), 100–18; Magnus, *Jahresbericht der D.M.V.* **52** (1943), 177–88; J. W. Miles, *Journal Appl. Phys.* **20** (1949), 760–71.

‖ See pp. 25, 28, where, however, a different convention concerning the time-factor is used.

It follows that a wave-function which represents waves divergent from a distribution of sources at a finite distance must vanish at infinity, being of the order of e^{-qr}/r when r is large. On the other hand, a wave-function which contains terms representing convergent waves must tend to infinity as r tends to infinity. The physical reason for this is evident, since the medium dissipates energy; q is usually called the attenuation constant.

When $k = p - iq$, Sommerfeld's conditions can therefore be replaced by the simpler condition, that a wave-function, which represents only divergent waves, and its first partial derivatives are uniformly bounded as r tends to infinity. Under this simple condition the exterior boundary-value problems for equation (1.21) have unique solutions, given by the formulae of the next section. In particular, a wave-function which satisfies this condition and has no singularity anywhere in space is null.

§ 1.3. The solution of boundary-value problems by means of Green's function

When sound waves with velocity-potential v_i, the time-factor $e^{i\omega t}$ being understood throughout, are incident on a perfectly reflecting body bounded by a closed surface S, the problem of reflection and diffraction consists in determining the velocity-potential v_s of the scattered waves. The function v_s satisfies the following conditions:

(i) it is a solution of $(\nabla^2 + k^2)v = 0$ with no singularities outside S;

(ii) it satisfies the radiation condition at infinity;

(iii) on the surface S, $\partial v_s/\partial n = -\partial v_i/\partial n$ where $\mathbf{n}$ is the outward normal unit vector.

This is a boundary-value problem in which the value of $\partial v_s/\partial n$ on S is given; it is usually called the exterior Neumann problem.

It is well known that a function which satisfies conditions (i) and (ii) is uniquely determined by the boundary values on S either of v or of $\partial v/\partial n$; in fact,

$$v(P_0) = \frac{1}{4\pi} \iint_S v(P) \frac{\partial G_1(P_0, P)}{\partial n}\, dS \tag{1.31}$$

and

$$v(P_0) = -\frac{1}{4\pi} \iint_S \frac{\partial v(P)}{\partial n} G_2(P_0, P)\, dS \tag{1.32}$$

where G_1 and G_2 are the Green's functions of the first and second kinds,† and dS is the element of surface at the integration point P.

These formulae are, however, of little use, not merely because the determination of the appropriate Green's function is often difficult, but because its determination is actually equivalent to the solution of a boundary-value problem of the very type we are considering. For instance, $G_2(P_0, P)$ is the total velocity-potential at P of the incident and scattered waves when the incident waves are due to a point source at P_0. If this diffraction problem were solved, then (1.32) would give the solution of all other diffraction problems with the same reflecting surface S.

A further difficulty arises in the discussion of the diffraction of electromagnetic waves by a perfectly reflecting surface S. When monochromatic waves with electric and magnetic vectors $\mathbf{d}^i$ and $\mathbf{h}^i$ are incident on S, we have to find the scattered field $\mathbf{d}^s$, $\mathbf{h}^s$ which satisfies the following conditions:

(i) the scattered field satisfies Maxwell's equations

$$ik\mathbf{d} = \operatorname{curl}\mathbf{h}, \quad ik\mathbf{h} = -\operatorname{curl}\mathbf{d}, \quad \operatorname{div}\mathbf{d} = 0, \quad \operatorname{div}\mathbf{h} = 0;$$

(ii) each component of $\mathbf{d}^s$ and $\mathbf{h}^s$ satisfies the radiation condition;

(iii) on the reflecting surface S,

$$\mathbf{d}^s \times \mathbf{n} = -\mathbf{d}^i \times \mathbf{n}, \qquad \mathbf{h}^s . \mathbf{n} = -\mathbf{h}^i . \mathbf{n},$$

since the tangential component of the total electric force $\mathbf{d} = \mathbf{d}^i + \mathbf{d}^s$ and the normal component of the total magnetic force $\mathbf{h} = \mathbf{h}^i + \mathbf{h}^s$ vanish on S.

The difficulty to which we referred is that, if v were a rectangular Cartesian component of $\mathbf{d}^s$ or of $\mathbf{h}^s$, the boundary conditions (iii) would not, in general, determine either v or $\partial v/\partial n$ on S. Hence, even if we could solve the scalar diffraction problems involved in the determination of the Green's functions, the formulae (1.31) and (1.32) would, in general, not suffice to determine the scattered electromagnetic field. Moreover, the difficulty is not avoided by using vector solutions of Maxwell's equations, such as the formulae of Larmor and Tedone, since these formulae require a complete knowledge of $\mathbf{d}^s$ and $\mathbf{h}^s$ on S. There is no special merit in a vector solution; for if the correct boundary values could be inserted in (1.31) and (1.32), the resulting expressions for $\mathbf{d}^s$ and $\mathbf{h}^s$ would automatically satisfy Max-

† See, for example, Bateman, *Partial Differential Equations of Mathematical Physics* (Cambridge, 1932), 140.

well's equations, since the boundary-value problems have unique solutions. Unfortunately, it is, in general, not possible to do this, since only the tangential component of $\mathbf{d}^s$ and the normal component of $\mathbf{h}^s$ are known on a reflecting surface.

§ 1.4. Reflection by a plane

When the reflecting surface is a plane, the Green's function formulae (1.31) and (1.32) become very simple, and can be made the basis of an analytical formulation of problems of diffraction by a plane screen in terms of integral equations. The formulae are as follows:

Let v be a solution of $(\nabla^2+k^2)v = 0$ which has no singularities in the half-space $x \geqslant 0$ and satisfies the radiation condition at infinity.† *Let $v = \Phi(y,z)$, $\partial v/\partial x = \Psi(y,z)$ when $x = 0$. Then, when $x_0 > 0$,*

$$v(x_0,y_0,z_0) = -\frac{1}{2\pi}\frac{\partial}{\partial x_0}\iint \Phi(y,z)\frac{e^{-ikR}}{R}dydz \tag{1.41}$$

and

$$v(x_0,y_0,z_0) = -\frac{1}{2\pi}\iint \Psi(y,z)\frac{e^{-ikR}}{R}dydz \tag{1.42}$$

where

$$R = +\sqrt{\{x_0^2+(y-y_0)^2+(z-z_0)^2\}}$$

and integration is over the whole plane $x = 0$.

The proof depends on Helmholtz's‡ formula. Let S be the surface bounding the hemisphere $r \leqslant a$, $x \geqslant 0$; and let (x_0,y_0,z_0) be a point inside S. Then

$$v(x_0,y_0,z_0) = \frac{1}{4\pi}\iint_S \left\{\frac{e^{-ikR}}{R}\frac{\partial v}{\partial n} - v\frac{\partial}{\partial n}\left(\frac{e^{-ikR}}{R}\right)\right\} dS$$

where R is the distance from (x_0,y_0,z_0) to the integration-point (x,y,z) on S. If, however, $x_0 < 0$, the value of this integral is zero.

The contribution of the curved part of S to the surface-integral is

$$\frac{1}{4\pi}\iint_S \left[\frac{r^2}{R^2}e^{-ikR}\left\{R\frac{\partial v}{\partial r}+(1+ikR)v\frac{\partial R}{\partial r}\right\}\right]_{r=a} d\omega$$

where $d\omega$ is the element of solid angle subtended by dS at the origin. But since $|e^{-ikR}| = e^{-qR}$ and since v and $\partial v/\partial r$ are uniformly bounded

† Either in the original sense if $k > 0$, or in the sense of § 1.2 if $k = p-iq$. The latter is used in the proof.

‡ See p. 24. Since the time-factor is now $e^{i\omega t}$, e^{-ikR} appears instead of e^{ikR}.

as $r \to \infty$, this integral evidently tends to zero as $a \to \infty$. Hence we have, when $x_0 > 0$,

$$v(x_0, y_0, z_0) = \frac{1}{4\pi} \iint \left[\frac{e^{-ikR}}{R} \frac{\partial v}{\partial n} - v \frac{\partial}{\partial n}\left(\frac{e^{-ikR}}{R}\right) \right]_{x=0} dydz \qquad (1.43)$$

where integration is over the whole plane $x = 0$. But, since $\mathbf{n}$ is the outward normal to S, it follows that

$$\frac{\partial v}{\partial n} = -\frac{\partial v}{\partial x} = -\Psi(y, z).$$

Also, since $R = +\sqrt{\{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2\}}$, we have

$$\frac{\partial}{\partial x}\left(\frac{e^{-ikR}}{R}\right) = -\frac{\partial}{\partial x_0}\left(\frac{e^{-ikR}}{R}\right).$$

Hence, if we insert the given boundary values in (1.43), we obtain

$$v(x_0, y_0, z_0) = -\frac{1}{4\pi} \iint \Psi(y, z) \frac{e^{-ikR}}{R} dydz - \frac{1}{4\pi} \frac{\partial}{\partial x_0} \iint \Phi(y, z) \frac{e^{-ikR}}{R} dydz, \qquad (1.44)$$

where R now denotes $+\sqrt{\{x_0^2+(y-y_0)^2+(z-z_0)^2\}}$.

If, however, x_0 is negative, the left-hand side of (1.44) has to be replaced by zero. Hence, if we suppose $x_0 > 0$ and write $-x_0$ for x_0 in the right-hand side of (1.44), we find that

$$0 = -\frac{1}{4\pi} \iint \Psi(y, z) \frac{e^{-ikR}}{R} dydz + \frac{1}{4\pi} \frac{\partial}{\partial x_0} \iint \Phi(y, z) \frac{e^{-ikR}}{R} dydz. \qquad (1.45)$$

The required results follow by adding and subtracting (1.44) and (1.45).

There is a corresponding theorem for the half-space $x \leqslant 0$, in which the signs of the expressions on the right-hand sides of (1.41) and (1.42) are changed.

When sound waves with velocity-potential $v_i(x, y, z)$ are incident on the positive side of the perfectly reflecting infinite plane $x = 0$, the velocity-potential $v_s(x, y, z)$ of the reflected waves satisfies the condition $\partial(v_i+v_s)/\partial x = 0$ on $x = 0$. Hence, if we put

$$\Psi(y, z) = -[\partial v_i/\partial x]_{x=0},$$

equation (1.42) will give v_s for $x_0 > 0$. This is not, however, of much importance since the method of images shows at once that

$$v_s(x, y, z) = v_i(-x, y, z).$$

Again, if electromagnetic waves $\mathbf{d}^i$, $\mathbf{h}^i$ are incident on the positive side of the perfectly reflecting plane $x = 0$, the electric and magnetic vectors $\mathbf{d}^s$, $\mathbf{h}^s$ in the reflected waves satisfy the boundary conditions

$$d_y^s = -d_y^i, \qquad d_z^s = -d_z^i, \qquad h_x^s = -h_x^i$$

on $x = 0$, so that (1.41) would determine d_y^s, d_z^s, and h_x^s, and the remaining components could then be found by Maxwell's equations; but again the method of images is simpler.

§ 2. Diffraction of sound by a plane screen

§ 2.1. The formulation of the problem as an integral equation†

Let us consider the case of an infinite perforated screen occupying the plane $x = 0$ and containing apertures of arbitrary shape and size, the screen itself being perfectly reflecting. Let 'monochromatic' sound waves be incident on the positive side of this screen, the velocity-potential of the incident waves being $v_i(x, y, z)$ in the absence of the screen. These waves are reflected and diffracted by the screen, the velocity-potential of the reflected and diffracted waves being $v_s(x, y, z)$. The total velocity-potential in the actual problem when the screen is present is

$$v(x, y, z) = v_i(x, y, z) + v_s(x, y, z). \tag{2.11}$$

The velocity-potential v_s has to be chosen to satisfy the radiation condition at infinity and the boundary-condition $\partial v/\partial x = 0$ on the material of the screen but not in the apertures.

Since the problem is linear we can superimpose the corresponding mirror-image problem, in which waves with velocity-potential $v_i(-x, y, z)$ are incident on the negative side of the screen; in this problem, the total velocity-potential is $v(-x, y, z)$. In the combined problem the boundary condition on the screen is satisfied automatically by $v_i(x, y, z) + v_i(-x, y, z)$; hence we can now remove the screen without altering the problem. But this implies that

$$v(x, y, z) + v(-x, y, z) = v_i(x, y, z) + v_i(-x, y, z). \tag{2.12}$$

This shows incidentally that if we can find v for $x > 0$, we can deduce at once the value of v for $x < 0$.

If we now let x tend to zero by positive values, (2.12) gives

$$v(+0, y, z) + v(-0, y, z) = 2v_i(0, y, z).$$

The values of v on the front and back faces of the screen are different,

† The corresponding general theorems for sound pulses are given by E. N. Fox, *Phil. Trans.* (A), **241** (1948), 71–103.

but v is continuous through each aperture. Hence, if $(0, y, z)$ is a point in one of the apertures,

$$v(0, y, z) = v_i(0, y, z),$$

that is, the total velocity-potential in an aperture is equal to the incident velocity-potential there. In other words, v_s vanishes in each aperture. The boundary-conditions satisfied by v_s are therefore of the following unusual type:

(i) $$\frac{\partial v_s}{\partial x} = -\frac{\partial v_i}{\partial x}$$

on both faces of the screen, and

(ii) v_s vanishes in each aperture.

This enables us to reduce the diffraction problem to the solution of an integral equation, in the following way.†

Let 'monochromatic' sound waves of velocity-potential $v_i(x, y, z)$ be incident on the positive side of an infinite perforated perfectly reflecting screen occupying the plane $x = 0$. Let the apertures in the screen be denoted by S_1, the screen itself by S_2. Then the total velocity-potential is given by

$$v(x_0, y_0, z_0) = v_i(x_0, y_0, z_0) + v_i(-x_0, y_0, z_0) - \frac{1}{2\pi}\iint\limits_{S_1} f(y, z)\frac{e^{-ikR}}{R}\,dydz \tag{2.13}$$

when $x_0 > 0$, and by

$$v(x_0, y_0, z_0) = \frac{1}{2\pi}\iint\limits_{S_1} f(y, z)\frac{e^{-ikR}}{R}\,dydz \tag{2.14}$$

when $x_0 < 0$, where

$$R = +\sqrt{\{x_0^2 + (y - y_0)^2 + (z - z_0)^2\}}.$$

The function $f(y, z)$ satisfies the integral equation

$$\iint\limits_{S_1} f(y, z)\frac{e^{-ikR_0}}{R_0}\,dydz = 2\pi v_i(0, y_0, z_0) \tag{2.15}$$

when $(0, y_0, z_0)$ is a point of S_1 and

$$R_0 = +\sqrt{\{(y - y_0)^2 + (z - z_0)^2\}}.$$

To prove this, we observe that, if we write

$$\left[\frac{\partial v_s}{\partial x}\right]_{x=0} = -\left[\frac{\partial v_i}{\partial x}\right]_{x=0} + f(y, z),$$

† See Rayleigh, *Phil. Mag.* **43** (1897), 259–72; *Proc. Roy. Soc.* (A), **89** (1913), 194–219 for the case of normally incident plane waves.

then, by condition (i), $f(y,z)$ vanishes on S_2 and is unknown on S_1. It follows from (1.42) that, when $x_0 > 0$,

$$v_s(x_0,y_0,z_0) = \frac{1}{2\pi}\iint\limits_{S_1+S_2}\left[\frac{\partial v_i(x,y,z)}{\partial x}\right]_{x=0}\frac{e^{-ikR}}{R}\,dydz - \\ -\frac{1}{2\pi}\iint\limits_{S_1} f(y,z)\frac{e^{-ikR}}{R}\,dydz.$$

Now the first term on the right-hand side of this equation is the value v_s would have in the half-space $x_0 > 0$ if $f(y,z)$ were identically zero, that is, if there were no apertures in the screen; and so this term is simply the velocity-potential $v_i(-x_0,y_0,z_0)$ of the waves reflected by a screen filling up the whole plane $x = 0$. Hence

$$v_s(x_0,y_0,z_0) = v_i(-x_0,y_0,z_0) - \frac{1}{2\pi}\iint\limits_{S_1} f(y,z)\frac{e^{-ikR}}{R}\,dydz. \quad (2.16)$$

The integral equation (2.15) follows at once, since v_s vanishes on S_1 by condition (ii).

Lastly, since the total velocity-potential in $x_0 > 0$ is given by $v = v_i+v_s$, equation (2.13) is an immediate consequence of (2.16), and (2.14) then follows from (2.12).

§ 2.2. Two-dimensional problems

If the incident waves have a velocity-potential $v_i(x,y)$ independent of z and if the apertures in the screen are all bounded by lines parallel to the axis of z, the motion is evidently the same in all planes perpendicular to Oz, and the total velocity-potential v is independent of z; the problem is then a two-dimensional one. In such a case, the theorem of § 2.1 simplifies since the integration with respect to z can be carried out by using the formula

$$\int\limits_{-\infty}^{\infty} e^{-ik\sqrt{(a^2+\zeta^2)}}\frac{d\zeta}{\sqrt{(a^2+\zeta^2)}} = \int\limits_{-\infty}^{\infty} e^{-ika\cosh\theta}\,d\theta = -\pi i H_0^{(2)}(ka),$$

which is valid when $a > 0$ and $k = p-iq$ where $q > 0$. This leads to the following theorem:

Let 'monochromatic' sound waves of velocity-potential $v_i(x,y)$ be incident on the positive side of an infinite perfectly reflecting screen lying in the plane $x = 0$. Let the apertures in the screen be bounded by straight lines parallel to the axis of z, so that the avertures cut the plane $z = 0$

in a set of straight lines L_1 lying on the axis of y. Then the total velocity-potential is

$$v(x_0, y_0) = v_i(x_0, y_0)+v_i(-x_0, y_0)+\tfrac{1}{2}i\int_{L_1} f(y)H_0^{(2)}(k\rho)\,dy \quad (2.21)$$

when $x_0 > 0$, and is

$$v(x_0, y_0) = -\tfrac{1}{2}i\int_{L_1} f(y)H_0^{(2)}(k\rho)\,dy \quad (2.22)$$

when $x_0 < 0$, where $\rho = +\sqrt{\{x_0^2+(y-y_0)^2\}}$.

The function $f(y)$ satisfies the integral equation

$$\int_{L_1} f(y)H_0^{(2)}(k|y-y_0|)\,dy = 2iv_i(0, y_0) \quad (2.23)$$

when $(0, y_0, 0)$ is a point of L_1.

§ 3. Diffraction of electromagnetic waves

§ 3.1. Two-dimensional electromagnetic problems

We proved in § 1.4 of Chapter IV that there are two types of two-dimensional solution of Maxwell's equations

$$ik\mathbf{d} = \operatorname{curl}\mathbf{h}, \quad ik\mathbf{h} = -\operatorname{curl}\mathbf{d}, \quad \operatorname{div}\mathbf{d} = 0, \quad \operatorname{div}\mathbf{h} = 0$$

for monochromatic electromagnetic waves. If the axes are chosen so that the field is the same in all planes perpendicular to the axis of z, the two types are as follows:

(i) A field polarized parallel to the axis of z

$$h_x = h_y = 0, \quad h_z = ikv, \quad d_x = \partial v/\partial y, \quad d_y = -\partial v/\partial x, \quad d_z = 0, \quad (3.11)$$

where $v(x,y)$ is a solution of $(\nabla^2+k^2)v = 0$. For such a field, $\partial v/\partial n$ vanishes on the surface of a perfect reflector. The function v satisfies the same conditions as the velocity-potential of sound waves, and the corresponding diffraction problem is that already discussed in § 2.2.

(ii) A field polarized perpendicular to the axis of z

$$d_x = d_y = 0, \quad d_z = ikw, \quad h_x = -\partial w/\partial y, \quad h_y = \partial w/\partial x, \quad h_z = 0, \quad (3.12)$$

where $w(x,y)$ is a solution of $(\nabla^2+k^2)w = 0$. For such a field, w vanishes on the surface of a perfect reflector, a type of boundary condition we shall discuss in the next section.

§ 3.2. The case when the wave-function vanishes on the screen

Let us consider again the case of an infinite perforated screen occupying the plane $x = 0$ and containing apertures of arbitrary shape and size, the screen itself being of such a material that the total wave-function w vanishes on the screen.† Let 'monochromatic' waves be incident on the positive side of this screen, the wave-function of the incident waves being $w_i(x,y,z)$ in the absence of the screen. These waves are reflected and diffracted by the screen, the wave-function of the reflected and diffracted waves being $w_s(x,y,z)$, so that the total wave-function in the actual problem when the screen is present is

$$w(x,y,z) = w_i(x,y,z)+w_s(x,y,z). \qquad (3.21)$$

The wave-function w_s is a solution of $(\nabla^2+k^2)w_s = 0$, and has to be chosen to satisfy the radiation-condition at infinity and the boundary-condition $w = 0$ on the material of the screen but not in the apertures.

Since the problem is linear, we can superimpose the corresponding mirror-image problem in which waves with wave-function

$$-w_i(-x,y,z)$$

are incident on the negative side of the screen; in this problem the total wave-function is $-w(-x,y,z)$. In the combined problem, the boundary condition on the screen is satisfied automatically by

$$w_i(x,y,z)-w_i(-x,y,z);$$

hence we can now remove the screen without altering the problem. But this implies that

$$w(x,y,z)-w(-x,y,z) = w_i(x,y,z)-w_i(-x,y,z), \qquad (3.22)$$

a relation which enables us to find w for $x < 0$ when w is known for $x > 0$ and which also shows, by making $x \to 0$, that w is continuous across the plane $x = 0$.

If we differentiate (3.22) with respect to x and then make x tend to zero by positive values, we find that

$$\left[\frac{\partial w}{\partial x}\right]_{x=+0}+\left[\frac{\partial w}{\partial x}\right]_{x=-0} = 2\left[\frac{\partial w_i}{\partial x}\right]_{x=0}$$

The values of $\partial w/\partial x$ on the front and back faces of the screen are

† A simple physical interpretation is that $w = \partial v/\partial x$ where v is the velocity-potential of sound-waves.

different, but $\partial w/\partial x$ is continuous through each aperture. Hence if $(0, y, z)$ is a point in one of the apertures,

$$\frac{\partial w}{\partial x} = \frac{\partial w_i}{\partial x}$$

at this point. This implies that $\partial w_s/\partial x$ vanishes in each aperture. The boundary-conditions satisfied by w_s are therefore

(i) $\partial w_s/\partial x$ vanishes in each aperture, and

(ii) $w_s = -w_i$ on both faces of the screen.

This leads to the following theorem:†

Let 'monochromatic' waves of wave-function $w_i(x, y, z)$ be incident on the positive side of an infinite perforated screen occupying the plane $x = 0$, the material of the screen being such that the total wave-function vanishes on the screen. Let the apertures in the screen be denoted by S_1, the screen itself by S_2. Then the total wave-function is given everywhere by

$$w(x_0, y_0, z_0) = w_i(x_0, y_0, z_0) - \frac{1}{2\pi} \iint\limits_{S_2} g(y, z) \frac{e^{-ikR}}{R} \, dydz \qquad (3.23)$$

where $R = +\sqrt{\{x_0^2 + (y-y_0)^2 + (z-z_0)^2\}}$.

The function $g(y, z)$ satisfies the integral equation

$$\iint\limits_{S_2} g(y, z) \frac{e^{-ikR_0}}{R_0} \, dydz = 2\pi w_i(0, y_0, z_0) \qquad (3.24)$$

when $(0, y_0, z_0)$ is a point of S_2 and

$$R_0 = \sqrt{\{(y-y_0)^2 + (z-z_0)^2\}}.$$

To prove this, we write

$$\left[\frac{\partial w_s}{\partial x}\right]_{x=+0} = g(y, z)$$

so that, by condition (i), $g(y, z)$ vanishes on S_1 and is unknown on S_2. Hence, by (1.42),

$$w_s(x_0, y_0, z_0) = -\frac{1}{2\pi} \iint\limits_{S_2} g(y, z) \frac{e^{-ikR}}{R} \, dydz, \qquad (3.25)$$

when $x_0 > 0$. If we make $x_0 \to +0$, the integral equation (3.24) follows from condition (ii). Also, since $w = w_i + w_s$, equation (3.23) is an immediate consequence of (3.25) when $x_0 > 0$; the result for $x_0 < 0$ then follows from (3.22).

† Cf. Rayleigh, loc. cit.

The corresponding two-dimensional result is:

Let 'monochromatic' waves of wave-function $w_i(x,y)$ *be incident on the positive side of an infinite perforated screen lying in the plane* $x = 0$, *the material of the screen being such that the total wave-function vanishes on it. Let the apertures in the screen be bounded by straight lines parallel to the axis of z, so that the screen cuts the plane* $z = 0$ *in a set of straight lines* L_2 *lying on the axis of y. Then the total wave-function is given everywhere by*

$$w(x_0, y_0) = w_i(x_0, y_0) + \tfrac{1}{2}i \int\limits_{L_2} g(y) H_0^{(2)}(k\rho)\, dy \tag{3.26}$$

where $\rho = +\sqrt{\{x_0^2+(y-y_0)^2\}}$.

The function $g(y)$ *satisfies the integral equation*

$$\int\limits_{L_2} g(y) H_0^{(2)}(k|y-y_0|)\, dy = 2iw_i(0, y_0) \tag{3.27}$$

when $(0, y_0, 0)$ *is a point of* L_2.

This last result can also be proved by a simple physical argument. Let us suppose that monochromatic waves in which h_x^i, h_y^i, and d_z^i are the only non-zero components of **d** and **h** are incident on the positive side of a perfectly reflecting screen in the plane $x = 0$; the incident field is then independent of z. The apertures in the screen are taken to be bounded by straight lines parallel to the axis of z. Then, as Poincaré pointed out many years ago, an alternating current-sheet flows in the screen in the direction of the z-axis. If the current-density is $I(y)$, this induced current gives rise to an electromagnetic field $\mathbf{d}^s$, $\mathbf{h}^s$, in which the only non-zero component of $\mathbf{d}^s$ is

$$d_z^s(x_0, y_0) = -ik \iint\limits_{S_2} I(y) \frac{e^{-ikR}}{R}\, dydz = -\pi k \int\limits_{L_2} I(y) H_0^{(2)}(k\rho)\, dy.$$

The total electric force is then

$$d_z(x_0, y_0) = d_z^i(x_0, y_0) - \pi k \int\limits_{L_2} I(y) H_0^{(2)}(k\rho)\, dy$$

when $y_0 > 0$. But since d_z vanishes on the screen, the current $I(y)$ satisfies the integral equation

$$d_z^i(0, y_0) = \pi k \int\limits_{L_2} I(y) H_0^{(2)}(k|y-y_0|)\, dy$$

when $(0, y_0, 0)$ is a point of the screen. This is the integral equation (3.27).

§ 3.3. Babinet's Principle

Two plane screens which together would just fill up a plane are said to be complementary; the apertures in the one screen are then congruent to the material of the other. When identical electromagnetic waves fall on complementary perfectly reflecting plane screens, complementary diffraction patterns are formed behind them. Babinet's principle asserts that the sum of the electric forces at congruent points in the two diffraction patterns is equal to the electric force at the same point if there were no screen. This form of the principle is often based on the classical Kirchhoff theory of diffraction.

An examination of the results of Rayleigh and others for special diffraction problems shows that Babinet's principle in this form is not even approximately true. This was noticed in 1941 by H. G. Booker, who observed that in a rigorous formulation of Babinet's principle it is necessary to assume that complementary fields are incident on complementary screens; and he proved the principle in this form by considering waves in a two-sheeted space.† This complementarity of incident fields and screens is closely associated with the complementary character of the theorems we have proved in §§ 2.1 and 3.2. In the former, the boundary-condition is the vanishing of the normal derivative of v on the screen, in the latter, the vanishing of w. But the integral equation of § 2.1 relates to the apertures S_1, that of § 3.2 to the material S_2 of the screen.

The rigorous form of Babinet's principle may be enunciated as follows:‡

Let the electromagnetic field in which $\mathbf{d}^i = \mathbf{F}$, $\mathbf{h}^i = \mathbf{G}$ *be incident on the positive side of a perfectly conducting plane screen in the plane* $x = 0$; *the holes in the screen are denoted by* S, *the metal of the screen by* S'. *Let the total field in* $x < 0$ *be* $\mathbf{d}^1$, $\mathbf{h}^1$.

Let the complementary field|| *in which* $\mathbf{d}^i = -\mathbf{G}$, $\mathbf{h}^i = \mathbf{F}$ *be incident on the positive side of the complementary perfectly conducting plane screen in the plane* $x = 0$; *the holes in the screen are* S', *the metal of the screen* S. *Let the total field in* $x < 0$ *be* $\mathbf{d}^2$, $\mathbf{h}^2$.

Then Babinet's principle is that

$$\mathbf{d}^1 + \mathbf{h}^2 = \mathbf{F}, \qquad \mathbf{h}^1 - \mathbf{d}^2 = \mathbf{G}.$$

† *Journal Inst. Elec. Eng.* Part III A, **93** (1946), 620–6.

‡ Copson, *Proc. Roy. Soc.* (A), **186** (1946), 116, where a proof of the general case will be found.

|| It will be recalled that Maxwell's equations are invariant under the transformation $\mathbf{d} \to \mathbf{h}$, $\mathbf{h} \to -\mathbf{d}$.

For simplicity we consider here only the two-dimensional case. In the first screen the apertures are bounded by straight lines parallel to the axis of z, and cut the plane $z = 0$ in a set of segments L on the axis of y; the metal of the screen cuts the same plane in a set of segments L', so that L and L' together fill up the axis of y. In the complementary screen, L' is the trace of the apertures, L of the metal of the screen.

The most general field which is the same in all planes perpendicular to the axis of z is

$$d_x = \frac{\partial v}{\partial y}, \qquad d_y = -\frac{\partial v}{\partial x}, \qquad d_z = ikw,$$

$$h_x = -\frac{\partial w}{\partial y}, \qquad h_y = \frac{\partial w}{\partial x}, \qquad h_z = ikv.$$

Since the field complementary to this is

$$d_x = \frac{\partial w}{\partial y}, \qquad d_y = -\frac{\partial w}{\partial x}, \qquad d_z = -ikv,$$

$$h_x = \frac{\partial v}{\partial y}, \qquad h_y = -\frac{\partial v}{\partial x}, \qquad h_z = ikw,$$

it is evident that we need consider only the case when w is identically zero. The general result will follow by superposition.

In the first problem the only non-zero components of the incident field are

$$d_x^i = \frac{\partial v^i}{\partial y}, \qquad d_y^i = -\frac{\partial v^i}{\partial x}, \qquad h_z^i = ikv^i.$$

It follows from § 2.2, that the only non-zero components of the total field behind the screen are

$$d_x^1 = \frac{\partial v^1}{\partial y}, \qquad d_y^1 = -\frac{\partial v^1}{\partial x}, \qquad h_z^1 = ikv^1. \tag{3.31}$$

The function v^1 is defined by

$$v^1(x_0, y_0) = -\tfrac{1}{2}i \int_L f(y) H_0^{(2)}(k\rho)\, dy \tag{3.32}$$

where $f(y)$ satisfies the integral equation

$$\int_L f(y) H_0^{(2)}(k|y-y_0|)\, dy = 2iv^i(0, y_0) \tag{3.33}$$

when $(0, y_0, 0)$ is a point of L.

In the complementary problem, the only non-zero components of the incident field are

$$h_x^i = \frac{\partial v^i}{\partial y}, \qquad h_y^i = -\frac{\partial v^i}{\partial x}, \qquad d_z^i = -ikv^i.$$

It follows from § 3.2 that the only non-zero components of the total field are

$$h_x^2 = \frac{\partial v^2}{\partial y}, \qquad h_y^2 = -\frac{\partial v^2}{\partial x}, \qquad d_z^2 = -ikv^2. \tag{3.34}$$

The function v^2 is defined by

$$v^2(x_0, y_0) = v^i(x_0, y_0) + \tfrac{1}{2}i \int_L g(y) H_0^{(2)}(k\rho)\, dy \tag{3.35}$$

where $g(y)$ satisfies the integral equation

$$\int_L g(y) H_0^{(2)}(k|y-y_0|)\, dy = 2iv^i(0, y_0) \tag{3.36}$$

when $(0, y_0, 0)$ is a point of L.

Comparing (3.33) and (3.36) we see that $f(y)$ and $g(y)$ are equal. Hence, by (3.32) and (3.35), we have $v^1 + v^2 = v^i$. Therefore, by (3.31) and (3.34), we have

$$d_x^1 + h_x^2 = \frac{\partial v^i}{\partial y},$$

$$d_y^1 + h_y^2 = -\frac{\partial v^i}{\partial x},$$

$$h_z^1 - d_z^2 = ikv^i,$$

when $x < 0$, the other components of $\mathbf{d}^1 + \mathbf{h}^2$ and $\mathbf{h}^1 - \mathbf{d}^2$ being zero. This proves the theorem in its two-dimensional form.

§ 4. Sommerfeld's Diffraction Problem

§ 4.1. The diffraction of plane sound waves

We shall now apply the integral equation method to the problem of the diffraction of plane sound waves with velocity-potential

$$v_i = e^{ikx\cos\alpha + iky\sin\alpha} \quad (-\tfrac{1}{2}\pi < \alpha < \tfrac{1}{2}\pi)$$

incident on the positive side of a rigid screen whose position is defined by $x = 0$, $y < 0$. The total velocity-potential behind the screen (i.e. where x is negative) is given by

$$v(x, y) = -\tfrac{1}{2}i \int_0^\infty f(t) H_0^{(2)}(k\rho)\, dt \tag{4.11}$$

where $$\rho = \sqrt{\{x^2 + (y-t)^2\}}.$$

The function $f(t)$ is the value of $\partial v/\partial x$ at $(0,t)$ where $t > 0$, and satisfies the integral equation

$$\int_{0}^{\infty} f(t)H_0^{(2)}(k|y-t|)\,dt = 2ie^{iky\sin\alpha} \quad (y > 0). \tag{4.12}$$

This equation, which was discovered by Magnus† in a slightly different connexion, is called a non-homogeneous Wiener-Hopf equation.‡ Although it appears to be of the form

$$\int_{-\infty}^{\infty} f(t)l(y-t)\,dt = g(y)$$

whose theory is well known, it is actually quite different and is much more difficult to solve; for $g(y)$ is given only when $y > 0$, and the solution $f(t)$ has to be identically zero when $t < 0$.

The integral equation (4.12) is solved by the use of generalized Fourier integrals,‖ which are analytic functions of a complex variable $w = u+iv$ and are defined, under certain conditions of integrability, by the equations

$$P_+(w) = \frac{1}{\sqrt{(2\pi)}} \int_{0}^{\infty} p(x)e^{iwx}\,dx \tag{4.13}$$

$$P_-(w) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{0} p(x)e^{iwx}\,dx. \tag{4.14}$$

The function $P_+(w)$ is regular in an upper half-plane $v > c$, whereas $P_-(w)$ is regular in a lower half-plane $v < c'$. If these half-planes have a common strip, it is convenient to write $P(w) = P_+(w)+P_-(w)$ in

† *Zeitschrift f. Phys.* **117** (1941), 168–79. See also Copson, *Quarterly J. Math.* (Oxford), **17** (1946), 19–34.

‡ The theory of the homogeneous Wiener-Hopf equation

$$\int_{0}^{\infty} f(t)l(y-t) = f(y) \quad (y > 0)$$

will be found in Paley and Wiener, *Fourier Transforms in the Complex Domain* (New York, 1934), 49–58; Titchmarsh, *Theory of Fourier Integrals* (Oxford, 1937), 339–42. We cannot give any reference to the theory of the non-homogeneous equation, though much the same ideas are involved.

‖ Titchmarsh, loc. cit. 4–6.

that strip. The function $p(x)$ can be expressed in terms of $P_+(w)$ and $P_-(w)$ at any point of continuity by the equation

$$p(x) = \frac{1}{\sqrt{(2\pi)}} \int\limits_{ia-\infty}^{ia+\infty} P_+(w)e^{-ixw}\,dw + \frac{1}{\sqrt{(2\pi)}} \int\limits_{ib-\infty}^{ib+\infty} P_-(w)e^{-ixw}\,dw, \tag{4.15}$$

where $a > c$, $b < c'$.

Let us write the equation (4.12) in the form

$$\int\limits_{-\infty}^{\infty} f(t)l(y-t)\,dt = g(y)+h(y)$$

for all y, where $f(y)$ and $g(y)$ are zero for $y < 0$, and $h(y)$ is zero for $y > 0$; this, of course, implies that $F_-(w)$, $G_-(w)$, and $H_+(w)$ are also zero. The kernel $l(y-t)$ is $H_0^{(2)}(k|y-t|)$. If we multiply this equation by e^{iwy} and integrate from $-\infty$ to $+\infty$, we obtain†

$$\begin{aligned} \sqrt{(2\pi)}\{G_+(w)+H_-(w)\} &= \int\limits_{-\infty}^{\infty} e^{iwy} \int\limits_{-\infty}^{\infty} f(t)l(y-t)\,dtdy \\ &= \int\limits_{-\infty}^{\infty} f(t) \int\limits_{-\infty}^{\infty} e^{iwy}l(y-t)\,dydt \\ &= \int\limits_{-\infty}^{\infty} f(t)e^{iwt}\,dt \int\limits_{-\infty}^{\infty} e^{iwz}l(z)\,dz \\ &= 2\pi F_+(w)L(w). \end{aligned}$$

If we assume that $k = p-iq$ where $q > 0$, the integral defining $G_+(w)$ is uniformly and absolutely convergent in any closed region in the half-plane $v > q\sin\alpha$. Carrying out the integration, we have

$$G_+(w) = \frac{1}{\sqrt{(2\pi)}} \int\limits_0^{\infty} 2ie^{iky\sin\alpha+iwy}\,dy = -\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{1}{w+k\sin\alpha};$$

the last expression provides the analytical continuation of $G_+(w)$ all over the w-plane.

The Fourier transform of the integral equation is therefore

$$2\pi F_+(w)L(w) + \frac{2}{w+k\sin\alpha} - \sqrt{(2\pi)}H_-(w) = 0, \tag{4.16}$$

† We shall content ourselves in this section only with the formal analysis, and refer the reader to Copson's paper for more rigorous details.

where $L(w)$ is the Fourier transform of the kernel,

$$L(w) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{iwy} H_0^{(2)}(k|y|)\, dy.$$

Since the integrand behaves like $-(2i/\pi)\log|y|$ near $y = 0$, this integral converges uniformly at $y = 0$ in any finite region of the w-plane. Moreover, since†

$$H_0^{(2)}(z) \sim \left(\frac{2}{\pi z}\right)^{\frac{1}{2}} e^{-i(z-\frac{1}{4}\pi)}$$

when $|z|$ is large and $-2\pi < \arg z < \pi$, the modulus of the integrand is asymptotically equal to

$$\left\{\frac{2}{\pi|ky|}\right\}^{\frac{1}{2}} e^{-vy+q|y|}$$

when $|y|$ is large, and so the integral converges uniformly and absolutely at the upper and lower limits in any closed region in the strip $-q < v < q$. Hence $L(w)$ is an analytic function, regular in the strip $-q < v < q$. By using a result due to Basset,‡ it can be shown that

$$L(w) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{1}{\sqrt{(k^2-w^2)}},$$

where the square root reduces to k when $w = 0$.

The next step is to find the half-planes in which $F_+(w)$ and $H_-(w)$ are regular, and the asymptotic behaviour of these functions. Now, when $y > 0$,

$$f(y) = \left(\frac{\partial v}{\partial x}\right)_{x=0} = \left(\frac{\partial v_i}{\partial x}\right)_{x=0} + \left(\frac{\partial v_s}{\partial x}\right)_{x=0}$$
$$= ik\cos\alpha\, e^{iky\sin\alpha} + \phi(y)$$

say. Hence we have

$$F_+(w) = \frac{ik\cos\alpha}{\sqrt{(2\pi)}} \int_0^{\infty} e^{i(w+k\sin\alpha)y}\, dy + \Phi_+(w).$$

The integration can be carried out when $v > q\sin\alpha$, and gives

$$F_+(w) = -\frac{1}{\sqrt{(2\pi)}} \frac{k\cos\alpha}{w+k\sin\alpha} + \Phi_+(w).$$

Now, by the radiation condition of § 1.2, $\phi(y)$ is bounded in any

† Watson, *Theory of Bessel Functions* (Cambridge, 1922), 198.
‡ Ibid. 388 (10).

interval $y \geqslant a'$; if we assume in addition that $|\phi(y)|$ is integrable over any finite interval $0 \leqslant y \leqslant a'$, we have, for $v \geqslant c > 0$,

$$|\Phi_+(w)| \leqslant \frac{1}{\sqrt{(2\pi)}} \int_0^{a'} e^{-cy} |\phi(y)|\, dy + \frac{1}{\sqrt{(2\pi)}} \int_{a'}^{\infty} e^{-cy} |\phi(y)|\, dy$$

$$< \frac{1}{\sqrt{(2\pi)}} \int_0^{a'} |\phi(y)|\, dy + \frac{A}{\sqrt{(2\pi)}} \int_{a'}^{\infty} e^{-cy}\, dy$$

where A is an upper bound of $|\phi(y)|$. The integral defining $\Phi_+(w)$ is therefore uniformly and absolutely convergent in $v \geqslant c > 0$; hence $\Phi_+(w)$ is regular in $v \geqslant c > 0$. It follows that $F_+(w)$ is an analytic function whose only possible singularity in $v > 0$ is a simple pole at $w = -k \sin\alpha$;† moreover $F_+(w)$ is bounded as $|w| \to \infty$ in $v \geqslant c > 0$. Actually we do not use all this information; it suffices to assume that $F_+(w)$ is regular in $v > a$, where $a = \max(0, q \sin\alpha)$, and is bounded as $|w| \to \infty$ in $v \geqslant c > a$. That $F_+(w)$ has a simple pole at

$$w = -k \sin\alpha,$$

with residue $-k \cos\alpha/\sqrt{(2\pi)}$, emerges in the solution.

Next, we observe that, if $y = -y' < 0$,

$$h(-y') = \int_0^{\infty} f(t) H_0^{(2)}\{k(y'+t)\}\, dt$$

where $$f(t) = ik \cos\alpha\, e^{ikt \sin\alpha} + \phi(t).$$

But since $$H_0^{(2)}\{k(y'+t)\} = O\{(y'+t)^{-\frac{1}{2}} e^{-q(y'+t)}\}$$

when y' is large, we have

$$h(-y') = o \int_0^{\infty} \{|k| \cos\alpha\, e^{qt \sin\alpha} + |\phi(t)|\} e^{-q(y'+t)}\, dt$$

$$= o(e^{-qy'})$$

as we should expect on physical grounds; hence if $y' \geqslant a' > 0$, there exists a constant A' such that

$$|h(-y')| < A' e^{-qy'}.$$

† Since $-\frac{1}{2}\pi < \alpha < \frac{1}{2}\pi$, this pole may lie in $v \leqslant 0$.

If we assume in addition that $|h(-y')|$ is integrable over any finite interval $0 \leqslant y' \leqslant a'$, we have for $v \leqslant c' < q$

$$|H_-(w)| = \frac{1}{\sqrt{(2\pi)}}\left|\int_0^\infty h(-y')e^{-iwy'}\,dy'\right|$$

$$\leqslant \frac{1}{\sqrt{(2\pi)}}\int_0^\infty |h(-y')|e^{vy'}\,dy'$$

$$< \frac{1}{\sqrt{(2\pi)}}e^{c'a'}\int_0^{a'} |h(-y')|\,dy + \frac{A'}{\sqrt{(2\pi)}}\int_{a'}^\infty e^{-(q-c')y'}\,dy'.$$

The integral defining $H_-(w)$ is therefore uniformly and absolutely convergent in $v \leqslant c' < q$, and so $H_-(w)$ is an analytic function which is regular in $v < q$ and bounded in $v \leqslant c' < q$.

The equation (4.16) is now of the form

$$\frac{F_+(w)}{\sqrt{(k^2-w^2)}} + \frac{1}{\sqrt{(2\pi)}(w+k\sin\alpha)} - \tfrac{1}{2}H_-(w) = 0. \tag{4.17}$$

Here $F_+(w)$ is regular in a half-plane $v > a = \max(0, q\sin\alpha)$, and $H_-(w)$ in a half-plane $v < q$, but $1/\sqrt{(k^2-w^2)}$ is regular only in a strip $-q < v < q$. The next step is the important one: under certain conditions, an analytic function $M(w)$, regular and non-zero in a strip $-q < v < q$, can be written in the form† $N_+(w)/N_-(w)$ where $N_+(w)$ is regular and non-zero in $v > -q$, $N_-(w)$ in $v < q$. In the present case, we have evidently

$$\frac{1}{\sqrt{(k^2-w^2)}} = \frac{1}{\sqrt{(k-w)}}\bigg/\sqrt{(k+w)},$$

where $1/\sqrt{(k-w)}$ is regular and non-zero in $v > -q$, $\sqrt{(k+w)}$ in $v < q$, the principal values of the square roots being taken.

† The proof is briefly as follows. If $0 < \gamma < q$, it can be proved by Cauchy's Theorem that the principal value of $\log M(w)$ is given by

$$\log M(w) = \frac{1}{2\pi i}\int_{-\infty-i\gamma}^{\infty-i\gamma} \frac{\log M(z)}{z-w}\,dz - \frac{1}{2\pi i}\int_{-\infty+i\gamma}^{\infty+i\gamma} \frac{\log M(z)}{z-w}\,dz$$

$$= P_1(w) - P_2(w)$$

when $-\gamma < v < \gamma$, provided that, as $|w| \to \infty$ in $-\gamma \leqslant v \leqslant \gamma$, $\{\log M(w)\}/w$ tends to zero uniformly with respect to v. The first term $P_1(w)$ is regular in $v > -\gamma$, the second $P_2(w)$ in $v < \gamma$; and $M(w) = e^{P_1(w)}/e^{P_2(w)} = N_+(w)/N_-(w)$.

If we multiply through equation (4.17) by $\surd(k+w)$, we obtain

$$\frac{F_+(w)}{\surd(k-w)}+\frac{\surd(k+w)}{\surd(2\pi)(w+k\sin\alpha)}-\tfrac{1}{2}H_-(w)\surd(k+w) = 0.$$

The first term is regular in an upper half-plane

$$v > a = \max(0, q\sin\alpha),$$

the last in a lower half-plane $v < q$. The middle term is regular in the strip $q\sin\alpha < v < q$; but it can be written in the form

$$\frac{\surd(k+w)-\surd(k-k\sin\alpha)}{\surd(2\pi)(w+k\sin\alpha)}+\frac{\surd(k-k\sin\alpha)}{\surd(2\pi)(w+k\sin\alpha)}$$

where the first term is now regular in $v < q$, the second in $v > q\sin\alpha$. Hence we have

$$\frac{F_+(w)}{\surd(k-w)}+\frac{\surd(k-k\sin\alpha)}{\surd(2\pi)(w+k\sin\alpha)}$$
$$= \tfrac{1}{2}H_-(w)\surd(k+w)-\frac{\surd(k+w)-\surd(k-k\sin\alpha)}{\surd(2\pi)(w+k\sin\alpha)}.$$

But now the left-hand side is regular in $v > a$, the right-hand side in $v < q$, and these two half-planes have a strip $a < v < q$ in common. Hence each side is the analytical continuation of the other; together they define an integral function $P(w)$, so that

$$\frac{F_+(w)}{\surd(k-w)}+\frac{\surd(k-k\sin\alpha)}{\surd(2\pi)(w+k\sin\alpha)} = P(w),$$

$$\tfrac{1}{2}H_-(w)\surd(k+w)-\frac{\surd(k+w)-\surd(k-k\sin\alpha)}{\surd(2\pi)(w+k\sin\alpha)} = P(w).$$

These equations, together with the known asymptotic behaviour of $F_+(w)$ and $H_-(w)$, show that $P(w)$ is $O(|w|^{-\frac{1}{2}})$ as $|w| \to \infty$ in $v \geqslant c > a$ and is $O(|w|^{\frac{1}{2}})$ in $v \leqslant c' < q$. Hence $P(w)$ is $O(|w|^{\frac{1}{2}})$ as $|w| \to \infty$ and is therefore, by the extension of Liouville's Theorem,† a polynomial of degree $\leqslant \frac{1}{2}$, and so is constant. But since $P(w)$ tends to zero as $|w| \to \infty$ in $v \geqslant c$, it is identically zero. This proves that

$$F_+(w) = -\frac{\surd(k-k\sin\alpha)\surd(k-w)}{\surd(2\pi)(w+k\sin\alpha)} \tag{4.18}$$

$$H_-(w) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}}\left\{1-\frac{\surd(k-k\sin\alpha)}{\surd(k+w)}\right\}\frac{1}{w+k\sin\alpha}.$$

† Titchmarsh, *The Theory of Functions* (Oxford, 1932), 85.

By (4.15), we have

$$f(t) = \frac{1}{\sqrt{(2\pi)}} \int_{ib-\infty}^{ib+\infty} F_+(w) e^{-itw}\, dw$$

where the path of integration lies in the half-plane of regularity of $F_+(w)$. From (4.18), we see that $F_+(w)$ is regular in $v > -q$, apart from a simple pole at $-k \sin\alpha$. Hence the required solution of Magnus's integral equation (4.12) is

$$f(t) = -\frac{1}{2\pi} \int_{ib-\infty}^{ib+\infty} \frac{\sqrt{(k-k\sin\alpha)}\sqrt{(k-w)}}{w+k\sin\alpha} e^{-itw}\, dw \tag{4.19}$$

where $q \sin\alpha < b < q$.

§ 4.2. Completion of the solution

In the diffraction problem discussed in § 4.1, the total velocity-potential behind the screen (i.e. where x is negative) was proved to be

$$v(x,y) = -\tfrac{1}{2}i \int_{-\infty}^{\infty} f(t) m(y-t)\, dt \tag{4.21}$$

where $$m(y) = H_0^{(2)}\{k\sqrt{(x^2+y^2)}\}.$$

The function $f(t)$, which vanishes for negative values of t, is nevertheless given for all values of t by the equation

$$f(t) = \frac{1}{\sqrt{(2\pi)}} \int_{ib-\infty}^{ib+\infty} F_+(w) e^{-iwt}\, dw$$

where $q \sin\alpha < b < q$. If we substitute this expression for $f(t)$ in (4.21) and invert the order of integration, we obtain

$$\begin{aligned} v(x,y) &= -\frac{i}{2\sqrt{(2\pi)}} \int_{ib-\infty}^{ib+\infty} \int_{-\infty}^{\infty} F_+(w) e^{-iwt} m(y-t)\, dt dw \\ &= -\frac{i}{2\sqrt{(2\pi)}} \int_{ib-\infty}^{ib+\infty} F_+(w) e^{-iwy} \int_{-\infty}^{\infty} e^{iwz} m(z)\, dz dw \\ &= -\tfrac{1}{2}i \int_{ib-\infty}^{ib+\infty} F_+(w) M(w) e^{-iwy}\, dw \qquad (4.22) \end{aligned}$$

where
$$M(w) = \frac{1}{\sqrt{(2\pi)}} \int\limits_{-\infty}^{\infty} e^{iwz} m(z)\, dz$$
$$= \frac{1}{\sqrt{(2\pi)}} \int\limits_{-\infty}^{\infty} e^{iwz} H_0^{(2)}\{k\sqrt{(x^2+z^2)}\}\, dz.$$

This integral represents an analytic function of w, regular in the strip $-q < v < q$; carrying out the integration, we have†

$$M(w) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{1}{\sqrt{(k^2-w^2)}} e^{-i|x|\sqrt{(k^2-w^2)}} \tag{4.23}$$

where the branch of $\sqrt{(k^2-w^2)}$ is taken which is equal to k when $w = 0$.

If we substitute the values of $F_+(w)$ and $M(w)$ given by (4.18) and (4.23) in (4.22) we obtain as the total velocity-potential in $x < 0$

$$v(x,y) = -\frac{1}{2\pi i} \int\limits_{ib-\infty}^{ib+\infty} e^{-i|x|\sqrt{(k^2-w^2)}-iyw} \frac{\sqrt{(k-k\sin\alpha)}}{\sqrt{(k+w)}} \frac{dw}{w+k\sin\alpha}; \tag{4.24}$$

and if we translate the path of integration to the position $v = c$ where $-q < c < q\sin\alpha$, taking account of the pole at $-k\sin\alpha$, this becomes

$$v(x,y) = e^{ikx\cos\alpha+iky\sin\alpha} - $$
$$-\frac{1}{2\pi i} \int\limits_{ic-\infty}^{ic+\infty} e^{-i|x|\sqrt{(k^2-w^2)}-iyw} \frac{\sqrt{(k-k\sin\alpha)}}{\sqrt{(k+w)}} \frac{dw}{w+k\sin\alpha}.$$

From this, we can deduce the velocity-potential in $x > 0$ by means of equation (2.12), which reduces here to

$$v(x,y)+v(-x,y) = e^{ikx\cos\alpha+iky\sin\alpha} + e^{-ikx\cos\alpha+iky\sin\alpha};$$

this gives

$$v(x,y) = e^{ikx\cos\alpha+iky\sin\alpha} +$$
$$+\frac{1}{2\pi i} \int\limits_{ic-\infty}^{ic+\infty} e^{-i|x|\sqrt{(k^2-w^2)}-iyw} \frac{\sqrt{(k-k\sin\alpha)}}{\sqrt{(k+w)}} \frac{dw}{w+k\sin\alpha}.$$

To sum up: *when plane sound waves with velocity-potential*

$$v_i = e^{ikx\cos\alpha+iky\sin\alpha} \quad (-\tfrac{1}{2}\pi < \alpha < \tfrac{1}{2}\pi;\ k = p-iq)‡$$

† The result is a particular case of equation (2) on p. 416 of Watson's *Theory of Bessel Functions*.

‡ The corresponding result for $k > 0$ can be found by making q tend to zero. The path of integration in (4.25) becomes the real axis suitably indented at $-k$ and $-k\sin\alpha$.

are incident on the positive side of a rigid screen whose position is defined by $x = 0$, $y < 0$, the total velocity-potential is

$$v = v_i + v_s,$$

the velocity-potential v_s of the reflected and diffracted waves being

$$v_s(x,y) = \pm\frac{1}{2\pi i}\int\limits_{ic-\infty}^{ic+\infty} e^{-i|x|\sqrt{(k^2-w^2)}-iyw}\frac{\sqrt{(k-k\sin\alpha)}}{\sqrt{(k+w)}}\frac{dw}{w+k\sin\alpha}, \tag{4.25}$$

where the upper or lower sign is taken according as x is positive or negative and $-q < c < q\sin\alpha$.

This solution can be identified with that given by Sommerfeld by deforming the path of integration and using Cauchy's theorem.

§ 5. Schwarzschild's Diffraction Problem

§ 5.1. Diffraction by two screens

Let us suppose that sound waves with velocity-potential v_i are incident on two perfectly reflecting screens, the one being bounded by a closed surface S_1 the other by a closed surface S_2. The velocity-potential v_s of the scattered and diffracted waves is such that the normal derivative of v_i+v_s vanishes on S_1 and S_2. Now a solution of $(\nabla^2+k^2)v = 0$ which has no singularities outside S_1 and satisfies the radiation condition, is uniquely determined by the values of its normal derivative on S_1; in fact

$$v(P_0) = -\frac{1}{4\pi}\iint\limits_{S_1}\frac{\partial v(P_1)}{\partial n}G_2^{(1)}(P_0,P_1)\,dS_1$$

where $G_2^{(1)}$ is the Green's function of the second kind for the region outside S_1. And similarly for S_2, we have

$$v(P_0) = -\frac{1}{4\pi}\iint\limits_{S_2}\frac{\partial v(P_2)}{\partial n}G_2^{(2)}(P_0,P_2)\,dS_2.$$

If we consider the pair of equations

$$v_2(P_0) = \frac{1}{4\pi}\iint\limits_{S_1}\left\{\frac{\partial v_1(P_1)}{\partial n_1}+\frac{\partial v_i(P_1)}{\partial n_1}\right\}G_2^{(1)}(P_0,P_1)\,dS_1, \tag{5.11}$$

$$v_1(P_0) = \frac{1}{4\pi}\iint\limits_{S_2}\left\{\frac{\partial v_2(P_2)}{\partial n_2}+\frac{\partial v_i(P_1)}{\partial n_2}\right\}G_2^{(2)}(P_0,P_2)\,dS_2, \tag{5.12}$$

the function $v_s = v_1+v_2$ satisfies all the conditions of the problem.

For it is a solution of $(\nabla^2+k^2)v = 0$ which has no singularities outside S_1 and S_2 and satisfies the radiation-condition; moreover when P_0 is on S_1, we have

$$\frac{\partial v_2(P_1)}{\partial n_1} = -\frac{\partial v_1(P_1)}{\partial n_1} - \frac{\partial v_i(P_1)}{\partial n_1},$$

so that the normal derivative of $v_i+v_1+v_2$ vanishes on S_1; and similarly on S_2. But the equations (5.11) and (5.12) involve the unknown normal derivatives of v_1 on S_1 and v_2 on S_2. These unknown functions satisfy a pair of simultaneous integral equations; for, if P_2 is a point of S_2, it follows from (5.11) that

$$\frac{\partial v_2(P_2)}{\partial n_2} = \frac{1}{4\pi} \iint\limits_{S_1} \left\{\frac{\partial v_1(P_1)}{\partial n_1} + \frac{\partial v_i(P_1)}{\partial n_1}\right\} \frac{\partial G_2^{(1)}(P_2, P_1)}{\partial n_2} \, dS_1, \tag{5.13}$$

and similarly, if P_1 is a point of S_1,

$$\frac{\partial v_1(P_1)}{\partial n_1} = \frac{1}{4\pi} \iint\limits_{S_2} \left\{\frac{\partial v_2(P_2)}{\partial n_2} + \frac{\partial v_i(P_2)}{\partial n_2}\right\} \frac{\partial G_2^{(2)}(P_1, P_2)}{\partial n_1} \, dS_2. \tag{5.14}$$

If we could solve these integral equations for $\partial v_1/\partial n_1$ and $\partial v_2/\partial n_2$, then the function $v_s = v_1+v_2$ would solve the problem of diffraction by the two screens S_1 and S_2.

The function $G_2^{(1)}(P_0, P)$ is the total velocity-potential at P when waves from a point-source at P_0 are incident on the screen S_1; it is the solution of the problem of diffraction by the screen S_1. Hence, if the problems of diffraction by S_1 and S_2 separately have been solved, equations (5.13) and (5.14) constitute an analytical formulation of the problem of diffraction by both screens simultaneously. For example, we could use the known solution of the problem of diffraction by a half-plane to reduce the problem of diffraction by an infinite slit in a perfectly reflecting plane to that of solving simultaneous integral equations; and the solution of these integral equations by successive substitution is, in fact, the solution obtained on more physical grounds by Schwarzschild in his 1902 paper already quoted.

§ 5.2. The integral equations of Schwarzschild's problem

As the determination of the normal derivatives of the Green's functions which appear in Schwarzschild's work on the problem of diffraction by an infinite slit is rather difficult, we shall not pursue that line but shall derive equivalent results from the theorems proved earlier in this chapter. We consider the problem of the diffraction of plane waves at a slit in a plane screen, the edges of the slit being perpendicular to the direction of propagation of the waves.

Let us choose the axes of coordinates so that the screen lies in the plane $x = 0$ and so that the edges of the slit are the lines $x = 0$, $y = 0$ and $x = 0$, $y = -h$; the wave-function of the incident waves is then

$$w_i = e^{ikx\cos\alpha+iky\sin\alpha}$$

where $-\frac{1}{2}\pi < \alpha < \frac{1}{2}\pi$. If we consider the case when the total wave-function w is required to vanish on the screen, we have a two-dimensional problem of the type considered in § 3.2, with solution

$$w(x_0,y_0) = e^{ikx_0\cos\alpha+iky_0\sin\alpha}+\tfrac{1}{2}i\left\{\int\limits_{-\infty}^{-h}+\int\limits_{0}^{\infty}\right\}f(y)H_0^{(2)}(k\rho)\,dy \tag{5.21}$$

where

$$\rho^2 = x_0^2+(y-y_0)^2.$$

The function $f(y)$, which is the unknown value of $\partial(w-w_i)/\partial x$ on the screen, satisfies the integral equation

$$\left\{\int\limits_{-\infty}^{-h}+\int\limits_{0}^{\infty}\right\}f(y)H_0^{(2)}(k|y-y_0|)\,dy = 2ie^{iky_0\sin\alpha}$$

when $y_0 > 0$ and when $y_0 < -h$; but the expression on the right-hand side of the integral equation is unknown for $-h < y_0 < 0$.

Since the integral equation really involves two unknown functions, viz. $f(y)$ for $y > 0$ and $f(y)$ for $y < -h$, we write $f(-y-h) = g(y)$, and obtain

$$\int\limits_0^\infty f(y)H_0^{(2)}(k|y-y_0|)\,dy+\int\limits_0^\infty g(y)H_0^{(2)}(k|y+y_0+h|)\,dy = 2ie^{iky_0\sin\alpha}$$

where $y_0 > 0$ or $y_0 < -h$. When $y_0 > 0$, the equation takes the form

$$\int\limits_0^\infty f(y)H_0^{(2)}(k|y-y_0|)\,dy+\int\limits_0^\infty g(y)H_0^{(2)}\{k(y+y_0+h)\}\,dy = 2ie^{iky_0\sin\alpha}. \tag{5.22}$$

But when $y_0 < -h$, we write $y_0 = -h-\eta_0$ where $\eta_0 > 0$; this gives

$$\int\limits_0^\infty f(y)H_0^{(2)}(k|y+\eta_0+h|)\,dy+\int\limits_0^\infty g(y)H_0^{(2)}(k|y-\eta_0|)\,dy = 2ie^{-ik(\eta_0+h)\sin\alpha}.$$

Replacing η_0 by y_0, we obtain a second integral equation

$$\int\limits_0^\infty g(y)H_0^{(2)}(k|y-y_0|)\,dy+\int\limits_0^\infty f(y)H_0^{(2)}\{k(y+y_0+h)\}\,dy = 2ie^{-ik(y_0+h)\sin\alpha}, \tag{5.23}$$

where $y_0 > 0$. The functions $f(y)$ and $g(y)$ are, therefore, the solutions of the simultaneous integral equations (5.22) and (5.23), though, in the case of normal incidence ($\alpha = 0$), $f(y)$ and $g(y)$ are identical and there is only one integral equation.

The integral equations, as they stand, are not of a known form, but they can be simplified by using Fox's formula†

$$H_0^{(2)}\{k(y+y_0+h)\} = \frac{1}{\pi}\int\limits_0^\infty \frac{e^{-ik(y+z+h)}}{y+z+h}\sqrt{\left(\frac{y+h}{z}\right)}\, H_0^{(2)}(k|z-y_0|)\, dz,$$

valid when $k > 0$ and when $k = p-iq$ with $q > 0$. This gives

$$\int\limits_0^\infty \{f(y)+G(y)\}H_0^{(2)}(k|y-y_0|)\, dy = 2ie^{iky_0\sin\alpha} \quad (y_0 > 0),$$

$$\int\limits_0^\infty \{g(y)+F(y)\}H_0^{(2)}(k|y-y_0|)\, dy = 2ie^{-ik(y_0+h)\sin\alpha} \quad (y_0 > 0),$$

where

$$F(y) = \frac{1}{\pi}\int\limits_0^\infty f(z)\frac{e^{-ik(y+z+h)}}{y+z+h}\sqrt{\left(\frac{z+h}{y}\right)}\, dz \quad (y > 0),$$

$$G(y) = \frac{1}{\pi}\int\limits_0^\infty g(z)\frac{e^{-ik(y+z+h)}}{y+z+h}\sqrt{\left(\frac{z+h}{y}\right)}\, dz \quad (y > 0).$$

The first pair are Wiener-Hopf equations of the type considered in § 4.1, and have solutions

$$f(y)+G(y) = \varpi(y,\alpha),$$
$$g(y)+F(y) = e^{ikh\sin\alpha}\varpi(y,-\alpha)$$

where

$$\varpi(y,\alpha) = -\frac{1}{2\pi}\int\limits_{ib-\infty}^{ib+\infty} \frac{\sqrt{(k-k\sin\alpha)}\sqrt{(k-w)}}{w+k\sin\alpha}\, e^{-iyw}\, dw$$

with $q\sin\alpha < b < q$. It remains then to solve the integral equations

$$f(y) = \varpi(y,\alpha)-\frac{1}{\pi}\int\limits_0^\infty g(z)\frac{e^{-ik(y+z+h)}}{y+z+h}\sqrt{\left(\frac{z+h}{y}\right)}\, dz, \qquad (5.24)$$

$$g(y) = e^{-ikh\sin\alpha}\varpi(y,-\alpha)-\frac{1}{\pi}\int\limits_0^\infty f(z)\frac{e^{-ik(y+z+h)}}{y+z+h}\sqrt{\left(\frac{z+h}{y}\right)}\, dz, \qquad (5.25)$$

where $y > 0$.

† E. N. Fox, *Phil. Trans.* (A), **241** (1948), 71–103, Appendix A.

A first approximation to the solution is

$$f(y) = \varpi(y,\alpha), \qquad g(y) = e^{-ikh\sin\alpha}\varpi(y,-\alpha).$$

This corresponds physically to taking the Sommerfeld formulae for the waves scattered by the two half-planes considered separately and neglecting the diffraction by each half-plane of the waves scattered by the other. A second approximation can then be obtained by substituting the first approximations in the right-hand sides of equations (5.24) and (5.25), and so on indefinitely. This process of successive substitution leads to series-solutions of the form

$$f(y) = \sum_0^\infty f_n(y), \qquad g(y) = \sum_0^\infty g_n(y)$$

where
$$f_n(y) = -\frac{1}{\pi}\int_0^\infty g_{n-1}(z)\,\frac{e^{-ik(y+z+h)}}{y+z+h}\sqrt{\left(\frac{z+h}{y}\right)}\,dz,$$

$$g_n(y) = -\frac{1}{\pi}\int_0^\infty f_{n-1}(z)\,\frac{e^{-ik(y+z+h)}}{y+z+h}\sqrt{\left(\frac{z+h}{y}\right)}\,dz$$

with
$$f_0(y) = \varpi(y,\alpha), \qquad g_0(y) = e^{-ikh\sin\alpha}\varpi(y,-\alpha).$$

These series can be shown to converge,† but the convergence is so slow that the solutions do not appear to be of much practical use in the case of greatest physical importance when k is real. So far, no one has found solutions of the integral equations in finite terms.‡

Very similar integral equations occur in Fox's solution of the problem of diffraction of sound pulses by an infinitely long strip; but in his work, ik is real and positive. As a result, the series-solution converges rapidly, and he shows that the third approximation leads to satisfactory numerical results.

§ 6. Approximate solutions

§ 6.1. Rayleigh's approximate solutions

The rigorous solution of diffraction problems by the integral equation method has, so far, proved feasible only when the Wiener-Hopf theory is applicable. The method can, however, be used to obtain

† Schwarzschild discussed this type of convergence problem for the case when k is real.

‡ The slit problem has been solved in series of Mathieu functions by Morse and Rubenstein, *Phys. Rev.* **54** (1938), 895–8.

approximate solutions, an idea which goes back to the work of Rayleigh.†

Let us consider the case when sound waves with velocity-potential v_i are incident on the positive side of a perfectly conducting screen in the plane $x = 0$, the aperture in the screen being S_1 and the origin of coordinates a point of the aperture. By the theorem of § 2.1, the total velocity-potential at a point (x, y, z) behind the screen $(x < 0)$ is

$$v(x,y,z) = \frac{1}{2\pi} \iint\limits_{S_1} f(y',z') \frac{e^{-ikR}}{R}\, dy'dz' \tag{6.11}$$

where $$R^2 = x^2+(y-y')^2+(z-z')^2.$$

The function $f(y,z)$ is the solution of the integral equation

$$\iint\limits_{S_1} f(y',z') \frac{e^{-ik\rho}}{\rho}\, dy'dz' = 2\pi v_i(0,y,z) \tag{6.12}$$

when $(0, y, z)$ is a point of S_1, where

$$\rho^2 = (y-y')^2+(z-z')^2.$$

Let us suppose that the aperture S_1 lies everywhere at a finite distance, and consider the effect at a point $P(-lr, -mr, -nr)$ behind the screen where $l^2+m^2+n^2 = 1$. If the distance r of P from the origin is large compared with the dimensions of the screen, we have $R = r+my'+nz'$ approximately, and equation (6.11) then reduces to

$$v(x,y,z) = A(m,n) \frac{e^{-ikr}}{r} \tag{6.13}$$

where $$A(m,n) = \frac{1}{2\pi} \iint\limits_{S_1} f(y,z) e^{-ik(my+nz)}\, dydz.$$

The wave-motion at great distances behind the screen is thus a spherical wave whose amplitude $A(m, n)$ varies with the direction of the radius vector OP.

If we assume in addition that the dimensions of the aperture are small compared with the wave-length $2\pi/k$, the formula for the amplitude simplifies to

$$A = \frac{1}{2\pi} \iint\limits_{S_1} f(y,z)\, dydz,$$

† *Phil. Mag.* **43** (1897), 259–72; *Proc. Roy. Soc.* (A), **89** (1913), 194–219. These papers will be found in Rayleigh's *Scientific Papers* **4** (1903), 283–96; **6** (1920), 161–86.

which is independent of direction. Moreover, in this case, v_i is sensibly constant over the aperture, and the integral equation becomes

$$\iint_{S_1} f(y',z')\,\frac{1}{\rho}\,dy'dz' = 2\pi v_i(0,0,0)$$

when $(0,y,z)$ is a point of S_1. The expression on the left-hand side of this equation is the electrostatic potential of a distribution of density $f(y,z)$ on S_1. But this potential has the constant value $2\pi v_i(0,0,0)$ on S_1; hence $2\pi A$ is simply the total charge on a conducting disk which is at potential $2\pi v_i(0,0,0)$ and has the size and shape of the aperture. Therefore $A = Mv_i(0,0,0)$ where M is the capacity of such a conducting disk. In particular, if the aperture is a circle of radius a, the velocity potential at a great distance behind the screen is approximately

$$v = \frac{2a}{\pi}\,v_i(0,0,0)\,\frac{e^{-ikr}}{r}. \tag{6.14}$$

When we turn to the corresponding problem in which the boundary condition is the vanishing of the total wave-function on the screen, the theorem of § 3.2 proves to be unsuitable since it involves integrals over the screen and not over the aperture. It is more convenient to use the expression

$$w(x,y,z) = \frac{1}{2\pi}\frac{\partial}{\partial x}\iint_{S_1} f(y',z')\,\frac{e^{-ikR}}{R}\,dy'dz' \tag{6.15}$$

for the wave-function at a point behind the screen. (Cf. § 1.4.) In this formula, $f(y,z)$ is the unknown value of w in the aperture, and has to satisfy two conditions:

(i) $f(y,z)$ vanishes on the boundary of S_1, since w is continuous and vanishes on the screen, and

(ii) $f(y,z)$ has to be chosen so that

$$\frac{\partial w}{\partial x} = \frac{\partial w_i}{\partial x}$$

on S_1, where $w_i(x,y,z)$ is the wave-function of the waves incident on the positive side of the screen.

It follows that, if the aperture lies entirely at a finite distance, the wave-function at a point P behind the screen at a distance r large compared with the wave-length is approximately

$$w = \frac{\partial}{\partial x}\left\{B(m,n)\,\frac{e^{-ikr}}{r}\right\}$$

where $(-l, -m, -n)$ are the direction-cosines of OP; the amplitude of the waves is

$$B(m,n) = \frac{1}{2\pi}\iint\limits_{S_1} f(y,z)e^{-ik(my+nz)}\,dydz.$$

To complete the approximate solution of the problem, it remains to determine $f(y,z)$ and $B(m,n)$. Now if we write $w = \partial\phi/\partial x$ where

$$\phi = \frac{1}{2\pi}\iint\limits_{S_1} f(y',z')\,\frac{e^{-ikR}}{R}\,dy'dz',$$

then ϕ is a solution of $(\nabla^2+k^2)\phi = 0$, and so the second condition becomes

$$\frac{\partial^2\phi}{\partial y^2}+\frac{\partial^2\phi}{\partial z^2}+k^2\phi = -\frac{\partial w_i}{\partial x} \tag{6.16}$$

on S_1. The expression on the right-hand side is a known function of y and z, and the problem reduces to a somewhat unusual two-dimensional boundary-value problem for the non-homogeneous equation (6.16); but this line of attack has, so far, proved of little use.

If, however, the dimensions of the aperture are small compared with the wave-length, the amplitude simplifies to

$$B = \frac{1}{2\pi}\iint\limits_{S_1} f(y,z)\,dydz$$

which is independent of the direction, and the motion behind the screen at a great distance from the aperture is the same as that due to a doublet at the origin with its axis perpendicular to the screen. The problem of determining $f(y,z)$ and B simplifies slightly; for $\partial w_i/\partial x$ is sensibly constant over the aperture and the term $k^2\phi$ in (6.16) can be neglected. We have therefore to find a function $f(y,z)$ which vanishes on the boundary of S_1 and is such that

$$\phi_0 = \frac{1}{2\pi}\iint\limits_{S_1} f(y',z')\,\frac{dy'dz'}{\rho}$$

satisfies on S_1 the equation

$$\frac{\partial^2\phi_0}{\partial y^2}+\frac{\partial^2\phi_0}{\partial z^2} = -C$$

where C is the value of $\partial w_i/\partial x$ at the origin. In the particular case when the aperture is a circle of radius a and centre O, it is possible to give an explicit solution; for it can be readily verified that

$$f(y,z) = \frac{2C}{\pi}\sqrt{(a^2-y^2-z^2)}$$

satisfies these conditions, and the amplitude is therefore

$$B = \frac{C}{\pi^2}\iint\limits_{S_1}\sqrt{(a^2-y^2-z^2)}\,dydz = \frac{2Ca^3}{3\pi}.$$

The wave-function behind the screen is thus given approximately by

$$\begin{aligned} w &= \frac{2a^3}{3\pi}\left(\frac{\partial w_i}{\partial x}\right)_0 \frac{\partial}{\partial x}\left(\frac{e^{-ikr}}{r}\right) \\ &= -\frac{2ika^3}{3\pi}\left(\frac{\partial w_i}{\partial x}\right)_0 \frac{x}{r^2}\,e^{-ikr} \end{aligned} \tag{6.17}$$

at a distance large compared with the wave-length.

Rayleigh† remarks that 'while in the first problem the wave divergent from the aperture is proportional to the first power of the linear dimension, in the present case the amplitude is very much less, being proportional to the cube of that quantity.' He observes further that from any solution it is possible to derive others by differentiation. If, for example, we take the value of v in the first problem and differentiate it with respect to x, we obtain a function which satisfies $(\nabla^2+k^2)w = 0$ and which vanishes on the screen. 'It would seem at first sight as if this could be no other than the solution of the second problem, but the manner in which the linear dimension of the aperture enters suffices to show that it is not so. The fact is that although the proposed function vanishes over the plane part of the wall, it becomes infinite at the *edge*, and thus includes the action of *sources* distributed there.'

Lastly, we note that Rayleigh showed that this difference between the effects of the two types of boundary condition also occurs in the two-dimensional problem of diffraction by a very fine slit. It turns out that, if polarized light is incident on a slit whose width is very small compared with the wave-length, 'there is a much more free passage when the electric vector is perpendicular to the slit than when it is parallel to the slit, so that unpolarized light incident upon the screen will, after passage, appear polarized in the former manner'.

† Loc. cit. Rayleigh obtains the amplitude without determining $f(y, z)$ by considering the corresponding problem for the motion of an incompressible fluid. See also Lamb, *Hydrodynamics* (Cambridge, 1916), 510–13.

§ 6.2. The variational principle of Levine and Schwinger

The diffraction of sound-waves is usually treated approximately on the assumption that the dimensions of the obstacle or aperture are small compared with the wave-length, whereas the exact opposite normally occurs in optics, and the results are accordingly very different in character. The development of short-wave radio made it necessary to investigate problems in which the obstacle or aperture has dimensions comparable with the wave-length, and neither of the classical approximations apply. In the absence of rigorous solutions, Levine and Schwinger† have devised a variational method based on the integral equations already considered but avoiding the need for solving the equations.

It follows from § 6.1 that, if plane waves of sound with velocity-potential $v_i = ae^{ik(\lambda x-\mu y-\nu z)}$, where $\lambda > 0$ and $\lambda^2+\mu^2+\nu^2 = 1$, are incident on the positive side of a screen in the plane $x = 0$, the total velocity-potential at a point P behind the screen is

$$v = \frac{e^{-ikr}}{2\pi r}\iint\limits_{S_1} f_{\mu,\nu}(y,z)e^{-ik(my+nz)}\,dydz \tag{6.21}$$

where $(-l, -m, -n)$ are the direction cosines of OP and the distance r from O to P is large compared with the wave-length. The function $f_{\mu,\nu}(y,z)$ satisfies the integral equation

$$\iint\limits_{S_1} f_{\mu,\nu}(y',z')\frac{e^{-ik\rho}}{\rho}\,dy'dz' = 2\pi ae^{-ik(\mu y+\nu z)} \tag{6.22}$$

when $(0,y,z)$ is a point of the aperture S_1. The waves at a great distance behind the screen are therefore spherical waves of amplitude

$$A(m,n;\,\mu,\nu) = \frac{1}{2\pi}\iint\limits_{S_1} f_{\mu,\nu}(y,z)e^{-ik(my+nz)}\,dydz. \tag{6.23}$$

The amplitude is a symmetrical function of the variables (m,n) and (μ,ν); for if we substitute for $e^{-ik(my+nz)}$ from the integral equation with (μ,ν) replaced by (m,n), we obtain

$$A(m,n;\,\mu,\nu) = \frac{1}{4\pi^2 a}\iint\limits_{S_1}\iint\limits_{S_1} f_{\mu,\nu}(y,z)f_{m,n}(y',z')\frac{e^{-ik\rho}}{\rho}\,dy'dz'dydz, \tag{6.24}$$

† *Phys. Rev.* **74** (1948), 958–74; **75** (1949), 1423–31. They consider the case when the wave-function is required to vanish on the screen, and use a rather different notation. The analysis in the former paper is rather obscure, as many of the integrals involved appear to diverge. See also J. W. Miles, ibid. **75** (1949), 695–6.

and hence $$A(m,n;\,\mu,\nu) = A(\mu,\nu;\,m,n).$$

This symmetry implies that the amplitude at $(-lr, -mr, -nr)$ due to incident waves with velocity-potential $ae^{ik(\lambda x-\mu y-\nu z)}$ is equal to the amplitude at $(-\lambda r,\ -\mu r,\ -\nu r)$ due to incident waves $ae^{ik(lx-my-nz)}$.

It follows from (6.23), (6.24) and the symmetry relation that

$$A(m,n;\,\mu,\nu)\iint\limits_{S_1}\iint\limits_{S_1} f_{\mu,\nu}(y,z)f_{m,n}(y',z')\,\frac{e^{-ik\rho}}{\rho}\,dy'dz'dydz$$

$$= a\iint\limits_{S_1} f_{\mu,\nu}(y,z)e^{-ik(my+nz)}\,dydz \iint\limits_{S_1} f_{m,n}(y',z')e^{-ik(\mu y'+\nu z')}\,dy'dz',$$

each side of the equation being equal to $4\pi^2 aA^2$. Let us now consider the variation δA produced by small variations $\delta f_{\mu,\nu}$ and $\delta f_{m,n}$ about the correct values $f_{\mu,\nu}$ and $f_{m,n}$ given by the integral equation. It is given by

$$\delta A(m,n;\,\mu,\nu)\iint\limits_{S_1}\iint\limits_{S_1} f_{\mu,\nu}(y,z)f_{m,n}(y',z')\,\frac{e^{-ik\rho}}{\rho}\,dy'dz'dydz$$

$$= -A(m,n;\,\mu,\nu)\iint\limits_{S_1}\iint\limits_{S_1}\{\delta f_{\mu,\nu}(y,z)f_{m,n}(y',z')+$$

$$+f_{\mu,\nu}(y,z)\delta f_{m,n}(y',z')\}\,\frac{e^{-ik\rho}}{\rho}\,dy'dz'dydz+$$

$$+a\iint\limits_{S_1}\delta f_{\mu,\nu}(y,z)e^{-ik(my+nz)}\,dydz\iint\limits_{S_1} f_{m,n}(y',z')e^{-ik(\mu y'+\nu z')}\,dy'dz'+$$

$$+a\iint\limits_{S_1} f_{\mu,\nu}(y,z)e^{ik(my+nz)}\,dydz\iint\limits_{S_1}\delta f_{m,n}(y',z')e^{-ik(\mu y'+\nu z')}\,dy'dz'.$$

The terms involving $\delta f_{\mu,\nu}$ on the right-hand side are

$$\iint\limits_{S_1}\delta f_{\mu,\nu}(y,z)\Bigg[ae^{-ik(my+nz)}\iint\limits_{S_1} f_{m,n}(y',z')e^{-ik(\mu y'+\nu z')}\,dy'dz'-$$

$$-A(m,n;\,\mu,\nu)\iint\limits_{S_1} f_{m,n}(y',z')\,\frac{e^{-ik\rho}}{\rho}\,dy'dz'\Bigg]\,dydz,$$

which vanishes for all $\delta f_{\mu\,\nu}$ in virtue of equations (6.22), (6.23) and the symmetry relation; and similarly for the terms involving $\delta f_{m,n}$. Thus $\delta A = 0$. Hence if

$$A(m, n;\, \mu, \nu) = \frac{a \iint\limits_{S_1} f_{\mu,\nu}(y, z) e^{-ik(my+nz)}\, dydz \iint\limits_{S_1} f_{m,n}(y', z') e^{-ik(\mu y'+\nu z')}\, dy'dz'}{\iint\limits_{S_1} \iint\limits_{S_1} f_{\mu,\nu}(y, z) f_{m,n}(y', z') \dfrac{e^{-ik\rho}}{\rho}\, dydzdy'dz'}, \tag{6.25}$$

and if $f_{\mu,\nu}$ and $f_{m,n}$ satisfy the integral equation, then A is stationary in the calculus of variations sense for small variations of $f_{\mu,\nu}$ and $f_{m,n}$.

Conversely, let us suppose that A, defined by (6.25), is stationary in the calculus of variations sense. It then follows that $f_{m,n}$ and $f_{\mu,\nu}$ must satisfy the integral equations

$$\begin{aligned} \iint\limits_{S_1} f_{m,n}(y', z') \frac{e^{-ik\rho}}{\rho}\, dy'dz' &= 2\pi\kappa a e^{-ik(my+nz)}, \\ \iint\limits_{S_1} f_{\mu,\nu}(y', z') \frac{e^{-ik\rho}}{\rho}\, dy'dz' &= 2\pi\kappa' a e^{-ik(\mu y+\nu z)}, \end{aligned} \tag{6.26}$$

where $(0, y, z)$ is a point of S_1, and κ and κ' are the constants defined by

$$2\pi\kappa = \iint\limits_{S_1} f_{m,n}(y', z') e^{-ik(\mu y'+\nu z')}\, dy'dz'/A(m, n;\, \mu, \nu),$$

$$2\pi\kappa' = \iint\limits_{S_1} f_{\mu,\nu}(y', z') e^{-ik(my'+nz')}\, dy'dz'/A(m, n;\, \mu, \nu).$$

Comparing (6.26) and (6.23), we see that the $f_{\mu,\nu}$ of (6.26) is κ' times the $f_{\mu,\nu}$ of (6.23); but since the expression on the right-hand side of (6.25) is homogeneous of degree zero in $f_{\mu,\nu}$, we may take κ' (and similarly κ) to be unity without loss of generality. Hence $f_{\mu,\nu}$ and $f_{m,n}$ satisfy the integral equation, and A is the amplitude of the waves at a large distance behind the screen.

We have thus proved that *the expression*† $A(m, n;\, \mu, \nu)$ *defined by* (6.25) *is the required amplitude if and only if it is stationary for small variations of* $f_{m,n}$ *and* $f_{\mu,\nu}$. This is the variational principle of Levine and Schwinger.

It is the amplitude A, rather than the function $f_{\mu,\nu}$, which is of

† It will be observed that A is proportional to the amplitude a of the incident waves, as would be expected.

physical importance; the great value of the variational principle is that it enables one to find the amplitude approximately by assuming for $f_{m,n}$ and $f_{\mu,\nu}$ physically plausible expressions involving certain arbitrary constants and then choosing these constants to make A stationary. In this way, Levine and Schwinger have solved numerically the problem of the diffraction of plane waves incident normally on a circular aperture for values of the characteristic parameter

$$ka = 2\pi(\text{radius of aperture})/(\text{wave-length})$$

between 0 and 10, and have found excellent agreement with the rigorous results obtained by C. J. Bouwkamp, in his 1941 Groningen dissertation, using expansions in terms of spheroidal wave-functions.

INDEX

PRINTED IN GREAT BRITAIN
AT THE UNIVERSITY PRESS, OXFORD
BY CHARLES BATEY, PRINTER TO THE UNIVERSITY