Paraxial-wave optics and relativistic front description. I. The scalar theory

E. C. G. Sudarshan*
Institute of Theoretical Physics, S-41296 Göteborg, Sweden

R. Simon†
Department of Physics, Indian Institute of Science, Bangalore 560012, Karnataka, India

N. Mukunda‡
Institute of Theoretical Physics, S-41296 Göteborg, Sweden
(Received 20 December 1982)

The scalar wave equation is analyzed in the relativistic front form, appropriate for paraxial-wave optics. The group-theoretical basis for this treatment is uncovered. The formal similarity of the propagation of paraxial beams through optical systems to the quantum mechanics of particles in two dimensions subject to harmonic impulses, and the role of the metaplectic group of Bacry and Cadilhac, are both traced back to the structure of the Poincaré group. Light rays are defined in this context as in statistical-wave optics, and the laws for their free propagation as well as transmission through lenses are derived.

I. INTRODUCTION

In the study of the passage of light through an optical system it is often advantageous to consider the paraxial approximation in which the beam travels along an "axis" and directions close to it. In such a context the illumination may be identified as a combination of axial pencils of light for which the opening angles are all rather small. If a typical angle is θ we designate those situations in which $\cos\theta$ may be well approximated by $1-\frac{1}{2}\theta^2$ as paraxial. Thin lenses lend themselves to considerable simplification for paraxial optics.

For light propagation in vacuum the wave numbers k_1,k_2,k_3 for light of frequency ω satisfy

$$k_3^2=\frac{\omega^2}{c^2}-k_1^2-k_2^2 .$$

If we have a paraxial pencil traveling along the positive x^3 direction, it would be a superposition of waves for each of which we have $k_1^2+k_2^2$ much less than k_3^2, and so we may deduce

$$k_3\cong\frac{\omega}{c}-\frac{k_1^2+k_2^2}{2\omega/c} .$$

Hence k_3 differs from its value ω/c for strictly axial propagation by terms quadratic in the transverse wave numbers k_1,k_2. Since in optics one is often interested in the propagation of quasimonochromatic radiation, we can see that as the wave travels along the axis the phase advance is not quite that for a strictly axial wave but differs from it by terms of the form

$$(k_1^2+k_2^2)\left[\frac{2\omega}{c}\right]^{-1} ,$$

which is reminiscent of the phase propagation in time of a Schrödinger wave function for a free nonrelativistic "particle" of mass $\hbar\omega/c^2$. On the other hand, the phase distortion introduced by a thin lens of focal length f ($\gtrless 0$ accordingly as the lens is convex or concave) is essentially of the form

$$-\frac{\omega}{2fc}(x^2+y^2) ,$$

provided the lens is placed centrally and normally on the axis. If we can find an exact description for the propagation of quasimonochromatic light which employs "wave functions" in planes normal to the axis, then a lens would appear as an object yielding a *harmonic* impulse to an otherwise *free* quantum-mechanical *particle.* The succession of harmonic impulses to free particles corresponding to a sequence of lenses can be easily computed. We would thus expect such a description to be easily visualizable in terms of our knowledge of quantum systems. Furthermore, the composition of several harmonic impulses separated by intervals of free particle propagation can be expressed in terms of a single magnification and a single harmonic impulse.

It would therefore be appropriate to search for such a description. It turns out that there is a transcription of relativistic wave propagation into the so-called "front form" in which such a description emerges naturally.[1] The usual idea of quasimonochromaticity is replaced by a new but closely related concept of "henochromaticity" (from the Greek "heis" for single as in henotheistic). In this paper we describe this approach to relativistic wave propagation and apply it to the problem of paraxial optics in successive stages of refinement. For convenience of presentation and since the ideas of the relativistic front description may be somewhat unfamiliar in the present context, we restrict ourselves in this paper to scalar waves.[2] In a succeeding paper we take up the case of the complete system of Maxwell's equations for vector waves.

We begin in Sec. II with a fairly detailed account of the front form of relativistic dynamics, adapted to the case of scalar wave propagation. We take care to point out the group-theoretical structures involved, in terms of the underlying Poincaré invariance of the wave equation, and to bring out the similarities as well as differences of the front form in relation to the usual split of space-time into physical space and physical time. The existence of a

(2 + 1)-dimensional Galilean group (or rather a central extension of it) embedded within the Poincaré group, which is revealed by the front form, is explained.[3] Section III explains the significance of the terms henochromatic and quasihenochromatic in the context of the front form, derives the conditions under which they are equivalent to one another, and to the corresponding more familiar concepts relating to temporal frequency of waves. With this preparation, we go on to deal with the propagation of paraxial beams, and show how the usual description of the effect of a lens on a monochromatic incident wave can be reexpressed in the front language. Since the independent front variables are combinations of the physical space and time coordinates, this reexpression has to be done with some care. We recall also at this point the very interesting considerations of Bacry and Cadilhac[4] on the relevance of the metaplectic group to these problems, and stress that the existence of this group can be traced back to the (2 + 1) Galilean algebra mentioned above. The composition of many lenses and their equivalence to a single lens plus a magnifier are traced. In Sec. IV we adapt the recently introduced notion of generalized rays of light in statistical wave optics,[5] to the case of paraxial beams in the front form.[6] In this limiting situation their propagation laws are much simpler than in the full three-dimensional case: We are able to explicitly solve the equation of motion for the generalized intensity of light rays and see that in free space these rays travel in straight lines.[7] The effect of a thin lens on a generalized light ray is shown to be a simple bending on encounter with the lens, in the quadratic approximation to the lens phase transformation. On the basis of this result a startlingly simple derivation of the fundamental thin lens formula is obtained. We also outline why "ray tracing" fails beyond the quadratic phase approximation in the action of a thin lens, not only because of caustics and astigmatism but because of the true wave behavior of pencils of generalized rays of light going beyond diffraction. The concluding section, Sec. V, summarizes the essential properties of the front form for this class of problems.

II. THE FRONT FORM AND PARAXIAL PROPAGATION

We use space-time coordinates x^μ with $x^0=ct$, and the spacelike metric $g_{00}=-1$. The free scalar wave equation

$$\partial_\mu\partial^\mu\psi(x)\equiv(\nabla^2-\partial_0^2)\psi(x)=0 \qquad (2.1)$$

is invariant under the action of the inhomogeneous Lorentz, or Poincaré, group $\mathscr{P}$. An infinitesimal element of $\mathscr{P}$ is specified by the six parameters $\omega^{\mu\nu}=-\omega^{\nu\mu}$ of an infinitesimal homogeneous Lorentz transformation, and an infinitesimal space-time translation vector a^μ:

$$x^\mu\to x'^\mu=x^\mu+\delta x^\mu,\quad \delta x^\mu=\omega^{\mu\nu}x_\nu+a^\mu\ . \qquad (2.2)$$

Its action on ψ can be expressed by saying that the functional form of ψ gets altered by the amount

$$\delta\psi(x)=-\delta x^\mu\partial_\mu\psi(x)\ . \qquad (2.3)$$

Following quantum-mechanical usage we can associate a generator G with this transformation in the following way:

$$G=\tfrac{1}{2}\omega^{\mu\nu}M_{\mu\nu}-a^\mu P_\mu\ ,$$
$$M_{\mu\nu}=-i(x_\mu\partial_\nu-x_\nu\partial_\mu),\quad P_\mu=-i\partial_\mu\ . \qquad (2.4)$$

Then Eqs. (2.2) and (2.3) can be expressed as the effects of applying G to x^μ and $\psi(x)$, respectively,

$$\delta x^\mu=-iGx^\mu,\quad \delta\psi(x)=iG\psi(x)\ . \qquad (2.5)$$

The ten differential operators $M_{\mu\nu},P_\mu$ reflect the structure of $\mathscr{P}$ through their commutation relations:

$$[M_{\mu\nu},M_{\rho\sigma}]=i(g_{\mu\rho}M_{\nu\sigma}-g_{\nu\rho}M_{\mu\sigma}+g_{\mu\sigma}M_{\rho\nu}-g_{\nu\sigma}M_{\rho\mu})\ ,$$
$$[M_{\mu\nu},P_\rho]=i(g_{\mu\rho}P_\nu-g_{\nu\rho}P_\mu)\ , \qquad (2.6)$$
$$[P_\mu,P_\nu]=0\ .$$

To avoid any misunderstanding, let us note the following points: Since we are dealing with a classical theory, the generator G has only a geometrical significance and not a dynamical one. It is therefore merely a convention to have the factors of i appear in the expressions above. The generators of $\mathscr{P}$ acquire the status of dynamical variables only when one develops a classical theory in its canonical Hamiltonian form, or goes into quantum mechanics. The homogeneous Lorentz generators $M_{\mu\nu}$ are dimensionless, while the translation generators P_μ carry dimension of inverse length. [In quantum mechanics one would essentially deal with $\hbar$ times the operators in Eq. (2.4), which are then physical angular momenta and energy momenta.]

The relativistic invariance of a dynamical system can be presented in several distinct forms, which at the classical level are equivalent to one another. These forms have been elaborated by Dirac,[1] and they differ from one another in the way in which the overall space-time development of the system is split into kinematical and dynamical parts. In the familiar instant form, we view the equal-time configuration of the system over all space as its state. The subgroup of Euclidean transformations generated by M_{jk},P_j, $j,k=1,2,3$, leaves a constant-time surface in space-time unchanged, and is kinematical. The generator P^0 changes one such surface into another and is dynamical; the boosts M_{0j} tilt such a surface and mix space and time coordinates, and so they are also dynamical.

We explain now in a preliminary way why the front form of relativistic description is particularly suited to paraxial wave propagation. A general positive-frequency plane wave solution of Eq. (2.1) is

$$\psi(x)=e^{ik\cdot x},\quad k^0>0,\quad |\vec{k}|=k^0\ . \qquad (2.7)$$

If this plane wave were strictly axial, i.e., propagating along the positive x^3 axis, its dependence on x^μ is only through the combination x^0-x^3:

$$k^3=k^0>0,\quad e^{ik\cdot x}=e^{-ik^0(x^0-x^3)}\ . \qquad (2.8)$$

This suggests the use of the light cone combinations $x^0\pm x^3$, $x_\perp\equiv(x^1,x^2)$ as independent space-time coordinates in place of the original x^μ. The general plane wave (2.7) has in these variables the form

$$e^{ik\cdot x}=\exp\{i[k_\perp\cdot x_\perp-\mathscr{M}(x^0-x^3)-\tfrac{1}{2}\mathscr{E}(x^0+x^3)]\}\ , \qquad (2.9)$$

where $k_\perp$ is the transverse (x^1-x^2) two-dimensional component of the wave vector $\vec{k}$, and $\mathscr{M}$ and $\mathscr{E}$ are defined as

$$\mathscr{M}=\tfrac{1}{2}(k^0+k^3),\quad \mathscr{E}=k^0-k^3\ . \tag{2.10}$$

Both $\mathscr{M}$ and $\mathscr{E}$ are inverse lengths, and the condition $k^2=0$ takes the (exact) form

$$\mathscr{M}\mathscr{E}=\tfrac{1}{2}k_\perp^2\ . \tag{2.11}$$

If this plane wave is paraxial, then by definition $|k_\perp|$ is much smaller than k^0 or k^3, so one has the system of inequalities

$$\mathscr{E}\ll|k_\perp|\ll\mathscr{M}\ . \tag{2.12}$$

The usefulness of the particular variables x^0-x^3,$x_\perp$, x^0+x^3 for a paraxial plane wave, or for a paraxial beam made up of a narrow superposition of such waves, is now clear: The function $\psi(x)$ experiences its most rapid variation when one changes x^0-x^3; with respect to $x_\perp$ it has a much slower variation, by the relative factor $|k_\perp|/\mathscr{M}$; and with respect to x^0+x^3 it has an even slower variation, by yet another factor of $|k_\perp|/\mathscr{M}$ as is clear from Eq. (2.11). These statements are valid, of course, under the assumption that the paraxial beam propagates roughly along the positive x^3 axis.

Let us now develop the front form of dynamics, paying attention to the group-theoretical aspects as well. The basic idea is to define a one-parameter succession of parallel hyperplanes or fronts in space-time, over each of which one of the two combinations $x^0\pm x^3$ stays constant. This one we shall call τ: It labels the fronts and plays the role of evolution parameter. The other combination of x^0 and x^3, denoted by σ, together with $x_\perp$, parametrize the points on each front and are to be thought of as "spatial" coordinates. It is known that when the generators of $\mathscr{P}$ are also rearranged in suitable fashion, a natural subalgebra with the structure of the Lie algebra of the Galilei group in (2 + 1) dimensions emerges.[8] The "Hamiltonian" in this subalgebra is that combination of P^0 and P^3 that causes changes in τ, i.e., moves a front, but leaves σ unchanged; while the role of "mass" is played by the other combination of P^0 and P^3 causing changes in σ and leaving τ invariant. We make our choices of σ and τ so that, as anticipated by the notation in Eq. (2.10), for paraxial beams propagating in the positive x^3 direction the "mass" is large and is essentially the temporal frequency, while the "Hamiltonian" is small. This requirement leads directly to the definitions

$$\sigma=x^0-x^3,\quad \tau=\tfrac{1}{2}(x^0+x^3)\ , \tag{2.13}$$

and to the front form of the wave equation (2.1):

$$\frac{\partial^2}{\partial\tau\,\partial\sigma}\psi=\tfrac{1}{2}\nabla_\perp^2\psi\ . \tag{2.14}$$

The front labeled by τ consists of all points in space-time at which τ has a given fixed value. There is a corresponding subgroup of $\mathscr{P}$ with the property that each of its elements takes a point on this front to another point on the same front, and so leaves the front as a whole unchanged. The independent generators of this subgroup are easily found by using Eqs. (2.4) and (2.5). We write $K_j=M_{0j}$,$(M_{23},M_{31},M_{12})=(J_1,J_2,J_3)$, using subscripts $a,b,\ldots,$ to go over the (transverse) values 1,2 and the two-dimensional antisymmetric symbol ϵ_{ab} with $\epsilon_{12}=1$.

The transformation of the wave equation (2.1) into the form (2.14) corresponds to the fundamental relativistic form being rewritten in terms of x_a, σ, and τ:

$$dx^jdx^j-(dx^0)^2=dx_a\cdot dx_a-2\,d\sigma\cdot d\tau\ . \tag{2.15}$$

Consequently the ten Poincaré generators can be reexpressed as

$$\begin{aligned}
&P_a=-i\frac{\partial}{\partial x_a},\quad J_3=-i\left[x_1\frac{\partial}{\partial x_2}-x_2\frac{\partial}{\partial x_1}\right],\\
&G_a=i\left[\tau\frac{\partial}{\partial x_a}+x_a\frac{\partial}{\partial\sigma}\right]=\tfrac{1}{2}(K_a-\epsilon_{ab}J_b)\ ,\\
&M=i\frac{\partial}{\partial\sigma},\quad K_3=i\left[\sigma\frac{\partial}{\partial\sigma}-\tau\frac{\partial}{\partial\tau}\right],\\
&F_a=i\left[\sigma\frac{\partial}{\partial x_a}+x_a\frac{\partial}{\partial\tau}\right]=(K_a+\epsilon_{ab}J_b)\ ,\\
&H=i\frac{\partial}{\partial\tau}\ .
\end{aligned} \tag{2.16}$$

From among these ten generators we can obtain six generators P_a,J_3,G_a,M which preserve the front. This set of six generators can be augmented by the τ-dependent quantity

$$K_3-\tau(P_0-P_3)=i\sigma\frac{\partial}{\partial\sigma} \tag{2.17}$$

which also preserves the front. These seven generators constitute a (τ-dependent) stabilizer for the fronts; this means that *the subgroup in $\mathscr{P}$ leaving the "front τ" invariant varies with τ*: We shall therefore write $\mathscr{P}(\tau)$ for this subgroup. (This is to be contrasted with the situation in the instant form, where the subgroup of $\mathscr{P}$ preserving a constant time surface does not change with time; it is the "fixed" Euclidean subgroup generated by J_j,P_j.) The three remaining generators of $\mathscr{P}$, namely $\frac{1}{2}(P^0-P^3)$ and $\frac{1}{2}(K_a+\epsilon_{ab}J_b)$, generate transformations that move or tilt the "front τ."

The commutation relations of the seven generators of $\mathscr{P}(\tau)$ can be read off from the basic ones (2.6). The variation of $\mathscr{P}(\tau)$ with τ is by conjugation by elements in $\mathscr{P}$ of the form e^{iaP_3}. There is however a six-parameter subgroup $\mathscr{P}_0\subset\mathscr{P}(\tau)$ that stays constant and is common to all the $\mathscr{P}(\tau)$: It is generated by J_3, G_a, P_a, and M. In the present realization (2.4), these generators have the forms

$$\begin{aligned}
&J_3=-i(x_1\partial_2-x_2\partial_1),\quad G_a=i\left[\tau\partial_a+x_a\frac{\partial}{\partial\sigma}\right],\\
&P_a=-i\partial_a,\quad M=i\frac{\partial}{\partial\sigma}\ .
\end{aligned} \tag{2.18}$$

If we now view the three "spatial" variables σ,x_a as coordinates on each front, we see that J_3, P_a, and M generate transformations with τ-independent effects on these variables, while this is not so in the case of G_a. Specifically, for fixed parameters ξ_a, the element $e^{i\xi_aG_a}\in\mathscr{P}_0$ acts on σ and x_a as follows:

$$e^{i\xi_aG_a}(\sigma,x_a)=(\sigma-\xi_ax_a+\tfrac{1}{2}\tau\xi_a\xi_a,x_a-\tau\xi_a)\ . \tag{2.19}$$

Therefore, one can identify a subgroup $\mathscr{P}_1\subset\mathscr{P}_0$ whose elements act in a τ-independent way on σ and x_a within each front: It is four dimensional and is generated by J_3, P_a, and M. Thus from the group-theoretical analysis we see that the front form involves the chain of subgroups $\mathscr{P}(\tau)\supset\mathscr{P}_0\supset\mathscr{P}_1$: The first is the τ-dependent stability group of a front; the second is that part of $\mathscr{P}(\tau)$ that is constant in τ; the third is the part that acts in a τ-independent way on σ,x_a within each front.

The Galilei group $\mathscr{G}^{(2)}\subset\mathscr{P}$, acting on a fictitious (2 + 1)-dimensional space-time and associated with the front form, arises as follows: We adjoin to the six generators J_3,G_a,P_a,M of $\mathscr{P}_0$ the single generator $H\equiv P^0-P^3$ which moves a front parallel to itself. The nonvanishing commutators among these seven operators are

$$[J_3,G_a]=i\epsilon_{ab}G_b,\ \ [J_3,P_a]=i\epsilon_{ab}P_b\ ,$$
$$[G_a,P_b]=i\delta_{ab}M,\ \ [G_a,H]=iP_a\ . \tag{2.20}$$

These relations follow directly from the basic ones (2.6) and do not depend on the particular realization (2.4) appropriate for the scalar wave equation. Thus they hold for any relativistic dynamical system. We recognize these commutation relations as corresponding to the Lie algebra of (a central extension of) the (2 + 1) Galilei group $\mathscr{G}^{(2)}$.[9] The importance of the existence of this structure within $\mathscr{P}$ for problems of paraxial wave optics will become clear in the sequel. In $\mathscr{G}^{(2)}$, J_3 and P_a generate rotations and translations respectively in a plane; G_a generate commuting Galilean boosts; M is the mass (here an operator), and H the "Hamiltonian." In the realization (2.4) we have, in addition to (2.18),

$$H=i\frac{\partial}{\partial\tau}\ . \tag{2.21}$$

One more item has to be dealt with before we can interpret the wave equation (2.1) as an evolution equation within the front form. Since we are dealing with waves traveling exactly with velocity c, there are solutions to Eq. (2.1) that cannot be treated at all in the spirit of an initial-value problem in the front form.[10] Such solutions are independent of σ and $x_\perp$, and are of the form

$$\psi(x)=f(\tau) \tag{2.22}$$

with f an arbitrary function of its argument. These solutions are constant in "space," i.e., over each front, and vary only with τ. (It is impossible to construct analogous solutions in the instant form other than $\psi=$const.) In wave number space, these are exactly antiaxial waves traveling in the negative x^3 direction: $k^3=-k^0$, $k_\perp=0$. It is easily seen that all the generators of $\mathscr{P}(\tau)$ annihilate every solution of this kind, since none of them involves $\partial/\partial\tau$. We shall explicitly avoid such solutions in our work. In fact we shall go further and restrict attention to "analytic signal" solutions $\psi(\sigma;x_a;\tau)$ in the front form: When Fourier analyzed with respect to σ, the spectrum of values of $\mathscr{M}=\frac{1}{2}(k^0+k^3)$ will be required to be *strictly positive.*[11] *Such solutions are automatically analytic signals* in the usual sense as well, since $\mathscr{M}>0$ implies $k^0>0$. Even more than that is true: On such solutions, the Galilean Hamiltonian is also non-negative.

This restricted manifold of "analytic signals" is carried into itself by all elements of $\mathscr{G}^{(2)}$ and also by the transformations generated by K_3. [Elements of $\mathscr{P}$ outside this set do not have this property and can mix "analytic signals" with solutions of the form (2.22).] The wave equation [(2.1) and (2.14)] can now be stated as an equation of motion in τ:

$$i\frac{\partial}{\partial\tau}\psi(\sigma;x_\perp;\tau)=H\psi=\frac{P_aP_a}{2M}\psi\ . \tag{2.23}$$

As long as the henochromatic approximation ($M\simeq$const) is obeyed this is an exact result: The unwieldy radical $[(k^0)^2-\vec{\nabla}^2]^{1/2}$ familiar from the instant form gives way to a much more manageable nonrelativistic looking but yet exactly relativistic expression in the front form. Propagation of the wave along the τ variable is formally the same as of a free nonrelativistic "quantum-mechanical" particle in two space ($x_\perp$) and one time (τ) dimensions. (Note the absence of $\hbar$, the dimensions of H, M, P_a, etc.)

We will analyze the properties of paraxial waves in some detail in the next section, but here we make some remarks which help in visualizing what is going on. For a paraxial solution $\psi(\sigma;x_\perp;\tau)$ with a small spread in $\mathscr{M}$, the dominant dependence of ψ is on the "spatial" coordinate σ; in comparison, the dependence on the remaining "spatial" coordinates $x_\perp$ is much weaker, and that on τ is even weaker. Indeed, for exactly axial waves of the form $\psi=f(\sigma)$ there is no τ dependence at all (or any dependence on $x_\perp$ either). There are, of course, no analogous nontrivial time-independent solutions in the instant form, just as there were no space-independent ones.

The Poincaré invariance of the wave equation, combined with its linearity, implies the following: If we apply any of the generators $M_{\mu\nu},P_\mu$ of $\mathscr{P}$, or any function $F(M_{\mu\nu},P_\mu)$ of them, to a solution ψ, the result is another solution. Of course, F must not carry any explicit dependence on σ, $x_\perp$, or τ. One might expect from this that paraxial solutions are carried into similar solutions if we apply (functions of) the generators of $\mathscr{P}(\tau)$ to them. This is not quite true, for two reasons. First, the generator $K_3-\tau(P_0-P_3)$ in (2.17) changes the value of M but leaves a henochromatic analytic signal henochromatic and analytic.

We now consider applying (functions of) the generators (2.18) of $\mathscr{P}_0$ to a paraxial solution and inquire if the result is again paraxial. Here it is trivially obvious that this will be so if we use only the generators J_3,P_a,M of $\mathscr{P}_1$ which have a τ-independent effect on the "space coordinates" $\sigma,x_\perp$. With G_a the situation is more subtle. A general paraxial solution ψ can be expanded in the form

$$\psi(\sigma;x_\perp;\tau)=\int d\mathscr{M}\int d^2k_\perp\mathscr{A}(\mathscr{M};k_\perp)\times e^{ik_\perp\cdot x_\perp-i\mathscr{M}\sigma-i\mathscr{E}\tau}\ ,$$
$$\mathscr{E}=k_\perp^2/2\mathscr{M}\ , \tag{2.24}$$

where it is understood that the integration on $\mathscr{M}$ is over a narrow band, and only small values of $k_\perp$ (in relation to $\mathscr{M}$) appear. (These conditions will be stated more precisely in the next section.) Then the action of G_a can be seen using Eq. (2.16) to be as follows:

$$G_a\psi(\sigma;x_\perp;\tau)=\int d\mathscr{M}\int d^2k_\perp\left[i\mathscr{M}\frac{\partial}{\partial k_a}\mathscr{A}(\mathscr{M};k_\perp)\right]\times e^{ik_\perp\cdot x_\perp-i\mathscr{M}\sigma-i\mathscr{E}\tau}\ . \tag{2.25}$$

For sufficiently smooth weight functions $\mathscr{A}(\mathscr{M};k_\perp)$, this shows that any *finite* polynomial in the G_a will preserve the paraxial nature of ψ. But with the exponential $e^{i\xi_a G_a}$ for instance, which is an element of $\mathscr{P}_0$ not contained in $\mathscr{P}_1$, we have

$$e^{i\xi_a G_a}\psi(\sigma;x_\perp;\tau)=\int d\mathscr{M}\int d^2k_\perp \mathscr{A}(\mathscr{M};k_\perp-\mathscr{M}\xi_\perp)\times e^{ik_\perp\cdot x_\perp-i\mathscr{M}\sigma-i\mathscr{E}\tau} . \tag{2.26}$$

The paraxial nature of ψ will in this case be preserved *only for* $|\xi_\perp|\ll 1$. Analogous restrictions can be worked out if one wishes to apply more complicated expressions, as for instance $\exp(i\theta G_a G_a)$, to ψ and wants to preserve the paraxial form of ψ.

Finally, it is characteristic of the Galilei algebra that it always supplies "canonical conjugates" to the translation generators P_a in a direct way.[12] From Eqs. (2.20) we have

$$\left[\frac{G_a}{M},P_b\right]=i\delta_{ab},\quad \left[\frac{G_a}{M},\frac{G_b}{M}\right]=0 . \tag{2.27}$$

This is true for the scalar wave equation as well as for the vector case, which we shall examine in Paper II. It is in fact a general feature of any relativistic dynamical system, and is seen on expressing relativistic invariance in the front form and exposing the existence of the $\mathscr{G}^{(2)}$ structure. For the scalar wave equation we see from Eqs. (2.18) that

$$\frac{G_a}{M}=x_a-\tau\frac{P_a}{M} , \tag{2.28}$$

so that x_a can be identified with G_a/M at $\tau=0$.

III. PROPAGATION AND TRANSMISSION BY LENS SYSTEMS

We now want to investigate the effect of a thin lens on a paraxial wave described in the front form. Physically, we consider a lens located at a *fixed* position along the x^3 axis in *real* space for all *real* time. This situation must be transcribed faithfully in terms of the front variables σ,τ. As a preliminary, we define various types of exact analytic signal solutions to the wave equation and examine the connections between them, then turn to the lens problem.

Let us say that an analytic signal is henochromatic and paraxial if it is of the form

$$\psi_{\rm I}(\sigma;x_\perp;\tau)=e^{-i\mathscr{M}_0\sigma}\int d^2k_\perp \mathscr{A}(k_\perp)\exp[i(k_\perp\cdot x_\perp-k_\perp^2\tau/2\mathscr{M}_0)] , \tag{3.1}$$

where $\mathscr{M}_0$ (>0) is fixed, and the function $\mathscr{A}(k_\perp)$ is nonzero only for values of $|k_\perp|$ much less than $\mathscr{M}_0$, i.e., only in the domain

$$|k_\perp|\lesssim\Delta k\ll\mathscr{M}_0 . \tag{3.2}$$

Superposing a narrow band of such waves, a quasihenochromatic paraxial analytic signal is one of the form

$$\psi_{\rm II}(\sigma;x_\perp;\tau)=\int d\mathscr{M}\int d^2k_\perp \mathscr{A}(\mathscr{M};k_\perp)exp[i(k_\perp\cdot x_\perp-\mathscr{M}\sigma-k_\perp^2\tau/2\mathscr{M})] , \tag{3.3}$$

where now $\mathscr{A}(\mathscr{M};k_\perp)$ is nonzero only when

$$\begin{aligned}&\mathscr{M}_0-\Delta\mathscr{M}\lesssim\mathscr{M}\lesssim\mathscr{M}_0+\Delta\mathscr{M},\quad \Delta\mathscr{M}\ll\mathscr{M}_0 ,\\ &|k_\perp|\lesssim\Delta k\ll\mathscr{M}_0 .\end{aligned} \tag{3.4}$$

A monochromatic paraxial analytic signal in the usual sense is of the form

$$\psi_{\rm III}(x^0;x_\perp;x^3)=e^{-i\bar{k}^0x^0}\int d^2k_\perp \mathscr{A}(k_\perp)\exp\{ik_\perp\cdot x_\perp+i[(\bar{k}^0)^2-k_\perp^2]^{1/2}x^3\} , \tag{3.5}$$

where $\bar{k}^0$ (>0) is fixed and $\mathscr{A}(k_\perp)$ is nonvanishing only for

$$|k_\perp|\lesssim\Delta k\ll\bar{k}^0 . \tag{3.6}$$

Finally, on superposing a narrow band of such waves, we arrive at a quasimonochromatic paraxial wave in the usual sense:

$$\psi_{\rm IV}(x^0;x_\perp;x^3)=\int dk^0\int d^2k_\perp \mathscr{A}(k^0,k_\perp)\exp\{ik_\perp\cdot x_\perp-ik^0x^0+i[(k^0)^2-k_\perp^2]^{1/2}x^3\} \tag{3.7}$$

with $\mathscr{A}(k^0,k_\perp)$ nonzero only when

$$\begin{aligned}&\bar{k}^0-\Delta k^0\lesssim k^0\lesssim\bar{k}^0+\Delta k^0,\quad \Delta k^0\ll\bar{k}^0 ,\\ &|k_\perp|\lesssim\Delta k\ll\bar{k}^0 .\end{aligned} \tag{3.8}$$

An analytic signal of type II is of course also one of type IV, by mere identification of integration variables, namely

$$\mathscr{M}=\tfrac{1}{2}\{k^0+[(k^0)^2-k_\perp^2]\} . \tag{3.9}$$

Up to terms of order $\Delta k/\mathscr{M}_0$, one can also identify $\bar{k}^0$ and Δk^0 with $\mathscr{M}_0$ and $\Delta\mathscr{M}$, respectively. On the other hand, for suitable limits on σ and τ, one can with good accuracy replace a quasihenochromatic wave of type II by a henochromatic wave of type I, both being paraxial[4]:

$$|\sigma| \ll \frac{2\pi}{\mathcal{M}_0}\left[\frac{\mathcal{M}_0}{\Delta\mathcal{M}}\right], \; |\tau| \ll \frac{4\pi}{\mathcal{M}_0}\left[\frac{\mathcal{M}_0}{\Delta\mathcal{M}}\right]\left[\frac{\mathcal{M}_0}{\Delta k}\right]^2 ,$$
$$\psi_{\mathrm{II}}(\sigma;x_\perp;\tau) \approx e^{-i\mathcal{M}_0\sigma}\int d^2k_\perp \left[\int d\mathcal{M}\,\mathcal{A}(\mathcal{M};k_\perp)\right]\exp\left[i\left[k_\perp\cdot x_\perp - \frac{k_\perp^2\tau}{2\mathcal{M}_0}\right]\right] . \tag{3.10}$$

Note that the limits on σ and τ are quite different, because a paraxial wave depends very differently on them. In the same spirit, we can search for conditions under which a henochromatic paraxial signal of type I is equal, with good accuracy, to a monochromatic paraxial signal of type III. By a simple analysis similar to that which leads to Eq. (3.10), we find that the controlling quantity here is, naturally, Δk:

$$|\sigma| \ll \frac{8\pi}{\mathcal{M}_0}\left[\frac{\mathcal{M}_0}{\Delta k}\right]^2, \quad |\tau| \ll \frac{16\pi}{\mathcal{M}_0}\left[\frac{\mathcal{M}_0}{\Delta k}\right]^4 ,$$
$$\psi_{\mathrm{III}}(x^0;x_\perp;x^3) \approx e^{-i\bar{k}^0\sigma}\int d^2k_\perp \mathcal{A}(k_\perp)\exp[i(k_\perp\cdot x_\perp - k_\perp^2\tau/2\bar{k}^0)] . \tag{3.11}$$

Again the condition on σ is the more stringent one. This case is particularly interesting for the following reason: The two functions ψ_{I} and ψ_{III} are both exact solutions of the wave equations, but one is simple in the front form and the other is simple in the instant form. If $\Delta k = 0$, they are of course the same, a single plane wave. If $\Delta k > 0$, these two exact solutions are equal to one another, with good accuracy, if $|\sigma|$ and $|\tau|$ are limited as in Eq. (3.11). We may identify $\mathcal{M}_0$ and $\mathcal{A}(k_\perp)$ occurring in ψ_{I} with $\bar{k}^0$ and $\mathcal{A}(k_\perp)$ occurring in ψ_{III}, respectively, in stating their approximate equality. This relationship can also be expressed in the following more convenient form. Let us introduce the functions $\phi_{\mathrm{I}},\phi_{\mathrm{III}}$ in cases I and III by

$$\psi_{\mathrm{I}}(\sigma;x_\perp;\tau) = e^{-i\mathcal{M}_0\sigma}\phi_{\mathrm{I}}(x_\perp;\tau) ,$$
$$\phi_{\mathrm{I}}(x_\perp;\tau) = \int d^2k_\perp \mathcal{A}(k_\perp)\exp[i(k_\perp\cdot x_\perp - k_\perp^2\tau/2\mathcal{M}_0)] , \tag{3.12a}$$

$$\psi_{\mathrm{III}}(x^0;x_\perp;x^3) = e^{-i\bar{k}^0x^0}\phi_{\mathrm{III}}(x_\perp;x^3) ,$$
$$\phi_{\mathrm{III}}(x_\perp;x^3) = \int d^2k_\perp \mathcal{A}(k_\perp) \times \exp\{ik_\perp\cdot x_\perp + [(\bar{k}^0)^2 - k_\perp^2]^{1/2}x^3\} . \tag{3.12b}$$

Then, when σ and τ obey the conditions stated in Eq. (3.11), we have the two equivalent assertions (remember $\mathcal{M}_0 = \bar{k}^0$)

$$\phi_{\mathrm{I}}(x_\perp;\tau) \approx e^{i\mathcal{M}_0[(\frac{1}{2})\sigma - \tau]}\phi_{\mathrm{III}}(x_\perp;\tau - \tfrac{1}{2}\sigma) , \tag{3.13a}$$

$$\phi_{\mathrm{III}}(x_\perp;x^3) \approx e^{i\mathcal{M}_0x^3}\phi_{\mathrm{I}}(x_\perp;x^3 + \tfrac{1}{2}\sigma) . \tag{3.13b}$$

These statements are nontrivial, and must be properly interpreted, in the following sense: One may be tempted to argue, in the case of (3.13a) say, that since the same combination $\frac{1}{2}\sigma - \tau$ appears both outside and inside ϕ_{III} on the right-hand side, and since we claim there is no surviving σ dependence, there can then be no τ dependence either. But this is not so. A simple but careful analysis confirms that because the limits on the two variables are quite different, *within these limits* the right-hand side is sensibly independent of σ but dependent on τ. In a similar way, the right-hand side of (3.13b) is sensibly independent of σ but not of x^3.

The conditions for $\psi_{\mathrm{I}} \approx \psi_{\mathrm{II}}$ in Eq. (3.10) involve both $\Delta\mathcal{M}$ and Δk. They are in general quite different from the conditions in Eq. (3.11) leading to $\psi_{\mathrm{I}} \approx \psi_{\mathrm{III}}$, which involve just Δk. We shall restrict ourselves to paraxial analytic signals in the front form which are such that whenever σ and τ obey the former conditions, they obey the latter also. The extra requirement needed to ensure this turns out to be the same in respect of σ and of τ, and it is

$$\frac{\Delta k}{\mathcal{M}_0} \lesssim 2\left[\frac{\Delta\mathcal{M}}{\mathcal{M}_0}\right]^{1/2} . \tag{3.14}$$

We will assume this is obeyed. It means that for a given degree of paraxial behavior (nonaxial wave vectors) in the wave, we have a *lower limit* on $\Delta\mathcal{M}$ or an *upper limit* on the precision with which the frequency may be specified. Granted this we can say the following: (at least) for all those values of σ and τ for which a quasihenochromatic paraxial signal can be treated as being a henochromatic one, we can go further and regard it as being approximately monochromatic as well. We comment briefly on the need for this condition at the end of this section.

We are now able to undertake the propagation of a quasihenochromatic paraxial wave through a thin lens. We will impose the restrictions (3.10) on σ and τ, so that the wave can be approximated well by a henochromatic ϕ_{I}. Both before and after the encounter with the lens, the equation of axial propagation is then given by (2.23) with the operator M replaced by its mean value $\mathcal{M}_0$. Let the lens be centrally located at a position $x^3 = a$ on the axis, its plane normal to the axis. We assume it is circular, has thickness Δ_0, refractive index n, and focal length f ($\gtrless 0$ accordingly as it is a converging or a diverging lens). Then it is well known that its effect on a paraxial monochromatic wave incident from the left, $x^3 < a$, is to introduce a phase transformation of amount[13]

$$e^{i\varphi(x_\perp)}, \quad \varphi(x_\perp) = \mathcal{M}_0(n\Delta_0 - x_\perp^2/2f) . \tag{3.15}$$

This expression for φ is to be used only for values of $x_\perp$ such that

$$|x_\perp|^2 \le 2(n-1)f\Delta_0 = 2\Delta_0R_1R_2/(R_1+R_2) \tag{3.16}$$

which corresponds to the edge of the lens; here R_1,R_2 are the positive radii of curvature of the spherical surfaces of the lens (assuming the case $f > 0$ and both sides convex). Beyond this point the phase shift is just the amount $\mathcal{M}_0\Delta_0$ due to free propagation. An incoming wave ϕ_{III} is changed by the lens to an outgoing wave $\phi'_{\rm III}$:

$$\phi'_{\rm III}(x_\perp;a+\tfrac{1}{2}\Delta_0)\approx e^{i\varphi(x_\perp)}\phi_{\rm III}(x_\perp;a-\tfrac{1}{2}\Delta_0)\ . \qquad (3.17)$$

We assume, of course, that $\phi'_{\rm III}$ is also paraxial; this will mean that the values of $x_\perp$ considered must be much smaller than R_1,R_2.

We will express this effect of the lens as a transformation of the henochromatic amplitude $\phi_{\rm I}(x_\perp;\tau)$ appropriate to the front form. Pending this, and remembering that the free propagation is given exactly by Eq. (2.23),

$$i\frac{\partial}{\partial\tau}\phi_{\rm I}(x_\perp;\tau)=-\frac{\nabla_\perp^2}{2\mathcal{M}_0}\phi_{\rm I}(x_\perp;\tau)\ , \qquad (3.18)$$

we see the following: The encounter of the wave $\phi_{\rm I}$ with the lens is like the encounter of a two-dimensional "quantum-mechanical" free nonrelativistic particle of "mass" $\mathcal{M}_0$ with an instantaneous harmonic impulse. If we were able to specify both position and "momentum" of the particle simultaneously, such an impulse would change the "momentum" discontinuously by an amount proportional to its position. For this change in "momentum" we may write

$$P'_\perp=P_\perp-\mathcal{M}_0 x_\perp/f\ . \qquad (3.19)$$

In the next section we will see how to realize the physical content of this operator equation in terms of the trajectories of generalized light rays.

For the moment suppress the transverse variables $x_\perp$. In the x^0-x^3 plane the space-time location of the lens is given by the line $x^3=a$. Switching to the front variables σ and τ, this world line of the lens becomes

$$\tau=\tfrac{1}{2}\sigma+a\ . \qquad (3.20)$$

The region left (right) of the lens in space $x^3<a$ ($x^3>a$) is the half plane $\tau<\frac{1}{2}\sigma+a$ ($\tau>\frac{1}{2}\sigma+a$) in the front variables. We now combine Eq. (3.13) with Eq. (3.17) and see that the encounter of the henochromatic paraxial signal $\phi_{\rm I}$ with the lens must, in the first instance, be described as follows: In the region $\tau<\frac{1}{2}\sigma+a$ we have the freely propagating incident wave $\phi_{\rm I}(x_\perp;\tau)$; across the line $\tau=\frac{1}{2}\sigma+a$ we have a phase transformation

$$\phi_{\rm I}\to\phi'_{\rm I}(x_\perp;\tfrac{1}{2}\sigma+a)=e^{i\varphi(x_\perp)}\phi_{\rm I}(x_\perp;\tfrac{1}{2}\sigma+a)\ . \qquad (3.21)$$

Beyond this line the free propagation law (3.18) takes over again and determines $\phi'_{\rm I}(x_\perp;\tau)$ for $\tau>\frac{1}{2}\sigma+a$. But now under the condition (3.14) which allows us to replace $\phi_{\rm I}$ by $\phi_{\rm III}$ via Eq. (3.13a), the precise value of σ is not relevant, as long as the limits on σ and τ are obeyed. We may as well take $\sigma=0$ both in Eq. (3.13a) and in (3.21) above. *Provided a henochromatic paraxial wave can be treated as a monochromatic one, the lens located in space at $x^3=a$ appears, to good approximation, as effectively located at $\tau=a$ in the front form.* The incident and refracted waves are then related at the lens by

$$\phi'_{\rm I}(x_\perp;a)=e^{i\varphi(x_\perp)}\phi_{\rm I}(x_\perp;a)\ . \qquad (3.22)$$

The complete wave function $\psi_{\rm I}$ contains also the factor $e^{-i\mathcal{M}_0\sigma}$ which goes through unaffected.

If we have a succession of thin lenses at positions $a_1,a_2,\ldots,a_n$ with focal lengths $f_1,f_2,\ldots,f_n$, placed centrally and normal to the axis, and if we assume that at each stage the paraxial nature is maintained, then the front form propagation law for the system is

$$\phi_{\rm I}(x_\perp;\tau)=e^{-i(\tau-a_n)P_\perp^2/2\mathcal{M}_0}e^{i\varphi_n(x_\perp)}e^{-i(a_n-a_{n-1})P_\perp^2/2\mathcal{M}_0}e^{i\varphi_{n-1}(x_\perp)}\cdots e^{i\varphi_1(x_\perp)}e^{-ia_1P_\perp^2/2\mathcal{M}_0}\phi_{\rm I}(x_\perp;0)\ ,$$

$$\varphi_r(x_\perp)=\mathcal{M}_0(n\Delta_0^{(r)}-x_\perp^2/2f_r)\ ,\quad r=1,2,\ldots,n \qquad (3.23)$$

provided $\tau>a_n>a_{n-1}>\cdots>a_1>0$.

In place of a spherical lens if we had a cylindrical one, say with its axis in the y direction, then the phase transformation produced on passage through it is given by the one-dimensional form of (3.15):

$$\exp[i\mathcal{M}_0(n\Delta_0-x^2/2f)]\ , \qquad (3.24)$$

and similarly for other orientations of the cylinder axis. In general for any thin lens in the paraxial approximation the phase transformation is of the form

$$\exp[-i\mathcal{M}_0(\alpha x^2+2\beta xy+\gamma y^2)] \qquad (3.25)$$

apart from a constant term.

Bacry and Cadilhac[4] have pointed out a very interesting group-theoretical connection that simplifies the computation of the effect of an axial system of thin lenses. They recognize that the operators x^2, P^2, and $(xP+Px)$ constitute a closed Lie algebra generating the metaplectic group on one canonical pair: This is the group of real linear transformations on x and P preserving the commutation relation between them.[14] The first two generate lens phase transformations and free axial propagation, respectively, while the third one generates scale changes. It follows that the composition of any number of thin circular lenses placed at arbitrary distances along the axis can always be represented as the product of three factors; hence any such system acting as in (3.23) is equivalent to a single lens, a magnifier, and a certain amount of axial propagation. Let us briefly indicate the foundations for this equivalence of lens actions and free propagations to elements of the metaplectic group. We use a mapping that uses dimensionless variables and so differs in detail from Ref. 4. The parameter $\mathcal{M}_0$ is related to the mean wavelength of the beam by

$$\mathcal{M}_0=2\pi/\lambda\ .$$

The dimensionless operators of interest, and their commutation relations corresponding to the Lie algebra of the metaplectic group, are

$$J_0=\tfrac{1}{4}(\lambda^2P_\perp^2+x_\perp^2/\lambda^2)\ ,\quad J_1=\tfrac{1}{4}(\lambda^2P_\perp^2-x_\perp^2/\lambda^2)\ ,$$
$$J_2=\tfrac{1}{4}(x_\perp\cdot P_\perp+P_\perp\cdot x_\perp)\ , \tag{3.26a}$$

$$[J_0,J_1]=iJ_2\ ,\quad [J_0,J_2]=-iJ_1\ ,$$
$$[J_1,J_2]=-iJ_0\ . \tag{3.26b}$$

These can be mapped into Pauli matrices with the same commutation relations in this way:

$$J_0\to-\tfrac{1}{2}\sigma_2\ ,\quad J_1\to\tfrac{1}{2}i\sigma_1\ ,\quad J_2\to\tfrac{1}{2}i\sigma_3\ . \tag{3.27}$$

Then free axial propagation over an interval d, represented by the action of an operator $U(d)$ on the amplitude $\phi_{\rm I}$, is mapped into an *upper* triangular matrix with positive off-diagonal element:

$$U(d)=\exp(-idH)=\exp(-idP_\perp^2/2\mathcal{M}_0)=\exp\left[-i\frac{d}{2\pi\lambda}(J_0+J_1)\right]\exp\left[-i\frac{d}{2\pi\lambda}\left[\frac{i\sigma_1-\sigma_2}{2}\right]\right]=\begin{bmatrix}1 & d/2\pi\lambda\\ 0 & 1\end{bmatrix}\ . \tag{3.28}$$

On the other hand, an axially symmetric lens of focal length f is represented by a *lower* triangular matrix. Omitting the constant term in the phase φ in Eq. (3.15) we have

$$L(f)=\exp\left[-i\frac{\mathcal{M}_0}{2f}x_\perp^2\right]=\exp\left[-\frac{2\pi i\lambda}{f}(J_0-J_1)\right]\to\exp\left[\frac{2\pi i\lambda}{f}\left[\frac{i\sigma_1+\sigma_2}{2}\right]\right]=\begin{bmatrix}1 & 0\\ -2\pi\lambda/f & 1\end{bmatrix}\ . \tag{3.29}$$

The operator J_2 generates magnifications and demagnifications which are mapped into positive diagonal matrices:

$$e^{2i\alpha J_2}\phi_{\rm I}(x_\perp;\tau)=e^{\alpha}\phi_{\rm I}(e^{\alpha}x_\perp;\tau)\ ,$$
$$e^{2i\alpha J_2}\to e^{-\alpha\sigma_3}=\begin{bmatrix}e^{-\alpha} & 0\\ 0 & e^{\alpha}\end{bmatrix}\ . \tag{3.30}$$

All these two-dimensional matrices are real unimodular.

As a simple application of these rules, consider the following arrangement. The incident beam encounters at first a convex lens of focal length $f>0$, then propagates freely through a distance d, and finally passes through a concave lens with focal length $-f$. The net effect on the amplitude is given by a product of three factors which can be mapped into a corresponding matrix:

$$L(-f)U(d)L(f)\to\begin{bmatrix}1-d/f & d/2\pi\lambda\\ -2\pi\lambda d/f^2 & 1+d/f\end{bmatrix}\ . \tag{3.31}$$

But this matrix can be rearranged as the following product:

$$\begin{bmatrix}1-d/f & d/2\pi\lambda\\ -2\pi\lambda d/f^2 & 1+d/f\end{bmatrix}=\begin{bmatrix}1 & d_1/2\pi\lambda\\ 0 & 1\end{bmatrix}\begin{bmatrix}e^{-\alpha} & 0\\ 0 & e^{\alpha}\end{bmatrix}\begin{bmatrix}1 & 0\\ -2\pi\lambda/f_1 & 1\end{bmatrix},$$
$$d_1=d(1+d/f)^{-1}\ ,\quad e^{\alpha}=1+d/f\ ,\quad f_1=f(1+f/d)\ . \tag{3.32}$$

This implies

$$L(-f)U(d)L(f)=U(d_1)e^{2i\alpha J_2}L(f_1)\ , \tag{3.33}$$

and so the given arrangement is equivalent to the following: The beam first meets a convex lens with focal length $f_1>f$ located at the same position as the original convex lens; it then immediately undergoes a magnification of amount $e^{-\alpha}$, then propagates freely over a distance $d_1<d$ and is then observed. If we consider $f<0$, so that the first lens in the original arrangement is concave but the second one is convex, then provided $d<|f|$ we see that in the equivalent system we have $f_1>0,e^{\alpha}<1,d_1>d$. Thus in this case the system is a convex lens, then a magnification, then free propagation.

When the circular lenses are replaced by more general ones, the generators become ten in number:

$$x_ax_b\ ,\quad P_aP_b\ ,\quad \tfrac{1}{2}(x_aP_b+P_bx_a)\ \text{ for } a,b=1,2\ . \tag{3.34}$$

They generate the metaplectic group on two canonical pairs.[14] To synthesize the most general configuration we now need nine independent elements including one free propagation. When the lenses are not exactly centered or are not precisely normal to the axis the linear terms x_a,P_a also come in; such linear dependence on $x_\perp$ is characteristic of prisms. The description now requires the 15-parameter affine metaplectic group as discussed by Bacry and Cadilhac.[4]

The significance of the relativistic front form for these considerations should by now be clear. Through the $\mathcal{G}^{(2)}$ structure it guarantees the existence of conjugates G_a/M to the transverse momenta P_a. Once one has canonical pairs of operators, one is automatically led to the metaplectic group. Thus we may trace the origins of this group, exposed by the work of Bacry and Cadilhac,[4] to the basic Poincaré invariance of the wave equation. The importance of this statement, which may not be fully apparent with the scalar wave equation, will certainly come through when we analyze the vector case in paper II. The other point is that the free propagation in between lenses is exactly given by the Hamiltonian H of Eq. (2.23) quadratic in P_a and so in fact a generator of the metaplectic group. If we had wished to get something analogous in the instant form, we would have been compelled to treat the free propagation along the x^3 axis in an approximate way.[15]

In concluding this section we make the remark that if we do not adopt the condition (3.14), then in the range of values of σ and τ where a quasihenochromatic paraxial wave can be replaced by a henochromatic one, it may not follow that it is monochromatic as well. The notion of the location of the lens will be unavoidably complicated in the front language, and it will not at all be easy to make a clean separation between the "spatial" variable σ and the "temporal" τ.

IV. PENCILS OF RAYS IN PARAXIAL WAVE OPTICS

We have so far dealt with the description, propagation, and transmission through lenses of the wave amplitude in the front form. It is well known, however, that for discussing partial coherence and more generally the statistical properties of light, the appropriate theoretical object is not the wave amplitude ψ but the two-point correlation function Γ defined symbolically as[16]

$$\Gamma(1,2)\sim\langle[\psi(1)]^*\psi(2)\rangle \ .$$

The angular brackets denote a stochastic averaging of the quantity enclosed, over an ensemble suited to the situation being considered. It has more recently been shown that in terms of Γ one can introduce the notion of generalized pencils of light rays in statistical wave optics, in an exact sense.[5] The key idea is to apply to Γ the Wigner-Moyal transform familiar from the phase space description of quantum mechanics.[17] One then faithfully reproduces the typical wave properties of interference and diffraction in a new language. In this section we indicate how these ideas can be adapted to the wave description in the front form.

We assume we are dealing with an ensemble of analytic signals $\psi(\sigma;x_\perp;\tau)$, all of which obey the quasihenochromatic and quasiparaxial conditions. We limit all space-time variables σ,τ to the region where the approximation (3.10) holds, so that we are dealing effectively with a henochromatic paraxial ensemble, over which we may write uniformly

$$\psi(\sigma;x_\perp;\tau)\approx e^{-i\mathcal{M}_0\sigma}\phi(x_\perp;\tau) \ . \tag{4.1}$$

Then the two-point function measuring the correlation between the amplitudes at two points on the same front is defined as

$$\begin{aligned}\Gamma(\sigma_1,x_{1_\perp};\sigma_2,x_{2_\perp};\tau)&=\langle\psi(\sigma_2;x_{1_\perp};\tau)^*\psi(\sigma_2;x_{2_\perp};\tau)\rangle\\&\approx e^{i\mathcal{M}_0(\sigma_1-\sigma_2)}\Gamma^{(0)}(x_{1_\perp};x_{2_\perp};\tau)\ ,\\ \Gamma^{(0)}(x_{1_\perp};x_{2_\perp};\tau)&=\langle\phi(x_{1_\perp};\tau)^*\phi(x_{2_\perp};\tau)\rangle \ .\end{aligned}\tag{4.2}$$

Applying the Wigner-Moyal transformation to $\Gamma^{(0)}$ we obtain the Wolf function W:

$$W(x_\perp;p_\perp;\tau)=(2\pi)^{-2}\int d^2\xi_\perp e^{ip_\perp\cdot\xi_\perp}\times\Gamma^{(0)}(x_\perp+\tfrac{1}{2}\xi_\perp;x_\perp-\tfrac{1}{2}\xi_\perp;\tau) \ . \tag{4.3}$$

This function is real but not necessarily pointwise positive definite. It is interpreted as the (generalized) intensity of light rays with transverse direction $p_\perp$ at the transverse position $x_\perp$ on the front τ. The complete wave vector, to leading order, is $p^\tau=M_0,p_a,p^\sigma=-p_\tau=0$. Note that this transverse wave vector $p_\perp$ has dimension of inverse length.

The free propagation equation (3.18) for ϕ implies for $\Gamma^{(0)}$ the corresponding law

$$\frac{\partial}{\partial\tau}\Gamma^{(0)}(x_{1_\perp};x_{2_\perp};\tau)=\frac{i}{2\mathcal{M}_0}(\nabla^2_{2_\perp}-\nabla^2_{1_\perp})\,\Gamma^{(0)}(x_{1_\perp};x_{2_\perp};\tau) \ . \tag{4.4}$$

For the Wolf function W we then get

$$\frac{\partial}{\partial\tau}W(x_\perp;p_\perp;\tau)=-\frac{p_\perp}{\mathcal{M}_0}\cdot\nabla_{x_\perp}W(x_\perp;p_\perp;\tau) \ . \tag{4.5}$$

This *linear* partial differential equation has the immediate solution

$$W(x_\perp;p_\perp;\tau)=W\left[x_\perp-\tau\frac{p_\perp}{\mathcal{M}_0};p_\perp;0\right] \ . \tag{4.6}$$

We emphasize that this is an exact propagation law for generalized rays in the front description. It states that the *rays travel in empty space along straight lines* with the natural linear displacement in the transverse position $x_\perp$, as the front advances in space-time.

The effect of a thin lens on generalized rays of light also has a simple geometrical form, which we now consider. At this point we will invoke the condition (3.14), so that the henochromatic ensemble can be approximated by a monochromatic one. As explained in the preceding section, the front label τ can then be equivalently regarded as denoting spatial location x^3 in the axial direction. Recall first the elementary derivation in the wave picture of the image forming property of a convex circular lens.[18] Let the lens be at $x^3=0$ on the axis, and let a point object at the location $(a_\perp,-u)$ to the left of the lens emit a spherical wave with frequency $\mathcal{M}_0$. We assume of course $|a_\perp|\ll u$. Over the plane of the lens the incident wave has the form, omitting unimportant factors and in the paraxial approximation:

$$\begin{aligned}\psi_{\rm inc}(x_\perp;0)&\sim\exp\{i\mathcal{M}_0[u^2+(x_\perp-a_\perp)^2]^{1/2}\}\\&\sim\exp\left[\frac{i\mathcal{M}_0}{2u}(x_\perp-a_\perp)^2\right] \ .\end{aligned}\tag{4.7}$$

The lens phase function (3.15) applied to this incident wave gives at $x^3=0$ an outgoing wave of the form

$$\psi_{\rm out}(x_\perp;0)\sim\exp\left[i\frac{\mathcal{M}_0}{2}\left[\frac{(x_\perp-a_\perp)^2}{u}-\frac{x_\perp^2}{f}\right]\right]$$

$$\sim\exp\left[(-i)\frac{\mathcal{M}_0}{2v}\left[x_\perp+\frac{v}{u}a_\perp\right]^2\right], \quad (4.8)$$

where

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f}. \quad (4.9)$$

This is seen to be the paraxial approximation, over the plane $x^3=0$, to

$$\psi_{\rm out}(x_\perp;0)\sim\exp\{(-i\mathcal{M}_0)[v^2+(x_\perp+vu^{-1}a_\perp)^2]^{1/2}\}, \quad (4.10)$$

which is exactly the phase of a spherical wave converging to the point $(-vu^{-1}a_\perp,v)$ to the right of the lens (assuming $v>0$ for a real image). Thus the lens transformation function (3.15) contains within it the ability of the lens to focus rays to form point images of point objects, the lens law (4.9) and the correct magnification factors, all in the paraxial approximation.

Now the effect of the lens on an incident two-point function $\Gamma^{(0)}$ is to supply two phase factors (3.15) at the position $\tau=0$ of the lens:

$$\Gamma^{(0)}_{\rm out}(x_{1_\perp};x_{2_\perp};0)=\exp\left[\frac{i\mathcal{M}_0}{2f}(x_{1_\perp}^2-x_{2_\perp}^2)\right]\times\Gamma^{(0)}_{\rm inc}(x_{1_\perp};x_{2_\perp};0). \quad (4.11)$$

This causes the following change in the Wolf function:

$$\begin{aligned}W_{\rm out}(x_\perp;p_\perp;0)&=(2\pi)^{-2}\int d^2\xi_\perp e^{ip_\perp\cdot\xi_\perp}\Gamma^{(0)}_{\rm out}(x_\perp+\tfrac{1}{2}\xi_\perp;x_\perp-\tfrac{1}{2}\xi_\perp;0)\\&=(2\pi)^{-2}\int d^2\xi_\perp\exp[i(p_\perp+\mathcal{M}_0x_\perp/f)]\xi_\perp\Gamma^{(0)}_{\rm inc}(x_\perp+\tfrac{1}{2}\xi_\perp;x_\perp-\tfrac{1}{2}\xi_\perp;0)\\&=W_{\rm inc}(x_\perp;p_\perp+\mathcal{M}_0x_\perp/f;0),\end{aligned}$$

i.e.,

$$W_{\rm inc}(x_\perp;p_\perp;0)=W_{\rm out}(x_\perp;p_\perp-\mathcal{M}_0x_\perp/f;0). \quad (4.12)$$

We see here the promised physical realization of the operator transformation law (3.19): The incoming generalized ray at $x_\perp$ with direction $p_\perp$ is bent by the lens into a ray at $x_\perp$ leaving in the direction $p_\perp-\mathcal{M}_0x_\perp/f$. The image forming property is seen on combining this result with the rectilinear propagation law (4.6). Different generalized rays at different points $(x_\perp,0)$ to the *immediate* left of the lens, all originating from a common object located at $(a_\perp,-u)$ on the left with $|a_\perp|\ll u$, have varying incident directions $p_\perp$ given by

$$p_\perp=\mathcal{M}_0u^{-1}(x_\perp-a_\perp). \quad (4.13)$$

The lens alters the ray at $x_\perp$ into the new direction

$$p'_\perp=p_\perp-\frac{\mathcal{M}_0}{f}x_\perp=\frac{\mathcal{M}_0}{u}(x_\perp-a_\perp)-\frac{\mathcal{M}_0}{f}x_\perp$$

$$=\frac{\mathcal{M}_0}{v}\left[\frac{v}{u}(-a_\perp)-x_\perp\right], \quad (4.14)$$

where v is determined by Eq. (4.9). Thus all these generalized rays at various points $x_\perp$ to the *immediate* right of the lens are headed to the common point $(-vu^{-1}a_\perp,v)$, giving correctly the image location and magnification.

The qualitative picture obtained above for the effect of a thin circular lens on generalized rays is maintained for *any* thin lens, as long as one stays within the *quadratic approximation* to the phase function $\varphi(x_\perp)$ characterizing the lens. This is also the domain of applicability of the metaplectic group. Thus the lens could be off center, nonaxially symmetric, nonnormal to the axis, etc. In every case the change in the two-point function is given by

$$\Gamma^{(0)}_{\rm out}(x_{1_\perp};x_{2_\perp};0)=e^{i[\varphi(x_{2_\perp})-\varphi(x_{1_\perp})]}\Gamma^{(0)}_{\rm inc}(x_{1_\perp};x_{2_\perp};0), \quad (4.15)$$

where, since $\varphi(x_\perp)$ is no more than quadratic, we will necessarily have

$$\varphi(x_{2_\perp})-\varphi(x_{1_\perp})=\mathcal{M}_0(x_{1_\perp}-x_{2_\perp})_a\times[\lambda_a+\tfrac{1}{2}\mu_{ab}(x_{1_\perp}+x_{2_\perp})_b], \quad (4.16)$$

for some numerical vector λ_a and matrix μ_{ab}. Thus the change in the Wolf function is no more complicated than a point transformation:

$$W_{\rm inc}(x_\perp;p_\perp;0)=W_{\rm out}(x_\perp;p_\perp-\mathcal{M}_0\lambda_\perp-\mathcal{M}_0\mu x_\perp;0). \quad (4.17)$$

Each incoming generalized ray remains a single ray in this approximation, with a position-dependent change in direction on encounter with the lens, but with no change in position. However, except for the circular lens, in all other cases a point object does not produce a point image. For example, if we have a departure from axial symmetry and consider a lens with transformation function

$$\varphi(x_\perp)=-\frac{\mathcal{M}_0}{2}\left[\frac{(x^1)^2}{f_1}+\frac{(x^2)^2}{f_2}\right],\quad f_1\neq f_2 \quad (4.18)$$

this corresponds to $\lambda_a=0$ and

$$(\mu)=\begin{bmatrix}1/f_1 & 0\\ 0 & 1/f_2\end{bmatrix} \quad (4.19)$$

in Eq. (4.16). Consequently each incoming generalized ray is still bent in a definite way but by different amounts in the two planes, so all the outgoing rays are not headed for any common image point. This is the case traditionally referred to as primary astigmatism.

However as soon as we go *beyond quadratic terms* in the phase function $\varphi(x_\perp)$, qualitatively new phenomena appear and *ray tracing is no longer possible,* even for generalized rays. As the simplest case let us take

$$\varphi(x_\perp) = -\frac{\mathcal{M}_0}{2f}x_\perp^2 + \alpha(x_\perp^2)^2 . \tag{4.20}$$

It follows that an incident spherical wave front will cease to be spherical. For the two-point function $\Gamma^{(0)}$ the effect is

$$\Gamma^{(0)}_{\text{out}}(x_{1_\perp};x_{2_\perp};0) = e^{i[\varphi(x_{2_\perp})-\varphi(x_{1_\perp})]}\Gamma^{(0)}_{\text{inc}}(x_{1_\perp};x_{2_\perp};0) ,$$
$$\varphi(x_{2_\perp})-\varphi(x_{1_\perp}) = (x_1-x_2)\cdot(x_1+x_2)\left[\frac{\mathcal{M}_0}{2f} - \frac{\alpha}{2}[(x_1+x_2)^2+(x_1-x_2)^2]\right] . \tag{4.21}$$

Thus the dependence of the phase factor on $x_1 - x_2$ is no longer linear. This implies that $W_{\text{out}}(x_\perp;p_\perp;0)$ is no longer obtained from $W_{\text{inc}}(x_\perp;p_\perp;0)$ by a point transformation. The transformation is local in space in the sense that $W_{\text{out}}(x_\perp;p_\perp;0)$ is given by an integral of $W_{\text{inc}}(x_\perp;p'_\perp;0)$ with respect to $p'_\perp$, including other factors depending on $x_\perp, p_\perp, p'_\perp$. The interpretation in terms of incident and emergent generalized rays described by the respective Wolf functions continues to be possible, but ray tracing is not.

The treatment of this section generalizes Gaussian geometrical optics on the one hand and Fourier optics on the other to partially coherent fields of illumination.

V. CONCLUDING REMARKS

We recount here briefly the main advantages of using the front form for paraxial wave problems. It leads to a clean separation of the space-time variables x^μ into the combinations, $\sigma, x_\perp, \tau$ on each of which the amplitude ψ has a characteristically different degree of dependence. Here it must be mentioned that while the usual terms monochromatic and quasimonochromatic denote harmonic or near-harmonic dependence of ψ on the physical time x^0, in the front form the analogous terms henochromatic and quasihenochromatic refer to dependences on the "spatial" variable σ. Factoring away the exponential $e^{-i\omega t}$ is replaced here by factoring away $e^{-i\mathcal{M}_0\sigma}$. In the instant form we then deal with the evolution of the residual wave function with respect to x^3, and this is governed by the radical $(\omega^2/c^2-\nabla_\perp^2)^{1/2}$. In the front form, on the other hand, the evolution of the residual wave function is with respect to τ and that is given by $-\nabla_\perp^2/2\mathcal{M}_0$ exactly, free of radicals. But this must be tempered with the following remark, amplified by the analysis of Sec. III: In order to obtain a physically cogent description of the action of systems of lenses on paraxial beams in the front language, we have to connect up the properties of henochromatic beams with ordinary monochromatic ones in some way.

The identification of the transverse coordinates $x_\perp$ with $G_\perp/M$, where $G_\perp$ and M appear in the Galilei algebra, will be the basis of our treatment of vector waves in the succeeding paper. This will mean that the metaplectic group retains its significance for the full Maxwell equations.

ACKNOWLEDGMENTS

E.C.G.S. and N.M. wish to thank Professor Jan S. Nilsson for providing gracious hospitality and excellent working conditions during the concluding stages of this work. This work was supported by U. S. Department of Energy Contract No. DE-AS05-76ER03992.

*Permanent address: Department of Physics, Center for Particle Theory, University of Texas at Austin, Austin, Texas 78712.
†On leave of absence from the Department of Physics, The American College, Madurai, Tamil Nadu, India.
‡Permanent address: Centre for Theoretical Studies and Department of Physics, Indian Institute of Science, Bangalore 560012, Karnataka, India.

[1]P. A. M. Dirac, Rev. Mod. Phys. 21, 392 (1949).
[2]The front form has been used in elementary particle physics in connection with the infinite-momentum frame, as well as for quantum electrodynamics. See, for instance, L. Susskind, Phys. Rev. 165, 1535 (1968); J. B. Kogut and D. E. Soper, Phys. Rev. D 1, 2901 (1970).
[3]See in this context the references cited in Ref. 2 and also L. C. Biedenharn and H. van Dam, Phys. Rev. D 9, 471 (1974).
[4]H. Bacry and M. Cadilhac, Phys. Rev. A 23, 2533 (1981); M. Nazarathy and J. Shamir, J. Opt. Soc. Am. 72, 356 (1982).
[5]E. Wolf, Phys. Rev. D 13, 869 (1976); E. C. G. Sudarshan, Physica 96A, 315 (1979); Phys. Rev. A 23, 2802 (1981).
[6]An independent approach to this problem without use of front variables has been described by one of us elsewhere. See R. Simon (unpublished).
[7]This agrees with the findings of R. Simon, Ref. 6 above.
[8]See the references cited in Refs. 2 and 3.
[9]For descriptions of the (3+1) Galilei group and its algebra, see, for instance, J.-M. Levy-Leblond, J. Math. Phys. 4, 776 (1963); E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics—A Modern Perspective* (Wiley, New York, 1974), Chap. 19.
[10]This has been explicitly noted in Ref. 1.
[11]In the front form the term "analytic signal" characterizes the Fourier transform with respect to the "spatial" variable σ within the front, rather than with respect to τ.
[12]See, for instance, E. C. G. Sudarshan and N. Mukunda, Ref. 9 above.
[13]J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill,

New York, 1968), Chap. 5; A. Papoulis, *Systems and Transforms with Applications in Optics* (McGraw-Hill, New York, 1968), Secs. 1.2 and 9.5.

[14]For one canonical pair this group is essentially the two-dimensional matrix group SL(2,*R*). See in this connection N. Mukunda, Proc. Indian Acad. Sci. 88A, 69 (1979). For two canonical pairs we have the four-dimensional matrix group Sp(4,*R*) which plays an important role in the work of L. C. Biedenharn and H. van Dam, Ref. 3 above.

[15]This is done in R. Simon, Ref. 6 above.

[16]M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1964), Chap. X; J. R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968).

[17]E. C. G. Sudarshan, cited in Ref. 5 above, and Phys. Lett. 73A, 269 (1979).

[18]See, for instance, A. Papoulis, Ref. 13, Sec. 1.2.