A Variant of Special Relativity and Long-Distance Astronomy

(cosmology/redshift/microwave background/causality/symmetry)

I. E. SEGAL

Massachusetts Institute of Technology, Cambridge, Mass. 02139

Contributed by I. E. Segal, November 5, 1973

ABSTRACT The redshift, microwave background, and other observable astronomical features are deduced from two theoretical assumptions: (1) global space-time is a certain variant of Minkowski space, locally indistinguishable in causality and covariance features but globally admitting the full conformal group as symmetries although having a spherical space component; (2) the true energy operator corresponds to a certain generator of this group which is not globally scale-covariant, whereas laboratory frequency measurements are inevitably such and correspond to the conventional energy operator $\hbar/i\partial/\partial t$.

The broad acceptance of the expansion of the universe as a physically real phenomenon has been rooted in part in the apparent lack of an alternative explanation of the red shift. Since its discovery more than a half-century ago, many new observational phenomena have been uncovered, of which the microwave background radiation and the quasars appear to be particularly fundamental and striking. Nevertheless there seem to have been few attempts to rework the foundations of cosmology in a way which might tie together these phenomena in a scientifically more economical way. This is probably due more to the momentum of the theoretical studies based on the expansion theory than to its agreement with observation, which has been quite limited and increasingly equivocal.

At an intuitive theoretical level, in fact, the redness of the shift is suggestive of a nonconservation of locally observed energy. Conservation of energy in a conventional temporally homogeneous theory then requires a physical form for the dissipated energy, providing an interlocking explanation for the microwave background. The theoretical form of the energy operator is in fact naturally brought into question by the observation that, by virtue of the arbitrariness of the fundamental units of space and time, the frequency of light as measurable in a laboratory is inherently scale-covariant. Such appropriate transformation under scale changes is valid for the conventional theoretical operator representing frequency, namely $\hbar/i\partial/\partial t$. However, if the Cosmos is not Minkowskian, there is no reason to expect the true energy operator of the Cosmos to be scale-covariant—indeed, global scale transformations will in general not exist—although locally, in laboratory elementary particle measurements, the true operator H may deviate by unobservably little from the conventional operator H_0 .

If the difference $H - H_0 = H_1$ is positive, and if H and H_0 have negligibly different expectation values in localized particle states ψ , then by conservation of energy the apparent microscopic energy at a point of observation O of particles freely propagated from the point of emission P can not be greater than that at P . Specifically, the apparent energy E_O near O is given by the equation $H_0\psi \sim E_O\psi$ near O , while the

apparent energy E_P near P is given by the equation $\exp(-iH)H_0\exp(iH)\psi \sim E_P\psi$ near P , t being duration of the time interval of propagation from P to O . Thus if H and H_0 do not commute, a decrease in apparent energy, quite possibly significant over cosmological distances, is to be anticipated. The energy loss in this redshifting process would then be represented by a corresponding increase in the expectation value of the nonscale-covariant component H_1 of the total energy. In the real Cosmos, such radiation would be freely propagated until absorbed by matter, perhaps chiefly that in other galaxies. In view of the temporal homogeneity of the model, this residual radiation from the totality of luminous objects in the universe would exist in an equilibrium state, which on the basis of the assumptions of energy conservation and maximal entropy would have a black-body spectrum. The ergodicity required for the proper derivation of this law should be amply fulfilled by virtue of the stochastic character of the emission from and motions of galaxies.

The anomalous situation of quasars in the expansion theory would not necessarily be ameliorated by such a model, but the altered redshift-distance-luminosity relation in the model would at least provide an alternative to the main currently pursued hypotheses, neither of which seems entirely satisfactory, that quasars are extraordinarily bright, fast-moving, and rare at large redshifts, as indicated by the expansion theory, or have redshifts largely unrelated to their spatial positions, on the so-called local hypothesis.

A priori, it is quite conceivable that there is no such model which agrees with detailed observation and/or satisfies general physical constraints such as spatial homogeneity and isotropy; seemingly less likely but also conceivable is the existence of an embarrassing multitude of such models, dependent upon many parameters which are fitted from the data, with a corresponding loss of predictive power or possibility for definitive validation. This note deals with the theoretical physical foundations of a model which is a remarkably simple and symmetrical variant of special relativity, and is mathematically unique among models of the indicated type satisfying very general physical constraints. Relations between observable quantities are deduced, and their empirical validation briefly reported.

Relation to the Minkowski cosmos

Let $\tilde{M}$ denote the 4-dimensional manifold of all pairs (t, u) , where t is a real number and $u = (u_1, u_2, u_3, u_4)$ is a point on the sphere S of radius R in 4-dimensional euclidean space. Let a notion of causality be introduced in $\tilde{M}$ by taking as the future directions at any given point those corresponding to

the infinitesimal transformations $a\partial/\partial t + b_j \partial/\partial u_j$, with $a > 0$ and $a^2 > \sum_j b_j^2$; for such a transformation to act in S it is necessary in addition that $b_j u_j = 0$. It can be shown that this notion is global, in the sense that there is no closed time-like arc, the latter being defined as one whose forward tangent at each point is into the future.

Observe next that Minkowski space M can be imbedded in $\tilde{M}$ in a causality-preserving manner, which has additionally the feature that its coordinates (x_0, x_1, x_2, x_3) are tangential near the origin to (t, u_1, u_2, u_3) , up to terms of third order. The unique* mapping effecting this imbedding is in fact given by the equations

$$t = \tan^{-1} \left[x_0 / \left(1 - \frac{x^2}{4} \right) \right],$$

$$-\pi < t < \pi \quad (x^2 = x_0^2 - x_1^2 - x_2^2 - x_3^2);$$

$$u_j = ax_j \quad (j = 1, 2, 3), \text{ where } a = \left[\left(1 - \frac{x^2}{4} \right)^2 + t^2 \right]^{-1/2};$$

$$u_4 = a \left(1 + \frac{x^2}{4} \right).$$

The inverse mapping having the form

$$x_j = 2u_j / (\cos t + u_4)$$

is properly defined only on the open region M' in $\tilde{M}$ into which M maps, in which in particular $|t| < \pi$ and $u_4 > 0$.

The imbedding of M into $\tilde{M}$ is not only causal, but covariant. For any infinitesimal conformal transformation in M , there is a corresponding transformation defined everywhere in $\tilde{M}$, whose action in M' is just that carried over from M . Moreover, the finite transformations generated by these infinitesimal transformations in $\tilde{M}$ are globally causality-preserving in $\tilde{M}$.

Is it possible that the real cosmos is actually $\tilde{M}$ and only appears locally as M ? Equally with M , $\tilde{M}$ is serviceable as a theoretical basis for fundamental physics. In particular, quantum field theory would carry over with positive energies, causal propagation, and all basic formal features. Wave equations for zero-mass particles, e.g., the Maxwell equations, carry over without essential change, by virtue of their conformal invariance (1). Wave equations for massive particles can be shown to be appropriately adaptable to $\tilde{M}$, the difference from conventional relativistic equations involving only terms of order $1/R$. While the two cosmos are physically quite different in that the 3-“space” of $\tilde{M}$ is spherical, rather than euclidean as for M , this difference is not directly observable for physically reasonable values of R . The question arises, are there any empirical means to check the validity of the $\tilde{M}$ -hypothesis, as it may be designated?

Consider first the empirical implications for purely local phenomena. The differences between the x_j and the corresponding coordinates in $\tilde{M}$ are of third order in the distance. Simple estimates give, in fact: $|(u_j - x_j)/x_j| \lesssim \epsilon^2$, if ϵ is the distance in units of R . If, therefore, as is commonly believed $R \geq 10^8$ ly, the respective coordinates deviate by less than 1 part in 10^{15} out to distances (or corresponding times) of 1 ly, or of less than 1 part in 10^6 out to galactic distances. There is

* Unicity follows from the conformality of causality-preserving transformations, together with the fact that a conformal transformation which infinitesimally agrees with the identity up to terms of third order is everywhere the identity.

no apparent means to detect such differences in classical observations.

Consider, therefore, the relative highly precise frequency measurements in elementary particle phenomena. The physically appropriate energy operator in $\tilde{M}$ is $-i(\partial/\partial t)$; in Minkowski coordinates this takes the form

$$-i[1 - x^2/4](\partial/\partial x_0) + x_0 K/2],$$

where K is the infinitesimal scale transformation $x_j(\partial/\partial x_j)$. The difference between the energy operators is of second order in the distance in units of R , and so again appears negligibly small. (Quantitatively, e.g., if $R \geq 10^8$ ly and a hypothetical laboratory of size ≤ 10 km is employed in a direct measurement of the two energies, an expected difference of less than 1 part in 10^{39} would result, in the case of a photon represented by a plane wave.) For any localized elementary particle states, the expected difference between the two energy operators would be unobservably small.

Only phenomena of extremely large spatial or temporal extent appear potentially discriminatory between M and $\tilde{M}$, for physically reasonable values of R . Actually, for extremely long times the second order difference between the two energy operators should produce a significant effect; consider therefore the redshift.

The redshift-luminosity relation

The determination of a redshift is based on a measurement of length, the units of which are entirely conventional. If length and time standards are multiplied by fixed factors, the measurement procedure is such that the wavelength and frequency are respectively multiplied by directly corresponding factors. Such “scale covariance” appears to be an inherent observational limitation on the determination, e.g., of photon frequency, analogous to that involved in the indeterminacy principle, and not at all a matter of approximation or experimental error.

It is evident from the expression for the $\tilde{M}$ -energy $H = -i(\partial/\partial t)$ in terms of the Minkowski coordinates x_j that it is not scale covariant, but is the sum of the scale-covariant term $H_0 = -i(\partial/\partial x_0)$ and an antiscale-covariant term H_1 represented by a linear differential operator with quadratic coefficients. From the foregoing observation, it would appear that in principle the local measurement of elementary particle energies is appropriately represented by the operator H_0 rather than the total energy H .

Supplementing these observational considerations are the following formal ones: (1) the decomposition $H = H_0 + H_1$ is Lorentz covariant; (2) for any field theory in which the special relativistic energy H_0 is positive (e.g., the Maxwell field), the component H_1 , and hence also the total $\tilde{M}$ -energy H , are positive. It is straightforward in fact to verify (1). That (2) holds results from the following representation for $\partial/\partial t$, where units such that $R = 1$ are used:

$$\partial/\partial t = \partial/\partial x_0 - Q(\partial/\partial x_0)Q,$$

where Q denotes the operation of conformal inversion: $x \rightarrow 4x/x^2$, which on $\tilde{M}$ appears as the singularity-free transformation, $(t, t) \rightarrow (\pi - t, u)$. This shows that Q is continuously connected to time reversal, and is represented by an antiunitary operator Q' for any positive-energy field. For any hermitian operator of positive spectrum, say A , its transform $Q'AQ'$ will consequently have a negative spectrum, and taking A as the special relativistic energy, (2) follows.

Thus, photon frequency measurement being of the special relativistic energy H_0 , but free temporal propagation being generated by the $\tilde{M}$ -energy H , a redshift takes place as indicated in the first section. It is straightforward to derive the commutation relations between these operators and thereby compute the special relativistic energy $H_0(s) = \exp(-isH)H_0\exp(isH)$ after duration of a time s :

$$H_0(s) = (1/2)(1 + \cos s)H_0 + (1/2)(1 - \cos s)H_1 + (1/2)\sin sK$$

where K is the hermitian generator of scale transformation $-ix_j(\partial/\partial x_j)$. If now ψ is a plane wave photon state of frequency ν , it is easily verified that along any ray of propagation, $K\psi = H_1\psi = 0$, so that at the point of emission, $\langle H_0(s) \rangle = (1/2)(1 + \cos s)\langle H_0 \rangle$, signifying a redshift of $z = \tan^2(s/2)$. The dispersion σ in frequency, $\sigma^2 = \langle H_0(s)^2 \rangle - \langle H_0(s) \rangle^2$ is $\lesssim \sqrt{\nu}$ for observable redshifts, say $z < 10$. This is negligible, since ν is measured in units in which $R = 1$. (E.g., if $R \leq 10^8$ ly and ν corresponds to 5000 Å, then σ corresponds to $< 10^{-10}$ Å.)

The luminosity of an object at distance (polar angle) ρ is consequently reduced by a factor of $\sec^2(\rho/2)$ by virtue of the redshift and independently a factor of $\sin^2\rho$ by the appropriate variant of the inverse square law, whence L_{obs} varies as $z^{-1}(1+z)$. If the source has spectral index $\alpha \neq 1$, a factor of $(1+z)^{1-\alpha}$ must be inserted on the right side, as usual. An additional factor of $(1+z)^{-1}$ which occurs in the expansion theory because of the motion of the source away from the point of observation is naturally absent here. Further relations involving number counts and angular diameters follow by spherical geometry.

Energetics of the microwave background

From the equation $z = \tan^2(\rho/2)$, it is evident that $z \rightarrow \infty$ as $\rho \rightarrow \pi$, i.e., the redshift approaches totality as the propagation interval approaches a half-circuit of space. Within the Minkowski framework, the antipode $\rho = \pi$ is infinitely distant. The photon wave function near the antipode of the point of emission will be almost entirely delocalized, no longer necessarily approximately sharp in frequency, and highly depleted in special relativistic energy by virtue of the redshift. In short, it will not be observable as radiation from a discrete object but only as background radiation.

It is appropriate, therefore, to distinguish two components of the total radiation: (a) the "pristine," that from discrete objects and making less than a half-circuit of space before observation; (b) the "residual," all radiation, presumably largely from discrete objects, making more than a half-circuit of space prior to being observed. In view of the apparent transparency of intergalactic space, the residual radiation should typically make many circuits of space before being ultimately absorbed by matter, in, e.g., another discrete object, or by intergalactic matter, if any. The precise state of this residual radiation is sensitive to the precise form of the wave function of the emitted radiation, approximation by a plane wave or other exact frequency wave function being probably unreliable outside the region of finite Minkowski distance, and is affected by the largely unknown absorptive characteristics of the various aggregations of matter in space. Fortunately, general equilibrium statistical mechanical considerations permit a rough estimate of the relative energy of the pristine and equilibrium radiation, without a quantitative

analysis of the mechanisms by which the equilibrium is attained.

Neglecting all absorption except that by bright galaxies will give an upper bound on the energy of the residual radiation; and unless there exist very large amounts of matter in presently unknown form, this upper bound should give an order of magnitude estimate.

Approximating the galaxies by completely absorbing spheres of a fixed radius r , then 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 r subtended from the center, by 4π . For order of magnitude purposes, Ω is sufficiently accurately estimated by placing all the galaxies at the expected distance (on the basis of spatial homogeneity) of $(3/4)\tau$ from the point of emission, and neglecting overlapping solid angles. If μ denotes the number density of bright galaxies, the resulting extinction is consequently

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

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

The special relativistic component of the pristine radiation is the space average of $(1+z)^{-1}$ times the total pristine radiation P ; assuming spatial homogeneity, the distribution of z is $(2/\pi)z(1+z)^{-2}dz$, yielding a factor of $(1/2)$. On the other hand, the total residual radiation amounts to $P\sum_{n=1}^{\infty}\exp[-(16\pi/27)\mu r^2n] \sim [(16\pi/27)\mu r^2]^{-1}P$. Thus:

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

Turning now to the comparison with observation, it is natural to identify the observed microwave background radiation with the residual radiation, and, to an adequate approximation, the starlight background with the special relativistic pristine radiation. How well does the observed relative energy of $\sim 10^3$ conform to the theoretical prediction? Defining "bright" as of magnitude ≤ 13 at a redshift $cz = 2000$ km/sec, the catalog of de Vaucouleurs (2) indicates ~ 250 bright spiral galaxies in this redshift region, indicating $\mu \sim 1.4 \times 10^5$ (using now units with $R = c = 1$). The angular diameters given in the catalog correspond to a metric diameter of ~ 10 kpc on the $\tilde{M}$ -hypothesis, assuming $z = 0.005$ at a distance of 15 Mpc (which incidentally limits the age of pristine radiation to $\lesssim 10^9$ ly). The value $r = 5$ kpc $= 1.2 \times 10^{-5}$ then is indicated if the absorption in the Galaxy, ~ 0.3 csc b , is reasonably typical. The resulting theoretical prediction for the ratio of the energy of the microwave background to that of starlight is 1.7×10^5 , an excess over observation corresponding to a black-body temperature of ~ 2 times that observed. This seems fully comparable in precision to the accuracy of the prediction based on the "big-bang" hypothesis in view of the hypothetical parameters, such as the entropy density of the original universe, involved in the latter prediction. Faint galaxies and/or unknown forms of intergalactic matter could account for the discrepancy, as could also in part the probable underestimation of r resulting from the existence of significant dust exterior to the observed optical diameters of the galaxies.

Conclusion

In summary, the $\tilde{M}$ -hypothesis is theoretically tenable and provides an economical model for the crucial phenomena of extragalactic astronomy. It is further indicated on the theo-

retical side in being the only four-dimensional cosmos other than that of special relativity which satisfies very general constraints of causality and symmetry, such as are implicit in the foregoing† (3).

Empirically, the predictions of the hypothesis for observable quantities are in excellent agreement with all published observations on the m (or S) — z — N — θ relations for large or statistically unexceptionable samples of galaxies, quasars, and radio sources, as detailed in comprehensive studies submitted for publication. Indeed, in a phenomenological anal-

† To be quite precise, mention should be made of the cosmos locally identical to $\tilde{M}$ in which the sphere S is replaced by the elliptic space obtained by identification of antipodal points in S . The basic relations between observable quantities are unchanged by this modification, and observational discrimination between these two models would not presently be possible.

ysis conducted jointly with J. F. Nicoll of complete data (2) on low-redshift galaxies, the exponent n in a redshift-distance power law was unequivocally indicated as $n \sim 2$, whether based on the $m - z$, $\theta - z$, $N(<z)$ relations, or the Schmidt V/V_m test; and the value $n \sim 1$ was contraindicated. Consequently, it is submitted that there appears to be no apparent compelling scientific reason not to base extragalactic astrophysics on the $\tilde{M}$ - (in preference to the expansion) hypothesis.

I thank the authors of ref. 2 for providing a magnetic tape of their data. This research was supported in part by the National Science Foundation.

1. See Bateman, H. (1910) *Proc. Lond. Math. Soc.* **8**, 223–264, 469–488; Cunningham, E. (1910) *Proc. Lond. Math. Soc.* **8**, 77–98; Gross, L. (1964) *J. Math. Phys.* **5**, 687–695.
2. de Vaucouleurs, G. & de Vaucouleurs, A. (1964) *Reference Catalogue of Bright Galaxies* (Univ. of Texas Press, Austin).
3. Segal, I. (1972) *Astron. Astrophys.* **18**, 143–148.