International Journal of Theoretical Physics, Vol. 19, No. 5, 1980

Causality and the Group Structure of Space-Time

A. J. Briginshaw

Department of Mathematics, The City University, London, England

Received July 25, 1979

This article contains a discussion of the principle of causality and its role in the derivation of the Lorentz transformations of special relativity; in particular, there is an exposition of a key theorem of E. C. Zeeman concerning the relationship between the causal group and the inhomogeneous Lorentz group. In addition there are several remarks on the interchangeability of the geometrical and algebraic modes of expression relating to space-time structure, and an explanation of the role of the conformal group in the description of space-time.

1. INTRODUCTION

In every area of mathematical physics there is, at any one time, a minor divergence of view between the mathematician and the physicist. The physicist is concerned that theory should accord with observation and physical intuition, and designs the mathematical model to accomplish this. The mathematician, on the other hand, having recognized some flavor of mathematical structure in the physical situation, has to look primarily at the internal consistency of the mathematical model; as one facet of this, he may well concern himself with aesthetic purity and logical economy as well, while, at the same time, he has in mind the question of giving a physical interpretation to procedures and results.

The initial stimulus in the development of relativistic physics came from theoreticians with a physical bent, as would befit the close of the nineteenth century—men such as Lorentz, Einstein, and Larmor. Their analysis of observational results concerned with the propagation of light was accomplished with whatever mathematical tools were available at the time: Euclidean and non-Euclidean geometries, the theory of continuous

0020-7748/80/0500-0329$03.00/0

groups, field theory and tensor analysis, the theory of differential equations, differential geometry. Subsequently, mathematicians such as Minkowski and Reichenbach became interested in the abstract structure known as space-time. Probably the most sustained attack on the problem was that of A. A. Robb (1936), who was an associate of Sir Joseph Larmor at Cambridge. Most of the studies concerning the geometrical aspects of space-time both from group-theoretic and axiomatic viewpoints owe something to Robb; see, for example, Synge (1965), Noll (1964), and Alexandrov (1967). In particular, the group-theoretic approach to the geometry has received a considerable boost in recent years by the publication of the paper "Causality Implies the Lorentz Group," by E. C. Zeeman (1964). The present work is intended to amplify and explain the background to Zeeman's paper, particularly with regard to a redundancy in the catalog of assumptions that are usually made in the derivation of the Lorentz transformations of special relativity (Pauli, 1958). Zeeman (1964) proves that it is not necessary to assume *a priori* that the transformations are linear[1] or affine[2]; instead, he asserts that the allowable mappings should preserve causality, a preservation that is only *part* of the requirement in the usual derivation. In geometrical parlance, the equivalent statement is that the *causal geometry* and the *Lorentz geometry* are the same; certain classes of affine line[3] being invariant. In the course of the present paper we sketch the proof of the crucial theorem of Zeeman (1964) as given in Nanda (1976). We shall then turn to reservations that the physicist may have concerning the assumptions on which the Zeeman result is based, as outlined in Flato and Sternheimer (1966), and we discuss the extent to which the conformal group is a more appropriate group than the Lorentz or Poincaré groups for a description of physics in the absence of gravitational effects.

2. SOME MAPS OF SPACE-TIME AND THEIR INVERSES

We can illustrate in an elementary way matters which may be obscured by a technical gloss of analytic, set-theoretic, or topological terminology by considering elements of the special conformal group of transformations of a two-dimensional Minkowski space, a group that we denote by S_2. Such mappings are compositions of inversions and translations, and the interest of such a group to physicists is immediate, for a closely related group in four-dimensional Minkowski space, namely, the

[1]Homogeneous linear, of the form $x'_i = \Sigma a_{ij} x_j$.

[2]Inhomogeneous linear, of the form $x'_i = b_i + \Sigma a_{ij} x_j$.

[3]Null, timelike, and spacelike lines; we explain the interchangeability of the geometrical and algebraic modes of expression at the beginning of Section 3.

full conformal group, C_4, is the invariance group for any sequence of electromagnetic phenomena. C_4 is the group generated by translations, inversions by reciprocal radii, Lorentz transformations, and positive dilations, and the fact that it is the invariance group for the electromagnetic field equations has been known since the work of Cunningham (1910) and Bateman (1910).

On the face of it, if we are concerning ourselves with electromagnetic phenomena, i.e., with massless particles, it is the conformal group that we invoke; on the other hand, if we are dealing with massive particles, in uniform relative motion, it is a much more restricted subgroup, the Poincaré group, that holds sway. This difference is puzzling, and it causes us to examine very carefully the physical principles on which the mathematical model is based. That said, we are, for the time being, only concerned with the special conformal maps for illustrative purposes.

Two-Dimensional Minkowski Space. Let V_2 be the vector space of ordered pairs of real numbers of the type (x_0, x_1) with the inner product $\cdot$ defined by $x \cdot y = x_0 y_0 - x_1 y_1$.

We notice that

$$\text{(i)} \qquad x \cdot y = y \cdot x, \qquad \forall\, x, y \in V_2$$

$$\text{(ii)} \qquad x \cdot (\lambda_1 y_1 + \lambda_2 y_2) = \lambda_1 x \cdot y_1 + \lambda_2 x \cdot y_2$$

$$\forall\, \lambda_1, \lambda_2 \in R, \qquad \forall\, x, y_1, y_2 \in V_2$$

However, $x \cdot x$ is not positive definite, the inner product is *not* that of a Hilbert space, and $x \cdot x$ is *not* a norm.

We shall say that $(V_2, \cdot)$ is the *orthogonal space* associated with V_2 and $\cdot$, and we refer to it as M_2, the *two-dimensional Minkowski space*.

Null Lines. At each point $a \in M_2$, there is a pair of lines $(x-a) \cdot (x-a) = 0$, which distinguish the regions where $(x-a) \cdot (x-a)$ is, respectively, positive and negative; the lines with this property are called the *null lines* at a, and we refer to them as N_a, so that N_a is the line pair $x_0 - a_0 = \pm (x_1 - a_1)$.

Inversions M_2. The map of points of M_2 given by $Ix = x / x \cdot x$ is called an *inversion* of M_2. We note that I is undefined on the null pair $x_0 = \pm x_1$, i.e., on N_0.

Special Conformal Maps. Following Engstrom and Zorn (1936), and Wess (1960), we say that $f = I \circ t_a \circ I$, where t_a represents a translation by the vector $a \neq 0$, is a *special conformal map* of M_2. It is clearly not a map of the whole of M_2, for f is undefined on $N_0 \cup N_{-a/a\cdot a}$, provided that[4] $a\cdot a \neq 0$.

Properties of *f*. (i) *Explicit formula for f.* If f is nonsingular, then it has the explicit form

$$fx = \frac{x + (x\cdot x)a}{1 + 2(x\cdot a) + (x\cdot x)(a\cdot a)} \tag{2.1}$$

(ii) *Inverse of f.* It is clear that f^{-1}, leucocyte given by

$$f^{-1}x = \frac{x - (x\cdot x)a}{1 - 2(x\cdot a) + (a\cdot a)(x\cdot x)}$$

is, where defined, the *inverse* of f, since $f^{-1} = I \circ t_{-a} \circ I$. However, f^{-1} is not defined everywhere, since all the points of the set $N_0 \cup N_{a/a\cdot a}$ are singular.

(iii) *Group Structure*. If we exclude the singular points, we can say that the collection of special conformal maps of (an appropriate part of) M_2 is an Abelian group with composition as the group operation.

For, let $f = I \circ t_a \circ I$; $g = I \circ t_b \circ I$; $h = I \circ t_c \circ I$; clearly $f \circ (g \circ h) = (f \circ g) \circ h$, $f^{-1} = I \circ t_{-a} \circ I$ is an element of the set, and $f \circ f^{-1} = f^{-1} \circ f = e$ (the identity map).

Also $f \circ g = I \circ t_{a+b} \circ I = g \circ f$, as required. We may now refer to the collection as the *group* S_2, although it must be emphasized that $f \circ g$ and $g \circ f$, etc., are only meaningful for the part of M_2 that excludes leucocyte $N_0 \cup N_{-a/a\cdot a}$, etc.

(iv) *f is One*: *One and Onto*. If we restrict our attention to the domain $D = M_2 -$ leucocyte $N_{a/a\cdot a} \cup N_{-a/a\cdot a}$, we notice that f has two further properties. Firstly, if $x, y \in D$ and $x \neq y$, then $fx \neq fy$, and we say that f is one: one (*injective*). Secondly, if $y \in D$ then $\exists\ x$ such that $fx = y$ (viz., $x = f^{-1}y$), and we say that f is onto (*surjective*). When both of these properties hold, a function is said to be a one: one correspondence or bijection.

(v) *Remark*. The usefulness of this rather detailed examination for the mapping f may not be clear to the physicist. However, the reason is compelling enough; for, when two space-time observers seek to compare notes, they must have a store of bijections with which to make their comparisons, and they need to agree on the domain D to which the comparisons refer. Otherwise, the comparisons would be subject to a degree of indeterminacy on the one hand and subject to one observer "recording" a singular value on the other.

(vi) *Continuity of f and f^{-1}*. For the domain D, both f and f^{-1} contain the quotients of continuous functions of two real variables and, as a result,

[4]If $a \neq 0$, but $a\cdot a = 0$, then the singular points comprise the set $N_0 \cup \{x : 1 + 2ax = 0\}$.

they are themselves continuous. However, when we discuss continuity, it should be remembered that we are referring to intervals associated with the metric topology of the real plane, for continuity of a function ϕ means that, for a given disk A of radius ε in D (i.e., an interval of the metric topology), $\exists\ \delta>0$ and a disk B of radius δ in D such that $\phi^{-1}A\subset B$.

In its most general form, continuity is a concept associated with a topological space, but it is one that we shall not attempt to explore in any depth. It is sufficient to note that when both ϕ and ϕ^{-1} are continuous functions with respect to a certain topology, we say that ϕ is a *homeomorphism* with respect to that topology. In the case of the special conformal group, as we have already stated, each $f\in S_2$ is a homeomorphism of D with respect to the (usual) (metric) Euclidean topology.

3. THE SPECIAL CONFORMAL GEOMETRY

There is a class of geometric entities associated with the vector space structure of V_2, sets such as $l=\{x=p+\lambda b;\ p,b\in V_2,\ \lambda\in R\}$, which are called lines or *affine lines*. We are interested in the behavior of these sets under $f\in S_2$, that is to say, we would like to know whether some or all affine lines are invariant figures of the *group geometry* associated with S_2. The information that we shall obtain in this section is not new information about the elements of S_2; it is a reformulation of the fact that $f\in S_2$ is *not* an inhomogeneous linear transformation, which is exactly what we would expect from a glance at formula (2.1). We shall distinguish two categories of affine line—null lines and nonnull lines—and we shall find that, with our usual reservation regarding singular points, null lines,[5] are invariant for S_2 but nonnull lines are not.

Theorem. Null lines are invariant figures for the inversion I.

Proof. Let l be the null line $x=p+\mu b$, where $b\cdot b=0$. If $x\in l$,

$$Ix=\frac{p+\mu b}{p\cdot p+2\mu p\cdot b}$$

$$=\frac{p}{p\cdot p}+\frac{\mu X}{p\cdot p\left[p\cdot p+2\mu p\cdot b\right]} \tag{3.1}$$

where $X=(p\cdot p)b-2(p\cdot b)p$ is null.[6] Thus Il is a null line. ■

Corollary. If $f\in S_2$, then $f=I\circ t_a\circ I$, and it follows that null lines are invariant figures for S_2.

[5]The line $x=p+\lambda b$ is null iff $b\cdot b=0$.
[6]$X\cdot X=(p\cdot p)^2(b\cdot b)-4(p\cdot p)(p\cdot b)^2+4(p\cdot b)^2(p\cdot p)=0$.

Remark. This theorem and corollary should take account of the singular points on l, thus the corollary should read: "If $f \in S_2$, and l is a null line in V_2, then $f[l \cap D]$ is a subset of a null line."

> *Theorem.* There is at least one affine line whose map under $f \in S_2$ is not an affine line.

Proof. Consider the line $l = \{x : x = (x_0, \xi_1), \xi_1 \text{ fixed}\}$. Let

$$X = Ix = \frac{x}{x \cdot x}$$

then

$$X \cdot X = \frac{1}{x \cdot x}$$

and

$$X_0 = \frac{x_0}{x \cdot x}; X_1 = \frac{\xi_1}{x \cdot x}$$

Thus

$$Il = \left\{ X : X \cdot X = \frac{X_1}{\xi_1} \right\}$$

which is an hyperbola; clearly $(t_a \circ I)l$ is also an hyperbola. Suppose that $(t_a \circ I)l = \{Y : Y \cdot Y = (Y + A) \cdot B\}$, and let

$$y = IY = \frac{Y}{Y \cdot Y}$$

then

$$y \cdot B = \frac{Y \cdot B}{Y \cdot Y} = 1 - \frac{A \cdot B}{Y \cdot Y} = 1 - (A \cdot B)(y \cdot y)$$

Thus $(I \circ t \circ I)l$ is also an hyperbola. ■

Remarks. This theorem formally confirms that $f \in S_2$ is not an inhomogeneous linear (*affine*) transformation of M_2, a fact that is not unexpected firstly because of the presence of singular points for each $f \in S_2$, and secondly because of the fractional nature of the transformation formula (2.1). We now turn to the four-dimensional Minkowski space M_4, and the corresponding special and general conformal groups, S_4 and C_4.

4. FOUR-DIMENSIONAL MINKOWSKI SPACE

Let V_4 be the vector space of quadruples of real numbers of the type (x_0, x_1, x_2, x_3), let the inner product $\cdot$ be defined by $x \cdot y = x_0 y_0 - x_1 y_1 - x_2 y_2 - x_3 y_3$, and let M_4 be the orthogonal space associated with V_4 and $\cdot$; we shall refer to M_4 as the *four-dimensional Minkowski space*. We also write $Q(x) = x \cdot x$ and say that Q is the *quadratic form* associated with M_4.

The Null Cone. The set of points for which $Q(x-a)=0$ is referred to as the *null cone* at a, and denoted by N_a. As before, the set N_a separates the points for which $Q(x-a)>0$ from those for which $Q(x-a)<0$; these sets contain, respectively, the *timelike* and *spacelike* points with respect to a. There is a similar description for the affine lines associated with V_4 which pass through a; thus $x=a+\mu b$ is *timelike*, *spacelike*, or *null* according to whether b is timelike, spacelike, or null.

Remark. In physical terms, a null line through a represents the history of a photon that includes the event a in its history, and a timelike line similarly represents the history of an unaccelerated material particle.

The Affine Geometry for V_4. We have a richer heirarchy of geometrical entities in the affine geometry of V_4 than in that of V_2. The collections

$$\{\{a\}:a\in V_4\},\qquad \{a+\lambda b:a,b\in V_4,\lambda\in R\},$$
$$\{a+\lambda_1 b_1+\lambda_2 b_2:a,b_1,b_2\in V_4,\lambda_1,\lambda_2\in R\}$$
$$\{a+\lambda_1 b_1+\lambda_2 b_2+\lambda_3 b_3:a,b_i\in V_4,\lambda_i\in R\}$$

which are, respectively, designated points, lines, planes, and hyperplanes, are all elements of the affine (usual) geometry for V_4. If we now look at the structure $(V_4,\cdot)$, i.e., M_4, we shall be interested in the qualities of maps of M_4 that preserve $\cdot$, i.e., *orthogonal* maps of M_4. In particular, we shall ask under what conditions an orthogonal map of M_4 preserves all the entities of the affine geometry of V_4. However, this question is simply a reformulation in geometrical terms of another in group theoretic terms: "Under what conditions is an orthogonal map of M_4 affine (inhomogeneous linear)?"

The Standard Transformation Groups. The notation for the standard groups is as follows (in each case the group operation is composition):

- L: full Lorentz group, the collection of linear[7] mappings that preserve $\cdot$.
- $L^{\uparrow}$: orthochronous Lorentz group, a subgroup of L, members of which preserve the sign of x_0.
- $L_+^{\uparrow}$: restricted Lorentz group, a subgroup of $L^{\uparrow}$, members of which preserve the inner orientations of x_1,x_2,x_3.
- P: full Poincaré group, the collection of affine[8] mappings that preserve $\cdot$.

[7]See footnote 1.
[8]See footnote 2.

$P^{\uparrow}$: orthochronous Poincaré group (as $L^{\uparrow}$).
$P^{\uparrow}_{+}$: restricted Poincaré group (as $L^{\uparrow}_{+}$).
P^*: augmented Poincaré group, the collection obtained by composing an element of P with a uniform positive dilation (magnification factor).
$P^{\uparrow *}, P^{\uparrow}_{+}{}^{*}$: augmented orthochronous/restricted Poincaré groups.
C_4: the full conformal group for M_4; this group is generated by inversions[9] and elements of P^*. It is the widest continuous group[10] for which Maxwell's equations are invariant.
S_4: the special conformal group for M_4 (as S_2).

Affine Subgroups of C_4. Clearly S_4 is a nonaffine subgroup of C_4; as for affine subgroups, it is well known that L, P, $P^{\uparrow}_{+}$, P^*, etc., are all subgroups of C_4. In fact, it was shown by Frank (1911) that P^* is the largest affine subgroup of C_4.

Now, it is evident that all the procedures that applied for C_2 will apply equally in the case of C_4, and we are left with the following situation: if a mapping f were given to be C^{∞}, and also one:one and onto for the whole of M_4, we have only one conclusion, $f \in P^*$. The question posed by Zeeman (1964) was what other criterion could replace the C^{∞} property and yet obtain a similar result, assigning f to the collection $P^{\uparrow}_{+}{}^{*}$. In our attempt to put Zeeman's paper in a less forbidding setting, the reader may temporarily have lost sight of its goal, which is to give minimal criteria from which to establish the Lorentz/Poincaré transformations of special relativity. The answer given by Zeeman is that it is the physical *principle of causality*, interpreted as the preservation of a certain partial ordering of M_4, allied with the global one:one and onto properties for f, that guarantees the affine nature of the transformation and the conclusion that $f \in P^{\uparrow}_{+}{}^{*}$. Then a further physical constraint (Pauli, 1958, p. 10) determines that the magnification factor is unity, and leaves us with the transformation group $P^{\uparrow}_{+}$. Certainly Zeeman does not assume that the transformations are inhomogeneous linear as do other authors (Pauli, 1958); indeed, it is worth emphasizing that far from stipulating the affinity[11] of the transformations, he does not assume *a priori* that they are C^{∞}, or even continuous.

Causal Automorphisms of M_4. There are three relations between pairs of elements of M_4 that are useful in the proof of the main theorem.

[9]For M_4, as for M_2, $Ix = x/x \cdot x$ is called an inversion by "by reciprocal radii."
[10]A continuous group contains only C^{∞} mappings, i.e., the mapping f and its derivatives of all orders have to be continuous.
[11]See footnote 2.

Following Nanda (1976), we shall denote them by $<$, $\ll$ and $<\cdot$, and define them by

$$x < y \text{ iff } Q(x-y) > 0 \text{ and } x_0 < y_0$$

$$x \ll y \text{ iff } Q(x-y) \geqslant 0 \text{ and } x_0 < y_0$$

$$x <\cdot y \text{ iff } Q(x-y) = 0 \text{ and } x_0 < y_0$$

The first two of these relations are partial orders, but the third is not, because it does not have the property of transitivity.[12] Without going into technical details, the idea behind the three relations is a simple one; not every pair of points is comparable,[13] but if x,y *are* comparable one or two of the relations can associate with the pair a precedence which is simply decided by the real number order of x_0 and y_0. None of the three relations purports to compare *every* pair of elements in M_4 (hence the name partial order), but if two elements x,y are comparable, it is easy to see which relations apply for the pair, and which element has precedence. It is clear that $x \ll y$ iff $x < y$ or $x <\cdot y$.

A bijection[14] f of M_4 for which both f and f^{-1} preserve $<$ (respectively, $\ll$) is called a $<$ *automorphism*[15] (respectively, $\ll$ automorphism) of M_4. It is clear that the collection of $<$ automorphisms of M_4 constitutes a group, with composition as the group operation; we shall refer to it as the *causality group* and denote it by G.

Remark. The selection of the causality group for special consideration has a clear echo in the arena of physical intuition and observation; it certainly looks natural enough, although, as we shall see, there is an objection to it on the grounds that a global condition is less realistic than a local one (see Flato and Sternheimer, 1966). The same source also objects to the global requirement of one: one ontoness that is required by Zeeman (1964). The tenor of these objections will lead us to reexamine the role of the group C_4, but for the time being we shall state the theorem of Zeeman and give an outline of the proof provided by Nanda (1976).

[12]If $x < y$ and $y < z \Rightarrow x < z$, $<$ is said to be *transitive*.

[13]If $Q(x-y) < 0$, we say that x and y are not comparable, and also that the vector $x-y$ is *spacelike*.

[14]f is one: and onto M_4.

[15]The prefix *auto-* simply means that f is a mapping from M_4 to itself; *-morphism* implies that some aspect of mathematical structure is being preserved.

Notation. Let $K(x) = \{x + u : Q(u) > 0, u_0 > 0\}$ represent the interior of the "future" null cone at x, and let $K'(x) = \{x\} \cup K(x)$. Similarly, $-K(y) = \{y - v : Q(v) > 0, v_0 > 0\}$ is the interior of the "past" null cone at y, with $-K'(y) = \{y\} \cup -K(y)$.

Theorem (Zeeman, 1964):

$$P^{\uparrow *} = G$$

Remark. It is clear that $P^{\uparrow *} \subset G$, for, by definition $f \in P^{\uparrow}$ preserves the precedence $x < y$, and a composition with a positive multiplier preserves the precedence, too. To establish the inclusion $G \subset P^{\uparrow *}$, Nanda (1976) proves a series of lemmas to the effect that if $f \in G$, then f maps *every* straight line of the usual (affine) geometry of M_4 to a straight line. This implies that f is an affine transformation, and a known result gives $f \in P^{\uparrow *}$. In the following, some details of proof are omitted to avoid confusion by the use of topological technicalities.[16]

> *Lemma 4.1, 4.2.* Let $f \in G$. Then f, f^{-1} are continuous[17] with respect to the Euclidean metric topology.

Proof. These results depend on the analog of a procedure in V_2, the import of which is that it does not matter, when we are gauging the continuity of a mapping from V_2 to itself, whether we deal with a collection of disks as the basic intervals or a collection of rhombuses.[18] ■

> *Lemma 4.3.* Let $f \in G$. Then f is a $\ll$ automorphism.

Proof. Here, f is given to preserve the relation of precedence in the interior of the null cone; because f is continuous we may infer[19] that f preserves the relation of precedence on the boundary of the null cone, too, so that, in all, f preserves $\ll$. ■

This is an analog of the following: let D_1, D_2 be the interiors of two rhombuses in V_2 (usual topology), and let f be defined and continuous on D_1, D_2 (including the respective boundaries). Let f map D_1 one:one onto D_2; if f preserves the order of points on every interior line through one

[16] See, for example, Mendelson, 1968.

[17] I.e., f is a homeomorphism with respect to the topology.

[18] Topological details omitted.

[19] The relation between the interior of the null cone, its closure and its boundary is somewhat analogous to that between the subsets $|z| < 1$, $|z| \leq 1$, $|z| = 1$ of the complex plane. In particular the properties that carry through from interior to closure are those that are transmitted by continuous maps and homeomorphisms. The formal proof is omitted.

corner of the rhombus, then f maps D_1 one:one onto D_2 and preserves the order of points on the sides which meet at the corner.

Lemma 4.4. Let $f \in G$ and let l be a null line through x. Then fl is a null line through fx.

Proof. Let y be an arbitrary point on l with $x < \cdot y$; then $x \ll y$, so that $fx \ll fy$, from Lemma 4.3. Now, if $fx < fy$ then $f^{-1}fx < f^{-1}fy$, i.e., $x < y$, so we must have $fx < \cdot fy$. Let $[x,y]$ denote the closed interval of the Euclidean topology on l. It is clear that $[x,y] = \overline{K'(x)} \cap \overline{[-K'(y)]}$.[20] Now f preserves both $\overline{K'(x)}$ and $\overline{[-K'(y)]}$, and we have

$$f[x,y] = \overline{K'(fx)} \cap \overline{[-K'(fy)]}$$

$$= [fx, fy] \subset fl$$

We may repeat the procedure for an arbitrary point $y \in l$ with $y < \cdot x$, and the proof is complete. ■

Now that we have established that f maps a null line to another null line it would seem inconceivable that $f \notin C_4$; this is good speculation but not such good logic, it is still possible that f is a C^0 function but not C^1, or C^1 but not C^2, etc.[21] What the theorem will establish is that, if f is a causal automorphism then f goes all the way to C^∞ and we may deduce that $f \in C_4$, then the one:one and onto properties imply that we are dealing with an affine subgroup of C_4. However, the proof proceeds directly with the next lemma.

Lemma 4.5. If $f \in G$, then f maps the tangent hyperplane to the cone $\overline{K'(x)}$ touching the null line l to the tangent hyperplane to the cone $\overline{K'(fx)}$ touching the null line fl.

Proof. (See Figure 1.) Let $A = \cup_{z \in l} \overline{K'(z)}$. Then $fA = \cup_{fz \in fl} \overline{K'(fz)}$, from Lemmas 4.3, 4.4. If ∂ indicates the boundary of a set in M_4 with the Euclidean topology, it is clear that $\Pi_l = \partial A$. Now f is a continuous map, and preserves interior, closure, and boundary,[22] then

$$f\Pi_l = f\partial A = \partial fA = \Pi_{fl}$$ ■

[20]The upper bar denotes closure with respect to the Euclidean topology.

[21]A C^0 function is continuous, a C^r function has continuous derivatives to the rth order.

[22]See footnote 19.

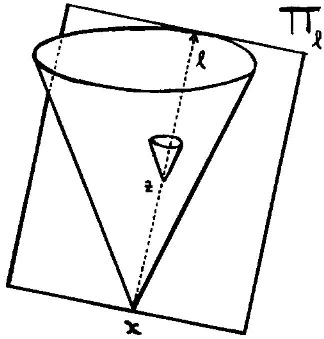

Fig. 1. Map of tangent hyperplane.

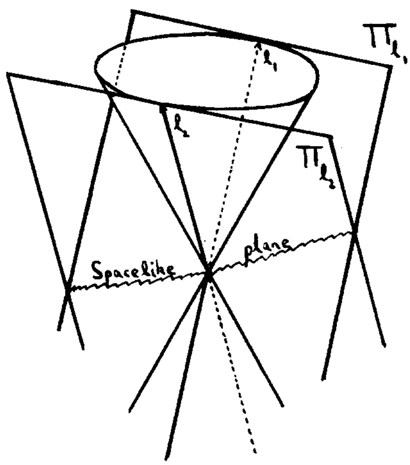

Fig. 2. Map of a spacelike plane.

Corollary 4.1. (See Figure 2.) The intersection of two tangent hyperplanes Π_{l_1}, Π_{l_2} to the null cone is a spacelike plane[23]; conversely every spacelike plane can be realized in this way. It follows from Lemma 4.5 that $f \in G$ maps each spacelike plane to a spacelike plane.

Corollary 4.2. (See Figure 3.) We observe that every spacelike line l through a point x can be realized as the intersection of two spacelike planes P_1, P_2; it follows from Corollary 4.1 that $f \in G$ maps spacelike lines to spacelike lines.

[23]A spacelike plane is the affine entity generated by two distinct spacelike lines.

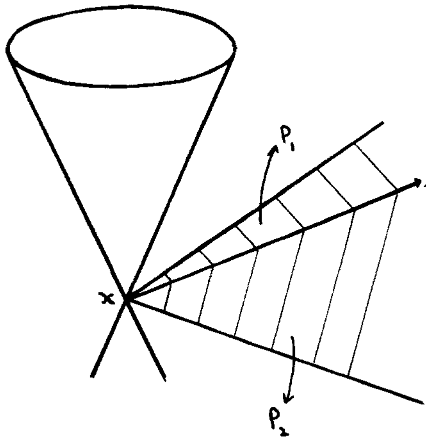

Fig. 3. Map of a spacelike line.

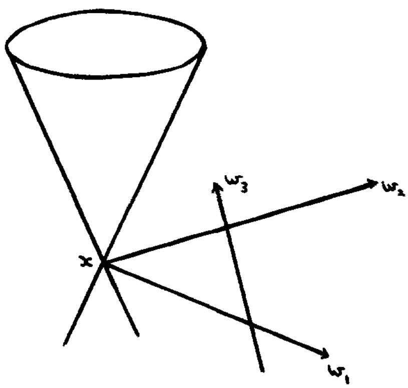

Fig. 4. Map of a plane generated by two spacelike lines.

Corollary 4.3. (See Figure 4.) Any plane that is generated by a pair of spacelike lines is mapped by $f \in G$ to a plane. To prove this assertion, we suppose that w_3 is a third spacelike line in M which intersects w_1 and w_2. It is clear from Corollary 4.2 that $f \in G$ maps the plane determined by w_1, w_2, and w_3 to the plane determined by fw_1, fw_2, and fw_3.

Corollary 4.4. (See Figure 5.) If $f \in G$, then f maps every timelike line through x to a straight line. The proof of this assertion follows by noticing that, if l is a timelike line, we can write $l = P_1 \cap P_2$, where P_1 and P_2 are planes, each of which is generated by two spacelike lines. The result then follows from Corollary 4.3.

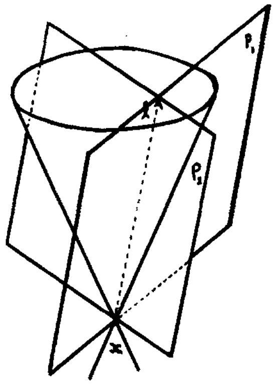

Fig. 5. Map of a timelike line.

Remark. Thus every line through x, whether spacelike, timelike, or null is mapped by an element $f \in G$ to a line through fx; indeed every affine entity is mapped to the same variety, plane to plane, hyperplane to hyperplane. The fundamental theorem of affine geometry applies and we have the following result.

> *Theorem 4.1.* If $f \in G$, then f is an inhomogeneous linear transformation of M_4.

Remark. Since $f \in G$ preserves null cones, the physical interpretation of each of which is the history of an electromagnetic wave front, we conclude that $f \in C_4$, and, adapting the result of Frank (1911), we are led to the conclusion that $f \in P^{\uparrow *}$. We have now established that $G \subset P^{\uparrow *}$, and the main theorem.

> *Theorem 4.2.* $P^{\uparrow *}$ and G coincide.

Remark. Zeeman (1964) has pointed out that Theorems 4.1 and 4.2 are valid for any Minkowski space M_n with dimension $n \geqslant 3$ but they are violated for $n = 2$. That is to say, this mathematical analysis does not cater for a universe in which there is one spatial dimension.

5. CAUSALITY AND THE CONFORMAL GROUP

We now return to a consideration of the conformal group, C_4; we already know that the elements of a certain subgroup of C_4, viz., S_4,[24] are not defined globally in M_4. We now ask whether, and in what fashion,

[24]As far as interpretation is concerned, attempts have been made to use S_4 as the group which is representative of a uniformly accelerated observer.

these maps violate the principle of causality. Of course, even the affine subgroup P^* contains causality-violating maps, since any $f \in P^* - P^{\uparrow *}$ necessarily violates the relations of precedence $<$ and $<\cdot$; but we are interested in the causality preservation or violation properties of S_4.

Violation of the Relation $<\cdot$. The group geometry for S_4 is exactly similar to that for S_2; in particular, we have the result that I maps a null line $l: x = p + \mu b$ into

$$Ix = \frac{p}{p \cdot p} + \alpha X$$

which is another null line, the result being formally the same as (3.1).

We now concentrate our attention on the parameter

$$\alpha = \frac{\mu}{p \cdot p[\,p \cdot p + 2\mu p \cdot b\,]},$$

for its behavior as a function of μ will decide whether the relation $<\cdot$ is preserved or reversed. We may write

$$\alpha = \frac{1}{2(p \cdot p)(p \cdot b)}\left(1 - \frac{p \cdot p}{p \cdot p + 2\mu p \cdot b}\right)$$

and it is clear that the behavior of α as a function of μ is illustrated in Figure 6, where $\mu^* = -p \cdot p / 2p \cdot b$ is the parameter of the singular point on l,

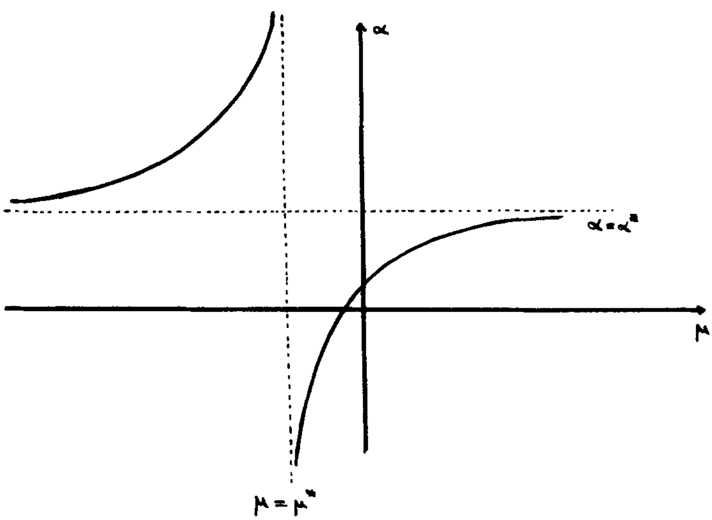

Fig. 6. Violation of $<\cdot$.

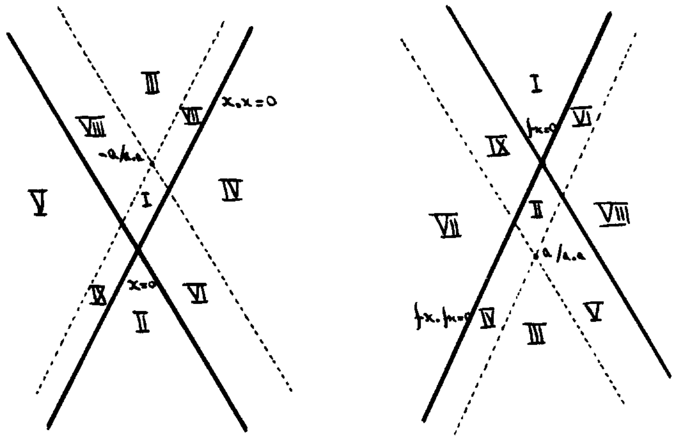

Fig. 7. Regions of local causality preservation.

and $\alpha^* = 1/2(p \cdot p)(p \cdot b)$. Now, if $\mu_1 < \mu_2 < \mu^*$, or if $\mu^* < \mu_1 < \mu_2$, then $\alpha_1 < \alpha_2$ and the relation $< \cdot$ is preserved. On the other hand, if $\mu_1 < \mu^* < \mu_2$, then $\alpha_2 < \alpha_1$, and the relation $< \cdot$ is violated.

If $f = I \circ t \circ I \in S_4$, the effect is to translate the first boundary of causality violation; this is then retained after the second I operation, but at the same time the second I operation will introduce a second causality violation boundary. In this way, M_4 is divided into nine domains of local causality preservation, which are exhibited in Figure 7, taken from Wess (1960). The numerals indicate corresponding regions of M_4 under the map $f \in C_4$; it is clear that $<$ and $< \cdot$ are preserved within and on the $x \cdot x = 0$ boundary of I.

Concluding Remark. As Flato and Sternheimer (1966) have pointed out, it may be too much to expect that a physically useful map of M_4 should be both bijective and causality preserving *globally*. They argue that a group generated by S_4 and $P^{\uparrow *}$ will preserve null lines *locally* and will preserve causality *locally*.[25] This, they say, is sufficient; perhaps it is as well to bear in mind a remark of W. K. Clifford,[26] made in 1873:

> The geometer of today knows nothing about the nature of actually existing space at an infinite distance; he knows nothing about the properties of this present space in a past or future eternity...but he knows...as of Here and Now; beyond his range is a There and Then of which he knows nothing....

[25]Nonnull lines are not preserved, even locally.

[26]Quoted in Eddington, p. 152.

REFERENCES

Alexandrov, A. D. (1967). *Canadian Journal of Mathematics*, **19**, 1119.

Bateman, H. (1910). *Proceedings of the London Mathematical Society*, **8**, 223.

Cunningham, E. (1910). *Proceedings of the London Mathematical Society*, **8**, 77.

Eddington, A. S. (1959). *Space, Time and Gravitation*, Harper and Row, New York.

Engstrom, H. T., and Zorn, M. (1936). *Physical Review*, **49**, 701.

Flato, M., and Sternheimer, D. (1966). *Comptes Rendues de l'Academie des Sciences, Paris*, **263**, 935.

Frank, P. (1911). *Annalen der Physik (Leipzig)*, **35**, 599.

Mendelson, B. (1968). *Introduction to Topology*, Allyn and Bacon, Boston.

Nanda, S. (1976). *Mathematical Proceedings of the Cambridge Philosophical Society*, **79**, 533. (1976).

Noll, W. (1964). *American Mathematical Monthly*, **71**, 129.

Pauli, W. (1958). *Theory of Relativity*, p. 9, Pergamon, New York.

Robb, A. A. (1936). *Geometry of Time and Space*, pp. 1–25, Cambridge University Press, Cambridge, England.

Synge, J. L. (1965). *Relativity: The Special Theory*, p. 58, North-Holland, Amsterdam.

Wess, J. E. (1960). *Nuovo Cimento*, **18**, 1086.

Zeeman, E. C. (1964). *Journal of Mathematical Physics*, **5**, 490.