DET KGL. DANSKE VIDENSKABERNES SELSKAB
MATEMATISK-FYSISKE MEDDELELSER, BIND XXII, NR. 8

A NOTE ON THE FOUNDATIONS OF GEOMETRICAL OPTICS

BY

NIELS ARLEY

KØBENHAVN
I KOMMISSION HOS EJNAR MUNKSGAARD
1945

Printed in Denmark
Bianco Lunos Bogtrykkeri A/S

§ 1.

As is well-known all the laws of geometrical optics may be deduced from FERMAT's *principle*[1]

$$\delta\int_{P'}^{P} n ds = \delta\int_{\tau'}^{\tau} n \sqrt{\left(\frac{dx}{d\tau}\right)^2 + \left(\frac{dy}{d\tau}\right)^2 + \left(\frac{dz}{d\tau}\right)^2}\, d\tau = 0 \tag{1}$$

or from the mathematically equivalent *principle of* HUYGENS[1]. In (1) τ is an arbitrary parameter of the integration curve from P' to P, and n is the *ray index* given by

$$n = n_r = \frac{c}{v_r}, \tag{2}$$

c being the velocity of light in vacuo and v_r the *velocity of the ray*, i. e. of the energy current[2]. Thus the integral in (1) is simply c times the time-interval from P' to P. In the most general case, v_r and thus n are functions of the point, the direction, and the colour, i. e. the frequency of the light. In geometrical optics we only consider *monochromatic light*, i. e. we abstract from *dispersion phenomena*. If for a certain medium n is independent of the point, the medium is *homogeneous*, if n is independent of the direction, it is *isotropic*.

In most textbooks on optics[3] FERMAT's principle is only proved to hold true for a *finite* number of reflections and refractions in an isotropic medium with piece-wise constant index of

[1] This is shown most completely in CARATHÉODORY (1937). In this paper the theory is developed quite generally for an inhomogeneous, anisotropic medium by means of the methods due to HAMILTON. Furthermore, it also briefly outlines the history of the two famous principles.

[2] n is often erroneously stated as being the *index of refraction*, $n_n = \frac{c}{v_n}$, v_n being the *normal* or *phase velocity*. This statement is only true in the case of isotropic media, the normal and the ray directions as well as v_n and v_r coinciding only in this case.

[3] See e. g. the well-known textbook of DRUDE (1900) or the modernized version of this standard treatise: FÖRSTERLING (1928).

refraction. Furthermore, HUYGENS' principle is as a rule only stated but not proved[1].

Now it is an obvious task to deduce the two fundamental principles in the *general* case of an inhomogeneous, anisotropic, absorbing medium from MAXWELL's electro-magnetic theory of light, at the same time deducing the conditions for the validity of geometrical optics. It is the purpose of the present note to work out this programme, which surprisingly enough has not, as far as we know, been done before. We thereby proceed by combining a method due to SOMMERFELD and RUNGE[2] with the general theory of partial differential equations of the first order and their connection with the calculus of variations.[3] Although the following considerations do not, of course, yield any new results, they may perhaps be of some pedagogic interest.

§ 2.

Any propagation of light in an arbitrary non-ferromagnetic medium is governed by the four MAXWELL equations

$$\operatorname{rot} \boldsymbol{H} = \frac{4\pi}{c} \sigma \cdot \boldsymbol{E} + \frac{1}{c} \varepsilon \cdot \dot{\boldsymbol{E}} \tag{1}$$

$$\operatorname{rot} \boldsymbol{E} = -\frac{\mu}{c} \dot{\boldsymbol{H}} \tag{2}$$

$$\operatorname{div} \varepsilon \cdot \boldsymbol{E} = 0, \ \operatorname{div} \mu \boldsymbol{H} = 0, \tag{3}$$

in which $\varepsilon = \varepsilon(x, y, z)$ and $\sigma = \sigma(x, y, z)$ are symmetric tensor-functions (indicated by a dot in all products) and $\mu = \mu(x, y, z)$ a scalar function (no dots in the products). The main equations

[1] See e. g. the most modern textbook on optics: BORN (1933). In LANDÉ (1928) HUYGENS' principle is deduced from that of FERMAT, but this principle itself has not been proved.

[2] SOMMERFELD and RUNGE (1911). This method is also described in BORN's textbook, in PLANCK (1927) and JENTZSCH (1927).

[3] See e. g. COURANT-HILBERT (1937) or CARATHÉODORY (1935). The method of SOMMERFELD and RUNGE is in fact only a special case of a general method for obtaining the characteristic equation belonging to an arbitrary partial differential equation of the second order with n variables, see COURANT-HILBERT (1937) chap. VI § 10.1. Furthermore, the same method has been used by DIRAC (1930), p. 120, to deduce the corresponding transition from wave to classical mechanics. In this calculation a term $-i\hbar \sum_r \frac{\partial^2 S}{\partial q_r^2}$ has, however, erroneously been omitted; this term is just important because the condition that it may be neglected imposes the restriction that the radius of curvature of the "path" is to be large compared with the wave-length (cf. p. 8 below).

are (1) and (2), while (3) only imposes certain conditions on the direction of the field-vectors $\boldsymbol{E}$ and $\boldsymbol{H}$. In fact it follows from (1) and (2) that if (3) is satisfied at the initial time, (3) will be satisfied for all times. We eliminate $\boldsymbol{H}$ between (1) and (2) by taking the time derivative of (1) and rot of (2), obtaining by means of the vector identities

$$\left.\begin{aligned} \operatorname{rot} \varphi \boldsymbol{A} &= \varphi \operatorname{rot} \boldsymbol{A} + [\operatorname{grad} \varphi \times \boldsymbol{A}] \\ \text{and} \qquad \operatorname{rot}\operatorname{rot} \boldsymbol{A} &= -\Delta \boldsymbol{A} + \operatorname{grad}\operatorname{div} \boldsymbol{A} \end{aligned}\right\} \tag{4}$$

the general *wave equation*

$$\Delta \boldsymbol{E} - \operatorname{grad}\operatorname{div} \boldsymbol{E} + \left[\frac{\operatorname{grad} \mu}{\mu} \times \operatorname{rot} \boldsymbol{E}\right] - \frac{4\pi\mu}{c^2}\sigma \cdot \dot{\boldsymbol{E}} - \frac{\mu}{c^2}\varepsilon \cdot \ddot{\boldsymbol{E}} = \boldsymbol{0}. \tag{5}$$

Having calculated $\boldsymbol{E}$ from (5) and the boundary conditions we obtain $\boldsymbol{H}$ from (2). We now assume the time variation of $\boldsymbol{E}$ to be strictly harmonious (otherwise we only need expand $\boldsymbol{E}$ in a Fourier series) i. e. we put

$$\boldsymbol{E} = \boldsymbol{F} e^{i\omega t}, \quad \omega = 2\pi\nu, \tag{6}$$

which inserted in (5) gives the time independent wave equation, being a system of 3 simultaneous, linear, homogeneous, partial differential equations of the second order,

$$\lambda_0^2\left(\Delta \boldsymbol{F} - \operatorname{grad}\operatorname{div} \boldsymbol{F} + \left[\frac{\operatorname{grad} \mu}{\mu} \times \operatorname{rot} \boldsymbol{F}\right]\right) - \frac{4\pi\mu i}{c}\lambda_0 \sigma \cdot \boldsymbol{F} + \mu\varepsilon \cdot \boldsymbol{F} = \boldsymbol{0}. \tag{7}$$

Here

$$\lambda_0 = \frac{c}{\omega} \tag{8}$$

is the vacuum wave-length divided by 2π.

Now geometrical optics is just characterized as that branch of optics in which we may consider λ_0 as infinitely small, i. e. strictly speaking obtained by making the limit $\lambda_0 \to 0$. In order to obtain a solution of (7) which will approximately describe a *light-ray*, i. e. a wave propagation which vanishes outside a very narrow region which in the limit $\lambda_0 \to 0$ may be considered as a geometrical curve, we put $\boldsymbol{F}$ approximately equal to a homogeneous plane wave

$$\boldsymbol{F} = \boldsymbol{A}(x, y, z) \exp\left[\frac{i}{\lambda_0} S(x, y, z)\right], \tag{9}$$

in which the amplitude $\boldsymbol{A} = \boldsymbol{A}(x, y, z)$ is a slowly varying vector-function, being approximately constant along the light-ray and zero outside it, and the phase function $S = S(x, y, z)$, describing the wave fronts, deviates but little from the linear function $n(\boldsymbol{s}\cdot\boldsymbol{r}) = n(s_x x + s_y y + s_z z)$. (In order to obtain the *exact* solution we had, as is well-known, to superpose plane waves with slightly different wave vectors $\boldsymbol{s}$). Using the well-known vector identities (4),

$$\left.\begin{aligned} \operatorname{grad}(\varphi\psi) &= \varphi \operatorname{grad}\psi + \psi \operatorname{grad}\varphi, \\ \operatorname{div}(\varphi \boldsymbol{A}) &= \varphi \operatorname{div}\boldsymbol{A} + (\boldsymbol{A}\cdot\operatorname{grad}\varphi) \end{aligned}\right\} \tag{10}$$

and

$$\operatorname{grad}(\boldsymbol{A}\cdot\boldsymbol{B}) = (\boldsymbol{A}\cdot\operatorname{grad}\boldsymbol{B}) + (\boldsymbol{B}\cdot\operatorname{grad}\boldsymbol{A}) + [\boldsymbol{A}\times\operatorname{rot}\boldsymbol{B}] + [\boldsymbol{B}\times\operatorname{rot}\boldsymbol{A}]$$

we obtain (using furthermore the fact that rot grad $\equiv 0$)

$$\left.\begin{aligned} \operatorname{rot}\boldsymbol{F} &= \frac{i}{\lambda_0}\exp\left[\frac{i}{\lambda_0}S\right][\operatorname{grad}S\times\boldsymbol{A}] + \exp\left[\frac{i}{\lambda_0}S\right]\operatorname{rot}\boldsymbol{A} \\ \Delta\boldsymbol{F} &= \operatorname{div}\operatorname{grad}\boldsymbol{F} = \left(\frac{i}{\lambda_0}\right)^2\exp\left[\frac{i}{\lambda_0}S\right](\operatorname{grad}S)^2\boldsymbol{A} + \\ &+ \frac{i}{\lambda_0}\exp\left[\frac{i}{\lambda_0}S\right]\left(2(\operatorname{grad}\boldsymbol{A}\cdot\operatorname{grad}S) + \boldsymbol{A}\Delta S\right) + \exp\left[\frac{i}{\lambda_0}S\right]\Delta\boldsymbol{A} \\ \operatorname{grad}\operatorname{div}\boldsymbol{F} &= \left(\frac{i}{\lambda_0}\right)^2\exp\left[\frac{i}{\lambda_0}S\right](\boldsymbol{A}\cdot\operatorname{grad}S)\operatorname{grad}S + \\ &+ \frac{i}{\lambda_0}\exp\left[\frac{i}{\lambda_0}S\right]\Big(\operatorname{div}\boldsymbol{A}\operatorname{grad}S + \left(\boldsymbol{A}\cdot\operatorname{grad}(\operatorname{grad}S)\right) + \\ &+ (\operatorname{grad}S\cdot\operatorname{grad}\boldsymbol{A}) + [\operatorname{grad}S\times\operatorname{rot}\boldsymbol{A}]\Big) + \exp\left[\frac{i}{\lambda_0}S\right]\operatorname{grad}\operatorname{div}\boldsymbol{A}. \end{aligned}\right\} \tag{11}$$

(11) inserted in the wave equation (7) finally, after division with $\exp\left[\frac{i}{\lambda_0}S\right]$, gives

$$\left.\begin{aligned} &\left(-(\operatorname{grad}S)^2 + \operatorname{grad}S(\operatorname{grad}S\cdot) + \mu\varepsilon\cdot\right)\boldsymbol{A} + i\lambda_0\Big((\operatorname{grad}S\cdot\operatorname{grad}\boldsymbol{A}) + \\ &+ \boldsymbol{A}\Delta S - \operatorname{grad}S\operatorname{div}\boldsymbol{A} - \left(\boldsymbol{A}\cdot\operatorname{grad}(\operatorname{grad}S)\right) - [\operatorname{grad}S\times\operatorname{rot}\boldsymbol{A}] + \\ &+ \left[\frac{\operatorname{grad}\mu}{\mu}\times[\operatorname{grad}S\times\boldsymbol{A}]\right] - \frac{4\pi}{c}\sigma\cdot\boldsymbol{A}\Big) + \lambda_0^2(\Delta\boldsymbol{A} - \operatorname{grad}\operatorname{div}\boldsymbol{A}) = \boldsymbol{0}, \end{aligned}\right\} \tag{12}$$

in which grad S(grad $S\cdot$) denotes the tensor product of grad S with itself. If now λ_0 is a very small number, we see in fact that (9) is a solution of the wave equation (7) if S is a solution of the equation

$$\left(-(\operatorname{grad} S)^2 + \operatorname{grad} S\,(\operatorname{grad} S\cdot) + \mu\varepsilon\cdot\right)\boldsymbol{A} = \boldsymbol{0}. \tag{13}$$

This equation is, however, as an equation for $\boldsymbol{A}$, a linear homogeneous vector equation. Consequently the necessary and sufficient condition for its having other solutions than $\boldsymbol{A} = \boldsymbol{0}$ consists in the determinant vanishing, i. e.

$$\left|-(\operatorname{grad} S)^2 + \operatorname{grad} S\,(\operatorname{grad} S\cdot) + \mu\varepsilon\cdot\right| =$$

$$\left.\begin{vmatrix} \mu\varepsilon_{xx} - (\operatorname{grad} S)^2 + \left(\dfrac{\partial S}{\partial x}\right)^2, & \mu\varepsilon_{xy} + \dfrac{\partial S}{\partial x}\dfrac{\partial S}{\partial y}, & \mu\varepsilon_{xz} + \dfrac{\partial S}{\partial x}\dfrac{\partial S}{\partial z} \\ \mu\varepsilon_{yx} + \dfrac{\partial S}{\partial y}\dfrac{\partial S}{\partial x}, & \mu\varepsilon_{yy} - (\operatorname{grad} S)^2 + \left(\dfrac{\partial S}{\partial y}\right)^2, & \mu\varepsilon_{yz} + \dfrac{\partial S}{\partial y}\dfrac{\partial S}{\partial z} \\ \mu\varepsilon_{zx} + \dfrac{\partial S}{\partial z}\dfrac{\partial S}{\partial x}, & \mu\varepsilon_{zy} + \dfrac{\partial S}{\partial z}\dfrac{\partial S}{\partial y}, & \mu\varepsilon_{zz} - (\operatorname{grad} S)^2 + \left(\dfrac{\partial S}{\partial z}\right)^2 \end{vmatrix} = 0, \right\} \tag{14}$$

in which (x, y, z) is an arbitrary *Cartesian* coordinate system. This equation is just the so-called *characteristic equation* belonging to the wave equation (7). It is a partial differential equation of the first order and the degree 6 (viz. equal to the product of the order of and the number of equations in the wave equation). We shall, however, see in a moment that the degree of (14) is in fact only 4. Before doing so we shall state under what conditions the terms in (12) with λ_0 and λ_0^2 may be neglected in comparison with the first term. They are, obviously, since the dimensionless quantity $\operatorname{grad} S$ is of the order of magnitude 1, the following:

$$\left.\begin{aligned} &\text{(a)}\quad \lambda_0\left((\operatorname{grad} S\cdot\operatorname{grad}\boldsymbol{A}) - \operatorname{grad} S\operatorname{div}\boldsymbol{A} - [\operatorname{grad} S\times\operatorname{rot}\boldsymbol{A}]\right) \ll \boldsymbol{A} \\ &\qquad\text{i. e.}\quad \lambda_0\left|\frac{\partial A_i}{\partial x_k}\right| \ll |\boldsymbol{A}|\ \ (i, k = 1, 2, 3) \end{aligned}\right\} \tag{15}$$

$$\left.\begin{aligned} &\text{(b)}\quad \lambda_0\left(\Delta S - (\operatorname{grad}(\operatorname{grad} S)\cdot)\right)\boldsymbol{A} \ll \boldsymbol{A} \\ &\qquad\text{i. e.}\quad \lambda_0\left|\frac{\partial^2 S}{\partial x_i\partial x_k}\right| \ll 1\ \ (i, k = 1, 2, 3) \end{aligned}\right\} \tag{16}$$

$$\left.\begin{aligned} &\text{(c)}\quad \lambda_0\left[\frac{\operatorname{grad}\mu}{\mu}\times[\operatorname{grad} S\times\boldsymbol{A}]\right] \ll \boldsymbol{A} \\ &\qquad\text{i. e.}\quad \lambda_0\left|\frac{\partial\mu}{\partial x_i}\right| \ll \mu\ \ (i = 1, 2, 3) \end{aligned}\right\} \tag{17}$$

$$
\left.
\begin{aligned}
&\text{(d)} && \lambda_0 \frac{4\pi\mu}{c}\sigma\cdot\mathbf{A} \ll \mathbf{A} \\
& && \text{i.e.}\quad \lambda_0 \frac{4\pi\mu}{c}\sigma_i = \frac{1}{2}\left(\frac{\lambda_0}{d_i}\right)^2 \ll 1 \quad (i = 1, 2, 3),
\end{aligned}
\right\} \quad (18)
$$

in which[1]

$$
d_i = \frac{\lambda_0}{4\pi\varkappa_i} = \frac{c}{4\pi\sqrt{\nu\mu\sigma_i}} \tag{19}
$$

($\varkappa_i$: the coefficient of absorption in the direction of the i'th principal axis of σ, d_i: the corresponding length of penetration).

$$
\left.
\begin{aligned}
& && \lambda_0^2(\Delta\mathbf{A} - \operatorname{grad}\operatorname{div}\mathbf{A}) \ll \mathbf{A} \\
&\text{(e)} && \text{i.e.}\quad \lambda_0^2\left|\frac{\partial^2 A_i}{\partial x_k \partial x_l}\right| \ll |\mathbf{A}| \quad (i, k, l = 1, 2, 3).
\end{aligned}
\right\} \quad (20)
$$

(a) + (e) means that each component of the amplitude $\mathbf{A}$ *is to vary so slowly that both the first and the second variation of* $\mathbf{A}$ *in a distance of the order of magnitude of the wave-length* λ_0 *is negligibly small compared with* $|\mathbf{A}|$. *If we consider a sharply defined light ray,* $\mathbf{A}$ *will obviously vary strongly in a distance of the order* λ_0 *on the border between the ray and the shadow. In this region we shall, consequently, expect deviations from the laws of geometrical optics: the phenomena of diffraction.*

(b) means that the phase function S *is to deviate so little from linearity that its second derivatives are negligible compared with* $\frac{1}{\lambda_0}$, *i.e. the principal radii of curvature of the wave fronts given by* $S = const$ *are to be large compared with* λ_0. *This statement, again, is equivalent to two other conditions: (α) the radii of curvature of the light rays themselves are to be large compared with* λ_0. *(β) no points in which the light rays diverge or converge are to be considered (as in these points one or both of the principal radii of curvature vanish). Finally we see from (14) that,* S *being determined by the variation of the product* $\mu\varepsilon$, *our condition (16) demands that* $\mu\varepsilon$ *varies so slowly that its variation in a distance of the order* λ_0 *is negligible compared with* $\mu\varepsilon$ *itself:*

$$
\lambda_0\left|\frac{\partial(\mu\varepsilon_{ik})}{\partial x_l}\right| \ll \mu\varepsilon_{ik} \quad (i, k, l = 1, 2, 3). \tag{21}
$$

[1] Cf. e.g. Born (1933) p. 261.

(c) means that μ is to vary so slowly that its variation in a distance of the order λ_0 is negligible compared with μ itself. From (21) it then follows that the same applies to the dielectric tensor ε itself.

(d) means that the medium is to have so small a conductivity tensor σ that the principal lengths of penetration d_i are very large compared with the wave-length λ_0. This condition is, besides, obvious, as we could not otherwise speak of light rays at all, the rays being at once absorbed.

(f) Finally we see that geometrical optics is only valid so long as all questions regarding phenomena of intensity, polarization, and interference, and as already stated also of dispersion, may be disregarded.

§ 3.

Having stated the exact conditions for the validity of geometrical optics, we shall now investigate the equation § 2 (14), which contains all its laws. Firstly, we note that the equation is invariant against all coordinate transformations. Secondly, that the dielectric tensor ε being always *symmetric*, it may in each point be transformed on diagonal form, and to obtain simpler formulae it will, therefore, be convenient to transform to new generalized, orthogonal coordinates

$$\left.\begin{aligned} &(\xi, \eta, \zeta) = (\xi(x, y, z),\ \eta(x, y, z),\ \zeta(x, y, z)) \\ &ds^2 = dx^2 + dy^2 + dz^2 = g_\xi d\xi^2 + g^\eta d\eta^2 + g_\zeta d\zeta^2 \\ g_\xi = \left(\frac{\partial x}{\partial \xi}\right)^2 &+ \left(\frac{\partial y}{\partial \xi}\right)^2 + \left(\frac{\partial z}{\partial \xi}\right)^2, \quad g_\eta = \left(\frac{\partial x}{\partial \eta}\right)^2 + \left(\frac{\partial y}{\partial \eta}\right)^2 + \left(\frac{\partial z}{\partial \eta}\right)^2, \\ &g_\zeta = \left(\frac{\partial x}{\partial \zeta}\right)^2 + \left(\frac{\partial y}{\partial \zeta}\right)^2 + \left(\frac{\partial z}{\partial \zeta}\right)^2. \end{aligned}\right\} \quad (1)$$

(g_ξ, g_η, g_ζ are the only non-vanishing elements of the fundamental metric tensor g_{ik} of the (ξ, η, ζ) coordinate system). Here the transformation matrix $\left|\frac{\partial x^i}{\partial \xi^k}\right|$ is determined so that ε is in *each* point of the (ξ, η, ζ) coordinate system on diagonal form. In case the medium is homogeneous, the transformation becomes simply orthogonal with constant coefficients and $g_\xi = g_\eta = g_\zeta = 1$, otherwise it consists in general of non-linear functions of (x, y, z),

g_ξ, g_η, g_ζ being then functions of (ξ, η, ζ). Now $\frac{\partial S}{\partial \xi} = S_\xi$, $\frac{\partial S}{\partial \eta} = S_\eta$, $\frac{\partial S}{\partial \zeta} = S_\zeta$ are the covariant components of grad S in the (ξ, η, ζ) coordinate system, and the elements of the tensor in § 2 (14) the mixed components, $\mu\varepsilon^i_k - \delta^i_k (\text{grad } S)^2 + (\text{grad } S)^i (\text{grad } S)_k$. Expressed in terms of the covariant components S_ξ, S_η, S_ζ our equation § 2 (14) thus becomes, because in case of an orthogonal coordinate system we have $S^i = g^{ii} S_i = \frac{1}{g_{ii}} S_i$ (no summation over i),

$$\mu^3 \varepsilon_\xi \varepsilon_\eta \varepsilon_\zeta \begin{vmatrix} 1 - \dfrac{(\text{grad } S)^2 - \frac{1}{g_\xi} S_\xi^2}{\mu\varepsilon_\xi} & \dfrac{S_\xi S_\eta}{g_\xi \mu\varepsilon_\xi} & \dfrac{S_\xi S_\zeta}{g_\xi \mu\varepsilon_\xi} \\ \dfrac{S_\eta S_\xi}{g_\eta \mu\varepsilon_\eta} & 1 - \dfrac{(\text{grad } S)^2 - \frac{1}{g_\eta} S_\eta^2}{\mu\varepsilon_\eta} & \dfrac{S_\eta S_\zeta}{g_\eta \mu\varepsilon_\eta} \\ \dfrac{S_\zeta S_\xi}{g_\zeta \mu\varepsilon_\zeta} & \dfrac{S_\zeta S_\eta}{g_\zeta \mu\varepsilon_\zeta} & 1 - \dfrac{(\text{grad } S)^2 - \frac{1}{g_\zeta} S_\zeta^2}{\mu\varepsilon_\zeta} \end{vmatrix} =$$

$$\mu^3 \varepsilon_\xi \varepsilon_\eta \varepsilon_\zeta \left\{ \left(1 - \frac{(\text{grad } S)^2 - \frac{1}{g_\xi} S_\xi^2}{\mu\varepsilon_\xi}\right) \left(1 - \frac{(\text{grad } S)^2 - \frac{1}{g_\eta} S_\eta^2}{\mu\varepsilon_\eta}\right) \left(1 - \frac{(\text{grad } S)^2 - \frac{1}{g_\zeta} S_\zeta^2}{\mu\varepsilon_\zeta}\right) + \right.$$

$$+ 2 \frac{S_\xi^2 S_\eta^2 S_\zeta^2}{g_\xi g_\eta g_\zeta \mu^3 \varepsilon_\xi \varepsilon_\eta \varepsilon_\zeta} - \frac{S_\xi^2 S_\zeta^2}{g_\xi g_\zeta \mu^2 \varepsilon_\xi \varepsilon_\zeta} \left(1 - \frac{(\text{grad } S)^2 - \frac{1}{g_\eta} S_\eta^2}{\mu\varepsilon_\eta}\right) -$$

$$\left. \frac{S_\eta^2 S_\zeta^2}{g_\eta g_\zeta \mu^2 \varepsilon_\eta \varepsilon_\zeta} \left(1 - \frac{(\text{grad } S)^2 - \frac{1}{g_\xi} S_\xi^2}{\mu\varepsilon_\xi}\right) - \frac{S_\xi^2 S_\eta^2}{g_\xi g_\eta \mu^2 \varepsilon_\xi \varepsilon_\eta} \left(1 - \frac{(\text{grad } S)^2 - \frac{1}{g_\zeta} S_\zeta^2}{\mu\varepsilon_\zeta}\right) \right\} =$$

$$\mu^3 \varepsilon_\xi \varepsilon_\eta \varepsilon_\zeta \left\{ 1 - \left(\frac{(\text{grad } S)^2 - \frac{1}{g_\xi} S_\xi^2}{\mu\varepsilon_\xi} + \frac{(\text{grad } S)^2 - \frac{1}{g_\eta} S_\eta^2}{\mu\varepsilon_\eta} + \frac{(\text{grad } S)^2 - \frac{1}{g_\zeta} S_\zeta^2}{\mu\varepsilon_\zeta} \right) + \right.$$

$$\left. + (\text{grad } S)^2 \left(\frac{S_\xi^2}{g_\xi \mu^2 \varepsilon_\eta \varepsilon_\zeta} + \frac{S_\eta^2}{g_\eta \mu^2 \varepsilon_\zeta \varepsilon_\xi} + \frac{S_\zeta^2}{g_\zeta \mu^2 \varepsilon_\xi \varepsilon_\eta} \right) \right\} = 0 \quad \text{(covariant comp. of grad } S)$$

$$\mu = \mu(\xi, \eta, \zeta), \; \varepsilon_{\xi,\eta,\zeta} = \varepsilon_{\xi,\eta,\zeta}(\xi, \eta, \zeta), \; (\text{grad } S)^2 = \frac{1}{g_\xi} S_\xi^2 + \frac{1}{g_\eta} S_\eta^2 + \frac{1}{g_\zeta} S_\zeta^2. \quad (2)$$

Here $\varepsilon_\xi, \varepsilon_\eta, \varepsilon_\zeta$ are the principal dielectric constants in the (ξ, η, ζ) coordinate system. They are in the general case of inhomogeneous media functions of ξ, η, ζ. We see in fact, as already announced, that (2) is a partial differential equation for the phase function S, which is of the first order and the 4th degree. We note that writing each factor of the first term in the form $\left(1-\frac{(\mathrm{grad}\, S)^2}{\mu\varepsilon_\xi}\right)+\frac{S_\xi^2}{g_\xi \mu\varepsilon_\xi}, \cdots$ we would have obtained (2) in the form

$$\frac{\frac{1}{g_\xi} S_\xi^2}{(\mathrm{grad}\, S)^2-\mu\varepsilon_\xi}+\frac{\frac{1}{g_\eta} S_\eta^2}{(\mathrm{grad}\, S)^2-\mu\varepsilon_\eta}+\frac{\frac{1}{g_\zeta} S_\zeta^2}{(\mathrm{grad}\, S)^2-\mu\varepsilon_\zeta}=1. \quad \text{(covariant)} \quad (3)$$

If we multiply on both sides with $(\mathrm{grad}\, S)^2$ and next subtract $(\mathrm{grad}\, S)^2$ from both sides, (3) may also be written in the form

$$\frac{\mu\varepsilon_\xi \frac{1}{g_\xi} S_\xi^2}{(\mathrm{grad}\, S)^2-\mu\varepsilon_\xi}+\frac{\mu\varepsilon_\eta \frac{1}{g_\eta} S_\eta^2}{(\mathrm{grad}\, S)^2-\mu\varepsilon_\eta}+\frac{\mu\varepsilon_\zeta \frac{1}{g_\zeta} S_\zeta^2}{(\mathrm{grad}\, S)^2-\mu\varepsilon_\zeta}=0. \quad \text{(covariant)} \quad (4)$$

We note that the two last equations are just the generalization of the well-known *normal equation of* FRESNEL (in the covariant form).

In order to prove FERMAT's and HUYGENS' principles it is now only necessary to write (2) in Hamiltonian form and then calculate the corresponding Lagrangian function L, which function shall then turn out to be equal to the integrand in § 1 (1). By this procedure there is, however, a difficulty because L, in the case of an isotropic medium, becomes homogeneous of the first degree in $\dot{x}, \dot{y}, \dot{z}$, the ray index being in this case only dependent on x, y, z—and furthermore equal to the index of refraction. In this case the Legendre transformation from the variables $\dot{x}, \dot{y}, \dot{z}$ to the generalized momentum variables p_x, p_y, p_z, being necessary in order to write the equations in Hamiltonian form, becomes impossible[1]. In order to avoid this complication we shall, therefore, assume that all the light rays considered are of such a kind that we may as the parameter τ in § 1 (1) choose one

[1] COURANT-HILBERT (1937) chap. II § 9.

of the variables ξ, η, ζ themselves which will in practice always be so[1]. The equation for the phase function S, (2), containing only terms of the 0^{th}, 2^{nd} and 4^{th} degree, we may solve it with respect to S_ξ, assuming in what follows ξ to be the parameter. We thus in general obtain two different solutions (which may be shown always to be both real) corresponding to the well-known fact that we have in general two different kinds of rays, called the *ordinary* and the *extraordinary*, respectively, which coincide only in the case of isotropic media:

$$S_\xi + H^\pm(\xi, \eta, \zeta, S_\eta, S_\zeta) = 0 \tag{5}$$

$$H^\pm(\xi, \eta, \zeta, S_\eta, S_\zeta) = -\sqrt{-\frac{1}{2}\Phi \pm \sqrt{\frac{1}{2}\Phi^2 - \Psi}} \tag{6}$$

$$\Phi = \left(\frac{\varepsilon_\eta}{\varepsilon_\xi} + 1\right)\frac{g_\xi}{g_\eta} S_\eta^2 + \left(\frac{\varepsilon_\zeta}{\varepsilon_\xi} + 1\right)\frac{g_\xi}{g_\zeta} S_\zeta^2 - \mu\left(\varepsilon_\eta + \varepsilon_\zeta\right) g_\xi$$

$$\Psi = \frac{\varepsilon_\eta}{\varepsilon_\xi}\frac{g_\xi^2}{g_\eta^2} S_\eta^4 + \frac{\varepsilon_\zeta}{\varepsilon_\xi}\frac{g_\xi^2}{g_\zeta^2} S_\zeta^4 + \left(\frac{\varepsilon_\eta + \varepsilon_\zeta}{\varepsilon_\xi}\right)\frac{g_\xi^2}{g_\eta g_\zeta} S_\eta^2 S_\zeta^2 - \mu\left(\frac{\varepsilon_\eta \varepsilon_\zeta}{\varepsilon_\xi} + \varepsilon_\eta\right)\frac{g_\xi^2}{g_\eta} S_\eta^2 -$$
$$\mu\left(\frac{\varepsilon_\eta \varepsilon_\zeta}{\varepsilon_\xi} + \varepsilon_\zeta\right)\frac{g_\xi^2}{g_\zeta} S_\zeta^2 + \mu^2 \varepsilon_\eta \varepsilon_\zeta g_\xi^2 .$$

Here (5) is just of the form of a Hamilton-Jacobi partial differential equation for the phase function S, which function thus corresponds to Hamilton's principal action—or characteristic—function, ξ corresponding to the time variable. Next $H^\pm = H^\pm(\xi, \eta, \zeta, p_\eta, p_\zeta)$ given in (6) are the two Hamiltonian functions corresponding to the ordinary and the extraordinary ray, respectively. Furthermore, p_η, p_ζ are the generalized momentum variables defined as usual by

$$p_\eta = S_\eta, \quad p_\zeta = S_\zeta . \tag{7}$$

We note that the two signs of the first $\sqrt{}$ in (6) do not correspond to two different light propagations, but determine only the direction of the light rays. We have chosen the negative sign in order to make the normal and the ray directions point to the same

[1] Cf. also Carathéodory (1937) p. 9.

side of the wave fronts, i. e. have an angle between them less than $\frac{\pi}{2}$.

From the general theory of partial differential equations of the first order ([1] p. 11) it follows that the equation (5) has a complete solution (i. e. a solution containing three arbitrary constants of integration), which may be written in the form

$$\left.\begin{aligned} S^{\pm}(\xi,\eta,\zeta) &= \int_{P_0}^{P} L^{\pm}(\xi,\eta,\zeta,\dot{\eta},\dot{\zeta})\,d\xi + S(\xi_0,\eta_0,\zeta_0), \\ P &= (\xi,\eta,\zeta),\; P_0 = (\xi_0,\eta_0,\zeta_0). \end{aligned}\right\} \tag{8}$$

Here $S(\xi_0, \eta_0, \zeta_0)$ is an arbitrary function of P_0 giving the initial values of S on an arbitrary boundary surface $F(\xi_0, \eta_0, \zeta_0) = 0$. In case this surface is a wave front, we have $S(P_0) = \text{const}$ on this surface. (We note that $S(P)$ corresponds to the indefinite integral, $\int_{P_0}^{P} L d\xi$ to the definite integral in the case of differential equations with only one unknown.) Next it follows from the general theory that (5) is equivalent to the variation problem

$$\delta \int_{P'}^{P} L^{\pm}\,d\xi = 0. \tag{9}$$

The Lagrangian function L occurring in (8) and (9) is, furthermore, obtained by differentiating (8) along the integration curve and using (5) and (7)

$$L^{\pm}(\xi,\eta,\zeta,\dot{\eta},\dot{\zeta}) = -H^{\pm} + p_{\eta}\dot{\eta} + p_{\zeta}\dot{\zeta}. \tag{10}$$

Finally $\eta = \eta^{\pm}(\xi)$ and $\zeta = \zeta^{\pm}(\xi)$ occurring in (8) and (9) are the coordinates of the ordinary, respectively extraordinary light ray between P_0 and P. They are given uniquely as solutions of the Hamiltonian system of ordinary differential equations

$$\left.\begin{aligned} \dot{\eta} &= \frac{d\eta}{d\xi} = \frac{\partial H^{\pm}}{\partial p_{\eta}} & \dot{\zeta} &= \frac{d\zeta}{d\xi} = \frac{\partial H^{\pm}}{\partial p_{\zeta}} \\ \dot{p}_{\eta} &= \frac{dp_{\eta}}{d\xi} = -\frac{\partial H^{\pm}}{\partial \eta} & \dot{p}_{\zeta} &= \frac{dp_{\zeta}}{d\xi} = -\frac{\partial H^{\pm}}{\partial \zeta}. \end{aligned}\right\} \tag{11}$$

In the general case the principal dielectric constants $\varepsilon_{\xi}, \varepsilon_{\eta}, \varepsilon_{\zeta}$, the magnetic permeability μ as well as the elements of the fundamental metric tensor g_ξ, g_η, g_ζ of the (ξ, η, ζ) coordinate system are functions of ξ, η, ζ and the same, therefore, applies to $H^\pm$. If, especially, they are all constants, $H^\pm$ becomes independent of ξ, η, ζ, and (11) is then immediately solved to

$$\dot{p}_\eta = \dot{p}_\zeta = 0 \quad \text{i. e.} \quad p_\eta = \text{const} = c_1, \; p_\zeta = \text{const} = c_2$$
$$\dot{\eta} = \text{const} = f_\eta(c_1, c_2), \; \dot{\zeta} = \text{const} = f_\zeta(c_1, c_2),$$

i. e.

$$\left.\begin{aligned} \eta &= f_\eta(c_1, c_2)\, \xi + c_3 \\ \zeta &= f_\zeta(c_1, c_2)\, \xi + c_4 . \end{aligned}\right\} \quad (12)$$

In the generalized coordinates ξ, η, ζ the light rays are thus straight lines if $\varepsilon_\xi, \varepsilon_\eta, \varepsilon_\zeta, \mu, g_\xi, g_\eta, g_\zeta$ are constants. This is just the case for a *homogeneous* medium. Here the transformation $(x, y, z) \to (\xi, \eta, \zeta)$ may, as already mentioned, simply be taken as a linear, orthogonal transformation, i. e. a rotation of the Cartesian coordinate system. Thus the light rays are also straight lines in the coordinates (x, y, z) of real space. For *inhomogeneous* media the light rays, however, in general are curved lines[1].

Now (8) just expresses Huygens' principle, as it follows from this equation that if $S^\pm = \text{const} = c$ is one wave front, then we obtain the neighbouring wave fronts $S = c \pm dc$ by drawing through each point of the first wave front a light ray and on each of these rays mark out the distance $\pm ds$ given by

$$\left.\begin{aligned} ds &= \sqrt{g_\xi d\xi^2 + g_\eta d\eta^2 + g_\zeta d\zeta^2} = \sqrt{dx^2 + dy^2 + dz^2} = \frac{dc}{n} \\ n &= n^\pm(\xi, \eta, \zeta, \dot{\eta}, \dot{\zeta}) = \frac{L^\pm}{\dfrac{ds}{d\xi}} = \frac{L^\pm}{\sqrt{g_\xi + g_\eta \dot{\eta}^2 + g_\zeta \dot{\zeta}^2}} . \end{aligned}\right\} \quad (13)$$

[1] We note that a treatment of geometrical optics based on eqs. (5)–(6), which is due to Hamilton, shows, as already stressed by Hamilton himself, a very close analogy with classical mechanics. It was just this analogy which led Schrödinger (1926) to his discovery of wave mechanics by his idea of interpreting the mechanical principal action function of Hamilton as a phase function of the de Broglie waves (or in mathematical language, of interpreting the Hamilton-Jacobi eq. as the characteristic equation of a wave equation). In fact, starting with the H-J eq. and carrying out in the opposite order the calculations (with $\lambda_0 \to \hbar$) which led us from the wave eq. § 2 (7) to the H-J eq. § 3 (5), just leads to the time dependent Schrödinger wave eq. Consequently, conditions corresponding to those in § 2 for the validity of geometrical optics, viz. small wave-length, apply to the validity of classical mechanics.

We may now interpret the phase function $S^{\pm}(P, P') = \int_{P'}^{P} n ds = c \int_{P'}^{P} dt$, in which the integral is taken along a light ray between two arbitrary points P' and P, as a geodetic distance, equal to the time interval, between the points P' and P. (S is then called the characteristic function or the *eiconal*). Two wave fronts given by $S = c'$ and $S = c$, respectively, may then be said to be geodetic parallel surfaces corresponding to the fixed distance $S(P, P') = c - c'$, P' being an arbitrary point of the first wave front and P the point of intersection between the ray through P' and the second wave front; for it follows ([1] p. 11) that (α) $S(P, P')$ is independent of P' and (β) $\delta S(P, P') = 0$ for each fixed P' and P varying in the second wave front. Hence, if we plot a geodetic 'sphere' with radius $c - c'$ about each P', i. e. a surface—called a *ray surface*—having the constant geodetic distance $c - c'$ from P', then, for each of these surfaces, we also have $\delta S(P, P') = 0$. This fact, however, just means that the second wave front is the envelope of all the geodetic 'spheres', which is exactly HUYGENS' principle. We note, furthermore, that because of the property $\delta S = 0$ the light rays may in a generalized sense be said to be transverse to the wave fronts. They are, therefore, also called the transversal curves—not to be confused with the orthogonal, or normal, trajectories, the direction of which are given by grad S.

Introducing n from (13) in (9) and comparing with § 1 (1), we see that in order to establish FERMAT's principle it now only remains to prove that n defined by (13) is identical with the ray index defined in § 1 (2). We treat isotropic and anisotropic media separately.

(A). Isotropic media.

Such media are characterized by the dielectric tensor reducing to a scalar, i. e. $\varepsilon_\xi = \varepsilon_\eta = \varepsilon_\zeta = \varepsilon = \varepsilon(x, y, z)$. Consequently we can put $(\xi, \eta, \zeta) = (x, y, z)$, i. e. our fixed Cartesian coordinates. In this case the ordinary and the extraordinary rays are seen to coincide, since our equations reduce to the following

$$\left.\begin{array}{l} \left.\begin{array}{l} \Phi = 2(S_y^2 + S_z^2 - \mu\varepsilon) \\ \Psi = (S_y^2 + S_z^2 - \mu\varepsilon)^2 \end{array}\right\} \text{ i. e. } \frac{1}{4}\Phi^2 - \Psi = 0 \\ \text{i. e. } H^+ = H^- = H = -\sqrt{-\frac{1}{2}\Phi} \end{array}\right\} \quad (14)$$

$$S_x + H = S_x - \sqrt{\mu\varepsilon - S_y^2 - S_z^2} = 0 \quad \text{i. e.} \quad (\text{grad } S)^2 = \mu\varepsilon \tag{15}$$

$$\dot{y} = \frac{dy}{dx} = \frac{\partial H}{\partial p_y} = \frac{p_y}{\sqrt{\mu\varepsilon - p_y^2 - p_z^2}}, \quad \dot{z} = \frac{dz}{dx} = \frac{\partial H}{\partial p_z} = \frac{p_z}{\sqrt{\mu\varepsilon - p_y^2 - p_z^2}} \tag{16}$$

$$L = -H + p_y\dot{y} + p_z\dot{z} = \sqrt{\mu\varepsilon - p_y^2 - p_z^2} + \frac{p_y^2 + p_z^2}{\sqrt{\mu\varepsilon - p_y^2 - p_z^2}} = \frac{\mu\varepsilon}{\sqrt{\mu\varepsilon - p_y^2 - p_z^2}} \tag{17}$$

$$\frac{ds}{dx} = \sqrt{1 + \dot{y}^2 + \dot{z}^2} = \sqrt{\frac{\mu\varepsilon}{\mu\varepsilon - p_y^2 - p_z^2}} \tag{18}$$

$$n = n\,(x, y, z) = \frac{L}{\frac{ds}{dx}} = \sqrt{\mu\,(x, y, z)\;\varepsilon\,(x, y, z)}\,. \tag{19}$$

(19) is, however, just the MAXWELL relation for the index of refraction $= \frac{c}{v}$, q. e. d. From (16) it next follows, by means of (15) and (7), that the normal direction, grad S, and the ray direction, $(1, \dot{y}, \dot{z})$, coincide:

$$\text{grad } S = (-H, p_y, p_z) = \sqrt{\mu\varepsilon - p_y^2 - p_z^2}\,(1, \dot{y}, \dot{z})\,. \tag{20}$$

Consequently, the index of refraction and the ray index coincide. (15) together with (19) and (20) may finally be written in the well-known form[1]

$$\text{grad } S = n\boldsymbol{s} = \sqrt{\mu\varepsilon}\,\boldsymbol{s}, \quad \boldsymbol{s} = \frac{\text{grad } S}{|\text{grad } S|}. \tag{21}$$

(B). Anisotropic media.

In this case it would be extremely laborious to determine $n^{\pm}$ directly from (13) and verify that it is a solution of the well-known *ray equation* of FRESNEL (in the contravariant form) obtained from the normal equation (4) by making the transformation $\boldsymbol{s} \to -\mathfrak{s}$, $n_n \to \frac{1}{n_r}$, $\varepsilon_i \to \frac{1}{\varepsilon_i}$ (together with $\mathfrak{s}_i = g_i \mathfrak{s}^i$) [2]

$$\frac{g_\xi (\mathfrak{s}^\xi)^2}{n_r^2 - \mu\varepsilon_\xi} + \frac{g_\eta (\mathfrak{s}^\eta)^2}{n_r^2 - \mu\varepsilon_\eta} + \frac{g_\zeta (\mathfrak{s}^\zeta)^2}{n_r^2 - \mu\varepsilon_\zeta} = 0 \quad \text{(contravariant)} \tag{22}$$

[1] BORN (1933) p. 48.
[2] BORN (1933) p. 225.

in which

$$\mathfrak{s} = (\mathfrak{s}^{\xi}, \mathfrak{s}^{\eta}, \mathfrak{s}^{\zeta}) = \frac{1}{\sqrt{g_{\xi} + g_{\eta}\,\dot{\eta}^2 + g_{\zeta}\,\dot{\zeta}^2}}(1, \dot{\eta}, \dot{\zeta}) \quad \text{(contravariant)} \qquad (23)$$

is a unit vector in the ray direction and $\mathfrak{s}^{\xi}, \mathfrak{s}^{\eta}, \mathfrak{s}^{\zeta}$ its contravariant components in the ξ, η, ζ coordinate system. We may, however, make a short cut, observing that the index of refraction $n_n = \frac{c}{v_n}$ is equal to the rate of change of the phase function S in the direction of the normal given by $\boldsymbol{ds} =$ const. grad S

$$dS = (\text{grad S}, \boldsymbol{ds}) = n_n ds, \qquad (24)$$

whereas the ray index $n_r = \frac{c}{v_r}$ is equal to the rate of change of S in the ray direction given by $\boldsymbol{d\mathfrak{s}} =$ const. $(1, \dot{\eta}, \dot{\zeta})$

$$dS = (\text{grad}\, S, \boldsymbol{d\mathfrak{s}}) = n_r d\mathfrak{s}. \qquad (25)$$

Considering two neighbouring wave fronts S and $S + dS$, we have $ds = d\mathfrak{s} \cos(\boldsymbol{s}, \boldsymbol{\mathfrak{s}}) = d\mathfrak{s}\,(\boldsymbol{s} \cdot \boldsymbol{\mathfrak{s}})$, and the well-known relation follows:

$$n_r = n_n (\boldsymbol{s} \cdot \boldsymbol{\mathfrak{s}}). \qquad (26)$$

From (24), (5) and (7) we now have that the covariant components of grad S still satisfy (21) with $n = n_n$:

$$\text{grad}\, S = (S_{\xi}, S_{\eta}, S_{\zeta}) = (-H, p_{\eta}, p_{\zeta}) = n_n (s_{\xi}, s_{\eta}, s_{\zeta}). \text{ (covariant)} \quad (27)$$

Introducing (27) in (4), $n_n = n_n^{\pm}(\xi, \eta, \zeta, s_{\xi}, s_{\eta}, s_{\zeta})$ is thus seen to be a root in FRESNEL's normal equation (in the covariant form)

$$\frac{\frac{1}{g_{\xi}} s_{\xi}^2}{\frac{n_n^2}{\mu\varepsilon_{\xi}} - 1} + \frac{\frac{1}{g_{\eta}} s_{\eta}^2}{\frac{n_n^2}{\mu\varepsilon_{\eta}} - 1} + \frac{\frac{1}{g_{\zeta}} s_{\zeta}^2}{\frac{n_n^2}{\mu\varepsilon_{\zeta}} - 1} = 0. \quad \text{(covariant)} \qquad (28)$$

Introducing (27) in (10) then just gives, with the use of (26), for the integrand in (9) of either sort of ray

$$\left.\begin{aligned} L d\xi &= (-H + p_{\eta}\,\dot{\eta} + p_{\zeta}\,\dot{\zeta})\, d\xi = (n_n s_{\xi} + n_n s_{\eta}\,\dot{\eta} + n_n s_{\zeta}\,\dot{\zeta})\, d\xi = \\ &= n_n (s_{\xi}\mathfrak{s}^{\xi} + s_{\eta}\mathfrak{s}^{\eta} + s_{\zeta}\mathfrak{s}^{\zeta}) \sqrt{g_{\xi} + g_{\eta}\,\dot{\eta}^2 + g_{\zeta}\,\dot{\zeta}^2}\, d\xi = \\ &= n_n (\boldsymbol{s} \cdot \boldsymbol{\mathfrak{s}})\, d\mathfrak{s} = n_r d\mathfrak{s}, \end{aligned}\right\} \qquad (29)$$

q. e. d. $n_n^{\pm}$ satisfying the normal equation (28) it follows as usual that $n_r^{\pm}$ satisfies the ray equation (22).

In conclusion we see that FERMAT's principle has now been proved for both the ordinary and the extraordinary ray in the most general case of a non-ferromagnetic, absorbing, inhomogeneous, anisotropic medium under the one assumption that the magnetic permeability scalar μ and the dielectric tensor ε are *continuous* functions of x, y, z. If, however, we have a discontinuity surface, i. e. two neighbouring media with different values of μ and ε, a case which is of the utmost practical importance, then we have—for either sort of ray—in the first medium a certain Hamiltonian $H^{(1)}$ and Lagrangian $L^{(1)}$ and in the second medium another one, $H^{(2)}$ and $L^{(2)}$, respectively. Now in both media the phase function S satisfies the HAMILTON-JACOBI equation (5) and is given by (8). Next it follows from the general theory of partial differential equations[1] that S is uniquely given from the equation and the boundary conditions, i. e. the functions $F(\xi_0, \eta_0, \zeta_0) = 0$ and $S(\xi_0, \eta_0, \zeta_0)$ in (8). Thus, if the values of S are given on a certain surface $F(\xi_0, \eta_0, \zeta_0) = 0$ in the first medium, the values of S on the discontinuity surface will be determined by the equation. Next, these values of S being the boundary values in the second medium, we see that the behaviour of S is uniquely determined in *both* media and that S is a continuous function given by (8) throughout. From this fact, which is equivalent to the general validity of HUYGENS' principle also for discontinuously varying ε and μ, follows the well-known law of refraction (as well as that of reflection) in the general case of both media being inhomogeneous and anisotropic, the directions of the light rays $(1, \dot{\eta}, \dot{\zeta})$ being simply given on both sides of the discontinuity surface by (11). It may be shown[2] that the general law of refraction may equally well be obtained from postulating FERMAT's principle to hold true also for discontinuity surfaces. Consequently, this principle is also generally valid.

It may be of interest to note that in the case of *isotropic* media these facts may be seen quite elementarily. It follows, namely, in this case from (21) that

$$n\,(\boldsymbol{s}\cdot\boldsymbol{dr}) = (\operatorname{grad} S\cdot\boldsymbol{dr}) = \frac{\partial S}{\partial x}\,dx + \frac{\partial S}{\partial y}\,dy + \frac{\partial S}{\partial z}\,dz = dS \qquad (30)$$

[1] COURANT-HILBERT (1937) chap. II.

[2] CARATHÉODORY (1937) § 23.

is a total differential, and as S is, furthermore, *continuous* throughout, we therefore have

$$\oint_C n\,(\boldsymbol{s}\cdot\boldsymbol{dr}) = 0 \tag{31}$$

for an arbitrary, closed curve C independent of, whether or not C passes any discontinuity surface[1]. From (31) it follows firstly, in the same way as in electro-dynamics, that

$$(n\boldsymbol{s})^{(1)}_{\mathrm{tg}} = (n\boldsymbol{s})^{(2)}_{\mathrm{tg}} \tag{32}$$

i. e. the continuity of the component $(n\boldsymbol{s})_{\mathrm{tg}}$ of the vector $n\boldsymbol{s}$ in the tangent plane of the discontinuity surface. (32) is, however, just the law of refraction, which states (α) that $\boldsymbol{s}^{(1)}$ and $\boldsymbol{s}^{(2)}$ are both lying in the plane of incidence and (β) that $n_1 \sin\psi_1 = n_2 \sin\psi_2$.

Secondly, it follows from (31), if C is a light ray from P' to P and K an arbitrary different curve between P' and P, that

$$\int_C n ds = \left|\int_K n\,(\boldsymbol{s}\cdot\boldsymbol{dr})\right| \leqq \int_K n dr \qquad (|\boldsymbol{s}|) = 1). \tag{33}$$

The curve integral of n is thus a minimum along the light ray C from P' to P and the first variation of $\int n ds$ consequently vanishes along C, which is just FERMAT's principle §1 (1). We note, however, that $\delta \int n ds = 0$ does not always mean $\int n ds = \min.$; it may also, as is well-known, mean a relative minimum or even a maximum, e. g. by reflexions from certain curved surfaces. In fact the proof in (33) is only valid under the assumption that the light ray field $n\boldsymbol{s}$ is a unique vector field, but in the case of reflexions this does not hold, two or more rays passing through each point. Nevertheless it is possible to interpret the correct formulation of FERMAT's principle §1 (1) in terms of a minimum statement[2]:

A curve C may then and only then be the path of a light ray, if every point P of C is an inner point of at least one subcurve of C with the following property: the integral $\int n ds$ along this subcurve between its endpoints P' and P'' has a lower value than the same integral taken along a different curve having the same endpoints P' and P'' and lying in a certain narrow neighbourhood of C.

We see that by this formulation the above proof in (33) is also valid for reflexions.

Summary.

As is well-known all the laws of geometrical optics may be deduced from FERMAT's principle or the equivalent principle of HUYGENS. It is the purpose of the present note to deduce these principles from MAXWELL's electro-magnetic theory of light in

[1] We note that (31) here follows directly from the theory, whereas in BORN (1933) p. 50 it is deduced from the law of refraction.

[2] CARATHÉODORY (1937) p. 11.

the general case of non-ferromagnetic, absorbing, inhomogeneous, anisotropic media, at the same time obtaining the exact conditions for the validity of geometrical optics. In § 1 the problem is stated. In § 2 the equation for the phase function is deduced together with the conditions mentioned. In § 3 the phase equation is written in Hamiltonian form and FERMAT's and HUYGENS' principles are deduced together with the generalizations of the normal and ray equations of FRESNEL. It is seen that the n occuring in FERMAT's principle is the ray index and *not* the index of refraction, as seldom stressed in the literature. Finally the case of discontinuity surfaces is discussed.

Institute of Theoretical Physics of the
University of Copenhagen.

List of References.

BORN 1933, Optik, Berlin.

CARATHÉODORY 1935, Variationsrechnung und partielle Differentialgleichungen erster Ordnung, Leipzig.

— 1937, Geometrische Optik, Erg. d. Math. Bd. 4 Hft. 5, Berlin.

COURANT-HILBERT 1937, Methoden d. math. Physik, Bd. II, Berlin.

DIRAC 1930. The Principles of Quantum Mechanics, Oxford.

DRUDE 1900, Lehrbuch der Optik, Leipzig.

FÖRSTERLING 1928, Lehrbuch der Optik, Leipzig.

JENTZSCH 1927, Die Beziehungen der geometrischen Optik zur Wellenoptik, Hdbuch d. Physik Bd. 18, Berlin.

LANDÉ 1928, Optik, Mechanik und Wellenmechanik, Hdbuch d. Physik Bd. 20, Berlin.

PLANCK 1927, Einf. in die theor. Physik Bd. 4 Optik, Leipzig.

SCHRÖDINGER 1926, Quantisierung als Eigenwertproblem (2. Mitt.), Ann. d. Physik (4) **79**, 489.

SOMMERFELD und RUNGE 1911, Anwendung der Vektorrechnung auf die Grundlagen der geometrischen Optik, Ann. d. Physik (4) **35**, 277.

Indleveret til Selskabet den 4. November 1944.
Færdig fra Trykkeriet den 23. August 1945.