General Relativity and Gravitation, Vol. 2, No. 1 (1971), pp. 7–21.

Topology and Cosmology

G. F. R. ELLIS

Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge

Read on 18 May 1970 at the Gwatt Seminar on the Bearings of Topology upon General Relativity

This paper is intended simply to review some topics concerning topology and cosmology (the word 'topology' being interpreted rather loosely). It is divided into three main sections: a discussion of topological properties one might expect in *any* reasonable model of the universe; a discussion of properties of some exact solutions which might serve as reasonable simple cosmological models; and some comments on properties one might expect in more realistic universe models.

The essence of the study of cosmology is trying to determine not merely *possible* space-time structures, but ones that *probably exist* in the universe; this leads to the related problem of attempting to verify what structure the universe actually has. I shall try to emphasize this viewpoint.

1. *Plausible Properties of Space-time*

We consider first properties one might expect any reasonable space-time to have.

The usual model of space-time is a C^∞ *4-dimensional connected Hausdorff manifold* M on which there is a *pseudo-Riemannian metric* g which is *at least* C^2, *piecewise* C^4. (I shall take the normal form to be $(-+++)$.) The existence of the metric on M is implied by the local validity of Special Relativity. One assumes in addition *field equations* determining the metric g on M (which I shall assume are the Einstein field equations), suitable equations determining the behaviour of the matter content of space-time (e.g. Maxwell's equations, the Weyl equation, etc.) and hence equations of state relating the components of the energy-momentum tensor; these equations of state may often be represented by suitable 'energy conditions', representing the positive definiteness of energy rather than more detailed properties of the matter.

It is no restriction taking the manifold to be C^∞ rather than, say, C^5, since the atlas of any C^1 manifold contains an analytic subatlas (Whitney, [*1*]). One usually takes the metric to be C^∞; this is not an essential restriction, since (even disregarding quantum difficulties) one cannot measure the differentiability class of the metric because of the finite accuracy of any measurement. The metric determines a unique Christoffel connection on (M, g); this implies (Geroch, [*2*]) that M has a countable basis, and so (Kobayashi and Nomizu, [*3*] p. 271) that M is paracompact.

The existence of a Lorentz metric g on M is equivalent to the existence of a line element field on M (Steenrod, [*4*], pp. 204–207); this always exists if M is non-compact, but if M is compact it admits a line element field if and only if the Euler–Poincaré characteristic $\chi(M)=0$ (Markus [*5*], Fierz and Jost [*6*]). In fact the space-times we consider will be non-compact. This is because if space-time is compact then (Bass and Witten [*7*], Geroch [*8*] Kronheimer and Penrose [*9*]) there exist closed timelike lines in M, violating the usual concept of causality. In fact, one cannot regard the existence of closed timelike lines as necessarily disproved by experiment; for example it would be difficult to prove or disprove the existence in the universe of closed timelike lines whose proper length was $\gtrsim 10^9$ years. (Although there would then be closed timelike lines of arbitrarily small proper length, these would in general involve large integrated accelerations.) However, if they existed it would be impossible to maintain the simple concept of the evolution of the universe (or of systems contained in the universe); the Cauchy problem would change to a complicated eigenvalue problem in which almost all 'initial data' was non-admissible. One will therefore usually assume that space-time satisfies one of the causality conditions (cf. e.g. Penrose [*10*]); perhaps the most plausible of these is that of *stable causality* (Hawking [*11*]) i.e. the condition that there exist no closed timelike curves in M even if g is slightly perturbed. It then follows that M must be non-compact. Two further consequences are (i) that the past light cone of any point is part of the boundary of its past (Kronheimer and Penrose [9], Penrose [10]) except where the past light cone lies within the past of the point†, and (ii) that if space-time is time-orientable, then a cosmic time function exists, i.e. there exists a function t such that the surfaces $\{t=\text{constant}\}$ are a family of closed imbedded spacelike 3-surfaces without boundary, and t increases along every future-directed timelike or null curve (Hawking, [*11*].

In fact one would expect that space-time was *time-orientable* since this follows if each observer can determine a unique forward arrow of time by thermodynamic and electrodynamic experiments (clearly such an assignment of a + direction of time would be continuous). One would further expect that *space-time is orientable*, since (Geroch [*12*], Zeldovich and Novikov [*13*]) arguments based on the CPT invariance of field theory, together with the fact that interactions which violate C, P and CP have been observed, imply that M is either both space and time orientable or both space and time non-orientable; in either case, space-time is orientable. It follows on combining these arguments that space-time is space-orientable. Thus one can safely conclude that the space-sections $\{t=\text{constant}\}$ defined

† The light-cone lies inside the past after self-intersections or conjugate points occur.

by a cosmic time function t are orientable 3-surfaces, that M admits a global system of oriented time lines, and that a non-zero totally skew tensor ϵ_{abcd} can be defined globally. Further, these arguments disallow the intriguing possibility (Markus, [*14*]) that a space-traveller might leave a right-hand glove at home and find on his return that the glove fits his left hand!

A further property one would usually demand is that *M should admit a spinor structure*, i.e. that one could define two-component spinor fields globally on M (Penrose [*10*], Geroch [*2*]) since one might in principle determine a spinor structure (see Aharonov and Susskind [*16*]) if the local properties of spin-$\frac{1}{2}$ particles are everywhere the same. The existence of a spinor structure implies various conditions (Lichnerowicz [*15*], Penrose [*10*], Bichteler [*17*], Geroch [*18*]) besides the existence of a Lorentz metric; in particular it implies (Penrose [*10*]) that M is space and time-oriented,† and (Geroch [*2*]) that if M is non-compact it admits a global field of orthonormal tetrads.

Finally, we note that any non-compact space-time M can be imbedded globally in a Euclidean space R^{89} with signature $(-, -, +, \ldots, +)$ (87 plus signs!) (Clarke [*19*]). Thus the original concept of a manifold as a subspace of a flat space extends to the space-time manifold (M, g) of every reasonable cosmological model.

2. *The Clifford-Klein Space-form Problem*

We now have a list of properties one might demand of any reasonable model of the universe. One way of proceeding further is to examine in detail simple solutions of Einstein's equations which are good models of the observed universe.

In fact, it is known (Hawking and Ellis [*20*], Ellis [*21*]) that‡ the spatially homogeneous, isotropic universe models are very good approximations to the observed universe back until the time of last scattering of the observed microwave background radiation. The argument is a purely local one, following from the isotropy of the radiation and from the Liouville equation governing the propagation of radiation in a curved space-time, and does not give any information regarding the connectivity of space-time (M, g) in the large.

Given any space-time (M, g) there exists a universal covering manifold $\tilde{M}$ with metric $\tilde{g}$ such that $\tilde{M}$ is simply connected and (M, g) is obtained from $(\tilde{M}, \tilde{g})$ by suitably identifying points in $\tilde{M}$ (we shall always label quantities in $\tilde{M}$ with a $\sim$, and corresponding quantities in M by the same symbol with no $\sim$). In fact one recovers (M, g) from $(\tilde{M}, \tilde{g})$ by identifying points equivalent under a group Γ of isometries of $\tilde{M}$ which acts freely and

† Space-time admits a spinor structure if it admits a spin structure (Lichnerowicz [*15*]) and is in addition time and space orientable.

‡ Assuming our galaxy is not at the centre of the universe.

properly discontinuously† (Wolf [22], Section 1.8 and Lemma 2.3.10), i.e. by forming the quotient space $M = \tilde{M}/\Gamma$ where Γ is a properly discontinuous discrete group of isometries without fixed points.‡ Thus given any solution of the field equations which is determined by local properties only, one can first form the universal covering space $(\tilde{M}, \tilde{g})$ and then determine the possible identifications Γ.

Four of the spatially homogeneous isotropic universe models are *homogeneous space-times*: namely the space-times of constant curvature K ($K=0$: Minkowski space F^4, diffeomorphic to R^4; $K=+1$: de Sitter space D^4, diffeomorphic to $S^3 \times R^1$; and $K=-1$: anti-de Sitter space A^4, whose universal covering space $\tilde{A}^4$ is diffeomorphic to R^4) and the Einstein static universe (diffeomorphic to R^4). The possible identifications in the first three spaces, resulting in the *relativistic space-forms*, have been studied in detail (see Auslander and Markus [23] for the case $K=0$, and Calabi and Markus [24] for the case $K=+1$ and $K=-1$; see also Schrodinger [25], Markus [14], Wolf [22]). Wolf has constructed a large class of *twisted Einstein worlds* by studying identifications possible in the Einstein static universe (private communication). However, we shall not pursue these cases further; the systematic Doppler shift observed in the spectra of distant galaxies implies (since we assume General Relativity to be valid) that the density of matter is decreasing and so that space-time is not locally homogeneous. Thus a reasonable way to proceed is to consider the space-times $(\tilde{M}, \tilde{g})$ which are locally the *expanding Friedmann (or Robertson-Walker) universes* (Robertson [27], Bondi [28]) with metric

$$ds^2 = -dt^2 + R^2(t)\, d\sigma^2$$

and in which the 3-spaces $\{t = \text{constant}\}$ with metric $d\sigma^2$ are simply connected complete 3-spaces of constant curvature $k=+1$, 0 or -1 (respectively a 3-sphere S^3, Euclidean space R^3 or hyperbolic space H^3 diffeomorphic to R^3), and then to consider possible identifications in these space-times. This gives a list of those space-times which satisfy locally restrictions of spatial homogeneity and isotropy, and have complete space-sections with the various possible (global) topological structures.

Because these universes are expanding, Γ must map each 3-space

† The group Γ acts *freely* if $\gamma(x)=x$ for any $x \in M$ when $\gamma \in \Gamma$ implies γ is the identity e, and acts *properly discontinuously* if every point $x \in M$ has a neighbourhood U such that $\{\gamma \in \Gamma : \gamma(U) \text{ meets } U\}$ is empty unless γ is the identity, and further whenever $x, y \in M$ are such that there is no $\gamma \in \Gamma$ with $\gamma(x)=y$, there are neighbourhoods U, U' of x, y respectively such that $\gamma(U) \cap U'$ is empty for all $\gamma \in \Gamma$.

‡ If one is to identify points in $(\tilde{M}, \tilde{g})$ equivalent under a diffeomorphism $\gamma : \tilde{M} \to \tilde{M}$, γ must clearly be an isometry. Unless it is the identity it must have no fixed points, for otherwise suppose p is left invariant but q is not. Then a continuous curve λ between p and q will have some point $r \in [p, q]$ on it such that $\gamma(r)=r$ but distinct points of λ in every open neighbourhood of r are not left invariant by γ and are identified with their images. Thus there can be no neighbourhood of r diffeomorphic to R^4.

$\{t = \text{constant}\}$ into itself;† in fact, one need consider only *one* such surface, the identifications in all other surfaces being determined from those in one surface because Γ commutes with the congruence of surface normals. Now any isometry $\gamma \in \Gamma$ which maps the metric h of each 3-space $\{t = \text{constant}\}$ into itself leaves the second fundamental form χ of each surface invariant (as $\chi_{ab} = f(t)h_{ab}$), so the problem reduces to the *Clifford-Klein space-form problem* (Killing [29], Klein [30], Rinow [31]; for all aspects of this problem, see Wolf's book [22], especially Theorems 2.4.9 and 2.4.10) of classifying all space-forms, that is all complete connected Riemannian 3-spaces (metric $(+++)$) of constant curvature. The restriction to connected 3-spaces is made because if there were disconnected components there would be no physical connection whatever between the corresponding disconnected parts of space-time. The restriction to complete 3-spaces is necessary because otherwise what one can do is highly arbitrary; for example, Clarke and Schmidt have given as an example the universal covering space of a flat 3-space from which a line has been removed. This space is locally flat and is diffeomorphic to R^3, but it is incomplete and inextendible.

A word of warning is in order here: identification of points in $\tilde{M}$ via Γ to produce M usually lowers the dimension of the group of isometries of the space. Thus one might wish to know which of the space-times we consider admits the full 6-dimensional group of isometries, and which of them are spatially homogeneous (i.e. admit *any* group of isometries G transitive on the 3-surfaces $\{t = \text{constant}\}$). We shall consider these special cases in our subsequent discussion. The essential point here is that if Γ is used to identify points in $\tilde{M}$ to produce M, this identification must be invariant under any isometries of M, so G must be the centralizer of Γ in the group $\tilde{G}$ of isometries of $\tilde{M}$. If for example a Killing vector field $\tilde{\mathbf{K}}$ generates a 1-parameter group of isometries of $\tilde{M}$ which does not commute with a non-trivial element $\gamma \in \Gamma$, then $\tilde{\mathbf{K}}$ does not project into a vector *field* on M, and (as γ can have no fixed points) the corresponding Killing vectors on M generate local 1-parameter groups of isometries but not complete isometry groups.‡ It follows that the dimension of G can be the same as that of $\tilde{G}$ only if Γ commutes with the isotropy group H_p of each point p, for otherwise p and p' would be identified where p was invariant under H_p but p' was not. It also follows that M is homogeneous if and only if the centralizer of Γ in $\tilde{G}$ (i.e. the set of all $g \in \tilde{G}$ such that $g\gamma = \gamma g$ for all $\gamma \in \Gamma$) is transitive on $\tilde{M}$ (Wolf [22], Theorems 2.4.17 and 1.8.19). Incidentally, it follows (Wolf [22], Theorem 2.7.5) that if M is homogeneous then every element γ of Γ is a Clifford translation of $\tilde{M}$, i.e. is a transformation $\gamma : x \to \gamma(x)$ such that the distance between x and its image $\gamma(x)$ is the same for all $x \in \tilde{M}$.

† Some universe models collapse to a second singularity and then there *do* exist fixed-point free isometries γ of $\tilde{M}$ which map different spaces $\{t = \text{constant}\}$ into each other; however the resulting space-times $\tilde{M}/\Gamma$ are not time-orientable.

‡ If $\tilde{g}_t$ is a 1-parameter group of isometries of $\tilde{M}$, identifying under γ means that $\tilde{g}_t(\gamma(p)) = \tilde{g}_t(p) = \gamma(\tilde{g}_t(p))$ for all t. Thus either $\tilde{g}_t$ and γ commute, or $\tilde{g}_t$ does not project into a group g_t of isometries of M (cf. footnote p. 10).

We consider first the case $k = +1$. The 3-spaces $\{t = \text{constant}\}$ in M are covered by S^3, and so are necessarily compact ('closed' or 'finite'). The space-times M are invariant under two simply transitive† isometry groups of Bianchi type IX, and these groups commute with each other (Ellis and MacCallum [32]). Each of these is in fact a group SU(2) locally isomorphic to the group SO(3) and diffeomorphic to the 3-sphere S^3 (which doubly covers SO(3)). One can identify the representative surface $\{t = \text{constant}\}$ with one of these groups, and can represent it by the unit sphere in R^4; however the connection with SO(3) is clearer if it is represented by two balls S_1, S_2 of radius π in R^3 with antipodal points on their surfaces identified. Then each point (α, θ, ϕ) of S_1 represents a rotation through an angle α about the (θ, ϕ) direction, and each point (α, θ, ϕ) in S_2 represents a rotation through an angle $\alpha + 2\pi$ about the (θ, ϕ) direction: rotations through 2π are *not* identified with the identity but rotations through 4π *are* identified with the identity. As a body is rotated through 4π about the (θ, ϕ) direction, a point representing its position traverses a straight line through the origin of S_1, through the origin of S_2 and back to the origin of S_1. Thus in this representation of the 3-surfaces $\{t = \text{constant}\}$ the orbits through the origins of S_1 and S_2 of 1-parameter subgroups of the group of isometries are the straight lines through these points (which are geodesics in these surfaces) and the metric at these points is the natural Euclidean metric (cf. Schmidt [33]). One regains SO(3) as the real projective space P^3 obtained by identifying points on the straight lines through the centres of S_1 and S_2 which are a distance 2π apart.

The particular usefulness of this representation lies in the fact that the discrete subgroups of the rotations SO(3) about any point in R^3 are well known: they are (Wolf [22], Section 2.6), (i) the cyclic group Z_m of rotations through an angle $2\pi/m$ about any given axis, (ii) the dihedral group D_m, i.e. the symmetry group of any regular m-gon lying in a 2-plane (which consists of Z_m plus reflections in certain planes perpendicular to the given plane); (iii) the polyhedral groups, namely the symmetry groups T of a regular tetrahedron, O of a regular octahedron and I of a regular icosahedron‡ in R^3. In an obvious way these finite rotation groups correspond to sets of points in the Bianchi IX group space which represent the action of the cyclic group Z_m, the binary dihedral group D^*_m, and the binary polyhedral groups T^*, O^* and I^* on the origin in $SU(2)$. Operating by the commuting simply transitive group one obtains corresponding sets of points for the same groups acting on any point in the surface $\{t = \text{constant}\}$ (the origin of the group was represented by an arbitrary point in this surface).

If one identifies the points in the space $\{t = \text{constant}\}$ corresponding under any of these groups, the resulting 3-space is homogeneous because the transformations used in the identifications (all from one of the Bianchi IX

† From now on, by 'transitive' we understand 'transitive on the 3-spaces $\{t = \text{constant}\}$'.

‡ A regular tetrahedron is a regular polyhedron with 4 vertices, 6 edges and 4 faces; an octahedron is that with 6 vertices, 12 edges and 8 faces; and an icosahedron that with 12 vertices, 30 edges and 20 faces. In each case the faces are equilateral triangles.

groups) commute with all the transformations in the other Bianchi IX group. In fact, one obtains in this way *all* the *homogeneous* space-forms with $k=+1$ (Wolf [*22*], Corollary 2.7.2); these are therefore (i) $M^3=S^3$, (ii) $M^3=P^3$, (iii) $M^3=S^3/Z_n$ $(n>2)$, (iv) $M^3=S^3/D^*_m$ $(m>2)$, (v) $M^3=S^3/T^*$, $M^3=S^3/O^*$ or $M^3=S^3/I^*$. This therefore gives all the spatially homogeneous expanding Robertson–Walker universes with $k=+1$. Clearly there are an infinite number of them (as n, m can take any value >2); only in the cases $M^3=S^3$, and $M^3=S^3/Z_n$ with n odd, are antipodal points not identified. The only spaces in which the full group of isometries is six dimensional are $M^3=S^3$ and $M^3=P^3$ (cf. Kobayashi and Nomizu [*3*], Theorem 1, p. 308); this follows because the isotropy group H_p of any point p leaves invariant precisely that point and its antipodal point, so the only identification not breaking this group is the identification of these two points.†

Further identifications are allowed if one drops the restriction that the resulting space be homogeneous; effectively what happens here is that one can combine elements of the type mentioned above from the two commuting simply transitive groups to obtain new discrete isometry groups Γ without fixed points.‡ One obtains (Wolf [*22*], Section 7.5, p. 224), besides S^3 and P^3, spaces in which the group Γ takes one of the forms $Z_n (n>2)$, $Z_u\times D^*_v$ (this can happen in two ways), $Z_u\times T^*_v$, $Z_u\times O^*$, and $Z_u\times I^*$, for certain values of u and v, where T^*_v are subgroups of T^*; all spherical space forms belong to one of these types.

Next we consider the case $k=0$. The 3-spaces $\{t=\text{constant}\}$ are invariant (Ellis and MacCallum [*32*]; but see§) under a simply transitive group of Bianchi type I (the *translations*) and, for each direction at a point, a 1-parameter family of simply transitive groups of Bianchi-Behr type VII_0 (generated by a *screw motion*, i.e. a translation in that direction accompanied by a rotation with pitch angle ϕ about the direction of translation, plus translations in the perpendicular directions). One can choose a discrete subgroup of any 1-parameter subgroup of any of these groups, and identify points equivalent under this discrete subgroup. One can also identify points equivalent under a '*glide reflection*', i.e. a reflection in a plane through the origin followed by translation parallel to the plane. Thus there are three simple types of isometry without fixed points (translation a given distance, a screw motion through a given distance or a glide translation a given distance; again one has in effect combined a translation with elements of the isotropy group of a point to obtain new isometries without

† There is a space of the type mentioned in the footnote on p. 10 in which the full isometry group is six-dimensional; this is the one in which (regarding the space-time as imbedded in five dimensions) space-time points antipodal about the imbedding centre are identified.

‡ In effect one identifies points equivalent under translations of a point combined with elements of the isotropy group of that point.

§ Note that an error occurs in [*32*], p. 128, where the statements (under the headings $R^*=0$ and $R^*<0$) 'a three-parameter family of groups' should read 'a two-parameter family of groups'.

fixed points). All the flat *non-compact* space-forms can be obtained by combining such isometries. (Wolf [22]. Theorem 3.5.1). The orientable space-forms of this kind are (a) type $\mathscr{E}$ ($\Gamma = I$, so $M^3 = R^2 \times R^1 = R^3$); (b) type $\mathscr{J}_1^\theta$ (Γ is generated by a screw motion through an angle θ; $\mathscr{J}_1^0 = (\text{cylinder}) \times R^1 = S^1 \times R^2$); (c) type $\mathscr{T}_1$ (Γ is generated by two independent translations; $M^3 = (\text{torus}) \times R^1$); and (d) type $\mathscr{K}_1$ (Γ generated by a translation and by a screw motion through an angle π in a direction perpendicular to the translation). The 4 non-orientable kinds of flat non-compact space-form, which include manifolds $M^3 = (\text{Möbius strip}) \times R^1$ and $M^3 = (\text{Klein Bottle}) \times R^1$, can be obtained similarly on including glide reflections in Γ.

The flat *compact* space-forms can also be thought of as built up by identifications under these operations, but it is perhaps easier to think of them as obtained by identifying opposite faces in possible crystal lattices. The orientable cases are (Wolf [22], Theorem 3.5.5) those obtained from a translation lattice (i) by identifying opposite sides, obtaining a torus T^3; (ii) by identifying opposite sides, one pair being rotated by π; (iii) by identifying opposite sides, one pair being rotated by $\pi/2$; (iv) by identifying opposite sides, with all pairs rotated by π (in cases (ii), (iii) and (iv) some of the lattice angles are necessarily right angles), and those obtained from the lattice made by translating a hexagonal plane lattice a certain distance perpendicular to the plane, by (v) identifying opposite sides with the top rotated by $2\pi/3$ with respect to the bottom, and (vi) by identifying opposite sides with the top rotated by $\pi/3$ with respect to the bottom. The 4 non-orientable cases can be obtained similarly (Wolf [22], Theorem 3.5.9). Clearly the precise identifications one gets depend on the sizes and angles assumed in the lattice used to determine the identifications.

Of these space-forms, the only ones which are homogeneous are those which are the direct product of a Euclidean space and a torus (Wolf [22], Theorem 2.7.1), i.e. are of one of the forms R^3, $R^2 \times S^1$, $R^1 \times T^2$ or T^3 in which identifications have been made using discrete subgroups of the Bianchi I (abelian) group. The only flat space-form invariant under a full G_6 isometry group is the original manifold R^3 (since the isotropy group of any point leaves no other point fixed; cf. Kobayashi and Nomizu [3], Theorem 1, p. 308). All flat space-forms belong to one of the 18 types mentioned above.

The remaining case is that of space-forms with $k = -1$. The classification of this case has not yet, as far as I am aware, been completed.† We shall consider just three possible kinds of identifications. The 3-space H^3 can be globally imbedded in Minkowski space F^4; for each direction at a point, it is invariant under a simply transitive group of Bianchi type V (translations of H^3) and a 1-parameter family of simply transitive groups of Bianchi-Behr type VII_h (a screw motion plus translations, cf. the case $k = 0$). One can again identify points under any discrete subgroup of any 1-parameter

† In fact, even the classification of 2-dimensional space-forms of constant negative curvature appears to be incomplete.

subgroup of these isometry groups, or under a glide reflection associated with the translations. Secondly, it is known (Ellis [*34*]) that the metric of a 3-space of constant negative curvature can be written locally in the form

$$d\sigma^2 = dr^2 + \cosh^2 r \, d\tau^2 \tag{2.1}$$

where $d\tau^2$ is the metric of a 2-space S of constant negative curvature. Now there are infinitely many such complete connected 2-surfaces (or 2-dimensional space-forms); for example Klein [*30*] shows how one can obtain compact complete 2-surfaces of constant negative curvature diffeomorphic to a 2-sphere with any number of handles attached. Taking any such 2-surface S, one obtains from (2.1) a complete 3-surface of constant negative curvature where $M^3 = R \times S$. Finally an elegant paper by Löbell [*35*] shows how one can find a set of 14-sided figures (2 sides are regular rectangular hexagons and 12 are rectangular pentagons) which fit together to fill H^3 just once.† Then one can proceed to construct compact 3-spaces of constant negative curvature by suitably identifying sides (cf. the procedure in the case $k = 0$); Löbell shows how to construct infinitely many different compact (orientable or non-orientable) 3-spaces in this way.

Thus although one does not have a complete solution to the space-form problem with $k = -1$, one knows there are very many such space-forms. However the only homogeneous such space-form is H^3 itself (Wolf [*22*], Theorem 2.7.1; cf. Yano and Bochner [*37*], theorem 2.10) and so *a fortiori* this is the only 3-space of constant negative curvature admitting a full G_6 group of isometries.

The complexity of these topological structures contrasts strikingly with the simplicity of the local metric properties of these space-times. In fact the ways in which points can be identified in space-times with rather more complicated metrics are rather simpler. Thus the next simplest realistic cosmological models after the spatially isotropic and homogeneous universes, are those which are spatially homogeneous but not isotropic, and the simplest of these are those space-times filled with a perfect fluid in which there is a group of isometries transitive on 3-surfaces orthogonal to the fluid flow vector. With the exception of the Kantowski–Sachs spaces [*38*] of class I, all such spaces have a subgroup of isometries simply transitive on the homogeneous 3-surfaces. The covering manifold $\tilde{M}^3$ in the Kantowski–Sachs spaces is $S^2 \times R^1$; for all the other such spaces $\tilde{M}^3$ is R^3, except for those invariant under a group of Bianchi type IX when $\tilde{M}^3$ is S^3 (Schmidt [*33*]). Now if the group is 4-dimensional then any point is invariant under a 1-parameter isotropy group, but in general the group is 3-dimensional and each point is invariant only under a discrete isotropy group (cf. MacCallum and Ellis [*39*], Schmidt [*33*]). Thus the space-form problem for these spaces is simpler than the corresponding problem in the Robertson–Walker spaces for there are similar discrete 'translations' but fewer reflections and rotations which can be combined with these translations.

For example, in the Kantowski–Sachs case, one can (a) identify points

† Cf. the recent paper by Garner [*36*].

under a translation in the R^1-direction, or under a translation in this direction plus a rotation or reflection, (i.e. a screw motion or glide reflection in this direction), thus producing compact space-times. One can also (b) identify antipodal points in each 2-sphere S^2; however, this is the maximum identification allowed in any two-sphere (cf. Wolf [22], Theorem 2.5.1). Thus the only other identifications allowed are (c) a combination of (a) and (b). If the group is of Bianchi type I, the group Γ is that subgroup of the group Γ discussed above in the Robertson–Walker ($k=0$) case which preserves the second fundamental form of the surfaces $\{t=\text{constant}\}$. In general the second fundamental form is invariant only under discrete isotropies through an angle π about its principal axes. Thus the non-compact orientable manifolds can in general† only be of type $\mathscr{E}$, $\mathscr{J}_1{}^0$, $\mathscr{J}_1{}^\pi$, $\mathscr{T}_1$ and $\mathscr{K}_1$, while the compact orientable manifolds can in general only be of types (i), (ii) and (iv). Similarly if the group is type IX, one can make the same identifications as those listed above for homogeneous space-forms with $k=+1$, and in general can only supplement these identifications with certain additional identifications corresponding to glide reflections or screw motions through π.

The global properties of all of these spaces are easily obtained as $M=M^3\times R$ and $M^3=\tilde{M}^3/\Gamma$. The identifications we are discussing would have two particular effects—first, they would affect the occurrence of *particle horizons* in the universe, and this could be important (Misner [40]). Second, they would be *observable* in principle, as an observer would in general see each galaxy (including his own galaxy) simultaneously in several different positions in the sky.‡ However observational verification of whether this had actually been observed or not would be rather difficult to obtain, since in general each image of the same galaxy would be an image of the galaxy seen from a different direction and at a different stage of its evolution. (Cf. attempts to verify if the antipodal galaxy could have been observed in universes with $k=+1$ and $\Lambda>0$ (Solheim [41]).) Note that if identifications are made, one can in general see round the universe even if $\Lambda=0$, cf. Audretsch and Dehnen [42]).

It is clear from the above discussion that (cf. Heckmann and Schucking [26]) a Robertson–Walker universe model can have compact spatial sections even if $k=0$ or -1; thus the question of whether the universe is spatially finite (i.e. the 3-spaces $\{t=\text{constant}\}$ have a finite volume) or not, and whether there are a finite number of particles in the universe or not, cannot necessarily be settled simply by determining k for an idealized universe model. While some of the topological structures discussed here may seem rather unlikely, they cannot simply be dismissed out of hand; the only way to determine the topological structure one should give models of the observed universe is by direct observation.

† The exceptions are the cases where the expansion quadric is rotationally symmetric; then the space admits a group of isometries G_r where $r>3$.

‡ Except in the cases $M^3=R^3$ and $M^3=P^3$, these space-times would not be spatially isotropic.

3. *The Observable Universe*

While the idealized universe models discussed in Section 2 are both useful and interesting in their own right, they clearly do not correspond exactly to the (macroscopic) observed universe. In particular, although the universe is very like a Robertson–Walker universe at recent times, it was probably (or at least, conceivably) not like a Robertson–Walker universe at the early times for which we have no direct observational evidence (Ellis [21]). We can have no causal connection with space-time points outside our causal past†, so any deductions one may wish to draw about the global structure of the universe have to be inferred from evidence from this rather limited region of space-time.

Since one has reason to believe that at early times in the universe's history it was filled with a hot dense ionized gas, there is for every observer a surface $\mathcal{S}$ in space-time (say, the surface on which the optical depth reaches unity) which is a surface of 'last scattering' of light; he cannot obtain direct information about earlier times by optical or radio observations, since light emitted at earlier times will have been multiply scattered or absorbed by the intervening plasma. We will write E^- for the past light cone of our present position p in space-time, I^- for the (chronological) past of p, $\mathcal{C}^-$ for that part of E^- containing p and bounded by $\mathcal{S} \cap E^-$, and $\mathcal{J}^-$ for that part of I^- between p, E^- and $\mathcal{D} \equiv \mathcal{S} \cap I^-$. The part of the universe we can observe directly is then $\mathcal{C}^-$. The part of the universe we can hope to determine in considerable detail is $\mathcal{J}^-$.

In the Robertson–Walker universes discussed in Section 2, $\overline{\mathcal{C}^-}$ is compact as it is covered by the unit ball in R^3, its actual topology being determined by the identifications on $\overline{\mathcal{C}^-}$ under Γ. $\mathcal{D}$ is diffeomorphic to $\overline{\mathcal{C}^-}$, and $\mathcal{J}^-$ is topologically $[0, 1] \times \mathcal{C}^-$. In more realistic universe models, one might still expect $\mathcal{D}$ and that part c of $\mathcal{C}^-$ which forms the boundary of $\mathcal{J}^-$ to be compact. If further $\overline{\mathcal{J}^-}$ is compact (one might expect this if the distribution of radiation is sufficiently uniform) then a theorem of Geroch (Geroch [8], theorem 2; cf. also Kundt [43]) can be adapted to show that (since M is causal) $\bar{c}$ and $\mathcal{D}$ are homeomorphic and that $\mathcal{J}^-$ is topologically $(0,1) \times c$. Hence one could determine the topological structure of $\mathcal{J}^-$ at least, since one could (in principle, cf. remarks in Section 2) determine the topological structure of $\mathcal{C}^-$ by direct observations. However, the probable existence of isolated singularities in the universe suggests that this picture might be false in two different ways, if one examines the space-time structure on a fairly fine (macroscopic) scale.

First,‡ if a 'black hole' crossed $\mathcal{C}^-$, it would probably deform it into a cusp or even tear a hole in it so that $\overline{\mathcal{C}^-}$ was non-compact; our present understanding of the evolution of massive objects indicates that some of

† Although the constraint equations do (cf. Gauss' theorem in Newtonian theory) convey *some* information about space-time outside our causal past.

‡ One can obtain exact solutions corresponding to both these possibilities, namely the 'Swiss-Cheese' universes (cf. Rees and Sciama [46], Kantowski [47]). The high isotropy of background radiation places limits on the size and density contrast of such fluctuations.

these objects must evolve to a collapsed state fairly rapidly, so that some such collapsed objects are likely to exist in the universe (Hawking and Ellis [20], Penrose [10]). Secondly† fluctuations in a Robertson–Walker universe are of two types, the relatively increasing and the relatively decreasing (Sachs and Wolfe [44], Rees [45]). There appears to be no *a priori* reason why one kind should be more prevalent than the other; however following a 'relatively decreasing' perturbation back in time, $\delta\rho/\rho$ ($\delta\rho$, >0, is the fluctuation in the density ρ) increases and so such a fluctuation would fall inside its Schwarzschild radius and become singular before the rest of the universe had collapsed. Thus such fluctuations would be capable of deforming or tearing $\mathcal{S}$. However, it is perhaps unlikely that singularities of this kind would have formed within $\mathcal{I}^-$.

It may be taken as well established that a singularity must occur somewhere in I^- in the sense that necessarily at least one timelike geodesic from p into the past is incomplete (Hawking and Ellis [20], Kundt [48], Hawking and Penrose [49], Penrose [10]). The nature of the singularity is not yet clear (Geroch [40], Schmidt [51]) nor how extensive it is (e.g. what proportion of the timelike geodesics through p are incomplete). If some past timelike geodesics through p evade the singularity (i.e. if they are complete in the past time direction) then it is unlikely that there would be a Cauchy surface for I^-. If this were so, there would be no Cauchy surface in M (equivalently, M would not be globally hyperbolic, cf. Geroch [52]) and one could think of the singularity as having a timelike character.

We do not know at present of any kind of observation which could conclusively show whether this was so or not. A statement as to whether the singularity is probably timelike or not has therefore to rely on our knowledge of general features of the known solutions of Einstein's equations and of possible pathological space-times. However the existence of highly pathological space-time models cannot be taken to indicate that there exist similar regions in the actual universe; in the absence of observational evidence, one can only attempt to evaluate what properties probably exist in the physical universe by deciding what properties are generic (cf. S. Hawking's paper‡). Thus one need not take all pathological space-times as indicating probable properties of the physical universe but only reasonably pathological space-times! In fact, one may argue (Hawking and Ellis [20]) on the basis of the probable inhomogeneity of the early universe and of the nature of the Kerr and Reissner–Nordström solutions, that it is likely that the singularity in our past is timelike; then further probable consequences are that some points in the past I^- of p cannot be joined to p by timelike geodesics (cf. Seifert [53]; for example there are pairs of points in the Reissner–Nordström solution which can be joined by timelike lines but not by timelike geodesics), and that some points in I^- which can be joined to p by a timelike geodesic cannot be joined to p by a maximal timelike curve

† See footnote ‡ on p. 17.
‡ *GRG Journal* (1971) 1, 393.

(cf. Boyer [54]). Under these conditions it seems difficult to decide whether the boundary of $\overline{I^-}$ is likely to consist simply of E^- or not, and whether particle horizons are likely to exist in our past or not.

In general it seems difficult to make any precise statements whatever that are in some sense verifiable about the part of M outside I^-. (Note here the warning contained in Misner's paper [55] which shows that even analytic continuation is not necessarily unique; in fact, Geroch has shown that there are many incomplete inextendible analytic extensions of any space-time which has one analytic extension.) Thus in general one cannot hope to investigate the *global* structure of the universe† but only its *large scale* structure. However there are exceptional cases: suppose the Robertson–Walker universe which gives the best representation of the structure of the universe since $\mathscr{S}$ has compact spatial sections. Then it might be that ϵ^- intersects the world lines of all galaxies at present existing in the universe, i.e. the particle horizon may have been removed at some (relatively late) time because of the compactness of the spatial sections. In this case (one could in principle observe whether this occurs or not by seeing whether the same galaxies are observed in different directions or not; but cf. remarks in Section 2) one would have sufficient information on ϵ^- to determine the complete evolution of the idealized Robertson–Walker universe and its contents—by a slight deformation of ϵ^- one would obtain a Cauchy surface and would be able to determine the initial values of all physical fields on this Cauchy surface. Even if a closer look revealed that space-time did not possess a Cauchy surface, one would still have sufficient information to determine the evolution of a complete branch of the universe for a finite time (unless ϵ^- was 'torn' by collapsed objects, cf. remarks above).

Clearly it is an interesting question to ask whether the universe has any compact spatial sections.‡ In the Robertson–Walker case, $k = +1$ implies that such sections do exist, but as we have seen in Section 2 they can exist even if $k = 0$ or -1. The sufficient condition can be generalized to any space-time, since if there exists any complete spacelike 3-surface S in M such that for every unit vector $\mathbf{n}$ tangent to S, $R^3{}_{ab}n^a n^b \geqslant \epsilon > 0$ ($R^3{}_{ab}$ is the Ricci tensor of S) where ϵ is some constant, then S is compact and has diameter less than or equal to $\pi\sqrt{(2/\epsilon)}$ (see Milnor [58], Lemma 19.5). The Gauss–Codacci equations determine $R^3{}_{ab}$ from R_{ab} and from the second fundamental form of S; roughly speaking, these equations imply that if

† One might want to ask, for example, whether M is diffeomorphic to R^4 (i.e. can be covered by one single coordinate neighbourhood) or not. One feels reasonably sure this cannot be true.

‡ Cf. Einstein [56], Avez [57]. When such submanifolds exist one can apply standard theorems valid for compact manifolds, e.g. Bass and Witten [7]. The example of de Sitter space shows that families of complete, simply connected imbedded 3-spaces of constant curvature $k = +1$, 0 and -1 can all occur in the same space-time. Bardeen has given an example of a causal, complete and connected space-time in which there is a cosmic time function t such that some space-sections {t = constant} are compact and diffeomorphic to S^3, and some are compact and consist of m disconnected components where m is any positive integer.

there is sufficient energy-density present on S then this energy density causes this 3-surface to close up spatially. It has so far proved impossible to determine unambiguously by astronomical observations whether sufficient matter is present to close up spatially an exact Robertson–Walker model of the universe (cf. Sciama [59]) or not, so the somewhat more delicate argument needed in the case of an almost Robertson–Walker universe cannot at present be pushed to a firm conclusion either.

This section has discussed some of the qualitative questions about the physical universe one might hope to settle by observation in conjunction with suitable theoretical developments. Finally I should mention that all along I have been concerned only with macroscopic scales; much more complex problems could arise when one considers space-time structure on a fine enough scale (cf. Professor Wheeler's discussion, for example).

Acknowledgments

I should like to thank S. W. Hawking, C. J. S. Clarke and R. P. Geroch for helpful discussions, and M. A. H. MacCallum for useful comments on the manuscript.

References

1. Whitney, H. (1936). *Annals of Maths.*, **37**, 645, 865.
2. Geroch, R. (1968). *J. Math. Phys.*, **9**, 1739.
3. Kobayashi, S. and Nomizu, K. (1963). *Foundations of Differential Geometry*, Interscience, New York.
4. Steenrod, N. (1951). *The Topology of Fibre Bundles*, Princeton.
5. Markus, L. (1955). *Annals of Maths.*, **62**, 411.
6. Fierz, M. and Jost, R. (1965). *Helv. Phys. Act.*, **38**, 137.
7. Bass, R. and Witten, L. (1957). *Rev. Mod. Phys.*, **29**, 452.
8. Geroch, R. (1967). *J. Math. Phys.*, **8**, 782.
9. Kronheimer, E. and Penrose, R. (1967). *Proc. Camb. Phil. Soc.*, **63**, 481.
10. Penrose, R. (1968). 'Structure of Space-Time'. In *Battelle Rencontres* (Ed. C. de Witt and J. Wheeler). Benjamin, New York.
11. Hawking, S. (1968). *Proc. Roy. Soc. A*, **308**, 433.
12. Geroch, R. (1967). Ph.D. Thesis, Princeton University.
13. Zeldovich, Ya. B. and Novikov, I. D. (1967). *J.E.T.P. Letters*, **6**, 236.
14. Markus, L. (1965). *Rev. Roum. Math. Pures et Appl.*, **X**, 1089.
15. Lichnérowicz, A. In *Battelle Rencontres* (see [*10*]).
16. Aharonov, Y. and Susskind, L. (1967). *Phys. Rev.*, **158**, 1237.
17. Bichteler, K. (1968). *J. Math. Phys.*, **9**, 813.
18. Geroch, R. (1970). *J. Math. Phys.*, **11**, 343.
19. Clarke, C. J. S. (1970). *Proc. Roy. Soc. A*, **314**, 417.
20. Hawking, S. W. and Ellis, G. F. R. (1968). *Ap. J.*, **152**, 25. See also a forthcoming book to be published by the Cambridge University Press.
21. Ellis, G. F. R. (1969). Lectures at International School of Physics, Varenna, Corso XLVII (to be published).
22. Wolf, J. A. (1967). *Spaces of Constant Curvature*, McGraw-Hill, New York.
23. Auslander, L. and Markus, L. (1958). *Memoir* 30, American Math. Soc.
24. Calabi, E. and Markus, L. (1962). *Annals of Maths.*, **75**, 63.
25. Schroedinger, E. (1956). *Expanding Universes*, Cambridge University Press.

26. Heckmann, O. and Schucking, E. (1959). 'Newtonsche und Einsteinsche Kosmologie', *Handbuch d. Physik*, LIII, 489; (1962). 'Relativistic Cosmology'. In *Gravitation*. (Ed. L. Witten). Wiley, New York.
27. Robertson, H. P. (1933). *Rev. Mod. Phys.*, **5**, 62.
28. Bondi, H. (1960). *Cosmology*, Cambridge University Press.
29. Killing, W. (1891). *Math. Annalen*, **39**, 257.
30. Klein, F. (1928). *Vorlesungen uber Nicht-Euklidische Geometrie*, Springer.
31. Rinow, W. (1961). *Die Innere Geometrie der Metrischen Räume*, Springer.
32. Ellis, G. F. R. and MacCallum, M. A. H. (1969). *Comm. Math. Phys.*, **12**, 108.
33. Schmidt, B. (1969). *Comm. Math. Phys.*, **15**, 329.
34. Ellis, G. F. R. (1967). *J. Math. Phys.*, **8**, 1171.
35. Lobell, F. (1931). Ber. Verhand. Sächs. Akad. Wiss. Leipzig, *Math. Phys., Kl.*, **83**, 167.
36. Garner, C. (1970). *Proc. Roy. Soc. A*, **316**, 441.
37. Yano, K. and Bochner, S. (1953). 'Curvature and Betti Numbers', *Annals of Maths Study No. 32*, Princeton University Press.
38. Kantowski, R. and Sachs, R. K. (1966). *J. Math. Phys.*, **7**, 443.
39. MacCallum, M. A. H. and Ellis, G. F. R. (1970). *Comm. Math Phys.*, **19**, 31.
40. Misner, C. W. (1969). *Phys. Rev. Lett.*, **22**, 1071; (1970). Article in *Relativity*, Plenum Press.
41. Solheim, J. E. (1968). *Nature*, **217**, 41 and (1968). **219**, 45; see also Dehnen, H. and Hönl, H. (1968). *Naturwissenschaft*, **55**, 413; Reinhardt, M. Von, (1969). *Naturwissenschaft*, **56**, 323; Petrosian, V. and Ekers, R. D., (1969). *Nature*, **224**, 484.
42. Audretsch, J. and Dehnen, H. (1969). *Astron. und Astrophys.*, **3**, 252.
43. Kundt, W. (1967). *Comm. Math. Phys.*, **4**, 143.
44. Sachs, R. K. and Wolfe, A. M. (1967). *Ap. J.*, **147**, 73.
45. Rees, M. (1969). Lectures at Varenna, see [*21*].
46. Rees, M. and Sciama, D. W. (1968). *Nature*, **217**, 511.
47. Kantowski, R. (1969). *Ap. J.*, **155**, 89.
48. Kundt, W. (1968). 'Recent Progress in Cosmology', *Springer Tracts in Modern Physics*, **47**, Springer.
49. Hawking, S. W. and Penrose, R. (1970). *Proc. Roy. Soc. A*, **314**, 529.
50. Geroch, R. (1968). *J. Math. Phys.*, **9**, 450; (1968). *Annals of Physics*, **48**, 526; (1970). Article in *Relativity*, Plenum Press.
51. Schmidt, B. Gwatt conference talk.
52. Geroch, R. (1970). *J. Math. Phys.*, **11**, 437.
53. Seifert, H. J. (1967). *Comm. Math. Phys.*, **4**, 324; (1967). *Zs. fur Naturfor.*, **22a**, 1356; Ph. D. Thesis, Hamburg University (1968).
54. Boyer, R. H. (1964). *Nuovo Cimento*, **33**, 345.
55. Misner, C. W. (1967). In *Lectures in Applied Maths*, Vol. 8 (Ed. J. Ehlers, Am. Math. Soc.).
56. Einstein, A. In *The Principle of Relativity*, Dover Publications Reprint (New York).
57. Avez, A. (1962). *C.R. Acad. Sci. t.* **254**, 3984; Thesis, University of Paris Faculty of Science (1964). See also Demianski, M. and Woronowicz, S. (1967). *Bull. Acad. Polon. Sci.*, **XV**, 355.
58. Milnor, J. (1963). 'Morse Theory', *Annals of Maths Studies, No. 51*, Princeton University Press.
59. Sciama, D. W. (1969). Lectures at Varenna, see [*21*].