"Zum Übergang von der Wellenoptik zur geometrischen Optik in der allgemeinen Relativitätstheorie," Z. Naturforschg. **22a** (1967), 1328-1333.

On the transition from wave optics to geometrical optics in the general theory of relativity (*)

JÜRGEN EHLERS

The University of Texas at Austin, Texas, USA

(Z. Naturforschg. **22 a**, 1328-1332 [1967]; received on 1 June 1967)

Dedicated to PASCUAL JORDAN *in his 65th birthday*

Translated by D. H. Delphenich

The transition from the (covariantly-generalized) MAXWELL equation to the geometrical optics limit is discussed in the context of general relativity by adapting the classical series expansion method to the case of curved space-time. An arbitrarily-moving, ideal medium is also taken into account, and a close formal similarity between wave propagation in a moving medium in flat space-time and in an empty, gravitationally-curved space-time is established by means of a normal hyperbolic optical metric.

The geometrical optics approximation is, in the particular context of relativistic cosmology, an important tool for the relatively simple description of the propagation of light in gravitational fields [see, e.g., ([1,2,3])]. However, from the standpoint of the (covariantly-generalized) MAXWELLian theory, its basis is strictly beneath the corresponding basis for geometrical optics in the classical theory of media at rest in flat spaces [for the latter, see ([4,5]) and the references cited there]. In this publication, it will be shown how one can generalize the series developments of the usual classical theory in a natural way such that they can serve as the justification for geometrical optics in general relativity, and at the same time, one will also obtain a process for the determination of higher approximations and the estimation of error.

The generalization consists of:

a) Admitting arbitrary gravitational fields (non-flat metrics).
b) Allowing arbitrary, non-uniform motion for the matter.
c) Not restricting the time-dependency of the fields.

(*) This research has been sponsored by the Aerospace research Laboratories, Office of Aerospace Research, United States Air Force.

([1]) J. KRISTIAN and R. K. SACHS, Astrophys. J. **143** (1966), 379.
([2]) B. BERTOTTI, Proc. Roy. Soc. London A **294** (1966), 195.
([3]) D. ZIPOY, Phys. Rev. **142** (1966), 825.
([4]) M. BORN and E. WOLF, *Principles of Optics*, New York, 1959, esp. chap. III.
([5]) R. COURANT and D. HILBERT, *Methods of Mathematical Physics*, vol. 2, New York, 1962, esp. chap. VI, § 5.

We hope to show that here, as in the other parts of electrodynamics, the four-dimensional covariant treatment, which emphasizes world-geometric relations, is both more general and more transparent than the older methods.

One can also derive the laws of relativistic geometrical optics formally by studying the characteristic discontinuities of MAXWELL's equations, which takes place at a different level of generality ([6, 7, 8]).

In the stationary case – i.e., when $g_{\alpha\beta}$, u^α, n (see section 1) are invariant under a one-dimensional group with time-like trajectories – the laws that are formulated in section 3 will be quite similar to then classical ones; a FERMAT principle will then be true, as well ([7]).

1. Assumptions and basic equations

Let W be a (flat or curved) space-time manifold, and let g_{ab} be its metric tensor, with the signature (+ + + –). An arbitrarily-moving medium has the four-velocity u^a, $u_a u^a = -1$. Let the medium be isotropic, transparent, and non-dispersive, such that its electromagnetic properties are characterized by a real, scalar, dielectric constant ε and a corresponding permeability μ that can both be (continuously-differentiable) functions of the space-time point x^a.

Under these assumptions, the two electromagnetic bivectors F_{ab} [$\equiv (\mathfrak{B}, \mathfrak{A})$) and H_{ab} ($\equiv (\mathfrak{H}, \mathfrak{D})$)] will satisfy MAXWELL's equations:

$$F_{[ab,c]} = 0, \qquad H^{ab}{}_{;\,b} = 0 \tag{1}$$

and the state equations ([9]):

$$H_{ab}\, u^b = \varepsilon\, F_{ab}\, u^b, \qquad F_{[ab,}\, u_{c]} = \mu\, H_{[ab}\, u_{c]}\,. \tag{2}$$

The covariant derivative in (1), which is suggested by ";", refers to the RIEMANNian connection that is defined in W by g_{ab} .

GORDON ([10]) has remarked that eq, (2) can be simplified substantially when one introduces a second "optical metric":

$$\bar{g}_{ab} \equiv g_{ab} + (1 - 1 / n^2)\, u_a\, u_b\,, \tag{3}$$

in addition to the metric g_{ab} , such that the former metric will be independent of the motion and index of refraction $n \equiv \sqrt{\varepsilon\mu}$ of the medium. Eqs. (2) are equivalent to:

([6]) A. LICHNEROWICZ, Ann. Mathematica **4** (1960), 1.

([7]) P. M. QUAN, Contribution to the conference on the theory of relativity in Royaumont 1959; see *Les Théories Relativistes de la Gravitation*, Paris, 1962, pp. 165.

([8]) B. HOFFMANN (ed.), *Perspectives in Geometry and Relativity*, Bloomington 1966; contribution by J. EHLERS, pp. 127.

([9]) It is sometimes asserted that eq. (1) and (2) are valid in only uniformly-moving media. That is false, as the electron-statistical derivation of these equations shows: A. N. KAUFMANN, Ann. Phys. (New York) **18** (1962), 264.

([10]) W. GORDON, Ann. Phys. **72** (1923), 421.

$$H^{ab} = (1/\mu)\,\bar{F}^{ab}, \qquad \bar{F}^{ab} \equiv \bar{g}^{ac}\,\bar{g}^{bd} F_{cd}\,. \tag{4}$$

If we agree to regard $\bar{g}_{ab}$ as the metric of W from now on, and correspondingly define the manipulation of indices and covariant derivatives, and if we regard F_{ab} as field quantities that are independent of the metric then, with the abbreviation:

$$e^{u} \equiv (\varepsilon/\mu)^{1/4}, \tag{5}$$

we will get the basic equations:

$$F_{[ab,c]} = 0, \qquad (e^{2u} F^{ab})_{;\,b} = 0, \tag{6}$$

in place of (1), (2).

Since the optical metric is also normal hyperbolic, from (6), the propagation law for electromagnetic waves in moving media will be formally almost equivalent to those in a vacuum – namely, up to the factor e^{2u}. The optical metric is generally not flat, even for a medium that is embedded in MINKOWSKI space.

Eq. (6) are conformally invariant, like the vacuum MAXWELL equations. That means that really it is not the RIEMANNian metric g_{ab} itself that is necessary for the study of eq. (6), but only the conformal structure that it defines in W [see ([8]), contribution by F. A. E. PIRANI and A. SCHILD, pp. 291].

Eqs. (6) can be combined into one equation by means of the complex, self-dual bivector:

$$G \equiv e^{u}\,(F + i\,F^{*}), \qquad (i\,G^{*} = G), \tag{7}$$

namely:

$$\nabla \cdot G + \nabla u \cdot \bar{G} = 0. \tag{8}$$

In this, ∇ means the covariant differentiation operator, and a dot suggests the contraction over the neighboring indices. F^{*} is the (real) bivector that is dual to F, and $\bar{a}$ is the complex conjugate of a.

2. Locally-approximate plane waves

In order to study the transition to geometrical optics, we consider one-parameter families $G(x;\varepsilon)$ of bivector fields ([11]) of the form:

$$G(x;\varepsilon) = e^{iS(x)/\varepsilon}\sum_{\nu=0}^{\infty}\varepsilon^{\nu}K_{\nu}(x) + e^{-iS(x)/\varepsilon}\sum_{\nu=0}^{\infty}\varepsilon^{\nu}L_{\nu}(x)\,, \tag{9}$$

$$\nabla S \neq 0. \tag{10}$$

In this, $S(x)$ means a real scalar field, the $K_{\nu}(x)$ and $L_{\nu}(x)$ are complex-self-dual [cf., eq. (7)] bivector fields, and ε is a parameter that varies over an interval $0 < \varepsilon < e(x)$.

([11]) From now on, the dielectric constant will no longer enter explicitly, and we shall take ε to be the symbol of a "small" parameter.

The Ansatz (9) should express the idea that for a sufficiently small ε, the fields $G(x, \varepsilon)$ will be locally-approximate plane waves with the (rapidly-varying) phase S / ε, propagation vector $1 / \varepsilon \cdot \nabla S$, and the (slowly-varying) amplitudes $K_0 + \varepsilon K_1 + \ldots$, $L_0 + \varepsilon L_1 + \ldots$

Due to the second term in (8), both series are necessary in (9); one needs both terms, even *in vacuo* ($u = 0$), if one would like to consider all polarization states from the outset.

Replacing (9) with (8) formally yields a series of terms of the type:

$$\left.\begin{matrix} e^{iS/\varepsilon} \\ e^{-iS/\varepsilon} \end{matrix}\right\} \cdot \varepsilon^{\nu} \cdot \text{function of } x, \qquad \nu = -1, 0, 1, 2, \ldots$$

We would like to let $\{\nu\}$ denote the two equations that arise when one sets the factors of ε^{ν} equal to zero.

If the two series in (9) converge point-wise, and the series (in W) that is point-wise differentiated with respect to x converges uniformly in a neighborhood then the system of equation pairs $\{\nu\}$ will be equivalent to eq. (8). Namely, since the series converge, it will follow from:

$$e^{iS/\varepsilon} \sum_{\nu=-1}^{\infty} \varepsilon^{\nu} a_{\nu} = e^{-iS/\varepsilon} \sum_{\nu=-1}^{\infty} \varepsilon^{\nu} b_{\nu}$$

that:

$$\lim_{\varepsilon \to 0} (e^{2iS/\varepsilon} a_{-1}) = b_{-1} ,$$

which implies that $a_{-1} = b_{-1} = 0$ for $S \neq 0$, and correspondingly yields the vanishing of all terms by induction.

We now seek the equations that would result from the equation-pair $\{\nu\}$. They look like:

$$\nabla S \cdot K_0 = 0, \qquad \nabla S \cdot L_0 = 0, \tag*{$\{-1\}$}$$

$$\begin{aligned} \nabla \cdot K_{\nu} + i \nabla S \cdot K_{\nu+1} + \nabla u \cdot \bar{L}_{\nu} &= 0, \\ \nabla \cdot L_{\nu} - i \nabla S \cdot L_{\nu+1} + \nabla u \cdot \bar{K}_{\nu} &= 0, \end{aligned} \qquad (\nu = 0, 1, 2, \ldots). \tag*{$\{\nu\}$}$$

It follows from $\{-1\}$, using (10), in a well-known way [see, e.g., ([12]), chap. IX], that: The hypersurfaces S = const. will be null hypersurfaces:

$$(\nabla S)^2 = 0 \tag{11}$$

(viz., the *eikonal* equation), and K_0, L_0 are null bivectors to the eigenvector:

$$k \equiv \nabla S. \tag{12}$$

The consequences of $\{-1\}$ are exhausted with that.

([12]) J. L. SYNGE, *Relativity: The Special Theory*, Amsterdam, 1956.

Let a solution S of (11) be given. The null lines that belong to the vector field k then define a normal, geodetic congruence – viz., the ray congruence of S. From (11) and (12), k is constant along the rays:

$$k \cdot \nabla k = 0. \tag{13}$$

We extend k to a null vierbein $\{t, \bar{t}, k, m\}$ that is normalized such that one then has:

$$t \cdot \bar{t} = k \cdot m = 1, \tag{14}$$

and all remaining scalar products of the bein-vectors are zero. Let t and m, like k, be parallel-displaced along the rays:

$$k \cdot \nabla t = k \cdot \nabla m = 0. \tag{15}$$

The expansion Θ and shearing σ of the ray congruence will then be given by:

$$\Theta \equiv \tfrac{1}{2} \nabla \cdot k = \tfrac{1}{2} \Box S = t \cdot \nabla k \cdot \bar{t}, \tag{16}$$

$$\sigma \equiv \bar{t} \cdot \nabla k \cdot \bar{t}, \tag{17}$$

and their geometric and optical interpretations are well-known [see., e.g., ([13])].

The null-vierbein $\{t, \bar{t}, k, m\}$ determines a basis $\{V, U, M\}$ of the linear space of complex-self-dual bivectors, namely, [$(p \wedge q)_{ab} \equiv 2\, p_{[a}\, q_{b]}$, cf. ([13]), Appendix 4]:

$$V \equiv k \wedge l, \qquad U \equiv m \wedge t, \qquad M \equiv l \wedge t + k \wedge m. \tag{18}$$

Due to (13) and (15), these bivectors will be constant along the rays; the scalars that appear in the developments:

$$\begin{aligned} K_\nu &= a_\nu V + b_\nu U + c_\nu M, \\ L_\nu &= a'_\nu V + b'_\nu U + c'_\nu M \end{aligned} \tag{19}$$

will then determine the way that the amplitudes K_ν, L_ν vary along the rays completely.

With the use of the given auxiliary quantities, the information about K_0, L_0 that is contained in the equations $\{-1\}$ can be expressed as:

$$b_0 = b'_0 = c_0 = b'_0 = 0. \tag{20}$$

With these preparations, we turn to the equations $\{0\}$, $\{1\}$, … We substitute the development (19) into $\{\nu\}$ and multiply the resulting vector equations by the scalars t, $\bar{t}$, m, k, respectively. Equations then arise that take the following form [let $(\;)^{\cdot} \equiv k \cdot \nabla(\;)$]:

$$\dot{a}_\nu + \Theta\, a_\nu = A[c_\nu, \bar{b}'_\nu, \bar{c}'_\nu], \tag*{$\{\nu\, t\}$}$$

$$b_{\nu+1} = B[a_\nu, b_\nu, c_\nu, \bar{a}'_\nu, \bar{c}'_\nu], \tag*{$\{\nu\, t\}$}$$

([13]) P. JORDAN, J. EHLERS, and R. K. SACHS, Akad. Wiss. Lit. Mainz Abhandl. Math.-Nat. Kl. No. 1 (1961).

$$c_{\nu+1} = C[a_\nu, b_\nu, c_\nu, \bar{a}'_\nu, \bar{c}'_\nu], \qquad \{\nu\ m\}$$

$$\dot{c}_\nu + 2\,\Theta\, c_\nu = D[\bar{b}'_\nu, \bar{c}'_\nu]. \qquad \{\nu\ k\}$$

In these equations, A, B, C, D are linear forms in the given argument functions and their gradients whose coefficients are constructed from the basis vectors t, $\bar{t}$, k, m, and their first derivatives. Corresponding equations are valid for a'_ν, $b'_{\nu+1}$, $c'_{\nu+1}$, c'_ν.

Due to (20), $\{0\ k\}$ is fulfilled identically.

These equations can be solved in succession by the following process:

One gives an initial hypersurface $\mathfrak{A}$ that cuts each ray of the congruence precisely once (see Fig. 1).

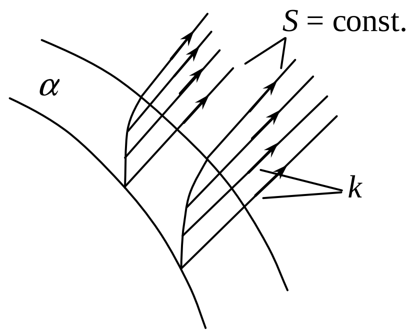

Figure 1.

In addition, one gives the values of the amplitudes a_ν, a'_ν for $\nu = 0, 1, \ldots$ on $\mathfrak{A}$. Due to (20), equations $\{0\ t\}$, i.e.:

$$\dot{a}_0 + \Theta\, a_0 = 0, \qquad \dot{a}'_0 + \Theta\, a'_0 = 0, \tag{21}$$

determine a_0 and a'_0 everywhere in the domain that is filled by the rays.

If $a_\nu, \ldots, c'_\nu$ are known already then the equations $\{\nu\ t\}$ will determine $b_{\nu+1}, \ldots, c'_{\nu+1}$ everywhere. In addition, $\{\nu\ m\}$ should be employed to determine $c_{\nu+1}$ and $c'_{\nu+1}$ on $\mathfrak{A}$, and thus the equation-pair $\{\nu+1\ k\}$ will allow one to calculate $c_{\nu+1}$, $c'_{\nu+1}$ everywhere. Ultimately, $\{\nu + 1\ t\}$ is coupled with the initial values $a_{\nu+1}$, $a'_{\nu+1}$.

If the series (9) that is defined by the coefficients thus-determined converges, and the term-wise differentiated series is locally uniformly-convergent then, by construction, G will satisfy the equation:

$$\nabla \cdot G + \nabla u \cdot \bar{G} = \lambda\, k, \tag{22}$$

in which λ is a scalar that vanishes along $\mathfrak{A}$. (In fact, $\{\nu\ m\}$ is satisfied only on $\mathfrak{A}$.) However, it follows from (22) that:

$$\nabla \cdot (\lambda\, k) = \dot{\lambda} + 2\,\Theta\lambda = \bar{\lambda}\dot{u}, \tag{23}$$

so λ is indeed equal to zero everywhere, and G fulfills MAXWELL's equations. (The c_ν thus satisfy the equations $\{\nu\ m\}$ everywhere, in such a way that the equations $\{\nu\ k\}$ are superfluous.)

Even when the series does not converge, one might expect that for sufficiently small e the partial sums in (9) will describe MAXWELL fields approximately.

3. The lowest geometrical optics approximation

If a solution $G(x;\varepsilon)$ of the form (9) (or an approximate solution in the form of a finite partial sum) is given then for sufficiently small ε:

$$e^{iS/\varepsilon} K_0 + e^{-iS/\varepsilon} L_0 \tag{24}$$

will be an arbitrarily good approximation for $G(x;\varepsilon)$.

Due to (19), (20), the behavior of (24) is determined completely by the eikonal equation (11) and the equations of propagation (15) and (21). If we let r denote a positive solution of:

$$\dot{r} = \Theta r \tag{25}$$

such that along a ray r^2 will vary like the cross-section of a thin ray bundle that surrounds the originally-stated ray ([13]) then we can write the field strength tensor F that corresponds to (24) in the form:

$$F = (1/r)\,(\mu/\varepsilon)^{1/4}\, k \wedge \Re\,(e^{iS/\varepsilon}[a_+,\, t + a_-\,\bar{t}\,]), \tag{26}$$

where a_+, a_- are now complex functions (whose amplitudes have positive and negative helicities) that are constant along the rays.

The energy-impulse tensor $T \equiv - F \cdot F$ that formally belong to (26) – which is constructed as if the optical metric were the "true" metric, and the field were a vacuum field – is:

$$T = k \otimes k\,(1/2\,r^2)\,\sqrt{\mu/\varepsilon}\;\cdot (|\,a_+|^2 + |\,a_-|^2 + 2\,\Re\,[a_+\,a_-\,e^{2iS/\varepsilon}]). \tag{27}$$

This tensor (precisely) satisfies the "conservation law":

$$\nabla \cdot (\sqrt{\varepsilon/\mu}\;\;T) = 0. \tag{28}$$

The equations that were given here include the combined generalization of the laws of classical geometrical optics of media at rest and the relativistic geometric optics in the gravitational vacuum that was announced in the introduction.

The last term in (27) drops out when one averages over a light period or makes the transition to a polarization mixture, and from (27) and (28), a geometrical-optical field will behave energetically like an incoherent photon gas with the "four-velocity" k and a density that one can read off from (27).

After one specializes eq. (26) to a vacuum, in which only r varies along the rays, a gravitational field will not change the polarization state of a ray (the degree of polarization of a mixture, resp.).

From (4), the MINKOWSKI energy-impulse tensor:

$$T_{Ma}{}^{b} \equiv F_{ac}\,H^{bc} - \tfrac{1}{4}\,\delta_a^b\,F_{cd}\,H^{cd} \tag{29}$$

is connected with the tensor (27) thus:

$$T_a{}^b = \mu\,T_{Ma}{}^b.$$

For any observer with the four-velocity v, one will then have $v^a\,T_{Ma}{}^b \sim k^b$, in such a way that the ray velocity will be:

$$\mathfrak{k} / k^4 = [\mathfrak{E} \times \mathfrak{H}] / (\mathfrak{E} \cdot \mathfrak{D}). \tag{30}$$

The result (30) rests upon the fact that the field propagates along the geodetic null-lines of the optical metric, which is a statement that first becomes meaningful for laterally-bounded ray bundles (and not for plane waves, strictly speaking). Whether or not $T_{Ma}{}^b$ is the physically-correct energy-impulse tensor is entirely irrelevant to (30); indeed, that difficult question can hardly be answered adequately with the primitive model of matter that was chosen here. (On that, confer ([14]), Appendix B. This opinion contradicts that of M. von Laue and W. Pauli; confer ([15]), Supplementary Note 11.)

For the observer above, one will also have:

$$T_a{}^b\,v_b \sim S_{,a} \sim T_{Mab}\,v^b,$$

so the spatial part of $g^{ab}\,S_{,b}$ – viz., the wave normal – will be proportional to $\mathfrak{D} \times \mathfrak{B}$.

In conclusion, without going into the higher-order approximations, we would like to give:

$$b_1 = i\,\bar{a}_0'\,\dot{u} - 2i\,\sigma a_0 \tag{31}$$

as an example of one of the equations $\{0\ t\}$. From (18) and (19), the appearance of a b_1-term in (9) means a partial “continuous reflection” of the “principal wave” (26). From (31), the magnitude of b_1 depends upon just $\dot{u}$ and σ. Similarly, the remaining equations $\{v\}$ show that the corrections that relate to (26) will be become larger the faster that u varies in space-time and the more strongly that the phase hypersurfaces S = const. are curved, which is to be expected physically.

([14]) V. L. GINZBURG, *The Propagation of Electromagnetic Waves in Plasmas,* New York, 1964.
([15]) W. PAULI, *Theory of Relativity,* London, 1958.