Proc. Indian Assoc. Cult. Sci. 9 271–286 (1926)

On the total reflection of light

PROFESSOR C V RAMAN, F.R.S.

1. Introduction

It has long been known that the explanation of the phenomenon of total reflection of light on the principles of the wave-theory involves the existence in the second medium of a disturbance which penetrates beyond the boundary to a depth depending on the angle of incidence and diminishing exponentially as the perpendicular distance of the point of observation from the boundary is increased. Stokes* showed how the expression for this disturbance which he designated as a *superficial undulation* may be derived directly from the Fresnel formulae for the intensities of the reflected and refracted beams of light, and applied the same method to the investigation of the appearance of the central spot in Newton's Rings formed beyond the critical angle of incidence. A discussion of the problem on the principles of the electromagnetic theory is given in Drude's† Theory of Optics, where the question of the flow of energy in the second medium is also considered on the basis of Poynting's Theorem. That the superficial disturbance in the second medium must be a physical reality is indicated by the consideration that it is closely related to the changes of phase occurring in total reflection and that the same theory which predicts it also gives a quantitative explanation of the elliptic polarisation actually observed when light plane-polarised in any azimuth is totally reflected.‡ Further, the phenomena of Newton's Rings beyond the critical angle, already mentioned, and the fact that small particles placed in contact with the boundary in the second medium are observed to scatter light when viewed through a microscope are usually regarded as confirming the theory. Some doubt has however been thrown on the usual treatment in a recent theoretical paper§ by Sir Arthur Schuster who appears to

---

* *Math. Phys. Papers*, **2**, 57.

† English Translation, p. 299.

‡ So far as the writer is aware, no measurements of the *absolute* change of phase of the light-vector for the two principal components taken *separately* have been made for the case of total reflection at any angle. For the case of total reflection at grazing incidence however, Bevan has made observations by the method of Lloyd's interference-fringes (*Philos. Mag.*, Oct. 1907), which are in agreement with theory.

§ *Proc. R. Soc. London* **A107**, p. 15.

hold that the assumption made in the theory of an infinitely extended surface is essentially illegitimate. Moreover, the present writer has recently shown* that by using intense monochromatic light (the green or violet line of the mercury arc), and an ordinary spectroscope, the light emerging from the second face of a prism on which the light was incident at an angle greater than the critical angle could be readily observed. Photographs showing the disturbance emerging into the second medium were published, and they clearly indicated that the effect observed was due to the limitation of the aperture of the pencil incident on the surface and was thus primarily a phenomenon of diffraction. Similar effects were also observed with a Lummer-Gehrcke plate when light was incident within the plate at an angle greater than the critical angle. These effects clearly indicate that diffraction does play a part in the phenomena of total reflection, and it becomes necessary to consider the matter afresh in the light of the new experimental evidence now available. It is proposed in this paper to consider, *de novo*, the phenomena of total reflection from the point of view of diffraction theory.

2. Application of the Fresnel–Huyghens principle

In the general explanation of total reflection first given on the principles of the wave-theory by Huyghens, the elementary parts of the boundary between the two media are regarded as the source of secondary wavelets emerging into both media. That there is no refracted wave in the second medium though the boundary is fully illuminated is a consequence of the fact that no common envelope can be drawn to the wavelets emerging into it. There is little doubt that the more recondite phenomena accompanying total reflection may also be explained by following up Huyghens's original line of thought and applying the principle of interference. In particular, the disturbance existing in the second medium at points close to the boundary, and the diffraction effects arising from the finiteness of the illuminated area should both be capable of determination in this way.

The first step in such a treatment is the marking out of the Fresnel zones on the boundary between the two media. When this is of limited area and the point at which the effect is to be determined is far away from it, the Fresnel zones obviously become parallel rectilinear strips on the surface, and the determination of the integrated effect due to all the zones follows the ordinary methods of diffraction theory. We find in fact that the surface on which light is incident beyond the critical angle and is "totally" reflected sends out into the second medium streamers of light giving rise to diffraction-patterns in the usual way. These diffraction-patterns differ however from those of the ordinary kind in being

* *Philos. Mag.*, 6th Series, **50**, 812.

strongly asymmetrical in character and also "truncated," that is to say, they consist only of certain outlying and relatively faint parts of the diffraction-patterns associated with the forms of aperture used, the principal and relatively intense parts being absent. For, none of the Fresnel zones included within the area correspond to a pole or region of stationary phase. Since the light thus streaming into the second medium represents energy, the reflection occurring at the boundary technically ceases to be total, though practically, the departure from totality is negligibly small, unless the aperture is very small or the incidence is only slightly greater than the critical angle. The streamers of light emerging into the second medium, have, as in all cases of diffraction, their origin at the margins of the diffracting area. The front and rear parts of the boundary are, at distant points, equally operative. The observations of Dr Chuckerbutti* and those of the present writer already quoted on the effects observed with a surface bounded by parallel edges entirely agree with these indications of theory. One special feature which comes into prominence in these observations and which deserves to be emphasised is that the intensity of the diffraction-pattern is zero at all distant points lying in the plane of the boundary when produced, that is to say, at all points from which when viewed, the angular aperture of the illuminated area is zero. As we may move away from this plane, the diffraction-pattern steadily gains in intensity. This may be regarded as an effect due to the variation of the "obliquity-factor" of diffraction, and is, in fact, thus explained in the papers already quoted. It has the influence of altering enormously the relative intensities of the diffraction-bands and making them very different from those calculated in the usual way.

3. Disturbance at points close to the surface

The same method of treatment may be applied to the other case of interest, namely, the effect at a point in the second medium very close to the illuminated area. It is a fact of observation that for angles of incidence exceeding the critical angle, the illumination dies away very quickly as we move away from the surface and the chief interest is thus in determining the effect at points lying within a distance of a few wavelengths from it. At such small distances, the usual approximate methods of finding the effect due to the Fresnel zone and of integrating the same over the whole of the surface of resolution are not quite rigorous. Nevertheless, as will be shown below, they may be applied with success to the elucidation of the particular case under consideration. In fact, even by merely considering the geometrical form of the Fresnel zones, a considerable insight into the problem may be obtained.

---

* *Proc. R. Soc. London* **A99**, 503, 1921.

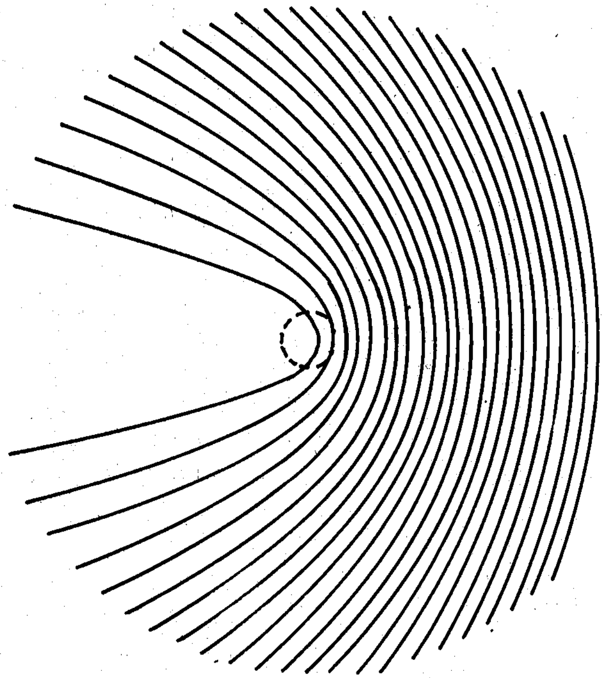

Figure 1. Fresnel zones on surface. Incidence = 45° = critical angle. Point of observation is on the surface.

The form of the Fresnel zones over the surface for any angle of incidence and for any point of observation may be readily mapped out in the following way. From the point of observation, a perpendicular may be dropped on the surface, and round its foot as centre, a set of circles spaced at half-period intervals from the point of observation are drawn. Crossing these are drawn a set of equidistant straight lines perpendicular to the plane of incidence and spaced at such intervals that the distance from one straight line to the next corresponds to a change of phase of the incident waves of half a period. The circles may be numbered, commencing from the centre outwards, 0, 1, 2, 3, 4, 5, etc. The straight line passing through the centre may be numbered 0, and those to the right of it, 1, 2, 3, 4, etc., and those to the left of it − 1, − 2, − 3, − 4, etc. The points of intersections of the circles and straight lines are then marked with the sum of the index-numbers corresponding to the particular circle and straight line cutting at each such point. These index-numbers represent the total difference of path between the secondary waves reaching the point of observation from the nearest element of the surface and from any other. Smooth curves may now be drawn free-hand or with the aid of a flexible steel strip through all the points having identical index-numbers. Very instructive diagrams may be obtained in this way for any specified angle of

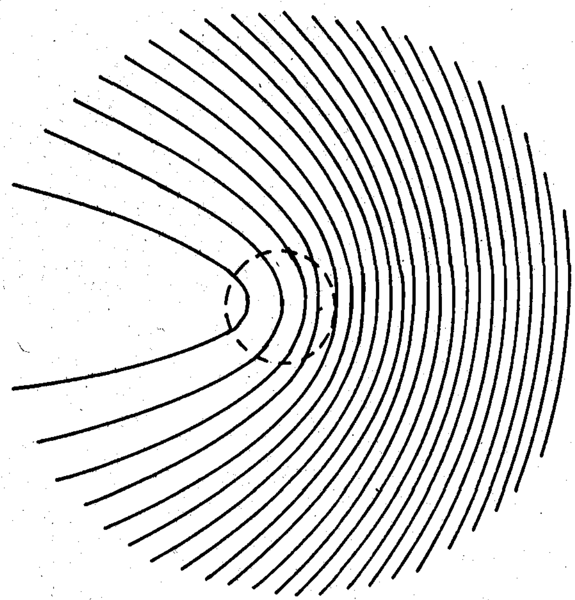

Figure 2. Fresnel zones on surface. Incidence = 45° = critical angle. Point of observation is λ above the surface.

incidence and for any assigned value of the perpendicular distance from the boundary, and they give accurately the form of the Fresnel zones.

The geometrical form of the curves drawn in this way shows the following general features. The zones for all angles of incidence in excess of the critical angle are approximately hyperbolic in form. The fact that they are not closed curves indicates that for no point of observation does the surface present any pole or region of stationary phase. The curvature of the lines is most marked for points of observation near the surface; as the distance is increased, the lines become more and more nearly straight. The spacing of the zones draws a striking diminution as we pass in the plane of incidence from negative to positive values of x, that is, from the left to the right of the foot of the perpendicular drawn from the point of observation. This change in the spacing is the more sudden, the smaller is the distance of the point of observation measured perpendicularly from the boundary. It is largest at the critical angle and diminishes with increasing angle of incidence. Figures 1, 2, 3 and 4 represent the form of the Fresnel zones for particular cases and illustrate the foregoing remarks. Figures 1 and 2 represent the case of incidence at the critical angle 45° and figures 3 and 4 for incidence at 60°. The refractive index μ is taken as 1.414. In figure 1 and figure 3 the point of observation is on the surface. In figure 2 it is λ above from the surface and in figure 4, 4λ above the surface. The centre of the smallest unit index circle drawn in each

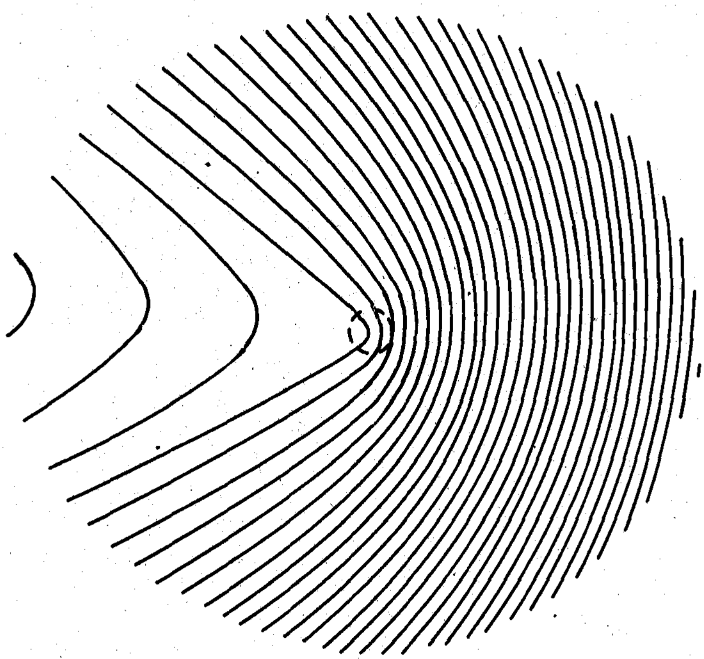

Figure 3. Fresnel zones on surface. Incidence = 60°. Point of observation is on the surface.

case is the foot of the perpendicular from the point of observation.

When we proceed to sum up the effects of the different Fresnel zones, taking into account the varying distances from the point of observation and the varying obliquities, we should obtain an idea of the way in which the residual effect observed in the second medium varies with the point of observation. Viewed in this way, it is seen that the penetration of the disturbance into the second medium in "total reflection" may always be regarded as a diffraction-effect. It is important however to determine what part of the effect arises from the outermost parts of the surface and what part from the area closest to the point of observation.

If the Fresnel zones had been uniformly spaced to the right and left of the foot of the perpendicular from the point of observation, they would have annulled each other's effects and given zero as the resultant disturbance. Actually however, as we have seen, there is a change in the spacing as we pass from left to right which is the more sudden, the closer we approach the surface between the two media. The summation over the Fresnel zones would therefore give a resultant effect which is the larger, the more nearly the point of observation approaches the surface. This effect arises from the part of the surface nearest the point of observation, and may be identified with the "superficial undulation" of Stokes and other writers. Since the change in the spacing of the Fresnel zones is most

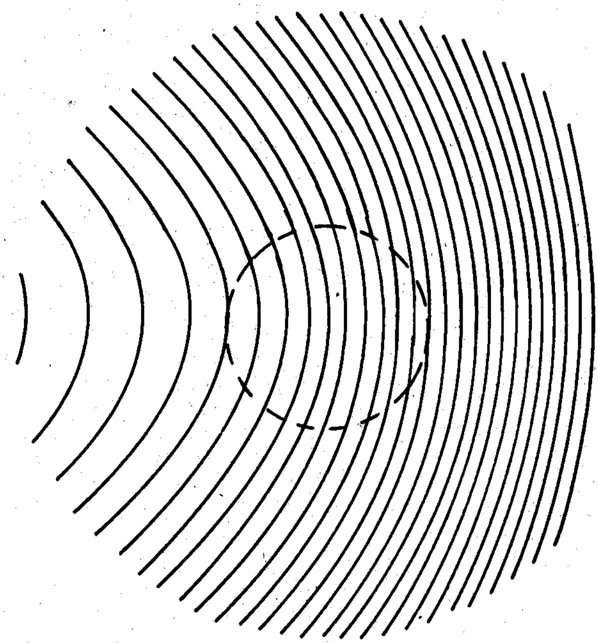

Figure 4. Fresnel zones on surface. Incidence at 60°. Point of observation 4λ above the surface.

marked when the incidence is just at the critical angle and diminishes rapidly as the incidence is increased, we should expect the resultant effect to diminish in the same way. This is in agreement with the "superficial undulation" formula. We have already seen that the obliquity-factor of diffraction becomes vanishingly small when the angular aperture of the surface as viewed from the point of observation approaches zero. When the point of observation is sufficiently close to the surface, the obliquity becomes practically 90° for all the elements of the surface except those nearest to it. It follows that when the integration is carried out over all the Fresnel zones, the marginal parts of the surface contribute nothing and may be neglected.

Analytical treatment of the problem

The preceding discussion indicates that the superficial undulation in the second medium is a diffraction-effect which arises, not from the margins of the illuminated area, but from the part of the area nearest the point of observation. This may be confirmed by mathematical analysis which indeed shows that the expected effect diminishes exponentially with the distance according to the law already derived from other considerations.

The diffraction integral expressing the effect at any point may easily be written down if we know the law of the secondary wave. The elementary disturbances arising from an area held obliquely to the wave-front have been expressed mathematically in Kirchoff's well-known formulation of Huyghens's principle. As has been remarked by various writers, however, Kirchoff's expression is not a unique solution of the problem, as an infinite number of formulae for the law of the secondary wave may be written down, all of which express correctly the disturbance in free space arising from specified light-sources. In our case, we are, moreover, dealing not with free space, but with the effects observed in the vicinity of a surface of separation between two media. The law of the secondary wave for this case has yet to be determined. For our present purpose, it is sufficient to proceed in the usual simple way and assume that the amplitude of the secondary wave is proportional to the area of the element from which it is sent out and inversely as the product of the wavelength and the distance of the element from the point of observation, and ignore all consideration of the obliquity factor. Let the surface be taken as coinciding with the xy plane, and the plane of incidence be taken as the xz plane. Further, let the point of observation be assumed to be on the Z-axis at a distance Z, from the origin, the latter being thus on the surface at the foot of the perpendicular drawn from the point of observation. Let r be the distance of an elementary area on the surface from the origin.

An element of area on the surface is $r dr d\theta$ and the resultant effect is

$$\text{Const.} \int_0^\infty \int_0^{2\pi} \frac{A}{\lambda(z^2+r^2)^{1/2}} \cos\frac{2\pi}{\lambda}\{Vt-(z^2+r^2)^{1/2}-r\mu\sin\phi\cos\theta+\varepsilon\} r dr d\theta,$$

where μ is the refractive index of the first medium, the second medium being assumed to be free space, ϕ is the angle of incidence on the surface, and θ is the angle which the radius vector r drawn on the surface makes with the plane of incidence. ε is the phase difference between the primary disturbance and the secondary waves to which it gives rise. The integral is assumed to be taken over a sufficiently extended area. It is obvious from physical considerations that the expression must give results which differ entirely in character according as

$$\mu\sin\phi \lessgtr 1,$$

that is, according as the incidence is less or greater than the critical angle. This agrees, as we shall see presently, with the actual results of integration.

Integrating with respect to θ and writing

$$2\pi/\lambda\cdot(Vt+\varepsilon)=\chi$$

for shortness, the expression reduces to the form

$$\text{Const.} \int_0^\infty \frac{2\pi A}{\lambda(r^2+z^2)^{1/2}} \cos\{\chi-2\pi/\lambda\cdot(r^2+z^2)^{1/2}\} J_o(2\pi/\lambda\cdot r\mu\sin\phi) r dr.$$

When z is put equal to zero, that is, on the surface itself, the expression reduces to

$$\text{Const.}\frac{2\pi A}{\lambda}\left[\cos\chi\int_0^\infty \cos 2\pi r/\lambda\cdot J_0(2\pi r/\lambda\cdot\mu\sin\phi)\mathrm{d}r\right.$$
$$\left.+\sin\chi\int_0^\infty \sin 2\pi r/\lambda\cdot J_0(2\pi r/\lambda\cdot\mu\sin\phi)\mathrm{d}r\right].$$

The integrals appearing within the square brackets are well-known standard forms, the values of which depend on whether $\mu\sin\phi$ is greater or less than unity. If $\mu\sin\phi<1$ the first integral vanishes and the second becomes equal to

$$(1-\mu^2\sin^2\phi)^{-1/2}$$

whereas if $\mu\sin\phi>1$, the second integral vanishes and the first becomes equal to

$$(\mu^2\sin^2\phi-1)^{-1/2}.$$

The case $Z=0$ corresponds to the surface of separation and in order that our result might reduce to the primary disturbance on the surface, the constants expressing the law of secondary wave must be suitably chosen. It is necessary to assume different values for them in the two cases:

If $\mu\sin\phi<1$, Const. $=\sqrt{1-\mu^2\sin^2\phi}$
and $\varepsilon=\pi/2$.

If $\mu\sin\phi>1$, Const. $=\sqrt{\mu^2\sin^2\phi-1}$
and $\varepsilon=0$.

We shall now substitute these values in the general expression, considering the two cases separately.

Case I. Incidence less than the critical angle and $\mu\sin\phi<1$.

The expression for the light-disturbance given above involves the evaluation of two integrals, namely

$$\int_0^\infty \sin 2\pi/\lambda\cdot(r^2+Z^2)^{1/2}\cdot J_0(2\pi/\lambda\cdot r\mu\sin\phi)(r^2+Z^2)^{-1/2}r\mathrm{d}r$$

and

$$\int_0^\infty \cos 2\pi/\lambda\cdot(r^2+Z^2)^{1/2}\cdot J_0(2\pi/\lambda\cdot r\mu\sin\phi)(r^2+Z^2)^{-1/2}r\mathrm{d}r.$$

Using the well-known formulae

$$J_{-1/2}(x)=\sqrt{\frac{2}{\pi x}}\cos x,\quad\text{and}\quad J_{1/2}(x)=\sqrt{\frac{2}{\pi x}}\sin x$$

the two integrals under consideration are found to be special cases of a very general type of integral involving products of Bessel functions which has been discussed by Sonine.* We shall however adopt a different method of evaluation. Lamb† has proved the following formula:

$$\int_0^\infty -\exp[\alpha(\xi^2-\eta^2)^{1/2}]\cdot J_0(\beta\xi)\cdot(\xi^2-\eta^2)^{-1/2}\xi d\xi = \frac{\exp[-i\eta(\alpha^2+\beta^2)^{1/2}]}{(\alpha^2+\beta^2)^{1/2}}.$$

In this relation, write $\xi = r$ and $\eta^2 = -Z^2$. Also put

$$\beta = 2\pi/\lambda\cdot\mu\sin\phi \quad \text{and} \quad \alpha^2 = -(2\pi/\lambda)^2.$$

The equation then stands thus:

$$\int_0^\infty \exp[-2\pi i/\lambda(r^2+Z^2)^{1/2}]\cdot J_0(2\pi/\lambda\cdot r\cdot\mu\sin\phi)\cdot(r^2+Z^2)^{-1/2}r dr$$

$$= -i\cdot\lambda/2\pi\cdot(1-\mu^2\sin^2\phi)^{-1/2}\cdot\exp[-iZ\cdot 2\pi/\lambda\cdot(1-\mu^2\sin^2\phi)^{-1/2}].$$

Separating the real and imaginary parts we have

$$\int_0^\infty \sin 2\pi/\lambda\cdot(r^2+Z^2)^{1/2}\cdot J_0(2\pi/\lambda\cdot r\cdot\mu\sin\phi)\cdot(r^2+Z^2)^{-1/2}r dr$$

$$= \lambda/2\pi\cdot(1-\mu^2\sin^2\phi)^{-1/2}\cdot\cos\{2\pi/\lambda\cdot Z\cdot(1-\mu^2\sin^2\phi)^{1/2}\}$$

and

$$\int_0^\infty \cos 2\pi/\lambda\cdot(r^2+Z^2)^{1/2}\cdot J_0(2\pi/\lambda\cdot r\cdot\mu\sin\phi)\cdot(r^2+Z^2)^{-1/2}r dr$$

$$= -\lambda/2\pi\cdot(1-\mu^2\sin^2\phi)^{-1/2}\cdot\sin\{2\pi/\lambda\cdot Z\cdot(1-\mu^2\sin^2\phi)^{1/2}\}.$$

These results are confirmed by comparison with the general formulae given by Sonine and Nielsen. Substituting the values of the integrals in the expression for the disturbance in the second medium, we find that the latter reduces to

$$A\cos 2\pi/\lambda\cdot(Vt - x\mu\sin\phi - Z\sqrt{1-\mu^2\sin^2\phi})$$

which is of the same form as the ordinary expression for the refracted wave.

Case II. Incidence at more than the critical angle, and $\mu\sin\phi > 1$.

With the same substitutions as before, Lamb's formula now reads thus:

$$\int_0^\infty \exp[-2\pi i/\lambda\cdot(r^2+Z^2)^{1/2}]\cdot J_0(2\pi/\lambda\cdot r\cdot\mu\sin\phi)(r^2+Z^2)^{-1/2}r dr$$

$$= \lambda/2\pi\cdot(\mu^2\sin^2\phi-1)^{-1/2}\exp[-Z\cdot 2\pi/\lambda\cdot(\mu^2\sin^2\phi-1)^{1/2}].$$

---

* *Math. Annalen, Band* **16**, p. 1. See also Nielsen, Cylinderfunction, 1904.

† *Philos. Trans. R. Soc. London* **A203**, 1904, p. 5.

Separating the real and imaginary parts, we have

$$\int_0^\infty \sin\cdot\frac{2\pi}{\lambda}\cdot(r^2+z^2)^{1/2}\cdot J_0(z\pi/\lambda\cdot r\cdot\mu\sin\phi)(r^2+z^2)^{-1/2}r\mathrm{d}r=0$$

and

$$\int_0^\infty \cos\frac{2\pi}{\lambda}(r^2+z^2)^{1/2}\cdot J_0 2\pi/\lambda\cdot r\cdot\mu\sin\phi)\cdot(r^2+z^2)^{-1/2}r\mathrm{d}r$$

$$=\lambda/2\pi\cdot(\mu^2\sin^2\phi-1)^{-1/2}\exp\left[\cdot-Z\frac{2\pi}{\lambda}(\mu^2\sin^2\phi-1)^{1/2}\right].$$

Substituting these values in the expression for the disturbance in the second medium when total reflection is occurring, we find that the latter reduces to

$$A\cdot\cos\frac{2\pi}{\lambda}\cdot(Vt-x\mu\sin\phi)\exp\left[-Z\frac{2\pi}{\lambda}(\mu^2\sin^2\phi-1)^{1/2}\right].$$

Our investigation thus leads to precisely the same law of exponential decay as that derived from the Fresnel formulae for the superficial disturbance in the second medium, and the view that the latter is a diffraction-effect arising from the immediately contiguous part of the surface is thus fully substantiated.

In evaluating the diffraction integral, the area of the surface was taken as infinite, and we found that the case $\mu\sin\phi=1$ marks a point of discontinuity at which the phase of the secondary waves alters suddenly by quarter of a period, and their amplitude becomes very large. This circumstance and the form of the Fresnel zones drawn in figures 1 and 2 show that when the incidence is exactly at the critical angle, the finite extent of the surface cannot be ignored and must be taken into account for a more exact discussion. When however, the incidence is increased beyond the critical angle, the marginal portions of the area cease to be of importance in determining the observed effect at points not far from the surface. The further discussion of the phenomena at or very near the critical incidence and close to the surface on the basis of the integrals already given is a problem worthy of investigation which must however be deferred for the present.