III

Physical theory

1. The Cosmos

Ultimately, any assertion about time, energy, and physical states must be transcribed into objective, experimentally verifiable statements, on whose validity the original assertion primarily depends for its own validation. However, it is often not possible, or desirable, to proceed in a purely logical-positivistic style, in which a physical theory is described solely in terms of predictions of the results of fully specified experiments. It is widely accepted that a general theoretical superstructure may be needed, or at any rate desirable, for a variety of reasons. Among these are economy and clarity of formulation, simplification of the means of correlation of the given physical theory with others, better adaptability to modifications which may prove desirable in other physical contexts, etc.

For these reasons, it seems desirable, and indeed perhaps necessary, to present our theory of the Cosmos from a viewpoint which is so fundamental and conceptually elementary that it may appear unfamiliar, and possibly overmeticulous. It seems especially important to approach the matter conservatively, because we attempt at the same time both to extend the direction in which special relativity departs from classical mechanics, and also to change the energy operator in quantum mechanics. It might appear desirable to separate these two developments, but they are logically very closely related, as will be seen later.

2. Postulational development

We now consider what may be deduced about the Cosmos on the basis of the following very general and broadly accepted assumptions.

Assumption 1 The Cosmos is a four-dimensional manifold.

Comment This means that in the vicinity of any point p of the Cosmos M, there is a four-dimensional coordinate system. It is of course a matter of the most elementary physical experience that in the vicinity of any observer, space–time events have a linear temporal order, and a three-dimensional position.

It might be objected that this assumption may eliminate singularities which could be significant from a general relativistic standpoint. The basic answer to this is that if these singularities be deleted, then the remaining region of space time is a regular manifold to which our considerations should then apply. Having settled the nature of this underlying regular manifold, one could then examine the adjunction of hypothetical singularities.

The second answer is that we wish to operate on as direct a level of experience as possible. Singularities in the space–time structure are theoretical possibilities of a definitely idealistic nature; their concrete analytical description involves delicate questions of the separation of physically essential aspects from matters of parametrization; until these have been materially clarified, it will be impossible to give operational meaning to an assertion as to the physical existence or nonexistence of space–time singularities.

Assumption 2 The Cosmos is endowed with a notion of causality.

Specifically: (a) at each point of the Cosmos there is given a convex cone of infinitesimal future directions, in the tangent space to the manifold at each point; (b) the future can never merge into the past, i.e., no curve that always points into the future can be closed.

Comment The existence of a sense of the infinitesimal future is a psychological fact; the physical meaning and implications of the notion of future are well developed in special relativity theory, and need not be repeated here. Bridgman has emphasized the logical independence of causality and the law for the addition of velocities in special relativity. This independence is indeed substantiated by the existence of space–time models that are globally acausal, while locally Minkowskian (cf. below). In such a model, the velocity of light would appear constant in all frames in the immediate vicinity of any observer, and the usual addition formula would hold, etc., but time would "wind back" on itself in the long run. This is counter to intuition, thermodynamics, and general physical ideas; while certain microscopic physical phenomena may well be cyclical in time, cyclicity of the Cosmos as a whole is generally implausible. We shall assume—without prejudice to future possibilities—that this is not the case.

The assumption of the convexity of the future cone in the tangent space at each point is not a matter of mere technical convenience, but is indicated by general conceptual considerations. One such consideration is that any displacement of the Cosmos which is the resultant of a succession of displacements *into the future* should itself be a displacement into the future. In particular, if X and Y are any two infinitesimal generators of one-parameter displacement groups into the future, in the sense that the group, denoted e^{tX}, generated by X carries each point p of the Cosmos M into a point $q = e^{tX}(p)$ which is temporally preceded by p when $t > 0$—in the sense that there is an arc from p to q whose forward tangent at every point is in the future direction—then $(e^{tX/n}e^{tY/n})^n$ should be a displacement of M into the future. But as $n \to \infty$, this displacement tends to $e^{t(X+Y)}$, a one-parameter group whose generator is $X + Y$. Thus, if X and Y are in the future direction, so also is $X + Y$, which means that the set of infinitesimal future directions at each point is closed under addition. Together with the evident fact that tX is always in the future direction if X is such and $t > 0$, this means that the set of all infinitesimal future directions at each point is convex.

Implicitly employed here is the extremely rudimentary assumption that a limit of points in the Cosmos which are preceded (or simultaneous with) a given point, is again such. This also means that the infinitesimal future at each point should be a *closed* set, in the mathematical sense. (For the most part in Chapter III we take for granted, unless otherwise indicated, such elementary points of mathematical regularity, and refer to Chapter II for formulations which are mathematically fully detailed.)

Assumption 3 The Cosmos admits stationary observers.

Comment It is difficult to see how any physical laws of the usual sort could be effectively discussed or verified without stationarity. For the dynamics, which form the crucial content of a complete physical theory, describe the change in state from one instant to another. Without a time-independent notion of state, such a description is evidently vacuous. Finally, in order to have a physical time-independent description of states, it seems necessary to have a stationary observer or equivalent operational means of labeling states.

Let us be quite explicit about the meaning of stationarity and of Assumption 3. We introduce the notions of "timelike" and "spacelike" in the fashion made possible by Assumptions 1 and 2, along customary physical lines.† Having done this, we can define a "forward displacement" as an admissible motion—i.e. a transformation of M onto itself which preserves causality, i.e., carries the totality of future directions at one point into those at the corresponding point—with the property that it carries each point into one that it (strictly) precedes (i.e., is the terminus of a strictly timelike arc originating at the original point). The concept of stationarity is then a relative notion; specifically, it is with respect to a one-parameter group of forward and backward displacements, or *temporal group*. The latter is defined as a family T_t of admissible motions of M, t being an arbitrary real parameter, such that

$$T_t T_{t'} = T_{t+t'} \qquad (t, t' \text{ arbitrary real numbers}),$$

$$T_t \quad \text{is a forward displacement for} \quad t > 0.$$

For example, in Minkowski space, if a denotes any fixed vector in the future cone, the family T_t: $x \to x + ta$ is of this nature. Conversely, in a Minkowski space of dimension greater than two, every temporal group has this form.

Thus, in order to have an effective notion of stationarity, it seems necessary that such an underlying temporal group of transformations be defined on the Cosmos. In general, however, a theoretical cosmos satisfying Assumptions 1 and 2 will admit no such group; indeed, in general there will be no admissible displacements. Assumption 3 thus carries first of all the implication that a temporal group exists. Beyond this, however, the usual notion of observer carries with it the implication that a space–time event is split by the observer into "space" and "time" components in a definite way. *An observer stationary with respect to the given temporal group* can be defined consonantly with this notion as one for whom this splitting into space and time components is unaffected by temporal evolution, as defined by the given group. More specifically, to each point p of the Cosmos, the observer assigns two components, t and x, where the time component t is a real number, while the space component x ranges over a three-dimensional manifold S. *A stationary observer*, with respect to the given temporal group, can now be defined as one such that the associated group T_t carries the point (t', x) of the Cosmos into the point $(t' + t, x)$ (more precisely, carries the point of the Cosmos with time and space components (t', x) according to his observation into the point with components $(t' + t, x)$). It can be shown that in Minkowski space this concept of stationary observer is equivalent to that of Lorentz frame, as one would expect. That is, there is a mutual correspondence between stationary observers and Lorentz frames, every such observer being associated with a Lorentz frame in the fashion earlier indicated.

† For further material on these notions and/or mathematical details, see Chapters I and II.

From an operational point of view, the "observer" is, in large part, in the present context simply this splitting of the Cosmos into time and space components; apart from such objective features, his existence is largely metaphysical. The splitting of space–time into space and time components relative to a complete local observational framework is partly a theoretical analysis and partly an empirical deduction from experience at a fundamentally more rudimentary level than a global theory of space–time. It is not merely a matter of anthropomorphic psychology, which as evidenced by the theory of relativity interacts nontrivially with theoretical ideas on the nature of space and time, but has a close relation to the concept of "stationary state" which is crucial in modern physics. Virtually all dynamics can be formulated as a description of transformations from one (at least approximately) stationary state to another; in particular, a temporal evolution group is required for an objective means of parametrization of states which can be correlated with experience. The parameters employed effectively define "space," particularly at extreme distances, the connection with the anthropomorphic notion of space being physically explicit only at the moderate-macroscopic level. In other words, "space" is defined by the condition that stationary state labels are (primarily) quantities (functions, vector fields, or operators) defined on space, together with the boundary condition that at middle distances, it coincides with the anthropomorphic notion. It is also limited by the conception that fundamental interactions are *local*, when expressed in spatial terms. The existence of degrees of freedom for elementary particles, which have not yet been correlated with geometrical space–time features (e.g., isotopic spin) does not essentially change these matters, since the only effect is to adjoin an "internal space" to space–time, which does not affect the physical space–time splitting.

The physically crucial notions *time* and *energy* are essentially immediate deductions from the formalism; they are defined relative to a given observer. The first component t in the space–time splitting is the observer's time; equivalently, it is the parameter of the one-parameter group T_t. The relation between the time of a space–time event, and time as the parameter of a temporal evolution group is simply that if $F(p)$ denotes the time of the space–time point p, relative to the given observer, for any t, $F(T_t p) = F(p) + t$, where T_t is the temporal evolution group associated with the observer. The temporal group invariance thus permits the correlation of time as an index of serial order with time as duration, an identification which is essential for real physics. At the same time, it uniquely specifies, apart from the choice of scale and zero point, the time parameter. The energy, on the other hand, is simply the conjugate or dual variable to time, i.e., the infinitesimal generator of temporal evolution. The situation thus is fundamentally more structured than in theories in which time appears primarily as an index of serial order—as, e.g., in general relativity as usually presented. This additional structure is of course not a technical burden, but rather an essential requirement for concrete physical interpretation. Without temporal invariance there is no conservation of energy—indeed, the very concept of energy becomes ambiguous; thus, despite intensive study, the precise formulation and properties of energy in general relativity appear to remain somewhat ambiguous.

Classically, Hamiltonian dynamics apply as readily to the space S as to the usual Euclidean configuration space. Quantum mechanically, the development likewise proceeds in entirely analogous fashion to the usual one. The dynamical variables are operators; the temporal evolution is defined, in, e.g., the case of a finite number of degrees of freedom, by a one-parameter group of unitary operators $U(t)$, the infinitesimal self-adjoint generator of which is the energy operator. This operator represents $-i(\partial/\partial t)$, which thereby defines the "energy" for the observer in question. Different observers will of course have different energy operators; depending on the geometry of their respective space–time splittings, these different energies may or may not be conjugate (in which case the eigenvalues are identical) or nonconjugate (in which case the eigenvalues are in general distinct; this theoretical possibility will be exemplified below).

Assumption 4 Space is homogeneous and isotropic.

It is entirely possible to conceive physical, and to give mathematical, examples of cosmos not satisfying this condition. However, it is intuitive, and substantiated at both macroscopic and microscopic levels. Moreover, as already suggested, it is physically essential to have some objective means of labeling particle states. The usual notions of angular and linear momenta, which have been found effective for this purpose, derive from the existence of just those symmetries which are here postulated. In all events, this postulate has been implicit in theoretical astronomy since the time of Cusanus.

Let us be quite explicit about what it means. Relative to any admissible observer, there is a splitting of space–time M into time and space components T and S; symbolically, $M = T \times S$, signifying that each point p in M corresponds to a pair (t, u), where the "time" t is in T, the range of time values (normally the real line), and the "spatial position" u is in S, the "space" of the observer. This splitting of space–time into one-dimensional time and three-dimensional space components is far from arbitrary; it is subject to the restrictions:

(a) For each fixed position in space u_0, the curve (t, u_0), where t varies over T, should be timelike. Indeed, this should be a maximal timelike curve, in the sense that any point which is timelike relative to (before or after) each point of the curve must already be on the curve.

(b) For each fixed time t_0, the submanifold (t_0, u), where u ranges over S, should be spacelike. Indeed, this should be a maximal spacelike submanifold, in the sense that any point which is spacelike relative to every point of the submanifold is already in the submanifold.†

These restrictions (a) and (b) are quite rudimentary and are totally independent of symmetry considerations. But as already noted, the correlations of the notion of observer with realistic physics leads naturally to the requirement of temporal homogeneity, without which one lacks a well-determined notion of energy. The requirement of spatial homogeneity is similar; it is not as fundamental as that of temporal homogeneity, but is tantamount to the intuitively plausible assumption that the laws of physics are independent of the physical location and orientation of axes. Philosophically speaking, it is undoubtedly possible to pursue physical theory without this assumption, but it would be extremely difficult to arrive at laws that were both nontrivial and definite.

Recently, difficulties in reconciling extragalactic astronomical observations with the expanding-universe model have led to proposals for limiting this postulate as regards the distribution of galaxies, if not for "empty" space itself. The work of G. de Vaucouleurs (1972) is representative of the observational background for such proposals, but can also (although not so construed by de Vaucouleurs) be interpreted as evidence against the expanding-universe model. In our view, the latter interpretation is more natural, and in fact, it will later be shown that the discrepancies studied by de Vaucouleurs may be resolved with a spatially homogeneous nonexpanding model (cf. also Sandage *et al.*, 1972).

In addition, the conventional theoretical microscopic picture—elementary particle analysis—is based on spatial homogeneity. The use of "linear momenta" as quantum numbers for particles is precisely tantamount to the assumption that spatial homogeneity is valid at the microscopic level.

For all these reasons, spatial homogeneity appears to be a quite reasonable postulate, in the simple form analogous to that of temporal homogeneity:

For any two points P and Q of "space" S, there exists a spatial transformation of the Cosmos, i.e., a smooth transformation $P \to P'$ of the cosmos M into itself, which preserves causality (i.e., carries relatively timelike and/or spacelike points into the same), and is spatial in the sense that any point P

† It would be just about as natural to define maximality in slightly different ways, e.g., in the spatial case as the absence of any spacelike submanifold of which the given one is a proper subset. However, all of the examples and applications of these notions treated here are maximal in both senses.

and its transform P' are relatively spacelike; and, for any time t, the point P corresponding to (t, p) is carried into the point P' corresponding to (t, p'), which carries P into Q.

Having dealt with spatial homogeneity, it is now a simple matter to deal with the similar notion of spatial isotropy, for which there are both ultra-macroscopic and microscopic forms of evidence at least as strong as those for spatial homogeneity. In mathematical terms, spatial isotropy means that given any two spatial directions at a point of the Cosmos, there exists a physically admissible transformation on the Cosmos (i.e., a causality-preserving smooth transformation) which carries one direction into the other. More precisely, if λ and λ' are any two tangent vectors at the point p of the space S, and if t is any time, then there exists a spatial transformation of the Cosmos which leaves the point P corresponding to the pair (t, p) fixed, and whose action on the tangent space at P carries λ into a nonzero multiple of λ'.

Applied to Minkowski space, these concepts naturally reproduce the usual ones. The only admissible (more specifically, "covariant") observers in the foregoing sense which are applicable to *all* of Minkowski space are obtained by a representation of the space in terms of pairs $(t, \mathbf{x})$, where t is a real number and $\mathbf{x}$ is a real vector, t being the temporal and $\mathbf{x}$ the spatial component, in the usual way. However, although this is the only *global* covariant observer, there are quite different *local* covariant observers. These cannot in general be extended to all of Minkowski space without encountering singularities; cf. below.

The concept of Lorentz frame is, in the case of Minkowski space, equivalent to that of observer in the present sense, except that the latter notion leaves unspecified a distance scale. Later, it will be shown how a specification of the distance scale may be accomplished on the basis of the present assumptions, without the presupposition of a given metric.

Example Let S be any three-dimensional Riemannian manifold, admitting an isotransitive group of isometries. (Here *isotransitive* means transitive both on points and directions at points; i.e., given any two points and directions at the points, there is a transformation in the group mapping the one point into the other, and the first given direction into the second.) Take as cosmos $M = R^1 \times S$, *and define as the cone* $C(t, q)$ *at any point* (t, q) *of* M the set of all tangent vectors of the form $a(\partial/\partial t) + \lambda$, where $a \geqq 0$, and λ is any tangent vector to S at q of length at most a. One then obtains an admissible cosmos, i.e., the foregoing assumptions are satisfied.

Spaces S of the indicated type have been completely classified by Tits (1957). Likewise classified are the four-dimensional Lorentzian manifolds (i.e., having a given pseudo-Riemannian structure whose fundamental form is of type (1, 3)) admitting certain types of transitivity, in an important work by Tits (1960). Physically speaking, the desiderata employed by Tits are of a qualitative relativistic nature. The physical process of observation, and the relation to symmetries defining the energy, etc., as considered here, supplement Tits' desiderata, and are materially restrictive. Thus, de Sitter space satisfies the cited qualitative relativistic desiderata, but admits no temporal translation group of the type earlier indicated (and thereby no natural definition of energy which results in a positive energy). Indeed, there are only three Lorentzian manifolds which satisfy Assumptions 1–5 on the Cosmos. The "universal space" $\tilde{M}$, consisting of the universal covering manifold of the conformal compactification $\bar{M}$ of Minkowski space M, has for its group of causality-preserving symmetries, one which is locally identical to $SO(2, 4)$ and so of dimension 15. Minkowski space can be regarded as an open dense submanifold of $\bar{M}$ which is covered infinitely often by $\tilde{M}$. Finally, the (two-fold) covering space $S^1 \times S^3$ of $\bar{M}$, consisting of the direct product of a circle and the surface of a sphere in four-dimensional space, contains an open submanifold M' whose causal symmetry group corresponds to the subgroup $SO(2, 3)$ of $SO(2, 4)$. Thus, $\tilde{M}$ is universal also in the sense that the other two cosmos are simply derivable from it, and their causal symmetry groups are essentially subgroups of that of $\tilde{M}$. The cosmos represented by M' has only locally, not globally, the property that the region of influence of compact regions in space are compact; and the theoretical redshift–distance relation is unaffected if M' is used in place of M below; it thus appears as a slightly complicated variant of $\tilde{M}$ and will not be further considered here. It should perhaps be mentioned, however, that the causal structures in M and M' may be defined by metrics admitting ten-parameter isometry groups, while that in $\tilde{M}$ admits, at most, a seven-parameter group. Since Maxwell's equations are well-defined and invariant under the full causal symmetry group on all of the manifolds, the isometry groups of special metrics play no apparent physical role in the analysis of photon propagation; and the deviation from isometry in the case of $\tilde{M}$ is of order R^{-1}, where R is describable as the radius of the universe (cf. below) and so surely unobservable, even if physically meaningful in a local macroscopic theory.

The causal cones in the Lorentzian manifold case are defined by equations of second order. This is natural from the standpoint of general relativity, but there appears otherwise to be no inherent observational or physical reason why the causal cones should be of this special type. There exist simple models for which they are not quadratic, but satisfy all of the assumptions except that of four-dimensionality. The models $R^1 \times SU(n)$ discussed in Chapter II admit a quite satisfactory unique notion of causality, and Gårding (1947) has given an effective treatment of analogues to the Maxwell and Dirac equations in closely related spaces. Of course, the groups $SU(n)$ admit invariant Riemannian metrics, which thereby determine causal orientations on $R^1 \times SU(n)$ in the manner earlier indicated; however, this Lorentzian structure is much less invariant than the non-Lorentzian one (for $n > 2$), and in particular, unlike the latter, there is in general no symmetry in the theory which will transform one given timelike direction into another (cf. Assumptions 5 and 6). It is not yet determined whether any such nonquadratic models exist in dimension 4, but it appears unlikely.

To indicate how such models fit into the present scheme, it is appropriate to generalize the example just given by permitting the space S to be essentially Finslerian, rather than Riemannian. More specifically, we assume that there is given in the tangent space to S at each point q a closed convex body $K(q)$, containing 0 in its interior. Let $n(\lambda)$ denote the corresponding norm function for tangent vectors, i.e., $n(\lambda)$ is the largest nonnegative number s such that $s\lambda$ is in $K(q)$. One may then define the causal structure in M in the same way, except that $n(\lambda)$ is used in place of the length of λ. Thus, given any three-dimensional isotransitive Finsler manifold, there is a corresponding cosmos. *Conversely*, every admissible cosmos arises in this way from such a manifold, as indicated in Chapter II.

3. Physical observers

From the standpoint of rudimentary causality and homogeneity considerations, all of the foregoing models for the Cosmos thus appear equally good. We might now further refine these considerations and obtain additional plausible physical restrictions. For example, there are some further features of Minkowski space which might reasonably be postulated:

Assumption 5 Any given timelike direction at a point p is tangential to the forward direction of some admissible observer.

That is, the Cosmos M can be split into time and space, $M = R^1 \times S$, in such a way that p is represented by the point $(0, u_0)$, u_0 in S, and that the given timelike direction is represented by $\partial/\partial t$, t being the component in R^1. This assumption corresponds to the intuitive idea that there is no preferred direction in space–time, of temporal evolution, from which to observe the universe.

In a related vein, it would also be reasonable to postulate:

Assumption 6 Two different observers at the same point see the Cosmos in causally compatible ways, i.e., the transformation between their respective maps of the Cosmos should be causality-preserving.

It would be interesting to explore the consequences of these further assumptions, but we shall only remark here that both Minkowski space and universal space satisfy both of these assumptions. Rather than proceed in an increasingly abstract and somewhat philosophical line, it seems preferable to analyze in physical terms the process of measurement by which these models may be differentiated, and to which they are relevant. The space–time geometry itself is not necessarily directly observed; no apparent departures from a Euclidean model have been found by classical measurements. Rather, the geometry influences the analysis of microscopic (notably, elementary particle) and ultramacroscopic (notably, extragalactic) phenomena. It therefore seems physically more appropriate to correlate the geometry with what is observed in these extreme-distance realms. We shall begin with an analysis of the concept of "observer," and especially that of "local observer."

Since this may appear somewhat lengthy, we first summarize the salient points. Briefly, it will be found that:

(1) Minkowski and universal space are locally identical *as* causal manifolds.

(2) However, the natural clocks and physical observers in these two models are *not* equivalent; this means that there are two essentially distinct types of local clocks.

(3) If processes run and/or are observed by these inequivalent clocks, the lack of synchronization will be unobservable for times of the order of 1 yr, if the clocks are instantaneously synchronous. The asynchronization for times up to the order of 10^7 yr increases approximately quadratically with the time, and attains an observable level, in the form of the alteration it produces in the apparent frequency of a freely propagated photon. The relative shift $\Delta\nu/\nu$ is frequency independent.

One analytically simple means to represent the relation between Minkowski and universal spaces M and $\tilde{M}$, and particularly their admissible observers, is to utilize the well-known formulation of Minkowski space as the set of all 2×2 Hermitian matrices, and the relation of these matrices to the unitary 2×2 group, denoted $U(2)$. If Minkowski space M is coordinatized in the usual way by time and space coordinates (t, x, y, z), we may map M onto the space $\mathbf{H}(2)$ of all 2×2 complex Hermitian matrices by the transformation

$$F: (t, x, y, z) \to \begin{bmatrix} t + x & y + iz \\ y - iz & t - x \end{bmatrix}.$$

The crucial point here is that this mapping preserves causality, if a notion of temporal precedence is introduced into $\mathbf{H}(2)$ by the definition: H is "before" H' if $H' - H$ is a positive semidefinite matrix. (Any matrix, or point of Minkowski space, is considered to be both before and after itself, to simplify the terminology.)

Now the universal space $\tilde{M} = R^1 \times S^3$ can be usefully related to the special unitary group $SU(2)$ of all 2×2 unitary matrices of unit determinant in the following way. The most general matrix of $SU(2)$ has the form

$$\begin{pmatrix} a + ib & c + id \\ -c + id & a - ib \end{pmatrix},$$

with $a^2 + b^2 + c^2 + d^2 = 1$, i.e., with (a, b, c, d) on the three-dimensional sphere in 4-space defined by the equation $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$. This is a one-to-one correspondence between S^3 and $SU(2)$, in which the unit matrix I corresponds to the point $(1, 0, 0, 0)$ on the sphere, and rotations of the sphere leaving this point fixed correspond to the transformations $U \to V^{-1}UV$, V being a fixed unitary matrix, on $SU(2)$. Note also that the usual invariant Riemannian metric on S^3 corresponds to the unique metric on the group $SU(2)$ which is invariant under both right and left translations on this group, i.e., the transformations $U \to VU$ and $U \to UV$, where V is a fixed element of the group $SU(2)$.

Locally, $\tilde{M} = R^1 \times S^3$ may be made to correspond to the 2×2 unitary group $U(2)$ by the mapping $(t, p) \to e^{it}u$, where $p \to u$ is the mapping from S^3 into $SU(2)$ just indicated. Sufficiently near to the point $t = 0, p = (1, 0, 0, 0)$, this mapping is one-to-one and smooth. A crucial point is that it is also causality-preserving, if a local notion of temporal precedence is introduced into $U(2)$ by taking as the set of future directions at the unit matrix I, all those that are represented by positive semidefinite Hermitian matrices; and at any other point defining the future directions by translating in the group $U(2)$ from I to the point in question. (Because of the invariance of the set of positive semidefinite matrices under the transformations $H \to UHU^{-1}$, U unitary, it is immaterial whether right or left translations are used.)

To set up a local causality-preserving transformation between Minkowski and universal space, it therefore suffices to set up such a mapping between $\mathbf{H}(2)$ and $U(2)$, which carries $H = 0$ into $U = I$. The simplest is the Cayley transform:

$$\Lambda: H \to \frac{2I + iH}{2I - iH}.$$

(For the proof that this is causality-preserving, see Chapter II; here 2 may be replaced by an arbitrary nonvanishing constant, but the present normalization will be convenient later.) These mappings are also conformal, in as much as a causality-preserving transformation on a pseudo-Riemannian causal manifold whose future cone is defined by the given metric is always conformal.

Consider how a given Lorentz transformation in Minkowski space M

appears from the standpoint of a locally equivalent observer on unispace $\tilde{M}$. Temporal evolution in M is

$$H \to H + sI.$$

Under the correspondence Λ, this transforms into a complex mapping in $\tilde{M}$ which mixes up the space and time components. The same is true of the space translations in a fixed direction,

$$H \to H + sK, \qquad K \text{ fixed in } \mathbf{H}(2), \quad \operatorname{tr} K = 0.$$

Thus time and space displacement in Minkowski space do not at all correspond via the causal mapping Λ to natural time and space displacements in $\tilde{M}$.

The natural space displacements in $\tilde{M}$ are given by the six-parameter group of rotations of the sphere S^3. Equivalently, this group consists of left and right translations on $SU(2)$, i.e., the transformations $V \to WVW'$, where V ranges over $SU(2)$ and W and W' are arbitrary fixed elements of $SU(2)$. Only the rotations, say those leaving the identity matrix I fixed, which are the transformations $V \to U^*VU$, where U is a fixed element of $SU(2)$, correspond precisely to conventional spatial displacements, i.e., the normal space rotations about the origin in R^3.

It is indeed impossible to set up between these spaces a conformal equivalence that does not mix space and time components in one space or the other. Actually, the space $\tilde{M}$ occurs in this discussion primarily as a means of exposing the central fact that Minkowski space admits two conceptually equally valid types of local physical observers, which are essentially distinct. Each observer sees space–time as split into space and time components in accordance with the underlying causal structure; each of them admits local temporal and spatial symmetries, acting on the time and space components separately, and being tantamount to the usual hypotheses of conservation of energy, linear momentum, and angular momentum; each has a unique notion of temporal duration and spatial distance (within scaling); each admits transformation to accelerated observers (i.e., there is no preferred strictly timelike direction). As far as general physical considerations go, there is no significant basis to prefer the one type of local observer to the other.

It might be argued that the conventional splitting into space and time components is "simpler"; it is "flat," while the other is "curved" (as regards space); it is traditional. On the other hand, there is no direct observational basis for asserting that the Cosmos is Minkowskian for very large times and distances. If indeed it should conform to the global extension of the unconventional splitting, the latter splitting would have the advantage of being much more symmetrical than the conventional Minkowskian one; it admits a 15-parameter group of admissible transformations on local frames of reference, while in the case of Minkowski space this group is only 11-dimensional. It will be seen that M is in a natural way, associated with any given observer, essentially contained in $\tilde{M}$, denumerably often, and one could argue thereby that $\tilde{M}$ may be more appropriate for the description of very long-range phenomena.

However this may be, it appears that a conclusive physical preference for one type of observer over the other can only be based on quantitative observation. A priori, both types could be valid, in the sense that different phenomena require different types of observers for their simple description. Indeed, it is instructive to compute how the frequency as measured by the one type of observer appears relative to one of the other type, so chosen as to be locally (more precisely, infinitesimally) at rest relative to the first one.

In order to make this computation, it is convenient to use the formalism of conformal space developed by Veblen and others. This formalism is fundamental in the present theory and we pause in our physical discussion to develop simple aspects of it in a somewhat rounded way.

4. Conformal geometry and the unitary formalism

It will be just as easy to take an $(n + 1)$-dimensional Minkowski space M with coordinates $t, x_1, \ldots, x_n$ (in fact, the case $n = 1$ will be relevant and illuminating). Our first step is to define a closed (compact) space $\bar{M}$, the so-called conformal space, in which M is imbedded naturally. Roughly speaking, $\bar{M}$ is obtained from M by adding a light cone at infinity. This does not disturb the underlying symmetry of the space; the Lorentz group and scale transformations continue to act conformally on $\bar{M}$. Indeed, one gains in symmetry, in that conformal inversion is a nonsingular operation on $\bar{M}$, and together with the Lorentz and scale transformations, generates a 15-dimensional Lie group which acts in an entirely regular and conformal manner on $\bar{M}$. (This is only a preliminary step; the space $\bar{M}$ is acausal; however, the remedy will be conceptually simple, and $\bar{M}$ will remain as a fundamental object for many computations.)

Specifically, $\bar{M}$ is the space of all (projective) conformal spheres in M of zero radius, endowed with the natural Lorentzian structure (cf. Chapter II). Analytically, let

$$X = (t, x_1, \ldots, x_n), \qquad t = x_0, \qquad X^2 = t^2 - \mathbf{x}^2$$

$$\xi_{-1} = 1 - \tfrac{1}{4}X^2, \qquad \xi_{n+1} = 1 + \tfrac{1}{4}X^2, \qquad \xi_j = x_j .$$

Set $\Xi = (\xi_{-1}, \xi_0, \ldots, \xi_n, \xi_{n+1})$, $\Xi^2 = \xi_{-1}^2 + \xi_0^2 - \xi_1^2 - \cdots - \xi_n^2 - \xi_{n+1}^2$ for arbitrary ξ_j; let Λ denote the transformation Λ: $X \rightarrow \Xi$ (given in terms of X

by the foregoing equations); then $\Xi^2 = 0$ if $\Xi = \Lambda(X)$ for some X. Set $\boldsymbol{\varepsilon} = (\varepsilon_{-1}, \varepsilon_0, \ldots, \varepsilon_{n+1}) = (1, 1, -1, \ldots, -1)$, so that $\Xi^2 = \sum \varepsilon_i \xi_i^2$. Let W denote the $(n+3)$-dimensional vector space of all Ξ (with arbitrary real values for the ξ_j). Let $\tilde{W}$ denote the $(n+2)$-dimensional projective space of rays in W (i.e., vectors Ξ when proportional vectors are identified). Let Q denote the quadric in W defined by the equation $\Xi^2 = 0$, and let $\tilde{Q}$ denote the corresponding quadric in $\tilde{W}$. Let $\tilde{\lambda}$ denote the map

$$\tilde{\lambda}\colon X \to \tilde{\Xi},$$

where $\tilde{\Xi}$ denotes the ray through Ξ. Then our earlier observation is to the effect:

$$\tilde{\lambda} \quad \text{is a one-to-one mapping of} \quad M \text{ into } \tilde{Q}.$$

More specifically, one can recover the $t, x_1, \ldots, x_n$ from a point of $\tilde{Q}$ (other than the exceptional points which do not correspond to points of M) as follows. For any nonzero Ξ in W, let $\mathbf{u}$ denote the vector $(u_{-1}, u_0, u_1, \ldots, u_{n+1})$, with $u_j = \xi_j/(\xi_{-1}^2 + \xi_0^2)^{1/2}$; we set $\mathbf{u} = u(\Xi)$, and note that u is the same for all vectors on the same ray as Ξ; the definition $\tilde{u}(\tilde{\Xi}) = u(\Xi)$ is therefore unique, and $\tilde{\Xi} \to \tilde{u}$ is a well-defined mapping from Q into an $(n+2)$-dimensional space R^{n+2}. It is clear from the definition that

$$u_{-1}^2 + u_0^2 = 1 = u_1^2 + \cdots + u_{n+1}^2.$$

Thus this mapping from $\tilde{Q}$ into R^{n+2}, to be denoted Γ, actually maps $\tilde{Q}$ into the direct product of a circle S^1, coordinatized by u_{-1} and u_0, and a sphere S^n, coordinatized by $u_1, \ldots, u_{n+1}$. Conversely, any point of this direct product $S^1 \times S^n$, $\mathbf{u} = (u_{-1}, \ldots, u_{n+1})$, corresponds to a point of $\tilde{Q}$ via the mapping $\gamma\colon \mathbf{u} \to \tilde{\Xi}$, where $\Xi = (u_{-1}, \ldots, u_{n+1})$. The mapping γ is precisely two-to-one, for both $\pm\mathbf{u}$ correspond to the same point of $\tilde{Q}$; we say γ is a *twofold covering* of $\tilde{Q}$ by $S^1 \times S^n$.

Now suppose one is given the coordinates $(u_{-1}, \ldots, u_{n+1})$ of one of the two points of $S^1 \times S^n$ corresponding to a given point X in M. Then the u_j are given by the equations

$$\begin{aligned} u_{-1} &= k^{-1}(1 - \tfrac{1}{4}X^2), & u_{n+1} &= k^{-1}(1 + \tfrac{1}{4}X^2), \\ u_j &= k^{-1}x_j, & k &= \pm[(1 - \tfrac{1}{4}X^2)^2 + t^2]^{1/2}. \end{aligned}$$

The mapping $X \to \mathbf{u}$ is one-to-two, but is locally one-to-one. In the vicinity of the point $X = 0$, then, with k chosen to be positive (so that the point corresponding to $X = 0$ is $\mathbf{u} = (1, 0; 0, \ldots, 0, 1)$, we may recover X by the equations

$$x_j = 2u_j(u_{-1} + u_{n+1})^{-1} = 2\xi_j(\xi_{-1} + \xi_{n+1})^{-1}.$$

Now $S^1 \times S^3$ has a natural Lorentzian structure, defined by the form $du_{-1}^2 + du_0^2 - du_1^2 - du_2^2 - \cdots - du_{n+1}^2$; or, introducing the new time $\tau = \tan^{-1} u_0/u_{-1}$ near the point indicated, $d\tau^2 - ds^2$, where ds denotes the usual element of length on the sphere S^n. This Lorentzian structure is not invariant under the action of the Lorentz group in M, after transference to $S^1 \times S^n$ by the foregoing correspondences; however, it remains conformally invariant. More generally, any of the classical conformal transformations on M (which apart from the Lorentz transformations and the scale transformations are not everywhere defined on M, and develop singularities if an attempt is made to extend their domains of definition in M) act without singularities on $S^1 \times S^3$ and also on Q; and are conformal on these spaces, in each of which M is in a certain sense contained.

Moreover, the space $S^1 \times S^3$ admits a local notion of causality, according to which the one of the two convex cones defined by the Lorentzian form $d\tau^2 - ds^2$, which is in the direction of increasing τ, defines the future direction in space–time. Of course, it is always possible in a Lorentzian manifold to introduce a local notion of causality by a choice of one of the two cones near a point as the future; but it is not always possible, as it is here, to do so in a manner which varies continuously from point to point, throughout the global space–time manifold. Despite the latter feature, however, $S^1 \times S^3$ is not causal in the large; indeed, the curve

$$u_{-1} = \cos\tau, \qquad u_0 = \sin\tau, \qquad u_i = c_i \quad (\text{const}), \qquad 1 \leqq i \leqq n+1,$$

is in the forward timelike direction (as a function of τ), but is cyclic.

Thus, locally at any point in Minkowski space, we have the two distinct type of coordinates: t, x, y, z on the one hand, and the spherical coordinates u_j $(-1 \leqq j \leqq 4)$ on the other. There are many other parametrizations, of course; what is distinctive about the foregoing ones (or variants thereof) is their extensive covariance. To make this explicit, consider the actions of fundamental symmetries in terms of these parametrizations.

On the $(n+3)$-dimensional vector space Ξ, let L'_{ij} denote the vector field

$$L'_{ij} = \varepsilon_i \xi_i D_j - \varepsilon_j \xi_j D_i, \qquad \text{where} \quad D_i = \partial/\partial \xi_i.$$

These generate the group $O(n+1, 2)$ of linear transformations in Ξ leaving invariant the quadratic form Ξ^2. Consequently, they determine corresponding vector fields L_{ij} on the projective quadric $\tilde{Q}$. As shown in Chapter II, the conformal structure on $\tilde{Q}$ is invariant under the transformations generated by the L_{ij}, i.e., they are all infinitesimally conformal. Conversely, all globally defined infinitesimally conformal transformations on $\tilde{Q}$ are linear combinations of the L_{ij}. It is important to note that the full group of symmetries of Minkowski space, including all Lorentz transformations and changes of scale, extend from the space M to the larger space $\tilde{Q}$, and to its twofold covering $S^1 \times S^3$, remaining all the while causality-preserving (in particular, conformal); and no singularities are involved in this extended action.

It is useful and of interest to calculate the form of the generators L_{ij} when expressed in terms of the usual Minkowski coordinates; and on the other hand, to calculate the form of the Lorentz transformations in terms of the ξ_i or spherical coordinates on $S^1 \times S^3$. In addition, the natural symmetries of $S^1 \times S^3$ relative to the given splitting (i.e., transformations acting on S^1 alone, respectively acting on S^3 alone) will be of interest. These matters are more readily achieved and better understood if further parametrizations of space–time closely related to the given ones are developed. We shall therefore treat parametrizations in terms of $U(2)$ and the space of conformal null spheres in Minkowski space.

In Minkowski space M, a *Lorentz sphere* is a locus in M consisting of all vectors X such that

$$(X - X_0)^2 = \text{const} \qquad (X_0 \text{ a fixed vector});$$

if the constant $= 0$, one has a *null sphere.* Now a Lorentz sphere has an equation of the form

$$aX^2 - 2B \cdot X + c = 0$$

with $a \neq 0$ and B and c arbitrary; this sphere is a null sphere if and only if $ac - B^2 = 0$ (here B is a vector in Minkowski space). Evidently, there is a one-to-one correspondence between points of M and null spheres (a, B, c), where (a, B, c) is considered as a point of projective $(n + 2)$-dimensional space, and (a, B, c) is restricted to lie on the quadric $ac - B^2 = 0$. (More specifically, (a, B, c) and (a', B', c') are identified if there exists a nonzero constant λ such that $\lambda(a, B, c) = (a', B', c')$.)

In addition to the normal null spheres just indicated, there exist "ideal" null spheres given in the same way except that $a = 0$. That is, an ideal null sphere is a point (a, B, c) in projective $(n + 2)$ space, lying on the quadric $ac - B^2 = 0$, but not corresponding to a normal Lorentz null sphere. The totality of all normal and ideal null spheres is then in one-to-one correspondence with the indicated quadric. It is convenient on occasion to take this quadric in the alternative form

$$ac - B^2 = \xi_{-1}^2 + \xi_0^2 - \xi_1^2 - \cdots - \xi_{n+1}^2,$$

as is possible in view of the signature of the quadratic form $ac - B^2$ for suitable real linear combinations ξ_j of a, B, c; these will be chosen to lead to new coordinates that are infinitesimally synchronous with the Minkowski coordinates.

Specifically, we shall choose

$$a = \tfrac{1}{2}(\xi_{-1} + \xi_{n+1}), \qquad c = 2(-\xi_{-1} + \xi_{n+1}), \qquad B = (\xi_0, \ldots, \xi_n);$$

equivalently, the ξ_j may be expressed in terms of a, $B = (b_1, \ldots, b_n)$, and c as

$$\xi_{-1} = a - \tfrac{1}{4}c, \qquad \xi_{n+1} = a + \tfrac{1}{4}c, \qquad \xi_j = b_j \qquad \text{for} \quad 0 \leqq j \leqq n.$$

Proceeding now as earlier, we introduce the sphere coordinates u_j by the equation

$$(u_{-1}, u_0, u_1, \ldots, u_{n+1}) = (\xi_{-1}^2 + \xi_0^2)^{-1/2}(\xi_{-1}, \xi_0, \ldots, \xi_{n+1}).$$

Here we could use either sign on the indicated square root, but normalize to make the point $X = 0$ correspond to the point $(1, 0; 0, \ldots, 0, 1)$ in $S^1 \times S^n$. This gives the earlier indicated equations (where $X = (x_0, \ldots, x_n)$):

$$\begin{aligned} u_{-1} &= p(1 - \tfrac{1}{4}X^2), \\ u_j &= px_j \qquad (0 \leqq j \leqq n), \\ u_{n+1} &= p(1 + \tfrac{1}{4}X^2), \end{aligned}$$

where

$$p = [(1 - \tfrac{1}{4}X^2)^2 + x_0^2]^{-1/2}.$$

We could now introduce angular coordinates on the space $S^1 \times S^n$, but will define only the most important two, τ and ρ, by the equations $u_{-1} = \cos\tau$, $u_0 = \sin\tau$; $\tau = \tan^{-1}(u_0/u_{-1})$, which leave τ undefined modulo 2π; and

$$\cos\rho = u_{n+1}, \qquad x_1^2 + \cdots + x_n^2 = \frac{4\sin^2\rho}{(\cos\tau + \cos\rho)^2}, \qquad 0 \leqq \rho \leqq \pi.$$

In the particular case $n = 3$ which is physically crucial, it is possible to give a useful representation of the present space–time splitting in terms of the unitary group. Recall the representation

$$H = \begin{pmatrix} t + x & y + iz \\ y - iz & -x + t \end{pmatrix}$$

for the Hermitian matrix X corresponding to the given point (t, x, y, z) of M, and define

$$U = (1 + \tfrac{1}{2}iH)/(1 - \tfrac{1}{2}iH).$$

Then U is in the unitary group $U(2)$, and the point 0 in M corresponds to the unit matrix I in U. More specifically, U has the form

$$U = (1 - it - \tfrac{1}{4}d)^{-1}\begin{pmatrix} 1 + ix + \tfrac{1}{4}d & i(y + iz) \\ i(y - iz) & 1 - ix + \tfrac{1}{4}d \end{pmatrix},$$

where $d = t^2 - x^2 - y^2 - z^2$. In terms of U, τ may be recovered in the natural way as

$$\tau = \frac{1}{2i}\log \det U = \frac{1}{2}\arg \det U,$$

which is equivalent to the equation

$$\tau = \tan^{-1}[t/(1 - \tfrac{1}{4}d)]$$

given earlier. In addition, the u_j ($j = 1, 2, 3$) are essentially tangential near I to the x_j; more exactly,

$$u_j = x_j/[(1 - \tfrac{1}{4}d)^2 + t^2]^{-1/2} \qquad (j = 1, 2, 3)$$

with U in the form

$$U = (u_0 + iu_{-1})\begin{pmatrix} u_4 + iu_1 & -u_3 + iu_2 \\ u_3 - iu_2 & u_4 - iu_1 \end{pmatrix}.$$

These u_j are identical to the spherical coordinates previously indicated, for $S^1 \times S^3$. It follows that near the origin in Minkowski space,

$$u_j = x_j + \text{terms of third or higher order} \qquad (j = 0, 1, 2, 3).$$

We now turn to the consideration of how the various basic symmetries act or are represented in terms of these parametrizations—these are notably the L_{jk}, Lorentz transformations, conformal inversion, etc. To indicate the method employed, consider the question of the action on $U(2)$ of conventional temporal translation $t \to t + s$ in Minkowski space. We note first that the indicated mapping of M into $U(2)$, while it does not cover all of $U(2)$, omits just those elements of $U(2)$ corresponding to ideal null spheres, or in terms of $U(2)$ itself, precisely those unitaries U for which -1 is an eigenvalue. As far as the correspondence between M and $U(2)$ is concerned, this is a well-known fact about the Cayley transform $H \to (1 + \tfrac{1}{2}iH) \times (1 - \tfrac{1}{2}iH)^{-1}$. That the omitted points in $U(2)$ correspond in one-to-one fashion to the ideal points in $\tilde{Q}$ follows from the precisely two-to-one representation of points in $\tilde{Q}$ in terms of the u_j, in terms of which the general element in $U(2)$ may also be represented in a matching two-to-one fashion.

5. Causal symmetries and the energy

We turn now to the consideration of the symmetries acting on the two cosmos.† We aim to give explicit expressions for the relevant symmetries, in

† We use the term *unispace* (short for universal covering space) for the universal space with the foregoing physical interpretation. Thus unispace is conformally an infinite-sheeted covering of Minkowski space augmented by a light cone at infinity. Similarly, *unitime* refers to the natural time τ in this space.

terms of the various parametrizations. These are useful in computations, and in clarifying special relativistic formalism. Before going into the matter of explicit formulas, we enumerate the key qualitative results regarding symmetries.

(1) Every causal automorphism of Minkowski space M, i.e., every Lorentz transformation or scale transformation, or product of such, extends uniquely and without singularities to a corresponding transformation on conformal space $\bar{M}$. That is, the restricted conformal group does indeed act on $\bar{M}$, where "restricted conformal group" is defined as the 11-parameter group of transformations on Minkowski space consisting of products of Lorentz with scale transformations. The reversal operations (time/space/total) also continue to act, on all of $\bar{M}$, without singularities.

(2) In addition, conformal inversion, formally the transformation $Q\colon X \to 4X/X^2$, although singular on M, becomes an analytic everywhere-defined transformation on all of $\bar{M}$. Together with the action on $\bar{M}$ of the restricted conformal group, Q generates a 15-parameter Lie group of transformations on $\bar{M}$; this is the action of $O(n, 2)$ earlier derived.

(3) Space rotations around an observer in Minkowski space correspond to space rotations in $S^1 \times S^n$ around the corresponding point—i.e., the space rotations are essentially the same in the two models. However, the temporal generators (i.e., energies) are basically different. The unispace generator is *strictly greater than the Minkowski energy*, and differs from it essentially by εQEQ, where ε is a small constant and E is the Minkowski energy operator.

Consider now the two temporal evolution groups. The transformation $T\colon t \to t + s$ in Minkowski space can be represented as a transformation either on $S^1 \times S^n$, or in the case $n = 3$, as a transformation on $U(2)$. Taking the latter representation first, with the correspondence

$$H = \begin{pmatrix} t + x & y + iz \\ y - iz & t - x \end{pmatrix},$$

T carries $H \to H + sI$. The corresponding element U in $U(2)$ to H is

$$U = \frac{1 + \frac{1}{2}iH}{1 - \frac{1}{2}iH}; \qquad \text{whence} \quad H = -2i\frac{U - I}{U + I}.$$

Representing T as a transformation on $U(2)$, it follows by a simple computation that its action is

$$U \to \frac{(1 + \frac{1}{4}is)U + \frac{1}{4}is}{-\frac{1}{4}isU + (1 - \frac{1}{4}is)}.$$

This is a transformation in the standard action of the group $SU(2, 2)$ on $U(2)$. Any element of $SU(2, 2)$ can be represented in natural fashion in the form $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ where A, B, C, D are suitable 2×2 matrices; in these terms the standard action is

$$U \to (AU + B)(CU + D)^{-1}.$$

Denoting the Lie algebra of this group as $su(2, 2)$, this can be identified with the 4×4 skew Hermitian matrices relative to the Hermitian form $z_1 \bar{z}'_1 + z_2 \bar{z}'_2 - z_3 \bar{z}'_3 - z_4 \bar{z}'_4$. With this identification, the generator of Minkowski temporal translation is then represented by the matrix (where I denotes the 2×2 identity matrix)

$$\frac{d}{dt} \to \frac{1}{4}\begin{pmatrix} I & I \\ -I & -I \end{pmatrix}.$$

On the other hand, the unitime group is the group $U \to e^{is}U$, having the generator

$$\frac{d}{dt'} \to \frac{1}{2}\begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}.$$

These one-parameter groups are not only distinct, but nonconjugate, within the group $SU(2, 2)$, as shown by the difference between the spectra of the two generators, which are respectively continuous and discrete, in relevant representations (cf. Segal, 1967a). This is evident also from the fact that one of the matrices is singular and the other nonsingular. However, *their local actions at a fixed point of space are very close, differing only by terms of order s^3 and higher, where s is the elapsed time.* To show this, it is by homogeneity no essential loss of generality to choose the fixed point of space as the origin in Minkowski space, or as the unit I in $U(2)$. (The correspondence between M and $U(2)$ depends on the point of reference; however, that may be arbitrarily designated as the origin in Minkowski space, in view of the Lorentz invariance of M, $\tilde{M}$, and their relation.) The unitime group sends $I \to e^{is}I$; the special relativistic temporal group sends

$$I \to \frac{1 + \frac{1}{2}is}{1 - \frac{1}{2}is} I;$$

these differ by $O(s^3)$.

It is thus evident that the special relativistic energy appears relatively complicated in the $U(2)$ formalism; this is also the case in the $S^1 \times S^3$ formalism. On the other hand, the unienergy appears relatively complicated in the Minkowski space formalism. To compute the action of unitemporal evolution in the Minkowski picture, we must transfer the action $U \to e^{is}U$ to an action on the corresponding H, and hence to an action on (t, x, y, z). Now, unitemporal evolution sends

$$U \to e^{is}U.$$

Since

$$H = -2i\frac{U - I}{U + I}, \qquad H \to -2i\frac{e^{is}U - 1}{e^{is}U + 1};$$

now, expressing U in terms of H again, it results that

$$H \to H' = \frac{H + 2\tan(s/2)}{1 - \frac{1}{2}H\tan(s/2)}.$$

We now relate the Minkowski energy to conformal inversion and the unienergy, and compute the latter in this way. We define conformal inversion, to be designated Q, as the transformation

$$Q: X \to 4X/X^2;$$

this is not everywhere defined on M, but extends naturally to an everywhere-defined transformation on the larger space $\tilde{M}$, as follows. In terms of the coordinates (a, B, c), Q is the map $(a, B, c) \to (c, 4B, 16a)$, as follows on substitution in the equation for a Lorentz sphere. It results that on the ξ_j, Q acts as

$$Q: \xi'_{-1} = -\xi_{-1}; \qquad \xi'_j = \xi'_j \qquad (j \geqq 0),$$

and on the u_j,

$$u_{-1} \to -u_{-1}; \qquad u_j \to u_j \qquad (j \geqq 0).$$

Thus, in unispace, conformal inversion affects only the time, not the space component. It is only the identification of antipodal points in $S^1 \times S^n$ which gives the transformation its spatial character on $\tilde{M}$ or $U(2)$. For example, the transformation

$$(1, 0; 0, \ldots, 1) \to (-1, 0; 0, \ldots, 1)$$

effected by conformal inversion on the origin (observational location) is purely temporal; but $(-1, 0; 0, \ldots, 1)$ represents the same point of conformal space as does its multiple by -1, i.e., $(1, 0; 0, \ldots, -1)$, which can be considered to be the image under purely spatial inversion of the origin.

We should distinguish between proper unispace and the locally identical but acausal space $S^1 \times S^n$, which becomes conformal space upon identifying antipodal points. In contrast, the former space is $R^1 \times S^n$ and is mapped upon $S^1 \times S^n$ by the transformation $(t, u) \to (e^{it}, u)$, and thence upon conformal space, in which Minkowski space is properly contained. It is important to note that Q extends naturally not only from an improper transformation in M to a proper one in $\tilde{M}$, but also corresponds to proper transformations in both $S^1 \times S^n$ and $R^1 \times S^n$. In the case of $S^1 \times S^n$, the transformation is $Q^{(2)}$: $(\lambda, u) \to (-\lambda^{-1}, u)$ where λ is a complex number of absolute value 1 representing the S^1 component. The temporal character of conformal inversion as a transformation on $R^1 \times S^n$ is particularly clear: it is $Q^{(\infty)}$: $(t, u) \to (\pi - t, u)$. Here we have adopted the notational device of adding a superscript to designate the space, whether conformal space, its twofold covering space $S^1 \times S^n$, or its ∞-fold covering $R^1 \times S^n$. (There are similar coverings for each subgroup of the center Z of the group $\widetilde{SO}(2, 4)/Z_2$ of all causality-preserving transformations on $\tilde{M}$; Z is an infinite cyclic group which is generated by the transformation $(t, u) \to (t + \pi, -u)$. These additional coverings may be relevant to elementary particle considerations, but not directly to the astrophysical considerations of present concern. Hence they will be ignored in the following.)

It thus appears that conformal inversion differs from time reversal in unispace by the transformation on $R^1 \times S^n$, $(t, u) \to (t + \pi, u)$. The latter transformation is evidently contained in the one-parameter group $t \to t + s$, and so is continuously connected to the identity transformation. Since time reversal is represented by an antiunitary transformation in any positive energy particle model, it follows that *conformal inversion is represented by an antiunitary operator in any positive-energy particle model.*

In the case $n = 3$, it follows from the correspondence between $S^1 \times S^3$ and $U(2)$ that conformal inversion acts as follows on $U(2)$:

$$Q: U \to -\frac{U}{\det(U)}.$$

Let us now consider the transformation properties of the energies, Minkowski and universal, under Q. It is evident that the Minkowski temporal evolution generator $\partial/\partial t$ does not commute with Q. A straightforward computation gives

$$Q\frac{\partial}{\partial t}Q = -\frac{1}{4}(t^2 + x^2 + y^2 + z^2)\frac{\partial}{\partial t} - \frac{t}{2}\left(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}\right).$$

On the other hand, if τ denotes time in unispace, then $\partial/\partial\tau$, the generator of unispace temporal evolution, is carried by Q into $-\partial/\partial\tau$. Indeed, $\partial/\partial\tau = u_{-1}(\partial/\partial u_0) - u_0(\partial/\partial u_{-1})$; on sending $u_{-1} \to -u_{-1}$ and $u_0 \to u_0$, this is evidently reversed. This shows explicitly that in any particle model in which the unienergy is positive, conformal inversion must be represented by an antiunitary operator.

Let us now compute how $\partial/\partial\tau$ appears in terms of Minkowski coordinates t, x, y, z. To do this, note that since $\exp(a\, \partial/\partial\tau)$ sends $U \to e^{ia}U$, for $U \in U(2)$, the corresponding action on H is as earlier computed,

$$H \to \frac{H + 2\tan(a/2)}{1 - \frac{1}{2}H\tan(a/2)} = H + aI + (\tfrac{1}{4}a)H^2 + O(s^2).$$

Thus, $\partial/\partial\tau$ carries H into $I + \frac{1}{4}H^2$. Recalling now that

$$H = \begin{pmatrix} t + x & y + iz \\ y - iz & t - x \end{pmatrix},$$

it follows that

$$\frac{\partial}{\partial\tau} = \frac{\partial}{\partial t} + \frac{1}{4}(t^2 + x^2 + y^2 + z^2)\frac{\partial}{\partial t} + \frac{t}{2}\left(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}\right).$$

Comparing the latter equation with that for $Q(\partial/\partial t)Q$, it results that

$$\frac{\partial}{\partial\tau} = \frac{\partial}{\partial t} - Q\frac{\partial}{\partial t}Q.$$

Corollary In any particle or field model that is (a) invariant under the group $O(n, 2)$ (or any locally identical group), (b) such that the special relativistic energy is positive, the unienergy is also positive, and exceeds (in all states) the special relativistic energy.

Proof Let U be any unitary–antiunitary representation in Hilbert space of the group $O(n, 2)$, and conformal inversion Q, which in particular is represented by an antiunitary operator Q, as earlier noted. By the foregoing equation,

$$U\left(\frac{\partial}{\partial\tau}\right) = U\left(\frac{\partial}{\partial t}\right) - U\left(Q\frac{\partial}{\partial t}Q\right),$$

where we denote also by U the natural extension of U to infinitesimal operators of the group. Now

$$U\left(Q\frac{\partial}{\partial t}Q\right) = U(Q)U\left(\frac{\partial}{\partial t}\right)U(Q).$$

The assumption that the special relativistic energy is positive means that $i^{-1}U(\partial/\partial t)$ is a positive (as well as automatically Hermitian) operator. Since $U(Q)$ is antiunitary, it carries any one-parameter unitary group with positive generator into one with a negative generator; i.e., $i^{-1}U(Q)U(\partial/\partial t)U(Q)$ is a negative Hermitian operator. Thus the unienergy $-iU(\partial/\partial\tau)$ is the sum of the relativistic energy and the positive operator $iU(Q\ \partial/\partial t\ Q)$.

As a further example, let us compute some of the generators L_{ij} of the conformal group in terms of space–time coordinates x, y, z, t. The general procedure is as follows. For $k = 0, 1, 2, 3$,

$$L_{ij}x_k = L_{ij}\left(\frac{2\xi_k}{\xi_{-1} + \xi_4}\right) = 2(\xi_{-1} + \xi_4)^{-1}L_{ij}\xi_k - 2\xi_k(\xi_{-1} + \xi_4)^{-2}L_{ij}(\xi_{-1}),$$

whence in terms of the x_k,

$$L_{ij} = 2\sum_k [(\xi_{-1} + \xi_4)^{-1}L_{ij}\xi_k - (\xi_{-1} + \xi_4)^{-2}\xi_k L_{ij}(\xi_{-1} + \xi_4)]\frac{\partial}{\partial x_k}.$$

Evidently, $L_{ij}\xi_k = \varepsilon_i\,\delta_{jk}\,\xi_{-i}\varepsilon_j\,\delta_{ik}\,\xi_j$. Thus, e.g., setting $e = t^2 + x^2 + y^2 + z^2$,

$$L_{-1,0} = \left(1 + \frac{e}{4}\right)\frac{\partial}{\partial t} + \frac{t}{2}\left(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}\right),$$

$$L_{0,4} = \left(1 - \frac{e}{4}\right)\frac{\partial}{\partial t} - \frac{t}{2}\left(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}\right),$$

$$L_{-1,4} = -\left(t\frac{\partial}{\partial t} + x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}\right).$$

It should now begin to be visible how a theoretical analysis could be conducted by an observer, based on either the curved (unispace, $R^1 \times S^3$) or flat (Minkowski, $R^1 \times R^3$) local decompositions of space–time into space and time components. In both cases, all the fundamental physical laws on which conventional reduction and analysis of observation are based remain valid. Specifically, in either case:

(a) there is a proper global notion of causality in the Cosmos; and locally the notions are identical in the two models;
(b) conservation of energy, angular momentum, etc. are valid;
(c) Lorentz invariance holds; given any two future directions of a point, corresponding to relatively accelerated observers, they are related by a global causality-preserving transformation;
(d) the energy is positive;
(e) there are essentially unique measures of temporal duration and of spatial distance, which are invariant under the respective underlying symmetry groups (unicity within a scale factor);
(f) there is finite propagation velocity for the causal structure relative to the time and distance measures in (e).

Physically, only one of these analyses can be globally valid, however. For example, if global conservation of energy is valid in one analysis, it cannot be valid in the other, for the respective energy operators do not commute. At first glance, it might appear that empirical confirmation of special relativity precluded the empirical validity of the curved formulation of the local space–time splitting. It develops, however, that as regards direct measurements the local unispatial analysis differs negligibly from the special relativistic analysis; it is only at extreme distances that significant differences emerge. This arises basically from two circumstances: (a) the unispatial space–time splitting is not only tangent to the special relativistic space–time splitting at the observer's location, but has the remarkable feature of agreeing with it within terms of third or higher order; (b) the basic distance scale, set in the unispace theory by the radius of the universe, is such that the times and distances involved in classical local measurements are ultramicroscopic.

In order to determine the distance scale, we must evaluate in conventional units the "radius of the universe" R, i.e., the radius of the S^3 component of unispace. At the present time, this can be deduced only from redshift measurements; this is natural since no other measurements are known to involve very great distances. It will be seen that the two time scales differ only by at most ~ 1 part in 10^{19} in the course of a year (or less), and similarly for distance scales (scaled by the velocity of light); the difference is thus well beyond the limits of present experimental capabilities. It is only by indirect measurements at extreme distances that the difference between the two models of an observer is empirically perceivable.

6. The redshift

Let us then provisionally adopt the unispace cosmos, and seek to analyze free propagation over very long times, and its effect on the measurement of frequency of light. The determination of the wavelength of light is very much of a local matter in practice. Conventionally the frequency is represented by the operator $i^{-1}\,\partial/\partial t$. We now have at hand an alternative possibility $i^{-1}\,\partial/\partial\tau$, where τ is the time in unispace. In the absence of any observed phenomenon such as the redshift, it might perhaps seem equally natural to represent the frequency by this alternative; there would, however, then be no apparent means of determining the distance scale, i.e., of measuring τ in natural units. In addition, the whole procedure of local measurement in the vicinity of a fixed observer is based on flat geometry. These considerations give a certain preferential basis for the flat energy $i^{-1}\,\partial/\partial t$, independently of the results of observational extragalactic astronomy; but what is really strongly indicative, indeed conclusive, is the fact of the observed redshift, which could not exist in the unispace cosmos if the alternative representation for the observed frequency were valid. We are led thereby to postulate that anthropomorphically possible local measurements are represented theoretically by the *flat* rather than *curved* dynamical variables; while on the other hand, the "true" nonanthropomorphic dynamics and analysis is curved (in the fashion appropriate to the unispace cosmos) rather than flat. That is, we measure the flat dynamical variables; but the Universe in the large runs on the curved basis, which agrees only instantaneously with the flat one. This postulate is provisional, pending the derivation of an effective treatment of redshift laws, etc., from it. We now begin this treatment, which will be found to agree with observation.

Among the dynamical variables that are chiefly involved in measurements are the space and time coordinates: the energy, angular momentum, and other particle quantum numbers. Since there is no direct means of observation of extragalactic distances, and the effect on galactic distances and coordinates is negligible, there is no apparent present possibility of distinguishing the theory by measurements of the coordinates. The angular momentum is the same in both the flat and curved cosmos, as it develops. However, the most basic of the dynamical variables, the energy, is affected in an observable fashion.

Let H_0 denote the dynamical variable $-i(\partial/\partial t)$ which has been postulated to represent theoretically the result of a local measurement of frequency. We are particularly interested in the case of light, its frequency being measured by the usual optical methods. According to the unispace theory, this dynamical variable is not the true energy, but only appears to be so infinitesimally at the location of an observer. (However, there is a scale factor between the two energies, and if synchronous at one location the scale factor will differ from unity at other locations.) The special relativistic energy is, therefore, not conserved, just as the energy relative to one Lorentz frame is not conserved relative to the temporal development in another frame. In the latter case, the difference in energy is relatively gross, being for small times s of the order of const $\times$ s, and so should be readily observable. In the present case, however, the unispatial frame is tangent to the flat frame, and defines an identical Lorentz frame at the observer's location; it is only the nonlinearity of the relation between the two cosmos and their respective groups of temporal displacement which causes a discrepancy in the energies. As a result, the extent of nonconservation of H_0 is proportional to s^2, rather than to s, for small values of s, within terms of higher order.

In order to obtain an exact expression, it is necessary to compute explicitly the dynamical variable representing the frequency after passage of a time s, i.e., the operator $H_0(s) = e^{-isH} H_0 e^{isH}$, where H is the true (conserved) energy, given in the unispace cosmos as $-i(\partial/\partial\tau) = -iL_{-1,0}$. Thus

$$H_0(s) = e^{-sL_{-1,0}} \frac{1}{2i}(L_{-1,0} + L_{0,4})e^{isL_{-1,0}}$$

$$= \frac{1}{2i} L_{-1,0} + \frac{1}{2i} e^{-sL_{-1,0}} L_{0,4} e^{sL_{-1,0}}.$$

Now $L_{-1,0}$, $L_{0,4}$, and $L_{-1,4}$ generate a three-dimensional subgroup of $O(4, 2)$; in particular

$$[L_{-1,0}, L_{0,4}] = L_{-1,4}, \qquad [L_{-1,0}, L_{-1,4}] = -L_{0,4}.$$

It follows that

$$e^{-sL_{-1,0}}L_{0,4}e^{sL_{-1,0}} = A(s)L_{-1,4} + B(s)L_{0,4},$$

where $A(0) = 0$, $B(0) = 1$. To evaluate $A(s)$ and $B(s)$, we take first and second derivatives in the foregoing equation. It results that

$$-e^{-sL_{-1,0}}[L_{-1,0}, L_{0,4}]e^{sL_{-1,0}} = A'(s)L_{-1,4} + B'(s)L_{0,4};$$

evaluating the commutator, it follows that

$$-e^{-sL_{-1,0}}L_{-1,4}e^{sL_{-1,0}} = A'(s)L_{-1,4} + B'(s)L_{0,4},$$

whence $A'(0) = -1$, $B'(0) = 1$. Differentiating again, it follows similarly that

$$-e^{-sL_{-1,0}}L_{0,4}e^{sL_{-1,0}} = A''(s)L_{-1,4} + B''(s)L_{0,4},$$

which implies that $A''(s) + A(s) = 0 = B''(s) + B(s)$. It results that

$$A(s) = -\sin s, \qquad B(s) = \cos s,$$

whence

$$\begin{aligned} H_0(s) &= \frac{1}{2i}[L_{-1,0} - \sin sL_{-1,4} + \cos sL_{0,4}] \\ &= \frac{1+\cos s}{2i}\frac{\partial}{\partial t} + \frac{1-\cos s}{2i}Q\frac{\partial}{\partial t}Q \\ &\quad + \frac{\sin s}{2i}K, \qquad (K = -L_{-1,4}). \end{aligned}$$

Now consider how this change in the operator $i^{-1}\,\partial/\partial t$ representing frequency measurement, over the time interval of duration s, is reflected in a measurement of frequency of a light wave, initially of a fixed frequency ν. The wave function Ψ has then the feature that

$$\frac{1}{i}\frac{\partial\Psi}{\partial t} = \nu\Psi, \qquad \text{i.e.,} \qquad H_0(0)\Psi = \nu\Psi,$$

at the point of emission, which may be taken as the origin. At the later point P at which the frequency is measured, say of coordinates (s, u), where s is the unitime and u the position in space S^3, the observed frequency ν' will be given by the equation

$$H_0(s)\Psi = \nu'\Psi.$$

However, while ν will be an exact value for the frequency $H_0(0)$, i.e., Ψ is a stationary state for H_0, there is no a priori reason for ν' to be an exact value for $H_0(s)$; Ψ need not be an exact stationary state for $H_0(s)$. Thus the frequency will be shifted from ν to ν', while at the same time a certain dispersion, effecting a corresponding line broadening, may be introduced into the frequency measurement. In order to compute ν' and its dispersion, we must know explicitly the wave function Ψ. The simplest reasonable postulate is that it is a plane wave of frequency ν. Neglecting polarization, which is presently irrelevant, it has then the form

$$\psi(t, x) = e^{i\nu(t - \mathbf{x} \cdot \mathbf{k})},$$

$\mathbf{k}$ being fixed. Strictly speaking, this representation for the wave function is incomplete, in that it is given as a function of the special relativistic coordinates, which are not globally applicable throughout unispace. It will be seen later that it nevertheless extends in a natural and unique way to a wave function throughout the region accessible from the point of emission by a light ray. In the meantime, we shall ignore apparent questions as to behavior of the wave function at extremely remote points of the Cosmos, e.g., our antipode.

We wish first to determine $H_0(\tau)\psi$, at the point of observation, which we take to have the form (τ, u), where u is the spatial position in S^3. Since Maxwell's equations and the wave equation are conformally invariant, the properties of solutions are basically independent of whether they are analyzed from a flat or curved standpoint. In particular, light continues to be propagated along light rays of the conventional type, which however appear in unispace to have the form $(\sigma, u(\sigma))$, $0 \leqq \sigma \leqq \sigma_0$, where $u(\sigma)$ describes a great circle on S^3 with constant velocity, normalized to be 1. In particular, at the point of observation, $\tau > 0$ and the distance of u from the point in S^3 of emission, taken here as $(0, 0, 0, 1)$, is precisely τ. It is essentially the same to say that, in terms of flat coordinates, $t^2 = x^2 + y^2 + z^2$, except that this parametrization is valid only on part of unispace.

Consider first $K\Psi$ evaluated at P. Evidently

$$K\Psi = \left(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z} + t\frac{\partial}{\partial t}\right)\Psi = i\nu(t - \mathbf{x} \cdot \mathbf{k})\Psi.$$

However, $t = \mathbf{x} \cdot \mathbf{k}$ along the ray of propagation, in particular at the point P. Thus $K\Psi|_P = 0$. Next, consider $Q(\partial/\partial t)Q$ evaluated at P. As earlier determined, $Q(\partial/\partial t)Q$ is a linear combination with constant coefficients of the operators $(t^2 - x^2 - y^2 - z^2)\,\partial/\partial t$ and tK. Evidently, the first of these operators on application to Ψ yields zero along any light ray. The second vanishes at P on application to Ψ by the preceding paragraph.

Thus

$$H_0(\tau)\Psi|_P = \frac{1 + \cos\tau}{2i}\frac{\partial\Psi}{\partial t}\bigg|_P = \frac{1 + \cos\tau}{2}\nu\Psi\bigg|_P \qquad \text{i.e., at } P,$$

$$H_0(\tau)\Psi = \nu'\Psi \qquad \text{with} \quad \nu' = \frac{1 + \cos\tau}{2}\nu.$$

This means there is an expected frequency-independent redshift in the amount

$$z = \frac{1 - \cos \tau}{1 + \cos \tau} = \tan^2\left(\frac{\tau}{2}\right).$$

However, for such a redshift of a discrete frequency level to be observable, it is necessary that the dispersion in the expected frequency be relatively small (of the order of no more than some angstroms). A bound on the dispersion indicating that this is the case can be obtained without detailed computation in the following manner.

By general quantum phenomenological principles, the variance σ^2 of a dynamical variable X in a given state may be expressed as follows. Let E denote the expectation value functional for the state; i.e., for a given dynamical variable Y, $E(Y)$ is the expectation of Y in the state. Then

$$\sigma^2 = E(X^2) - E(X)^2.$$

Let us apply this to the state ψ and the dynamical variable $H' = H(\alpha)$. We have seen that near the point P of observation

$$H'\psi \sim v'\psi.$$

The variance σ^2 may then be computed from the equation

$$H'^2\psi \sim (v'^2 + \sigma^2)\psi$$

near P. Now $H'\psi = a\psi$, where a is a certain function on space–time; it follows that

$$H'^2\psi = aH'\psi + (H'a)\psi = a^2\psi + (H'a)\psi,$$

inasmuch as H' is a homogeneous first-order linear differential operator. It follows that

$$\sigma^2 = H'a = v[H_0(\alpha)(t - x \cdot k)]_P .$$

This shows that σ is of the order of $v^{1/2}$, and hence negligible relative to v for large v.

The frequencies involved here (e.g., for an observed wavelength of 21 cm or less) are indeed quite large, especially in the units here in question. In these units, π units of time are required for light to traverse the distance from any point of space S^3 to its antipode. Assuming this distance to $\gtrsim 10^9$ ly (cf. Chapter IV) and the frequency v to correspond to an emitted wavelength $\leqq 20$ cm, this gives $v^{1/2}/v \lesssim 10^{-12}$, an entirely negligible (unobservable) dispersion. This applies in fact to an arbitrary stationary state. Within the plane wave approximation, the dispersion vanishes, i.e., ψ is effectively a stationary state of H' near the point of observation. But it is not at all an eigenvector for H' throughout the Cosmos.

In fundamental principle, expectation values and dispersions are defined by integration over the entire Cosmos; in practice, one analyzes only the behavior of the wave function in the vicinity of the points of interest. To within the approximation represented by the use of plane waves rather than normalizable photon wave functions, this accurately reflects the observational situation. In effect, one cuts off the defining integrals at a distance, say of the order of 1 ly, which is far beyond the limits of direct observation, but sufficiently small that $H_0(\alpha)^n\psi \sim \nu'^n\psi$ $(n = 1, 2)$ out to this distance, within observational accuracy. Without such a cutoff, the defining integrals over the entire Cosmos would, in the case of plane wave, be divergent.

This cut-off may be made rigorous and the entire redshift computation carried out within the Hilbert space of normalizable photon wave functions at the cost of some analytical complication. Since the Hilbert space analysis is the basis of the correlation of the Heisenberg picture just adopted with the Schrödinger picture, it seems useful to develop it. To do so, the photon Hilbert space must be set up explicitly. Within the scalar approximation already adopted, a photon may be represented by a solution φ of the wave equation, of the form

$$\varphi(X) = \int_{K^2=0} e^{iX\cdot K} f(K)\,d\mu(K),$$

where $X = (x_0, x_1, \ldots, x_n)$, $K = (k_0, k_1, \ldots, k_n)$, $X \cdot K = x_0 k_0 - x_1 k_1 - \cdots - x_n k_n$, n is the number of space dimensions, and $d\mu(K) = dk_1 \cdots dk_n/|k_0|$. The inner product between two such wave functions is given by the equation

$$\langle\varphi_1, \varphi_2\rangle = \int_{K^2=0} f_1(K)\bar{f}_2(K)\,d\mu(K).$$

This inner product is invariant under all orthochronous conformal transformations and is uniquely determined, within a constant factor, by this property.

According to the Heisenberg form of quantum mechanics, the operator $H_0(s) = e^{-isH}H_0 e^{isH}$ representing the relativistic energy at time s has expectation value $\langle H_0(s)\varphi, \varphi\rangle$ and variance $\langle H_0(s)^2\varphi, \varphi\rangle - \langle H_0(s)\varphi, \varphi\rangle^2$ if the photon is in the state φ, normalized by the condition that $\langle\varphi, \varphi\rangle = 1$.

In the Schrödinger picture in which the state changes but the dynamical variables remain unchanged, the photon state after time s is $\varphi_s = e^{isH}\varphi$, and the expectation value and variance of the relativistic energy H_0 in this state are given by the expressions $\langle H_0\varphi_s, \varphi_s\rangle$ and $\langle H_0{}^2\varphi_s, \varphi_s\rangle - \langle H_0\varphi_s, \varphi_s\rangle^2$, which are equal to those earlier given by virtue of the unitarity of the operators e^{isH}. Thus, as is well known, the two pictures give physically equivalent results. The computation of the redshift and its dispersion evidently depends on the evaluation of the inner products $\langle H_0\varphi, \varphi\rangle$, $\langle H_1\varphi, \varphi\rangle$, $\langle K\varphi, \varphi\rangle$, $\langle H_0{}^2\varphi, \varphi\rangle$, $\langle H_0\varphi, H_1\varphi\rangle$, etc.

Real photons may equivalently be represented by positive frequency wave functions, i.e., complex-valued wave functions φ that satisfy the wave equation and have vanishing negative frequency components, instead of wave functions φ that are real in physical space. Since the Fourier transform $F(K)$ of the latter type of wave function is hermitian, $F(-K) = \bar{F}(K)$, it is determined by its positive frequency component, on which the orthochronous conformal group acts in the same way. A positive frequency wave function can not be localized in physical space, since it consists of boundary values of an analytic function; consequently the most direct form of representation of a recently emitted (and therefore localized) photon is in terms of a real wave function. The simplest such function that seems physically relevant is a cutoff plane wave in two space-time dimensions, the one-dimensional space being defined by the direction of motion of the photon. This is of the form

$$\varphi(x_0, x_1) = f(\nu(x_1 - x_0)), \qquad f(x) = \begin{cases} 1 + \cos x & |x| \le p \\ 0 & |x| \ge p \end{cases}$$

where p is of the form $p = (2r + 1)\pi$, r being an integer; this is simply a plane wave of frequency ν, which is cut off smoothly beyond $2r + 1$ oscillations.

Without explicit computation, it follows that the ν-dependence of the relevant integrals is as follows, after normalization of the wave function by division by $\langle\varphi, \varphi\rangle^{1/2}$:

$$\begin{array}{lll} \langle H_0\varphi, \varphi\rangle \propto \nu & \langle H_1\varphi, \varphi\rangle \propto \nu^{-1} & \langle K\varphi, \varphi\rangle \propto \nu^0 \\ \langle H_0{}^2\varphi, \varphi\rangle \propto \nu^2 & \langle H_1{}^2\varphi, \varphi\rangle \propto \nu^{-2} & \langle K^2\varphi, \varphi\rangle \propto \nu^0 \end{array}$$

By virtue of the scale on which ν is measured, according to which a typical value is $\gtrsim 10^{26}$ (the value for the 21 cm line, assuming the radius R of the universe is $\gtrsim 100$ Mpc, the figure resulting from conventional estimates of the distance and redshift of Virgo galaxies), together with the Schwarz inequalities: $|\langle H_0\varphi, H_1\varphi\rangle| \le \|H_0\varphi\|\,\|H_1\varphi\|$, etc., the only possible nontrivial contribution to a dispersion in redshift (i.e., a contribution of order ν) can come from the terms involving H_0.

Explicit computation of these terms gives the results: $\langle H_0\varphi, \varphi\rangle = \nu(1 + O(\log p/p)$; $\langle H_0{}^2\varphi, \varphi\rangle = \nu^2(1 + O(\log p/p)$. Thus the redshift dispersion is $\nu O(\log p/p)$ and vanishes for an infinite plane wave. For a cutoff plane wave which is $\gtrsim 1$ light-second in extent, and ≤ 21 cm in wavelength, the dispersion is $\lesssim 1$ part in 10^{17}, and hence quite remote from observation. It follows also that the superrelativistic component of recently emitted radiation within the galaxy is negligable. But old, delocalized radiation, no longer approximately an eigenstate of H_0—although necessarily resoluble into such—can be highly energetic, particularly the very low frequency components, since $\langle H_1\varphi, \varphi\rangle$ varies inversely with ν. Explicit computation shows that the proportionality factor has the form $\kappa p^2(1 + O(\log p/p))$, where κ is an absolute constant of order 1; i.e., the superrelativistic energy varies approximately directly with the square of the diameter of the region of support. This result is naturally to be compared with the approximately quadratic rate at which the superrelativistic component of the energy of a freely propagated photon builds up according to the redshift law earlier derived.

7. Local Lorentz frames

Having thus derived an apparent redshift, let us now elucidate the physical connection between the flat and curved dynamical observables. This is necessary in order to predict results of other types of measurement. A central hypothesis used in the derivation is that an anthropomorphically measurable local observable is represented theoretically by a *flat* dynamical variable at the point of observation P_0. These flat dynamical variables are mathematically definable in a large region of the Cosmos, far from the point P_0, but at other points they are mathematically distinct from the corresponding flat dynamical variables, expressible in constant coefficients in terms of the local anthropomorphic Lorentz frame. At any point P, the latter frame is the unique Lorentz frame—unicity only within a scale factor, an important point to be further discussed—which is tangential to the globally given curved unispatial frame. This is the unique such unispatial frame which at the point of observation P_0 is tangential to the Lorentz frame of measurement at P_0.

A stationary anthropomorphic observer at a point P' should then see events as taking place in this tangential Lorentz frame. These frames vary in a well-defined way with the point P, and from a conventional Minkowskian point of view are in relative motion. The latter motion is entirely virtual; the Cosmos is stationary from the curved observational standpoint, which is, however, anthropomorphically indirect (i.e., accessible via redshift–magnitude observations, etc., and their theoretical interpretation). In other terms, the true driving physics is cosmologically stationary, but the Cosmos may appear in motion due in part to the theoretical analysis employed and in part to inherent restrictions on the mode of observation enforced by anthropomorphic and/or microscopic limitations.

It is instructive to compute explicitly the relation between the anthropomorphic Lorentz frames at two different points of the Cosmos. Setting $\hbar = c = 1$ leaves open the distance scale in Minkowski space. In unispace, we employ the natural distance scale, that in which the radius R of space S^3 is unity. It will be convenient to define the anthropomorphic distance scale by a constant R which expresses the ratio between a local distance as measured in Minkowski space and as measured in unispace; this scale $R(P)$ may, for the present, vary with the point P of the Cosmos. Taking the point P_0 of observation as the origin in Minkowski space, as is no essential loss of generality, and setting R_0 for the local distance scale, the Minkowski coordinates x_j are related to the unispatial coordinates u_k as follows:

$$x_j = 2u_j R_0(u_{-1} + u_4)^{-1},$$

in the vicinity of the point P_0. In the vicinity of a different point P_1, the local Minkowski coordinates x'_j which are tangential to the unispatial coordinates at P'—i.e., the x'_j vanish at P_1, and $dx'_j = R_1\, du_j$ ($j = 0, 1, 2, 3$) at P_1—are nonlinearly related to the Minkowski coordinates x_j. Of particular interest is the case in which P_0 and P_1 are relatively lightlike, e.g., P_0 is the point of emission of light and P_1 is the point at which it is observed. By making a suitable Lorentz transformation, it can be assumed that the x_3 and x_4 coordinates of P_1 vanish, so that one is in an essentially two-dimensional spatio-temporal situation. This simplifies the discussion, and serves to illustrate the useful simple form of the general theory in which space–time is two dimensional.

In the two-dimensional case, unispace may be fully parametrized by the angles τ and ρ, defined by the equations

$$u_{-1} = \cos\tau, \qquad u_0 = \sin\tau, \qquad u_1 = \sin\rho, \qquad u_2 = \cos\rho.$$

The tangential Minkowski coordinates at the origin are (t, x) where

$$t = \frac{2R_0 \sin\tau}{\cos\tau + \cos\rho}, \qquad x = \frac{2R_0 \sin\rho}{\cos\tau + \cos\rho}.$$

A point P_1 that is lightlike relative to the origin has unispace coordinates of the form $\tau = \rho = \alpha$. Near this point, tangential Minkowski coordinates (t', x') are given by the equations

$$t' = \frac{2R_1 \sin(\tau - \alpha)}{\cos(\tau - \alpha) + \cos(\rho - \alpha)}, \qquad x' = \frac{2R_1 \sin(\tau - \alpha)}{\cos(\tau - \alpha) + \cos(\rho - \alpha)}.$$

From these equations it is evident that (t', x') are well-defined functions of (t, x); it will suffice here to give the Jacobian matrix $(\partial x'_j/\partial x_k)$, evaluated at the point $(0, 0)$. A simple computation shows that this has the form

$$\left(\frac{R_1}{R_0}\right)\frac{\sec^2\alpha}{2}\begin{pmatrix} 1 + \cos^2\alpha & \sin^2\alpha \\ \sin^2\alpha & 1 + \cos^2\alpha \end{pmatrix}.$$

This is the product of a scale transformation, via the factor $(R_1/R_0)\sec\alpha$ with the Lorentz transformation

$$\frac{1}{2}\sec\alpha\begin{pmatrix}1+\cos^2\alpha & \sin^2\alpha\\ \sin^2\alpha & 1+\cos^2\alpha\end{pmatrix}.$$

This can be interpreted as a *virtual* motion of velocity dependent upon α, accompanied by a *virtual* expansion with factor $(R_1/R_0)\sec\alpha$.

A natural means of determining the distance scale R is, in the present theory, based on the assumption that the redshift is entirely due to the indicated chronometric effect, apart from possible small deviations due to intrinsic velocities, local gravitational effects, etc.; the main assumption here is that the fundamental properties of matter are the same in all parts of the universe. This is a provisional assumption; conceivably these properties vary with time, and even in a stationary universe, variations in the ages of emitting objects could introduce thereby an effect on the observed redshift. It will be found, however, that there appears to be no observational evidence for a significant effect of this nature, in the sense that all of the observations discussed in Chapter IV are consistent with the simple hypothesis of a chronometric redshift. The distance scale is constant throughout the Cosmos, on the present assumption.

The relation between the canonical Lorentz frames at different points is then unique and indicates the usual rate of time dilation by the factor $1+z$. For if the local relativistic coordinates x_j near the point O of observation are normalized so that they vanish at O, the equation

$$t=\tan^{-1}(x_0(1-x_0^2/4))$$

gives the relation between the unitime t and the observational time x_0 at O. The unitimes near any other point at rest relative to the point of observation differs only in zero point from t; this is true in particular of the unitime t' at the point E of emission; it follows that $x_0 = 2\tan[(t'-t_0)/2]$, where t_0 is the zero-point difference. Now the unitime t' was synchronous with the observational time x_0' at E at the time of emission; noting that

$$dx_0=(1+x_0^2/4)dt'; \qquad x_0^2/4=\tan^2(t/2),$$

it follows that $dx_0=(1+z)dx_0'$ is the relation between emitted and observed rates. The result just derived may plausibly be applied also to the dilation of the interval between wave crests; this provides a heuristic classical derivation of the chronometric redshift–distance law which is due to H. P. Jakobsen.

Actually, the wave function of the emitted light is not known with sufficient accuracy to warrant definitive conclusions at this time regarding time dilation. Strictly speaking, it can not be an exact stationary state of either relativistic or unitemporal time displacement, since such states can not vanish outside of bounded spatial regions. Locally, it is within observational limits, a stationary state of both energy operators. The earlier derivation of the redshift–distance law for localized states of the form $\varphi(x, t) = f(t - x)$ applies equally well to states of the form $f(\tau - \rho)$. It is essential for the argument that the function f vanish outside of a small local region; if f is a complex exponential, there is no redshift, since the state is stationary for the total energy H, and the wave function is in fact normalizable, unlike the usual relativistic plane wave. Moreover the temporal characteristics of the emission process are relevant in the determination of higher-order time dilation effects, which are not necessarily small in view of the apparent extreme nonlinearity of some of the fluctuation processes strong enough to be observed at great distances. Thus, the $1 + z$ time dilation factor should be regarded as a rough overall indication, and each particular type of process should be examined on its own merits. Rust (1974) has given some theoretical and observational evidence, suggesting that the factor may be different for supernovae time lapses. A totally different case is that of short-period quasar variability; its qualitative increase with z does not as yet appear to differ markedly from a $1 + z$ law.

8. Cosmic background radiation

From the equation $z = \tan^2(\rho/2)$, it is evident that $z \to \infty$ as $\rho \to \pi$, i.e., formally the redshift approaches totality as the propagation interval approaches a half-circuit of space. However, within the Minkowski framework, the antipode is infinitely distant, so that this total redshifting requires an infinite time relative to the local flat clock at the point of emission. In addition, there are physical circumstances which quite significantly modify the formal indication for an infinite redshift.

First, the photon wave function near the antipode of the point of emission will be almost entirely delocalized, as well as highly redshifted. It will then interact appreciably with the effective plasma formed of all galaxies and possible intergalactic matter throughout all of space. The photon will no longer be freely propagated, but may be scattered or absorbed.

Second, quantum dispersion, of the order of $\nu^{-1/2}$, becomes significant when ν becomes very small, and the description of the freely propagated wave function as "redshifted" becomes oversimplified. The frequency of the photon is no longer approximately sharp. From a flat local point of view, it is low in (relativistic) energy, and relatively high in superrelativistic energy, the total unienergy being conserved. Similarly, the linear momentum is not conserved, although an analogous unimomentum is conserved. This lack of conservation of momentum applies to the direction, as well as to the magnitude, of the momentum vector. While the *expected* momentum vector has the same as the original direction of propagation, there is a stochastic component to the direction representing the quantum dispersion, with nonvanishing contributions orthogonal to the line of sight. The effect of free propagation on the linear momentum is in fact computable by the same analysis as in the case of the relativistic energy.

For all of these reasons, the analysis of highly redshifted radiation as if it were an effectively localized plane wave packet is quite inappropriate. For sufficiently high redshifts, it should not be at all observable as radiation from a discrete source, even if originating in one, but only as background radiation. Its propagation from this point onward is probably not accurately represented by the free unitemporal evolution of a solution of Maxwell's equations in Minkowski space.

While the local dynamics of all contributions to this background radiation must be extremely complex, the general considerations of equilibrium statistical mechanics lead to a conveniently simple conclusion. All radiation may be divided into two classes: (a) the "pristine," which has made less than a half-circuit of the universe since emission; (b) "residual," the remainder. The origin of this radiation is not important for general considerations, but in the chronometric theory there is no special reason not to postulate that it arises primarily from discrete objects, and for concreteness, this may be assumed. In view of the apparent transparency of intergalactic space, the residual radiation should typically make many circuits of space before ultimately interacting with matter. The infinite time available for this low-frequency, high-dispersion radiation to accumulate implies quantitatively that it may be highly energetic, but in any event qualitatively that it is distributed in accordance with Planck's law, i.e., having a blackbody spectrum. For this law follows directly from the conservation of energy and maximization of entropy. The conservation of the unienergy is the starting point in the chronometric redshift analysis of the propagation of free photons, and its extension to all dynamical processes in the universe is tantamount to temporal homogeneity and causality.† It is a very natural and almost inevitable postulate. The maximization of entropy is implied by ergo-

† It should perhaps be emphasized that the so-called steady state theory is not at all temporally homogeneous from the present standpoint since energy in this theory is essentially ad hoc and not intrinsically definable in terms of the geometrical structure of the Cosmos. It is only for a theory of this latter type, in the presence of suitable noncyclical diffusion of energy yielding the requisite ergodicity, that the Planck law follows.

dicity, or nondeterministic mixing, which should be amply fulfilled by virtue of the stochastic character of the emissions from and motions of galaxies.

Thus the residual radiation should appear in the form of an energetic background blackbody radiation. The largely unknown absorptive characteristics of the various aggregations of matter in space, as well as of the extent of this matter itself, preclude a direct estimate of the energy density of this radiation. However, an approximate upper bound on its temperature may be estimated in terms of the energy density of starlight, and certain galaxy parameters, by neglecting all absorption except that by bright galaxies. Unless there exist large amounts of matter in presently unknown form, this upper bound may reasonably be expected to give the correct order of magnitude of the temperature of the radiation.

For such order-of-magnitude estimates, it suffices to approximate the galaxies by completely absorbing spheres of a fixed radius r. The extinction in a short time τ of propagation is the quotient of the total solid angle Ω subtended by all the galaxies in the spherical region of radius τ subtended from the center, by 4π. Again, Ω is sufficiently accurately estimated by placing all the galaxies at the expected distance, on the basis of spatial homogeneity, of $(\frac{3}{4})\tau$ from the point of emission, and neglecting overlapping solid angles. If μ denotes the number density of bright galaxies, the resulting extinction is consequently

$$[4\pi(3\tau/4)^2]^{-1}[\mu(4/3)\pi\tau^3]\pi r^2 = (16\pi/27)\mu r^2\tau,$$

implying an extinction of $\exp[-(16\pi/27)n\mu r^2]$ in the course of n half-circuits of space.

The flat (special relativistic) component of the pristine radiation is the space average of $(1 + z)^{-1}$ times the total pristine radiation, say S. Assuming spatial homogeneity, the distribution of z is $(2/\pi)z(1 + z)^{-2}\,dz$, which when averaged over space gives a factor of $\frac{1}{2}$. On the other hand, summing over all possible numbers of circuits, the total residual radiation amounts to

$$P\sum_{n=1}^{\infty} \exp[-(16\pi/27)\mu r^2] \sim [(16\pi/27)\mu r^2]^{-1}P.$$

Thus the ratio of the energy of the residual to that of the special relativistic pristine radiation is $\sim 0.4\mu^{-1}r^{-2}$.

In making a comparison with observation, it would be natural to identify the observed microwave background radiation with the theoretical residual radiation, and, to an adequate approximation, the starlight background with the special relativistic pristine radiation. Although quantitative astronomical discussion is being left for Chapter IV, it may here be remarked that the results concerning the background radiation are in satisfactory agreement with observation.

One of the most striking features of the observed cosmic background radiation is its apparent strong isotropy. While unexpected from a Friedmann model standpoint, there is no reason for any local anisotropy within the universal cosmos framework, apart from peculiar motions. These are probably quite small for galaxies, indeed appear of an order $\lesssim 60$ km/sec (see Chapter IV). Even on a classical basis for analysis of radiation, such slight motions should not produce presently observable anisotropy. Taking into account the quantum dispersion additionally involved in a more exact treatment could only increase the threshold of peculiar motions which would produce observable anisotropy, quite possibly to a level well above that of the Sun. On the chronometric hypothesis, anisotropy in the background radiation appears unlikely to be observed until considerably greater precision of measurement is obtainable, if ever.

9. Special relativity as a limiting case of unispatial theory

Special relativity can be regarded as a limiting case of unispatial theory, as the radius R of the universe becomes infinity, in a sense indicated by Einstein, Minkowski, and Weyl. The radius R is actually a physical constant, and the mathematical content of the formation of the limit as $R \to \infty$ requires some clarification. It can, however, be defined by analogy with the familiar cases $c \to \infty$, which leads to Galilean relativity from special relativity, and $h \to 0$ which leads to classical from quantum mechanics. We shall be relatively explicit and show how the fundamental dynamical variables of the unispace theory converge to those of special relativity.

Consider first the space–time coordinates. The physically observed coordinates are not the x_j $(0 \leqq j \leqq 3)$ of the first part of this chapter, but rather the $x'_j = Rx_j$, R being the radius of the universe. These x'_j are to be compared with the Ru_j. We have

$$Ru_j - x'_j = O(1/R) \qquad \text{as} \quad R \to \infty.$$

For

$$\begin{aligned} Ru_j - x'_j &= Rx_j[(1 - \tfrac{1}{4}d)^2 + t^2]^{-1/2} - x'_j \\ &= x'_j\left\{\left[\left(1 - \frac{d'}{4R^2}\right)^2 + \frac{t'^2}{R^2}\right]^{-1/2} - 1\right\} \\ &= x'_j O(R^{-2}). \end{aligned}$$

An even sharper and uniform estimate holds when $t'^2 + x'^2 + y'^2 + z'^2 = e'$ is sufficiently small, as would be the case out to classical macroscopic distances, according to the estimates of R in Chapter IV. Thus the suitably scaled Minkowski and universal space–time coordinates agree to within a close approximation, for moderate e'.

Consider next the dynamical variables corresponding to generators of space–time symmetries. These are specifically the energy–momentum vector, the angular momenta, the boosts, and the infinitesimal scale transformation K. There are 11 generators here in all; seven of them are the same (i.e., have the same expression in terms of the x'_j and $\partial/\partial x'_j$) in both theories; the energy–momentum vector is the quartet which are distinct between the two theories. There are, in addition, four linearly independent infinitesimal symmetries of unispace, whose action in Minkowski space as $R \to \infty$ remains to be explored.

We have the

Theorem The suitably scaled 15 linearly independent generators L_{ij} of symmetries of unispace, when formulated as expressions in the x'_j and $\partial/\partial x'_j$, differ from the 11 generators of the group of global conformal transformations in Minkowski space by terms of order $1/R^2$, as $R \to \infty$.

Remark 1 It might appear anomalous that 15 vector fields converge to 11 vector fields. What happens is that two ordered sets, each consisting of four of the L_{ij}, converge to the same conformal vector fields in Minkowski space.

Remark 2 This result is independent, as are many in this chapter, of the dimensionality of space–time.

Proof We take the Minkowski energy–momentum vector in the usual form

$$\left(-i\frac{\partial}{\partial t'}, i\frac{\partial}{\partial x'}, i\frac{\partial}{\partial y'}, i\frac{\partial}{\partial z'}\right),$$

and define the (physically scaled) uni-energy–momentum vector to have the form

$$(-iR^{-1}L_{-1,0}, -iR^{-1}L_{-1,1}, iR^{-1}L_{-1,2}, iR^{-1}L_{-1,3}).$$

From earlier obtained expressions for the L_{ij}, it follows that:

(the univector) − (the special relativistic vector)

= (inverted relativistic vector),

where the latter is defined as the transform of the special relativistic energy–momentum vector through conformal inversion. That is, the vector

$$-iQ\left(\frac{\partial}{\partial t'}, -\frac{\partial}{\partial x'}, -\frac{\partial}{\partial y'}, -\frac{\partial}{\partial z'}\right)Q$$

$$= \frac{-L_{-1,0} + L_{0,4}, L_{-1,1} - L_{1,4}, L_{-1,2} - L_{2,4}, L_{-1,3} - L_{3,4}}{-i(2R)}$$

The jth component of this vector is

$$\pm i\left[\frac{t'^2 - x'^2 - y'^2 - z'^2}{4R^2}\frac{\partial}{\partial x'_j} - \frac{x'_j \varepsilon_j}{2R^2}\left(t'\frac{\partial}{\partial t'} + x'\frac{\partial}{\partial x'} + y'\frac{\partial}{\partial y'} + z'\frac{\partial}{\partial z'}\right)\right],$$

which is $O(R^{-2})$.

The angular momenta L_{ij} $(i, j = 1, 2, 3)$ have the form

$$-i\left(x_j\frac{\partial}{\partial x_i} - x_i\frac{\partial}{\partial x_j}\right) = -i\left(x'_j\frac{\partial}{\partial x'_i} - x'_i\frac{\partial}{\partial x'_j}\right),$$

which is independent of R, and has the same expression both in Minkowski and unispace. The same is true of the boosts

$$-iL_{0,j} = -i\left(x'_0\frac{\partial}{\partial x'_j} + x'_j\frac{\partial}{\partial x'_0}\right) \qquad (j = 1, 2, 3),$$

and the infinitesimal scale transformation $K = \sum_j x'_j\, \partial/\partial x'_j$.

In particular, $R^{-1}L_{-1,j}$ and $R^{-1}L_{j,4}$ both differ from $\varepsilon_j\, \partial/\partial x'_j$ by $O(R^{-2})$, and agree in the limit $R \to \infty$ with the conventional energy–momentum component $\varepsilon_j \partial/\partial x'_j$. The differences $L_{-1,j} - L_{j,4}$ thus are locally approximate absolute constants of the motion. As such they are locally approximately representable by a slowly-varying vector field, which physically would appear most naturally as potentially related to gravitational phenomena as in the Weyl–Veblen theory, but possibly related also to microscopic processes as internal quantum numbers. There is no clear connection with extragalactic observation, and these generators will not be treated further here.

The philosophy of the chronometric approach to elementary particle theory may be briefly indicated here, as a means of clarifying its coherence with both macro- and microphysics. Its basic premise is that while the cosmos as a whole is covariant with respect to $SU(2, 2)$, or more precisely its universal covering group, say G_{15}, the observable microcosmos is covariant only with respect to the scale-covariant subgroup, say G_{11}. The scale generator $-L_{-1,4}$ is, like the superrelativistic energy-momentum vector, locally an approximate absolute constant of the motion and thus determines a slowly-varying scalar field. This again is most naturally interpreted from a gravitational standpoint, in terms, for example, of the Nordstrom-like theory or, in combination with the vector field just indicated, the Weyl–Veblen theory. From an elementary particle standpoint, what may be important is the restriction of the G_{11} to the usual G_{10} Lorentz group brought about by the elimination of the scale generator; thereby conformal invariance is not at all inconsistent with the existence of massive particles. The G_{11} is invariantly specified as the subgroup of $\widetilde{SU}(2, 2)$ leaving invariant the "infinite" points relative to the observer in question. These are the points that are carried by the covering transformation from unispace $\tilde{M}$ to the conformal compactification $\bar{M}$ of Minkowski space M into points that are in $\bar{M}$ but not in M; i.e., the antipodal point, all points lightlike relative to it, and all transforms of these points by the center of G_{15}. An alternative mechanism for the introduction of massive particles is the restriction to the $SO(2, 3)$ subgroup, which includes time development, space rotations, etc., i.e., that consisting of transformations leaving unaltered the last coordinate ξ_4. This mechanism would probably lead to a countably infinite series of theoretical particles of discrete masses without the interventation of adjustment for scaling, which could conceivably prove to be unobservably small.

While these ideas are qualitative, they are nevertheless suggestive of relatively concrete models for the basic elementary particles. The leptons, for example, may most simply and naturally be correlated with the solution space of the Dirac equation in $S^1 \times S^3$, or in $U(2)$. The Z_4 central subgroup of $SU(2, 2)$ provides a quantum number that can naturally be expected to distinguish neutral from charged leptons. The vanishing of the parameter m in the Dirac equation does not imply the vanishing of the physical masses of all the elementary particles involved here (i.e., irreducible constituents of the indicated representation of $SU(2, 2)$ on its restriction to the extended Poincaré subgroup), due to the curvature of space and to the role of scaling.

Similarly the baryons are quite conceivably represented by the totality of spinor fields on $\tilde{M}$, or one of its locally isomorphic versions, having "real mass," i.e., the square of the Dirac operator has a nonnegative spectrum in the state corresponding to the field (more exactly, semibounded spectrum, since the curvature of space displaces the zero-point). On restriction to G_{11}, the representation of the G_{15} defined by these fields may well split into only a finite number of irreducible constituents. (I am indebted to B. Kostant for citations of analogous known group-representation-theoretic phenomena.) The scale then becomes an experimental constant, but the ratios of the masses of the constituents would be mathematically computable.

These computations are technically fairly advanced, but appear entirely feasible as a program for immediate development. Of course, it is always possible that nonlinear interaction effects are so large as to dominate completely the spectrum of apparent elementary particles and to reduce the implications of the present group-theoretic approach to a qualitative level. This is, e.g., the present position of the Heisenberg school, among others. However, the discreteness and unicity of the leptons and baryons are strongly suggestive that elementarity of particles is at least partially real rather than relative; and the mechanisms here proposed for the classification of these particles are considerably more unique analytically and definitively accessible quantitatively than those which depend on proposed nonlinear quantized fields. In addition, the—possibly surprising—usefulness in cosmology of rational methods, based on general considerations of causality and symmetry, etc., evidenced by the good agreement with observation of the chronometric theory, must be taken into consideration, however different physically the situations in elementary particle physics and extragalactic astronomy may appear to be.