On the finiteness and shape of the Universe

Warren D. Smith
WDSmith@fastmail.fm

August 13, 2004

Abstract — **Previous plausible theoretical assumptions about the cosmic 3-manifold, such as isotropy, orientability, and compactness, have been unable to reduce the number of candidate topologies to a finite set. We now consider several new possible assumptions inspired by relationships between microscopic physics and cosmic topology. The most important are**

1. **"No-twist" assumption that there does not exist a twisted closed geodesic (to allow photons to exist in momentum-polarization eigenstates),**
2. **At most one isotopy class of nonseparating surface exists (related to charge quantization and seems necessary to allow charge to exist),**
3. **Orthogonal and/or commuting smooth vector fields exist, either locally or globally (may be needed to generalize quantum mechanics to curved spaces).**

We also introduce "1-curvature homogeneity," a weakened version of the common "constant curvature" assumption. We show that various combinations of these assumptions *are* powerful enough to winnow the candidate topologies down to a finite set.

We also present new "reasons for the 3-dimensionality of the universe" and a new argument the universe is spatially finite.

Contents

1 Introduction

This paper is about the topology of the universe. It is presently difficult or impossible to determine that topology by observations. Probably the best such attempt was made by Cornish et al. [43], who did an enormous computer search for "circle in the sky" statistical anomalies (caused, essentially, by multiple views of the same stuff) in the cosmic microwave background (CMB) sky. They found nothing. This, they report, strongly suggests that the shortest closed geodesic in the universe is at least 78×10^9 light years long. But no matter how much observing and computing is done in future, they do not believe their techniques will be able to tell *anything* about the topology of the universe if its shortest closed geodesic exceeds 92×10^9 light years. Furthermore, if the universe is very large, then we dispute all other claims to date of success at restricting the topology of the universe by examining CMB fluctuation statistics.

But it is possible to make progress purely by thinking. Continuing the theme of the preceding two papers [151][150], we argue that

1. topology, combined with
2. some fairly straightforward postulates one might think would arise from microscopic physics,

can together produce surprisingly strong conclusions about the shape of the universe. These techniques remain equally valid regardless of the universe's size.

Unfortunately, it is presently unclear *which* of the postulates we propose are valid (and sometimes also unclear precisely how they should be phrased). We leave that decision to the reader. A chart at the end of the paper (§10) summarizes everything and allows the reader to choose his favorite assumption subset and see what conclusions result. Because some assumptions are more probable than others, there is also another chart there giving my personal estimates of the odds of each.

Some notes to the reader. §2 reviews both our old and our new arguments and results. After that come detailed individual discussions of our new arguments. Most sections end with a concise "summary" and should be readable largely independently.

Some of our arguments cannot be understood without preparatory material. Specifically, to understand §7 and especially §8 it helps to appreciate the "Thurston Geometrization Conjecture" (TGC), the notion of "k-curvature homogeneity," and the "Seifert fiber spaces," all described in §6. To best appreciate §8 one should have read the previous [151]. The main proof in §7 depends on notions of "ergodicity" of various kinds of "flows"; for which see §7.3. The material in §4-5.2 requires an appreciation of "commuting" vector fields and "orthogonal curvilinear coordinate systems"; we provide extensive reviews.

1.1 Notation

Our tensor notation imitates [113].

S^n is the n-dimensional manifold that is the surface of a sphere in $\mathbb{R}^{n+1}$. Superscripts (here n) and prefixes ("n-manifold," "n-sphere," "n-geometry") denote intrinsic dimension. The n-dimensional torus T^n is the cartesian product of n copies of S^1. The real and complex projective planes are $\mathbb{R}\mathrm{P}^2$ and $\mathbb{C}\mathrm{P}^2$; the latter is 4 (real) dimensional. We shall often use these names to mean "or any manifold diffeomorphic to these." We shall sometimes use M^2 to mean "an (unspecified) 2-manifold of constant curvature." A "geometry" is a *locally homogeneous* manifold, i.e. one in which neighborhoods of any two points are isometric. (Weakened notions such as "k-curvature homogeneous," are defined in §6.) A "hyperbolic" manifold means one with constant negative curvature. Some arguments favor a *flat 3-torus* T^3 (parallelipiped with opposite faces identified) as the shape of the universe. There are 4 variants of T^3 got by possibly allowing some of the parallelipiped sidelengths to be infinite. For brevity we shall denote the entire set of them as "T^3_∞."

A smooth k-manifold is "parallelizable" if there exist k everywhere-orthonormal smooth tangent vector fields on it. Two manifolds A,B are "homeomorphic" if some continuous function with continuous inverse maps each point of A to each point of B. (This function is the homeomorphism; if differentiable then it is a "diffeomorphism.") They are "homotopic" if they are homeomorphic and there is a continuous function $F(a,t)$ ("the homotopy") for $a \in A$ and $0 \le t \le 1$ with $F(a,0) = a$ and $F(a,1)$ being *the* homeomorphism from A to B, while $F(a,t)$ is a homeomorphism for each fixed t with $0 \le t \le 1$. Two objects A,B (e.g. submanifolds) embedded in a manifold M are "isotopic" if there is a homotopy from A to B with every intermediate-t version also being embedded.

2 Summary of arguments and results

We shall regard the universe as a smooth connected boundaryless Riemannian 3-manifold[1]. The question we address is: what can we say about its topology, geometry, and finiteness?

The unidirectionality of time and the chiral "left handedness" of microscopic physical phenomena (e.g. neutron decay [176]) leave little doubt that the universe is an **orientable** manifold.

The observed angular uniformity of the distance-ℓ sky (over a large range of particular distance scales ℓ) combined with the assumption of the non-specialness of the Earth's location forces [150] the universe to be a "**harmonic** manifold," which for 3-manifolds is the same thing as both an "Einstein manifold" and a manifold of constant curvature. The standard Friedmann-Robertson-Walker cosmical model [113] had of course postulated a universe of constant curvature since the 1920s (motivated by philosophical desires for symmetry and the desire to simplify the analysis), but [150] gave perhaps for the first time a theorem supplying the logical connection needed to justify that assumption via *experimental evidence*.

However, with this understanding comes the realization that that experimental evidence is *insufficient*. If angular sky uniformity is *not* perfect, then the universe could be a 3-manifold of slightly *non*-constant curvature, e.g. a non-isotropic universe such as $S^2 \times S^1$. This is still entirely compatible with present observational evidence *if* the size of the $S^2 \times S^1$ is large enough. No amount of sky-uniformity observational evidence alone could ever suffice to rule that out; the most it could do would be to force larger and larger size lower bounds[2]. Assuming the universe has such an anisotropic topology (but nowadays looks isotropic at the length scales at which we can observe it) might be quite useful, because it gives extra freedom to make the early universe do more things which exploit the anisotropy. Thus many people might wish either to drop the assumption of constant curvature entirely (which would devastate attempts to restrict the universe's topology) or merely to replace it with some weaker postulate, such as "1-curvature homogeneity" (defined in §6).

As far as present-day measurements can tell, the universe is exactly **charge-neutral**. The positive charges (the vast majority in the form of $+2/3$-charged up-quarks) are balanced by the negative charges (the vast majority of which are electrons, with charge -1, and down-quarks with $-1/3$ charges) to a precision of better than 1 part in 10^{34}, otherwise Coulombic repulsion would exceed gravitational attraction and no stars or galaxies (assuming they are made of the same particle-mixture as the universe as a whole) would exist[3]. A priori,

[1] Note that this ignores time. However, the previous papers [151][150] have addressed that issue to some extent, and we shall also in our §3 and §9. A $(3+1)$-manifold which is foliated by spacelike 3-manifolds has often been called "totally hyperbolic" in the general relativity literature.

[2] On the other hand if it somehow became known that the universe was small, it *then* might become possible to prove its isotropy.

[3] There perhaps is reason to suspect the balancing is better than 1 part in 10^{37}: otherwise either electrons or positrons would be repelled from galaxies, assuming galaxies had the same charge distributions as the universe as a whole.

this is very surprising. Consider the $\approx 10^{79}$ charged particles in the observable part of the universe. One would a priori have expected, with high probability, that exactly a 1/3 fraction of each type would have been chosen, to precision 1 part in 10^{39}, leading to an enormous negative charge. But what is found is a 1:2:1 ratio.

Why is the universe so neutral? If the universe is compact and Maxwell's equations hold and have a solution, then charge neutrality is logically forced by Gauss' divergence theorem. This explanation of an otherwise perhaps-inexplicable fact, suggests that the universe is **compact**.[4]

We present in §3 an independent and new argument, based on the nature and effect of mass fluctuations, apparently also proving that the universe must have fintie volume.

But merely demanding that the universe be a compact orientable 3-manifold of constant curvature – and the compactness and constant curvature assumptions both are debatable – would still leave us with an infinite set of candidate topologies. This falls far short of the goal of reducing the candidates down to a *small finite* set. The purpose of the present paper is to introduce and analyse three **new assumptions** powerful enough to accomplish that goal:

1. "No twist" postulate (related to existence of photon momentum-polarization eigenstates),
2. Postulate of the nonexistence of two nonisotopic complete nonseparating surfaces (related to charge quantization),
3. Postulate of the local or global existence of orthogonal and/or commuting vector fields (related to quantum mechanics in curved space).

Our first such powerful postulate [150] is that the universe should not contain any closed geodesic with the property that parallel translation along one cycle of it induces a "twist" different from an integer multiple of 2π radians. Because: if it did then electrons taking a hypothetical journey round the universe along that geodesic could not exist in a momentum-spin eigenstate. Similarly, photons could not exist in a momentum-polarization eigenstate. Both of these, apparently, would contradict experimental reality[5] (although admittedly all such "experiments" necessarily are of the gedanken sort). This **no-twist postulate** alone is powerful enough to rule out a large number of candidate topologies. In fact we shall see in §7 that it rules out *every* compact orientable 3-geometry except for T^3, S^3, $S^2 \times S^1$, $M^2 \times S^1$, and "symmetric $T^2|S^1$ bundles."

Now [150] consider the highly publicized analysis [158] of WMAP satellite measurements of the cosmic microwave background radiation, which suggested that the Earth may be sitting on a short **closed geodesic** pointing roughly toward Virgo. If so, then under the assumption of the "nonspecialness of the Earth's location," we would conclude that a *positive measure* subset of locations in the universe are sufficiently Earthlike as to have at least one, and at most a countably infinite number, of bounded-length closed geodesics passing through them. This and the previous "no-twist" assumption suffice [150] to eliminate *all* the constant-curvature candidate topologies except for T^3_∞. The only leap here is believing that the massaged WMAP data really shows a short closed geodesic; neither this data, nor (especially) this interpretation of it, are tremendously convincing at present.[6]

Nevertheless, this deduction of flatness happens to jibe with recent supernova and CMB-fluctuation measurements suggesting the universe has low |curvature| – at least at the length scales small enough for us to see, corresponding to light travel times of order $\lesssim 10^{10}$ years.

A different possible interpretation of the WMAP data would be that it is the first observation of the fact that our universe actually is *anisotropic*.

Topology was employed in an entirely different manner in [151] to consider the question "**why do all electrons have the same charge?**" An explanation once proposed by P.A.M.Dirac had been "at least one isolated magnetic monopole exists." This, Dirac showed, most convincingly by considering the behavior of solutions of his electron wave equation, would in various ways logically force quantization of charge. One may attack that explanation both on experimental grounds (no monopole has ever been found, despite many searches) and on theoretical ones: monopoles would necessarily have some extreme and extremely peculiar properties, arguably could never be created in isolation no matter how much energy was supplied [151], according to an argument of S.Weinberg [169] would be acausal, according to an argument of C.Fronsdal [61] would be logically forbidden, and indeed [151] described how point monopoles are forbidden by Dirac's equation itself.

Instead [151] proposed an alternative explanation of charge quantization, based on a small "topologically trapped magnetic field" going around some uncontractible loop on the universe 3-manifold. This field was postulated to have been created during the birth of the universe, and thereafter could never vanish. It was shown in [151] that this too would logically force quantization of charge. The charge quantum would be a topological invariant. The question then becomes: which topologies for the universe allow the charge-quantization argument of [151] to work? (And, indeed, if the universe is one of those topologies, then the onus would be on opponents of this idea, to explain why the trapped magnetic field should miraculously be exactly zero.) Smith showed that the argument works for *every* compact orientable 3-manifold topology (simplest examples: the flat 3-torus T^3 and $S^1 \times S^2$) *except* for the "rational homology spheres" (simplest examples: the 3-

[4] On the other hand, if some ultra-precise measurement ever reveals the universe is *not* exactly charge neutral, we would then know it is not compact.

[5] For example, the existence of photon momentum-polarization eigenstates seems essential to get the thermodynamic Planck blackbody distribution right.

[6] More recent drafts of [158] have stepped away from the "short geodesic interpretation" (which had been touted in newspaper articles at the time as the "bagel shaped universe") indeed arguing that searches for "circles in the sky" anomalies centered on the Virgo bidirection (ala [43]) have refuted that hypothesis [125]. Our view is that they have not refuted it (because the geodesic might be a little too long for those techniques to be applicable, but still short enough to cause the anomaly) but they *have* cast doubt on it (because the plausible length range now seems much more restricted).

sphere S^3, the Poincaré homology 3-sphere[7] and the Brieskorn homology 3-spheres[8].) It can be further argued [151] that, were the universe to contain *two or more* disjoint nonseparating surfaces (with uncontractible cycles, generically containing nonzero magnetic flux, passing through each, and neither distortible to merge into the other) that would force charges simultaneously to obey two *different* quantization conditions – which in the generic[9] case where the ratio of the two quanta was irrational would prevent any charges from existing at all, contradicting experiment.

This all suggests our second powerful postulate: that the universe must have exactly *one* such nonseparating surface thus explaining charge quantization while nevertheless generically permitting charges to exist[10].

This would permit the universe to be $S^1 \times S^2$, but, as §8 shows, eliminates a vast number of candidate topologies (including T^3, the apparently [151] simplest example!).

One may ask whether there is some topological reason for the **3-dimensionality of the universe**. One interesting property of the number "3" is that it is the largest dimension n such that every n-manifold may be triangulated. The fact that 2-manifolds always may be triangulated, i.e. converted into a homeomorphic piecewise-linear (PL) manifold, was shown by Radó in 1925 (and seems obvious for compact 2-manifolds once Moebius' classification [7] of 2-manifolds is known), and for 3-manifolds this was Moise's theorem of 1952 [115][20][116][154]. Freedman in 1982 constructed a nondifferentiable analogue of the complex projective plane ([60] §8.3 and 10.1) and this fake $\mathbb{CP}^2$ is a 4-manifold that cannot be triangulated as a combinatorial manifold. By combining this with work of A.J.Casson (see [3] and p.xvi of [4]) one obtains examples of topological[11] 4-manifolds that cannot be triangulated at all. In contrast topological, smooth, and piecewise linear (≤ 3)-manifolds are all essentially the same thing, and any two triangulations of a (≤ 3)-manifold are combinatorially equivalent – one may be obtained from the other by a finite sequence of local combinatorial modifications or "moves" [58]. This "Hauptvermutung" [136]. reduces (≤ 3)D topology to combinatorics. However, the Hauptvermutung is false in dimensions $n \geq 4$; there exist homeomorphic, but not PL-homeomorphic, n-dimensional PL-manifolds [90][136][60] for each $n \geq 4$. Finally, we remark that the question of deciding whether two n-manifolds are topologically equivalent is algorithmically *undecidable* if $n \geq 4$ (a fact [103] which is highly related to the undecidability of isomorphism for groups presented as generators and relations [134]) – but is decidable if $n = 2$ due to Moebius' classification of 2-manifolds [7]. The question is open in dimension $n = 3$, but conjecturally eventually a decision-procedure will be found [159].[12]

All of the above suggest that perhaps any attempt to do physics on general curved n-manifolds with $n \geq 4$ would be defeated by severe and fundamental pathologies, which, however, do not exist when $n = 3$. Although numerous arguments (surveyed in [151]) have been made before about why the universe "must" be 3D, these somehow seem more fundamental.

Here is another argument favoring 3-dimensionality. Although we do not know what the laws of physics might be in some other (or other-dimensional) universe, let us postulate that they admit solutions which lead to n mutually orthonormal smooth vector fields in the n-dimensional universe. For example, a "photon momentum-polarization eigenstate" wavefunction for a photon in the universe would lead to a vector at each point in space giving the photon's momentum direction, an orthogonal vector giving its electric field, and a third orthogonal vector indicating its magnetic field. (Actually the latter two vectors are better viewed not really as "vectors," but rather as "bidirections," but that still suffices for our purposes.)

Theorem 1 (Dimension 3 and parallelizability). *Every compact orientable 3-manifold admits a parallelization; but this is not true (even in the smooth case) if any other number $n \geq 2$ is substituted for "3."*

Proof: The fact that every compact orientable 3-manifold is parallelizable arises by pasting together known topology theorems [151] dating back to Seifert. The n-sphere S^n is not parallelizable [1] for any $n \geq 2$ besides $n = 3$ and $n = 7$. Although many 7-manifolds are parallelizable [e.g. S^7, T^7, and SO(5)/SO(3)], the compact orientable 7-manifold $\mathbb{R}P^5 \times S^2$

[7]The "Poincaré homology sphere" is a non-simply connected orientable 3-manifold with the same homology as the ordinary 3-sphere. It has as its fundamental group the "binary icosahedral group" of order 120. It is the only Seifert – and only known – homology sphere with (non-trivial) *finite* fundamental group. Threlfall and Seifert's construction of the Poincaré sphere is a solid regular dodecahedron in a curved space S^3 of constant curvature (it may be regarded as one of the 120 dodecahedral faces of the "120-cell" regular 4-polytope [46]) where opposite pentagon faces are identified by a right-helix turn of $\pi/5$ radians; see [162][167]. A different description by Montesinos ([117] p.IX-X) is that it is the space of possible positions of a regular dodecahedron (or icosahedron) inscribed in a sphere; equivalently, as a topological group it is SO(3)/A_5. But this manifold, and in fact *every* rational homology sphere with a 1-curvature homogeneous metric besides S^3 itself, features "twist" and hence presumably is forbidden as a cosmology. See §7. and §6.

[8]These are the manifolds $x^p + y^q + z^r = 0$, $|x|^2 + |y|^2 + |z|^2 = 1$ in $\mathbb{C}^3$ where p, q, r are coprime positive integers.

[9]Here "generic" means "a generic real-linear combination of trapped magnetic fields going around each possible kind of uncontractible cycle." A different kind of "generic" magentic field would be got by choosing one particular kind of uncontractible cycle (which always will consist of integer number of traversals around the basic cycles, i.e. the previous case except that "real-linear" is replaced by "integer-linear") and then choosing a magnetic field which goes around it (which then may be scaled by a single generic real). With this latter kind of "generic" magnetic field, every kind of orientable compact 3-manifold except for rational homology spheres will yield charge quantization with a unique charge quantum. But with the former kind, if there are two or more different types of uncontractible cycle we will simultaneously get two different, generically incompatible, charge quantization conditions.

[10]Also, the universe could have *zero* such surfaces with charge quantization then being explained by some non-cosmic mechanism such as monopoles. That would permit S^3 and the Poincaré dodecahedral space.

[11]I.e., "C^0" manifolds, obeying the fewest possible smoothness demands. For precise definitions of most such concepts, see [118].

[12]The combination of Perelman's putative proof of Thurston's geometrization conjecture (which involves a canonization procedure called "Ricci flow with surgery" which somewhat resembles an algorithm), the Hauptvermutung, and the fact that equivalence-testing procedures are known for Haken manifolds [82][76][159][105] (and hence knots and links [66]) and for the 3-sphere [142][160][104], all are grounds for optimism. For example, if only some upper bound, no matter how large a function of N (provided it were algorithmically computable), were known for the finite number of "moves" required to interconvert two triangulated 3-manifolds with N tetrahedra total, then a decision procedure would exist. It seems extremely likely that such a bound exists.

is *not* parallelizable. To show this, employ the "product formula" for Stiefel-Whitney classes on page 54 of [112], and see also pages 47 and 133 of that book. We have (in the notation of [112]) $w(S^2) = 1$ and $w(\mathbb{R}P^5) = w(\mathbb{R}P^5 \times S^2) = 1 + a^2 + a^4 \neq 1$, therefore $w(\mathbb{R}P^5 \times S^2)$ is not parallelizable. However since S^2 and $\mathbb{R}P^{2k+1}$ both are compact and orientable (although $\mathbb{R}P^{2k}$ are non-orientable, cf. p.52-53 of [112]), so is $\mathbb{R}P^5 \times S^2$. Q.E.D.

This would seem to grant favor to space-dimension 3, although not completely forcing it since many parallelizable n-manifolds still exist with $n \neq 3$.

In §4 we consider the postulate (which perhaps would be required in a future theory of quantum mechanics in curved spacetimes) that there must be n **commuting** spacelike vector fields in the universe. This would actually eliminate all compact orientable candidate topologies not diffeomorphic to the flat n-torus T^n. (And, more strongly, a flat T^n actually has n mutually everywhere-*orthogonal* commuting vector fields – a condition that also might be required in a future theory of curved space QM.) Indeed, if $n = 3$, merely requiring *two* commuting vector fields is enough to force the 3-manifold (if compact and orientable) to be diffeomorphic to a torus bundle over S^1. Among these, there are only a few 3-manifolds diffeomorphable to constant curvature (and all happen to be flat) and only one satisfies the no-twist postulate, namely T^3.

Now suppose (§5.2) we only allow n-manifolds with an **orthogonal curvilinear coordinate system (OCCS).** (This is equivalent to demanding the existence of n smooth mutually everywhere-orthogonal *and* commuting vector fields.) If we only demand that an OCCS exist *locally* in the neighborhood of any point on the manifold, then every (sufficiently smooth) n-manifold, $n \leq 3$, does the job, but $n \geq 4$ is (for generic points on generic smooth manifolds) forbidden. This again could be regarded as a "reason the universe must be 3-dimensional." If we demand a *globally* valid OCCS, then only T^2 and T^3 are permitted.

3 New argument for the spatial finiteness of the universe

The generalization of the Schwarzschild metric (representing the gravitational field in vacuum exterior to a spherically symmetric mass M) to allow also a charge Q and a cosmical constant Λ is

$$\mathrm{d}s^2 = -N\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{N} + r^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2) \qquad (1)$$

where

$$N(r) \stackrel{\text{def}}{=} 1 - \frac{2M}{r} + \frac{Q^2}{r^2} - \frac{\Lambda}{3}r^2. \qquad (2)$$

This is an exact spherically symmetric solution of the Einstein-Maxwell equations with Λ-term. The r coordinate is here defined "circumferentially," i.e. so that spheres $r = k$ have surface area $4\pi k^2$.

Assume $Q = 0$ (uncharged) for simplicity. Then this line element is asymptotic to the usual Schwarzschild black hole element when $r \to 0^+$, and to the usual de Sitter expanding-universe element when $r \to \infty$. If $M > 0$ and $\Lambda > 0$ and $9M^2\Lambda < 1$ then it has two[13] "horizons" located at the positive roots of $N(r)$. (Anyone falling inside the inner horizon can never re-emerge and is doomed to continue to fall into the central singularity at $r = 0$, since within this horizon r-decrease becomes the future-timelike direction. Anybody outside the outer horizon will never be able to enter it, due to the expansion of the universe being too rapid.)

However, if the mass M is increased to $M_c \stackrel{\text{def}}{=} 3\Lambda^{-1/2}$ then the two horizons merge into one (located at $r = 3M_c$), and if it is increased further then both horizons disappear and we get a solution of drastically different character, representing a contracting universe in which every point is doomed eventually to hit the $r = 0$ singularity. (This metric may also be interpreted as an expanding universe containing a naked singularity at $r = 0$ visible from every point. Which of these two interpretations is preferred depends on which direction of time is regarded as the "future" direction.)

Now let us use this knowledge to give a (new) argument for the finite nature of the universe. Consider the usual picture of a homogenous isotropic spatially-infinite eternally-expanding universe with positive (repulsive[14]) Λ, which ultimately will tend toward the de Sitter vacuum metric. This is presently the most popular and most believed model of the universe. Now consider a sphere of surface area $36\pi M_c^2$. Because the universe is spatially-infinite, and assuming the "creation" of the universe and the matter within it was statistical in nature, we would expect with probability 1 that some such sphere must exist somewhere containing mass exceeding any specified bound, in particular exceeding M_c. If so, then the character of the universe will differ completely from the usual picture, since contraction to a point will be foreordained or else there will be a naked everywhere-visible singularity, and (either way) homogeneity is impossible.

In a nutshell, the problem is that in an infinite universe, there presumably must be arbitrarily large local fluctuations, but a sufficently large local fluctuation is capable of changing the *global* character of the universe – thereby contradicting and destroying the original assumption of homogeneity. This argument is new[15].

Summary: This seems to us to constitute a proof-by-contradiction that a $\Lambda > 0$ homogeneous eternally-expanding universe *must* have bounded fluctuations of mass-in-fixed-size-balls, and (hence, under suitable statistical assumptions about the high tail of the mass density distribution) *must*

[13]If $M = 0$ then we just have de Sitter space which has just one horizon, at $r = \sqrt{3/\Lambda}$. If $\Lambda = 0$ we just have the Schwarzschild black hole with just one horizon at $r = 2M$.

[14]Recent evidence from type-Ia supernovae "standard candles," and even more recently from observations by the Chandra satellite of of Xrays from hot gas in galactic clusters, both suggest that the expansion of the universe is actually starting to *accelerate*, presumably indicating that $\Lambda > 0$.

[15]Instead of regarding our argument as forcing the universe, if eternally expanding, to be finite-volume, another possibility would be that we really do live in an infinite universe with $\Lambda > 0$, *but* it is not eternal and not homogeneous, as will someday be revealed when the existence of a giant black hole with mass $> M_c$ becomes apparent. But since (at least in the universe represented by EQ 1) such a black hole cannot be behind a horizon, if it existed then it ought to be apparent *now*, and it isn't.

have **finite 3-volume**. However, this is not a "proof" as mathematicians understand the word, although it is probably satisfactory to many cosmologists. That is because the facts we have used about the specific metric EQ 1 might not be true in more general metrics corresponding to universes containing inhomogenous matter instead of just vacuum[16].

We had prevously argued that the observed charge-neutrality of the universe would have been forced if the universe were spatially compact; that is an independent reason to believe in compactness.

Finally, the reader is warned that this fluctuation-based argument does not rule out infinite-diameter but simultaneously finite-3-volume 3-manifolds such as hyperbolic manifolds with "cusps." That kind of "cuspy universe" *is* strongly disfavored by the charge-based argument, since charge-neutrality could perfectly well be violated in the interior of such a universe – see Lovelock [100] for an exact Einstein-Maxwell model universe solution exhibiting a "charged cusp."[17]

4 Milnor Rank and commuting vector fields

This section defines and explains commuting and orthogonal vector fields on manifolds. Both this and the following section explore the idea that physics may require the universe to admit 3 commuting and/or orthogonal vector fields, either locally or globally, and try to understand what restrictive effects would follow from such requirements.

There are a great number of different ways to look at commuting vector fields; we shall expalin them this section. They are also related to curvilinear coordinate systems, another tremendous area which we survey in §5. We later shall also require an understanding of "gauge freedom," an oft-misunderstood area we review in §5.1. We hope our surveys of these areas are useful, but the reader should avoid getting lost in their details. That avoidance is possible because our main results are both stated as "theorems" and summarized at the end of both this and the next section.

The *Milnor rank* of a compact differentiable manifold M is the maximum number of C^2-smooth tangent vector fields, all linearly independent at each point of M, and mutually pairwise *commuting* everywhere.

Lemma 2 (Invariant Milnor-Rank). *Milnor rank is invariant under diffeomorphisms of the manifold.*
Before we can prove the foundational lemma 2, we first need to explain what it *means* for two vector fields to "commute." There are several ways to understand this: in terms of tensor notation, in terms of partial differential operators, geometrically, and algebraically. In all cases we define the notion of the "commutator" $C = [A, B]$ of two vector fields A, B, and say that A, B *commute* if their commutator is the zero vector field.

Tensors: In terms of the usual index and Einstein-summed tensor notation [101][130] with x^τ being the coordinate system, we may define the *commutator* of two vector fields A^μ and B^ν to be the vector field C^κ

$$C^\kappa = [A^\mu, B^\nu] \stackrel{\text{def}}{=} A^\mu \frac{\partial}{\partial x^\mu} B^\kappa - B^\nu \frac{\partial}{\partial x^\nu} A^\kappa. \tag{3}$$

There are several important things to note about this expression. First of all, it does not depend at all on the metric of the manifold. Second, by rewriting EQ 3 in terms of *covariant* deriviatives (indicated by semicolons)

$$C^\kappa = A^\mu B^\kappa_{;\mu} - A^\mu \Gamma^\kappa_{\mu\alpha} B^\alpha - B^\nu A^\kappa_{;\nu} + B^\nu \Gamma^\kappa_{\nu\gamma} A^\gamma \tag{4}$$

we note that due to the symmetry $\Gamma^\eta_{\alpha\beta} = \Gamma^\eta_{\beta\alpha}$ of the Christoffel symbols, they cancel, so that

$$C^\kappa = [A^\mu, B^\nu] = A^\mu B^\kappa_{;\mu} - B^\nu A^\kappa_{;\nu}. \tag{5}$$

Thus the commutator of two tensorial vector fields is a *tensor*. Hence if we change to a different coordinate system y^ϕ that would just multiply all the vectors, including the commutator, by the Jacobian of the transformation, so that the "commuting fields" statement that $C^\kappa \equiv 0^\kappa$ is *invariant* under changes of coordinates. Those two observations prove lemma 2. Finally, observe that if A^μ is multiplied by some scalar-valued factor S, then EQ 3 changes to

$$C^\kappa = [SA^\mu, B^\nu] = SA^\mu B^\kappa_{;\mu} - SB^\nu A^\kappa_{;\nu} - B^\nu A^\kappa S_{;\nu} \tag{6}$$

which in view of the last term is generally *not* the same as SC^κ_{old}, unless $S_{;\nu} B^\nu \equiv 0$, as happens (for example) if S is identically constant. The fact that generic positionally-dependent scaling is forbidden for commuting vector fields prevents us from orthonormalizing them by a positionally-dependent QR matrix factorization [65], i.e. by taking positionally-dependent linear combiantions of our vector fields, since taking linear combinations of the vectors is forbidden. Since linear combinations with *constant* coefficients are OK, though, we *can* cause our commuting linearly independent vector fields to be orthonormal at any *one* particular *point*. (Further, if they happen *already* to be mutually orthogonal everywhere, then we may, by taking an appropriate *constant* linear combination, cause our fields still to be mutually orthogonal everywhere, but rotated arbitrarily away from their original directions.) This *contrasts* with the common situation where one has n everywhere linearly independent smooth vector fields on a manifold (without anything being said about commutation). In that case, by doing a smoothly positionally dependent QR matrix factorization, we may convert those fields to being everywhere mutally orthonormal. Thus the words "linearly independent" and "orthonormal" are often *interchangeable* in discussions of smooth vector fields on manifolds, with the latter being preferred because it is a more useful statement. However, those words are *not* interchangeable if we demand commutation.

Differential operators: If a vector field A^μ is regarded as corresponding to the partial derivative operator $\partial_A \stackrel{\text{def}}{=} A^\mu \frac{\partial}{\partial x^\mu}$, then the partial differential operator corresponding to C^κ is

$$\partial_C = \partial_A \partial_B - \partial_B \partial_A. \tag{7}$$

[16] Indeed there would seem to be no hope of getting full rigor at least until longstanding fundamental problems about solution existence in general relativity are resolved. This whole issue deserves more thought.

[17] Admittedly, if the "charges of the cusps" were included in the reckoning then any such universe would have to be neutral overall; but the cusps are at ∞, i.e. not on the manifold. We should also mention that hyperbolic 3-manifolds with finite volume but cusps at ∞ are disfavored by the extension by Adams et al. (mentioned in our §7.4) of our theorem 17.

This again is a first order partial differential operator because the second order derivative terms cancel out. Our two vector fields on M *commute* precisely if their differential operators commute at each point of M.

In view of EQ 5 we could instead define these operators to be $A^\mu D_\mu$ where D is the *covariant* (rather than partial) derivative operator.

Geometric picture: The geometric meaning of the commutator of two vector fields is, in the limit $\epsilon \to 0$, the vector field that "closes the quadrilateral" in figure 4.1.

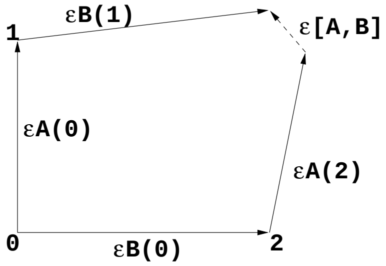

Figure 4.1. In the limit $\epsilon \to 0$, the dashed "missing link" closing the quadrilateral is $[A, B]$. For nonzero ϵ, the quadrilateral does not exactly close, but the error term is $O(\epsilon^3)$.

Algebraically: Vector fields under commutation form a Lie algebra, meaning that the commutator obeys the Jacobi identity

$$[A[B,C]] + [B[C,A]] + [C,[A,B]] = 0 \tag{8}$$

linearity

$$[mD + kA, B] = k[A,B] + m[D,B] \tag{9}$$

if k and m are *constant* scalars, and antisymmetry

$$[A,B] + [B,A] = 0. \tag{10}$$

Orthonormality for vector fields is completely logically unrelated to the demand that they commute. This fact may be confirmed algebraically via EQ 3 by considering the following two everywhere-orthonormal vector fields in the Cartesian xy plane

$$A = (-\sin x, \cos x), \quad B = (\cos x, \sin x) \tag{11}$$

whose commutator is

$$C = (1,0) \neq (0,0), \tag{12}$$

and by considering the fact that any two nonparallel constant vector fields in $\mathbb{R}^n$ commute. Also, we remark that the *Hopf fibration* (3 everywhere-orthonormal tangent vector fields on S^3 arising quaternionically from ix, jx, and kx where $|x| = 1$, see [151]) vector fields do not commute.

A **famous physics example** of commutation is the three "angular momentum operators" (which could, if one wished to be eccentric, be regarded as vector fields in xyz-space $\mathbb{R}^3$):

$$L_1 = y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}, \quad L_2 = z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z}, \quad L_3 = x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}. \tag{13}$$

Then $L_3 = [L_2, L_1]$ and the corresponding identities got by cyclically permuting the indices, also hold. Hence the algebra of operators $x_1L_1 + x_2L_2 + x_3L_3$ under commutations is the same as the usual algebra of 3-vectors under the cross product $\vec{z} = \vec{y} \times \vec{x}$:

$$[x_1L_1+x_2L_2+x_3L_3,\ y_1L_1+y_2L_2+y_3L_3] = z_1L_1+z_2L_2+z_3L_3. \tag{14}$$

Why might physicists care about commuting vector fields? In ordinary flat space nonrelativistic quantum mechanics, it is commonplace to employ three "momentum operators"

$$p_x = \frac{\partial}{\partial x}, \quad p_y = \frac{\partial}{\partial y}, \quad p_z = \frac{\partial}{\partial z} \tag{15}$$

(we have left out irrelevant constant factors of i and $\hbar$) which commute. Should physicists trying to generalize quantum mechanics to work in curved spaces deem it desirable for three such mutually commuting operators to exist everywhere – even in a curved universe – then we must demand that the universe have Milnor-rank 3.

Known results on Milnor-rank: Early work (mentioned in our previous paper [151]) showed that every 3-manifold has Milnor-rank at least 1; the 3-torus is the only compact 3-manifold with Milnor-rank 3; and that the Milnor-rank of S^3 (the 3-sphere) [99], and of $S^2 \times S^1$ [138] both are 1. S.P.Novikov later extended the lattermost result to show $\mathrm{MRank}(S^2 \times T^{n-2}) = n - 2$. Rosenberg, Roussarie, and Weil [141] then classified all compact boundaryless connected orientable 3-manifolds of Milnor-rank≥ 2, showing they were precisely the torus bundles over S^1, with the only one of Milnor-rank 3 being the 3-torus itself, and they indeed included a proof on their first page that *the n-torus T^n is the only compact boundaryless connected n-manifold with Milnor-rank n.* Finally, Arraut [9], building on work of Chatelet, Rosenberg, and Weil [38] extended this classification to allow replacing some of the vector fields by foliations.

Theorem 3 (The 3 constant curvature orientable 3-universes with rank≥ 2). *Let M be a compact orientable 3-manifold with Milnor-rank≥ 2. Assume further that it has constant curvature. Then that curvature is zero and M is parallelipiped with opposite faces identified,* but *with one pair of faces perhaps identified via a $90°$ or $180°$ (or no) twist.*

Proof: We know [141] that the theorem's first sentence forces M to be a torus bundle over a circle. From corrol. 4.6 p.449 of [145] we know that no such bundle can have a hyperbolic structure. Therefore we may restrict attention to compact 3-manifolds of positive and zero constant curvature. But these are completely classified [174]. Going through the classifications [174] we find that the only allowed compact 3-manifolds are flat ones which consist of parallelipipeds with opposite faces identified, *but* with one pair of faces perhaps identified via a $90°$ or $180°$ twist. All of these have zero curvature. Q.E.D.

Summary: If our universe is a compact 3-manifold containing 3 commuting vector fields (as might be required in future quantum gravity theories), then it must be T^3; even merely requiring it to contain 2 commuting vector fields and be orientable forces it to be one of only 3 particular flat 3-manifolds, of which only T^3 is twist-free.

5 Gauge freedom, n-tuply orthogonal foliations and commuting vector fields

In ordinary flat space quantum mechanics, in addition to the three momentum partial differential operators p_x, p_y, p_z, there are also three *non*differential operators x, y, z. Among the 21 unordered pairs among these 6 operators, all commute except for $[x, p_x]$, $[y, p_y]$, and $[z, p_z]$. The $[y, p_x]$ commutation relations (and their ilk) force p_x, p_y, p_z to correspond to mutually *orthogonal* vector fields. For example, in the flat xy plane, the differential operator $a\frac{\partial}{\partial x} + b\frac{\partial}{\partial y}$ commutes with multiplication by x if and only if $a = 0$. Similarly but more generally, in curved 3-space, we must have p_y and p_z (but not p_x) orthogonal to $\frac{\partial}{\partial x}$, p_x and p_z (but not p_y) orthogonal to $\frac{\partial}{\partial y}$, and p_x and p_y (but not p_z) orthogonal to $\frac{\partial}{\partial z}$. This forces p_x to be *proportional* to $\frac{\partial}{\partial x}$ (and similarly for p_y, p_z). If we then demand that $\{p_x, p_y, p_z\}$ be defined locally in the neighborhood of any particular point on a 3-manifold in a way not depending on the coordinate system (i.e. tensorially using the covariant derivative), we must make them arise from **3 orthogonal commuting vector fields**.

Perhaps it is desirable that all these properties still hold even in curved spaces. (However, the notion of a "position operator" no longer holds the same attraction in curved space quantum mechanics since it would seem to be useable only locally. If so, we perhaps only should require the *local* existence of 3 orthogonal commuting vector fields on our 3-manifold.) That leads to the **question:** *On which 3-manifolds do there exist 3 everywhere-orthogonal and nonzero commuting tangent vector fields?* There are two versions of this question, namely where the existence is inteneded to be global or merely local.

There turn out to be an astonishing number of equivalent questions. A 3-manifold has 3 orthogonal and commuting vector fields (locally or globally) if and only if it has an *orthogonal curvilinear coordinate system* (locally or globally), i.e a coordinate system in which the metric tensor $g_{\alpha\beta}$ is everywhere *diagonal*. This happens if and only if it has a *triply orthogonal surface system*, which happens if and only if it has 3 everywhere-orthogonal *foliations into 2-manifolds* (respectively locally or globally, in both cases).

To explain why these equivalences hold: Any pair among 3 commuting vector fields may be thought of as defining a family of 2-surfaces, i.e. a foliation, and these 3 kinds of surfaces must be orthogonal everywhere if the fields are. More generally in an n-manifold if there were n commuting vector fields we would get $\binom{n}{2}$ orthogonal foliations into 2-manifolds. Regarding the leaves in the kth among the 3 foliations as the surfaces $u_k =$ constant where each u_k is some function of position, we may employ u_1, u_2, u_3 as a new coordinate system on our 3-manifold. In this coordinate system the metric tensor must be diagonal due to orthogonality, and the gradient vectors of the u_k give us 3 orthogonal commuting vector fields back again.

Again, physicists might regard it as desirable for there to be an orthogonal curvilinear coordinate system, even in a curved universe, either locally or globally. So the question arises: *which* universes have that property?

We state the following results here, and prove them in §5.3.

Theorem 4 (Cotton-Darboux). *If $n \le 3$, then any n-manifold has an orthogonal coordinate system in a local neighborhood of any particular one of its points. However, if $n \ge 4$, this is false for a generic metric at a generic point.*

Theorem 4 could be regarded [151] as a "reason the universe must be 3-dimensional."

Theorem 5 (Torus). *Let $n \le 3$. Any compact n-manifold diffeomorphic to T^n whose metric tensor is close enough (in the Sobolev 2-norm involving both the metric tensor perturbation and its first derivative) to the metric tensor of flat space, has a globally valid orthogonal coordinate system. (If $n = 2$ then the Sobolev norm constraint may be dropped.) However (for any $n \ge 1$), a compact orientiable n-manifold not homeomorphic to T^n cannot have a globally valid orthogonal coordinate system.*

Theorem 5, which goes further, could be regarded (less convincingly) as a "reason the universe must be a 3-torus." Conjecturally, there is a strengthened combined version of the above theorems:

Conjecture 6. *A smooth 3-manifold has a globally valid orthogonal curvilinear coordinate system if and only if it is topologically equivalent to T^3. (I.e., the Sobolev norm constraint may be dropped entirely from theorem 5.)*

Before proving these theorems in §5.3, we first must explain something of the history of orthogonal curvilinear coordinate systems, an area very heavily studied during 1810-1920 but which thereafter (unfortunately) became moribund. Also, we must explain the important and oft-abused concept of "gauge freedom."

5.1 Gauge freedom in general relativity (GR) and differential geometry

Einstein's field equations of GR state that

$$G_{\alpha\beta} = \chi T_{\alpha\beta} \tag{16}$$

where $G_{\alpha\beta} = G_{\beta\alpha} = R_{\alpha\beta} - \frac{1}{2} R g_{\alpha\beta}$ is the Einstein tensor (here expressed in terms of the Ricci curvature tensor $R_{\alpha\beta}$ and the metric tensor $g_{\alpha\beta}$), $T_{\alpha\beta}$ is the stress-matter-energy-tensor, and χ is a physical constant arising from Newton's gravitational constant. Naively, in n-dimensional spacetime (where conventionally $n = 4$) this is a statement of $(n+1)n/2$ equalities (i.e. conventionally 10). There are to be solved for the 10 components of the metric tensor $g_{\alpha\beta} = g_{\beta\alpha}$. But because $G^{;\beta}_{\alpha\beta} = 0_\alpha$, expressing the automatically divergence-free nature of the Einstein tensor (which was the reason for Einstein's choice of "$\frac{1}{2}$") on the one hand, while $T^{;\beta}_{\alpha\beta} = 0_\alpha$ expresses the conservation of mass-energy on the other, really n equalities hold automatically. So really, there are only $(n-1)n/2$ equations constraining the metric tensor, leaving n leftover degrees of "gauge freedom." Thus whenever there is a solution of the Einstein equations, there in general will be an n-parameter *family* of solutions, all equivalent under "gauge transformations." The "infinitesimal gauge transformations" are

$$g_{\alpha\beta} \to g_{\alpha\beta} + (u_{\alpha;\beta} + u_{\beta;\alpha})\epsilon \tag{17}$$

in the limit $\epsilon \to 0$, where u_α is any tensorial vector field.

In differential geometry, it is always possible to employ different coordinate systems to describe the same metric. Consequently there are many different descriptions of the same manifold – and it is not trivial to determine whether two metrics are really the same! Again the group of equivalence transformations on metrics is infinitesimally generated by EQ 17.

It is common (and in fact essential in numerical general relativity) to impose n "gauge conditions," i.e. constraints on the metric, so that there will be only *one* solution of the Einstein equations. For example (when $n = 4$) one common choice is the "temporal gauge" that $g_{0\alpha} = g_{\alpha 0} = -\delta_{\alpha 0}$. Another common choice is the "Harmonic gauge"

$$\Gamma^\alpha \equiv g^{\mu\nu}\Gamma^\alpha_{\mu\nu} = (\ln\sqrt{g})^{,\alpha} = g^{\mu\nu}g^{;\alpha}_{\mu\nu} = 0^\alpha. \tag{18}$$

Still another possibility would be demanding that all eigenvalues of $g_{\alpha\beta}$ be ± 1.

A common impression gained from reading most physics books about GR is that these (or any set of n such) gauge demands all are imposable. The more careful physics books tell us they are merely imposable *locally* in the neighborhood of any generic point, but it might be impossible to make them hold globally on the manifold since attempting to do so may force coordinate singularities. Although the latter attempt to provide caution is admirable, it does not go far enough, since in fact, we have never seen a proof that n gauge demands need be imposable even locally. Undoubtably this is "usually" true for most points, most manifolds, and most reasonable-looking sets of gauge conditions, but there might be some exceptions. So really one always needs to justify any local existence claim of this sort by an ad hoc argument based on some existence theorem for PDEs.

Here are some examples. Weinberg [170] claimed on his pp.162-163 that "it is always possible to choose a coordinate system" in which the harmonic gauge conditions hold. That claim was either wrong or misleading. First of all, his proof was incomplete since he reduces the problem to solving certain system of partial differential equations, but does not prove that system always has a solution. Second, there is *no* single harmonic coordinate system on S^2, or S^1, or in fact on any compact Riemannian manifold whatever, by considering Weinberg's EQ 7.4.9 and the fact that the only scalar valued harmonic function on a compact Riemannian manifold is a constant. What *is* true, is that in any subset topologically equivalent to an open n-ball of a Riemannian n-manifold (and hence *locally*), there is a harmonic coordinate system. This is because the Dirichlet principle (which has been justified rigorously [53] by means of the "heat flow method") proves solution existence. Furthermore, on any Lorentzian $(n+1)$-manifold foliatable into spacelike n-manifolds ("timeslices"), in any subset S of such a spacelike-submanifold topologically equivalent to an open n-ball and its "future" (i.e. the set whose past is S) there exists a harmonic coordinate system. That is because Weinberg's system of partial differential equations is linear and hyperbolic and on each timeslice is solving an elliptic system which by the Dirichlet principle has a solution.

't Hooft (§9 of [80]) gave a similarly informal argument that, in many typical general relativistic scenarios, the temporal gauge cannot be imposed globally, but it can be imposed locally.

5.2 Orthogonal curvilinear coordinate systems (OCCS)

An orthogonal curvilinear coordinate system (OCCS) on a manifold is a coordinate system causing the metric tensor $g_{\alpha\beta}$ to be diagonal. Obviously any OCCS coordinate u may be replaced by a function $f(u)$ of it.

An OCSS $u_\alpha(\vec{x})$ must satisfy

$$\underbrace{\vec{\nabla}u_\alpha \cdot \vec{\nabla}u_\beta = 0 \text{ if } \alpha \neq \beta}_{\text{orthogonality}}, \quad \underbrace{\det\frac{\partial u_\alpha}{\partial x_k} \neq 0}_{\text{nondegeneracy}}, \quad \underbrace{H_\alpha^2 = |\vec{\nabla}u_\alpha|^2}_{\text{diagonal metric}}. \tag{19}$$

To any such coordinate system in $\mathbb{R}^n$ there corresponds a metric tensor $g_{\alpha\beta}$ which is diagonal with $g_{\alpha\alpha} = H_\alpha$.

The Riemann curvature tensor arising from that metric then must obey certain conditions. First of all, when α, β, μ, ν all are unequal $R_{\alpha\beta\mu\nu} = 0$ since that happens automatically if the metric tensor is diagonal.

If we are speaking of OCSS's *in* $\mathbb{R}^n$ (which was what all the 19th century investigators concerned themselves with) then, second, if $\alpha \neq \beta$, we have $R_{\alpha\beta\alpha\beta} = 0_{\alpha\beta}$ (these constitute $n(n-1)/2$ equations); and third, if α, β, μ all are unequal, then $R_{\alpha\beta\alpha\mu} = 0_{\alpha\beta\mu}$ (these constitute $n(n-1)(n-3)/2$ equations). These equations (called the Gauss-Lamé equations) are an overdetermined system of second order PDEs that arise simply because $\mathbb{R}^n$ is flat.

In the case $n = 3$, in §251 of [59] it is shown that if three "Lamé coefficients" H_1, H_2, H_3 exist satisfying the $6 = 3 + 3$ Gauss-Lamé PDEs, then they determine a unique (up to rotation and translation) triply orthogonal system of surfaces. It would also be possible to make alternate versions of the Gauss-Lamé equations valid in any particular curved manifold. That option unfortunately (due to the historical order of events) has largely remained unconsidered.

If the dimension n of the manifold is 1, every coordinate system is an OCCS. If $n = 2$, one may choose one coordinate u in an arbitrary (sufficiently smooth) way, then consider the vector field of normals to its level curves. The integral curves of this vector field (which exist, due to existence theorems for ODEs) give the second coordinate w. This argument shows that every 2-manifold has an OCCS locally; indeed that it has many of them, with one arising from each smooth-enough function u of 2 variables. A slicker argument accomplishing nearly the same goal is to use the known theorem that any 2-manifold is conformally flat. Applying the conformal transformation to an OCCS on the flat xy plane yields an OCCS of an arbitrary 2-manifold. This in fact yields a *global* OCCS for a 2-torus T^2 since any flat T^2 (parallelipiped with periodic boundary conditions) plainly has one, namely the Cartesian xy coordinate system. However, there is no global OCCS on any other orientable compact 2-manifold because there is no way to "comb the hairs" either on a sphere or g-holed torus with $g \geq 2$; and (by the Gauss-Bonnet theorem) there is no globally flat version of any compact orientable 2-manifold besides for T^2.

If $n \geq 3$ the problem becomes more difficult. The area of triply orthogonal families of surfaces in $\mathbb{R}^3$, was heavily studied during the years 1810-1920 by (in roughly chronological order) Dupin, Gauss, Lamé, Bouquet, Cayley, Salmon, Ribaucour, Combescure, Bianchi, Bôcher, Darboux, Cartan, Grönwall, and Cotton, who between them produced an enormous amount of work. E.g. the entire 3rd volume [18] of Bianchi's collected works (850 pages of Italian!) was devoted to (but hardly exhausted) this single subject. The orthogonal curvilinear coordinate systems in $\mathbb{R}^3$ in which the Laplacian is separable were completely classified by Lamé [119]; in all of them the surface families are either confocal quadrics or planes. Confocal quadric based systems had first been found by Dupin and Binet in 1810; there are also some systems based on quartic surfaces, due to Bôcher. Grönwall [69] classified systems in which one of the surface families consists of minimal surfaces, finding (a) the 3D familes got by taking a 2-dimensional family within a plane, and rolling that plane over a developable surface (this includes as a limiting case the rotationally symmetric systems; the 3rd surface family are the planes themselves, which of course are minimal surfaces) and (b) one additional quartic family[18].

The first realization that generic surfaces in $\mathbb{R}^3$ are *not* members of any triply orthogonal system of surfaces was apparently due to Bouquet in 1862; in §253 of [59] it is shown that in order for a surface to be a member, it is necessary and sufficient that it satisfy a certain 3rd-order PDE. The best attempt to summarize was Darboux's 1910 book [48] in French; shorter attempts (this time in English) were chapter XI ("triply orthogonal systems of surfaces") of Forysth's 1920 book [59], ch. XIV of Eisenhart's book [55], and chapter 2 of [36]. All of these sources written before 1925 were in pre-Einstein pre-tensorial notation[19] and hence in many places are far more difficult to understand than they need to be. After 1920, most active research in this area ceased.

The following theorem is due to Pierre Charles F. Dupin (1784-1873) in 1813 and J.Gaston Darboux (1842-1917), and is of interest for the purpose of understanding how OCCS's behave.

Theorem 7 (Dupin-Darboux). *A necessary and sufficient condition for the existence of a third system of surfaces, orthogonal to two given mutually orthogonal surface systems, is that the two families intersect along their lines of curvature.*
Definitions: The "lines of curvature" on a surface are the curves everywhere tangent to its principal curvature directions. A "system" of surfaces is a 1-parameter family of surfaces (2-manifolds); two families are "orthogonal" if at each point P, the surface from the first family which contains P, is orthogonal to the surface from the second family which contains P.

Proof: The "and sufficient" was added to Dupin's theorem by Darboux. This was stated and proven as theorems 10 and 11 in ch.4 vol.3 of Spivak's [152] and also is in [59] and [55], which also prove the related "Joachimsthal's theorem" that if two surfaces cut each other at a constant angle then if the curve is a line of curvature of one surface, it also is a line of curvature of the other. (One way to attack problems about two intersecting surfaces is "Darboux's method of the moving trihedron.") Although Spivak's proof only concerns surfaces in $\mathbb{R}^3$, it actually works on any smooth 3-manifold, if we read "curvature" as "extrinsic curvature" everywhere. However, the reader is *warned* that Darboux's "and sufficient" may only show existence on some open *subset* of the 3-manifold (or, for that matter, of $\mathbb{R}^3$). Q.E.D.

In §252 of [59] the theorem is proven that a triply orthogonal system exists containing any 3 assigned curves in the xy plane (no two cutting orthogonally) respectively inside each of its 3 kinds of surfaces. Hence "the utmost degree of generality that can be expected" for a triply orthogonal surface system is 3 arbitrary 2-variate functions. In fact, this level of generality *is* achieved [64].

Zakharov [178] has recently reexamined the OCSS area by making a connection to (his own version of) the theory of the "inverse scattering transform" concerning solitons. Zakharov's results seem interesting, although it is difficult to be sure exactly what they are since they are not stated as theorems. Zakharov also claimed that Bianchi [18] and Cartan [36] showed that for any $(n-1)n/2$ arbitrary 2-variate functions, there is an OCSS on $\mathbb{R}^n$, generalizing the above-mentioned results for $n = 2, 3$ to general n. But I have been unable to find this result in [18][36].

5.3 Proofs

Proof of theorem 4: The question is whether, for each point P on a n-manifold, under some change of coordinates, the metric tensor $g_{\alpha\beta}$ can be made diagonal in a neighborhood of P. The requirements that $g_{\alpha\beta} = 0$ if $\alpha > \beta$ are $(n-1)n/2$ equations, but there are only n available degrees of gauge freedom to satisfy them with. Since the former quantity is greater than the latter if $n \geq 4$, it is evidently impossible to diagonalize $g_{\alpha\beta}$, even locally, if $n \geq 4$. Another way to see this is to note that if $g_{\alpha\beta}$ is diagonal, then all the $(n-3)(n-2)(n-1)n$ entries of the Riemann curvature tensor $R_{\alpha\beta\mu\nu}$ having all four indices unequal, must be zero. If $n \leq 3$ this has no constraining effect at all because it is impossible for all four indices to be unequal. But if $n \geq 4$ it rules out generic points on generic manifolds. (For example with $n = 4$ consider $R_{\alpha\beta\mu\nu}$ having all 256 entries 0 except for $R_{1234} = 1$ and its index-permuted consequences.)

To prevent misunderstanding we remark that for *any* n, it always is possible to diagonalize $g_{\alpha\beta}$ at any *single* point P by a coordinate change based on the orthonormal eigendirections of $g_{\alpha\beta}(P)$ regarded as a symmetric $n \times n$ matrix. The problem is the impossibility of stitching such single point coordinate changes together compatibly over a nonzero-measure region.

If $n \leq 3$ then $(n-1)n/2 = n$ so that there are enough degrees of gauge freedom available to do the job, at least generically. But that is, while suggestive, not a complete proof. (To see why we say that, consider the analogous but simpler example of the statement that "k equations in k unknowns always have a solution" with the counterexample $x^2 = -1$ with $k = 1$. In contrast it *is* legitimate to say that "$k+1$ equations in k variables generically do *not* have a solution.") The fact that a

[18] Grönwall also showed that there is a rotationally symmetric system containing *two* families of minimal surfaces, namely the planes containing the axis of symmetry, and catenoids.

[19] The Riemann curvature tensor, and indeed the word "tensor," are nowhere to be found in [59], making it probably the last major differential geometry book with that property.

diagonalizing change of variables *is* always possible throughout a neighborhood of any point P need to be proved by using local existence theorems for solutions of PDEs. Fortunately this already was done (when $n = 3$, provided we are on an analytically-smooth manifold) in 1899, and is called the "Cotton-Darboux theorem" [44]. Chandrasekhar redid the proof in §12 of his [37], employing a 2+1 dimensional splitting and the Cauchy-Kovalevskaya PDE local-existence theorem in his argument. Finally, the $n = 1$ case is a triviality, and the $n = 2$ case is also much easier – the discussion in §5.2 suffices. Q.E.D.

Proof of theorem 5: If $n = 1$ the result is trivial and if $n = 2$ it is also quite easy; the discussion in §5.2 suffices. Finally, the fact that any manifold that is not a T^n cannot have n commuting vector fields (and hence cannot have a globally valid orthogonal coordinate system) was already known (see theorem 3). So assume $n = 3$ from here on.

Schoen [144][57] showed, and it later was redone more elegantly and simply by Rugang Ye [177], that any n-manifold with $n \le 3$ may be distorted by a flow on conformal transformations to a form with constant scalar curvature. (The corresponding statement in ≥ 4 dimensions is *false* because of an insufficiency of degrees of freedom.) Thus in particular any T^3 is conformally distortible to have scalar curvature everywhere *zero*. Obviously, a conformal transformation of an n-manifold preserves the existence (or nonexistence) of a n-tuply orthogonal surface system. So restrict attention to such *scalar-flat* T^3s from here on.

Let us also observe an important and especially simple special case. Consider the usual flat n-torus arising from some parallelipiped P with periodic boundary conditions, and consider any n linearly independent *constant* vector fields on it. Then the linear transformation whose inverse Jacobian matrix has columns corresponding to our vector fields will map that torus to a different flat torus, defined by a different parallelipiped P', in which the vector fields are now still constant, but now orthonormal. Because this transformation is merely a linear change of coordinates, it certainly is a gauge transformation.

Since any T^3 is diffeomorphic to the standard flat T^3, which we shall for convenience regard as $[0, 2\pi)^3$, we may work in the x, y, z coordinate system in that standard flat T^3. Let the metric tensor be $g_{\alpha\beta}$, which is some triply 2π-periodic (3×3)-matrix-valued function of x, y, z. Now obviously these three vector fields ∂x, ∂y, ∂z commute, but they in general are not orthogonal, i.e. $g_{\alpha\beta}$ in general is not diagonal.

For each $\alpha\beta$ we may consider the *average* (triple integrate $\mathrm{d}x\mathrm{d}y\mathrm{d}z$ each from 0 to 2π then divide by $(2\pi)^3$) value of $g_{\alpha\beta}$, call it $\widetilde{g}_{\alpha\beta}$. We may then perform the orthonormalizing linear eigen-change of coordinates appropriate to cause $\widetilde{g}_{\alpha\beta}$ to become the identity matrix $\delta_{\alpha\beta}$. Due to the linearity of both the transformation (which is just a pre- and post-multiplication by a *constant* orthogonal matrix and its transpose) and the averaging operator, in the new coordinates $g_{\alpha\beta}$ will be such that its average is $\delta_{\alpha\beta}$. Call this operation "orthogonalizing the average." Note that it leaves the Frobenius norm (sum of th4 squares of the enetries of a matrix) of $g_{\alpha\beta}$ invariant.

We shall (next paragraph) find a vector field u_α such that the infinitesimal gauge transformation (EQ 17) it induces always *reduces* the magnitude of the *variation* in each off-diagonal term of $g_{\alpha\beta}$ by a factor $1 - \epsilon$ while preserving scalar-flatness. By repeatedly applying these transformations, interspersed with orthogonalizations of the average, then, we ultimately converge (in a limit $\epsilon \to 0^+$ and an infinite number of transformations) to a metric such that the three off-diagonal terms of $g_{\alpha\beta}$ are *constants*, which in fact must be *zero*. Consequently, we have shown there is a gauge transformation that diagonalizes $g_{\alpha\beta}$, proving the theorem.

To find u_α, expand u_1, u_2, u_3, g_{12}, g_{13}, and g_{23} into triple Fourier series of the form

$$\sum_{j,k,\ell\ge0} C_{jk\ell} \frac{\sin}{\cos}(jx) \frac{\sin}{\cos}(ky) \frac{\sin}{\cos}(\ell z). \tag{20}$$

Note our abbreviated notation: for each integer triple (j, k, ℓ) there are 8 terms in the Fourier series arising from the possible choices of sin's or cos's, each having its own coefficient $C_{jk\ell}$. Our goal is to choose u_1, u_2, and u_3 (by specifying all their Fourier coefficients) so that $u_{\alpha;\beta}+u_{\beta;\alpha}$ is proportional to $-g_{\alpha\beta}$ for each $\alpha \ne \beta$. But this task, for each $(j, k, \ell) \ne (0, 0, 0)$, is[20] simply a matter of solving a 24×24 linear system ($24 = 3 \cdot 8$; there are 3 kinds of u_α and 8 Fourier coefficients for each) to find the 24 unknown $C_{jk\ell}$ Fourier coefficients for u_1, u_2, and u_3 in terms of the 24 known $C_{jk\ell}$ Fourier coefficients for g_{12}, g_{13}, and g_{23}. Unfortunately, due to the Christoffel-symbol terms inherent in the covariant derivatives, the 24×24 systems actually are not independently soluble, so really, we must solve an $\infty \times \infty$ system. But if the metric tensor $g_{\alpha\beta}$ is a small perturbation of the metric tensor $\eta_{\alpha\beta}$ of flat space:

$$g_{\alpha\beta} = \eta_{\alpha\beta} + \epsilon h_{\alpha\beta} \tag{21}$$

then the Christoffel symbol is

$$\Gamma^\mu_{\alpha\beta} = \frac{\epsilon}{2}\eta^{\mu\nu}(h_{\alpha\mu,\beta} + h_{\beta\nu,\alpha} - h_{\alpha\beta,\nu}) + O(\epsilon^2) \tag{22}$$

so that in the flat space limit $\epsilon \to 0$ the $\infty \times \infty$ linear system splits into independently soluble 24×24 systems. Each of the 24×24 determinants may be seen, with the aid of a symbolic manipulation system, to be the nonzero integer

$$256(k-\ell)^4(k+\ell)^4(j-\ell)^4(j+\ell)^4(j-k)^4(j+k)^4+O(\epsilon^2). \tag{23}$$

The only problem is that the determinant in EQ 23 *is* zero if (ond only if) two or more among $\{j, k, \ell\}$ are equal. There is an easy way to understand why: any such symmetrical Fourier term cannot yield an unsymmetrical metric perturbation, e.g. if $j = k$ then we get *equal* g_{xz} and g_{yz} perturbations of $jk\ell$ Fourier type, so that postulating unequal ones would yield no solution of the linear system.

There are thus essentially two kinds of potentially "bad" perturbations of the metric away from that of flat space, in the limit $\epsilon \to 0^+$, namely those with all three Fourier indices equal:

$$\mathrm{d}s^2 = \mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2+\epsilon(A\mathrm{d}x\mathrm{d}y+B\mathrm{d}y\mathrm{d}z+C\mathrm{d}x\mathrm{d}z)\sin x \sin y \sin z \tag{24}$$

[20] We do not care about $(j, k, \ell) = (0, 0, 0)$, since our goal is to reduce the *variation* in the off-diagonal $g_{\alpha\beta}$; we do not care about what happens to the constant terms C_{000} in their Fourier series.

or just one pair:

$$ds^2 = dx^2 + dy^2 + dz^2 + \epsilon(A dx dy + B dy dz + C dx dz) \sin x \sin y \quad (25)$$

where x, y, z each are 2π-periodic coordinates and A, B, C are constants. More precisely, there are more than these two: we can change the frequencies by e.g. using $\sin 3x$ instead of $\sin x$ (which has no effect on our argument, so we choose frequencies 1 and 0 to maximize simplicity), and we can change phases by using cosines instead of sines, e.g. consider

$$ds^2 = dx^2 + dy^2 + dz^2 \quad (26)$$
$$+\epsilon(A \cos x dx dy B \sin x dy dz + C \cos x dx dz) \sin y$$

instead of EQ 25. There are less than 24 different kinds of such variations – a reduction in work arises by noting one of the cosines may always be chosen to be a sine by virtue of an overall phase shift (which has no logical effect on validity). The important thing is that each of our two prototypical bad kinds of metrics is *forbidden* because they (respectively) have sc alar curvatures

$$R = 4\epsilon[A \cos x \cos y \sin z + B \sin x \cos y \cos z + C \cos x \sin y \cos z]$$
$$+O(\epsilon^2) \quad \text{and} \quad R = \epsilon A \cos x \cos y + O(\epsilon^2) \quad (27)$$

which (contrary to our initial assumption) are *nonzero*. Concentrating on the latter (and its variants which lead to sines in place of some of the cosines), observe that it is not possible to choose any nonzero linear combination of them which is zero everywhere. Of course, the scalar curvature *is* a *linear* operator (if $\pm O(\epsilon^2)$ terms are ignored) of the metric perturbation. Further, no combination of other (unsymmetrical or other-frequency) Fourier terms can null out a symmetric Fourier term.

We conclude from these facts (after more, but similar, analysis of the different cases; such case analysis is simplified by realizing that only one among $\{A, B, C\}$ need be regarded as nonzero at any time, and then that changing sin's to cos's is merely a phase shift that makes no qualitative difference) that any such "bad" metric perturbation *must* cause the scalar curvature to be nonzero somewhere, contrary to our initial assumption. So they cannot happen (or anyway the cases when they do are ignorable); only the "good" perturbations, whose off-diagonal elements are reducible in Sobolev norm by a gauge transformation, are permitted. Q.E.D.

This preceding proof depended in at least two ways on the assumption that the perturbation of the T^3 metric away from flatness was small, e.g. that ϵ was sufficiently small. Consequently we do not get a proof of conjecture 6, whose point was to avoid any such near-flatness assumption[21].

5.4 Foliations

At the beginning of this section we mentioned a connection to the existence of foliations of our 3-manifold into 2-manifolds. That makes the following theorem [163] relevant:

Theorem 8 (Thurston's foliation theorem). *A smooth n-manifold has a C^∞ foliation into $(n-1)$-dimensional submanifolds if and only if its Euler characteristic is zero. (This is a global statement. Consequently every smooth n-manifold has such a C^∞ foliation locally.)*

Summary: If the universe has an orthogonal curvilinear coordinate system, even locally, and if this is still true even after an arbitrary small perturbation, then it must be (≤ 3)-dimensional. If an orientable compact 3-manifold universe has a global orthogonal curvilinear coordinate system, or equivalently n commuting orthogonal vector fields, or equivalently 3 orthogonal foliations into 2-surfaces (all globally valid) then it must be T^3. Requiring only one such global foliation (and demanding it be C^∞) is equivalent to demanding that the Euler characteristic be 0.

6 TGC, curvature-homogeneity, & Seifert fiber spaces

The Thurston geometrization conjecture (TGC) [5][24][111][164][145][73][168] asserts that every smooth compact orientable 3-manifold[22] can be split in a canonical way by disjoint 2-spheres and 2-tori into pieces which may be diffeomorphed to have one of 8 particular geometric structures listed in Thurston's 8 geometries theorem below.

Here is the plan of this section. In §6.1-6.4 we shall describe the TGC more precisely after first doing some preparatory work about Lie groups and decompositions. §6.2 introduces the notion of "k-curvature homogeneity" and points out that Thurston's 8 geometries theorem may instead be regarded as being about 1-curvature homogeneous 3-manifolds. Seifert fiber spaces are important in the Thurston classification; §6.5 describes them.

This all will not only serve as a review of the TGC but also presents some new views and some warnings preventing the reader from falling into some mental traps.

6.1 Thurston's 8 geometries

A "geometric structure" on a manifold is a complete locally homogeneous Riemannian metric.

Theorem 9 (Thurston's 8 geometries). *There are exactly 8 maximal, simply connected, 3-dimensional geometric structures which have transitive automorphism groups with compact point-stabilizers. An annotated list of them is below.*

1. The three spaces H^3, E^3, and S^3 of constant (negative, zero, and positive) curvature.

2. The homogeneous but non-isotropic product geometries $S^2 \times E$ and $H^2 \times E$;

3. The universal cover $\widetilde{\mathrm{SL}_2\mathbb{R}}$ of $\mathrm{SL}_2\mathbb{R}$ (i.e. the multiplicative group of 2×2 real matrices with determinant 1). This can be equipped with the following metric invariant under left-multiplications. Let

$$C = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix} \quad (28)$$

[21] Perhaps the methods used to prove theorem 8 would enable progress.

[22] We shall assume orientability throughout this section; the non-orientable case is treated in [24].

be the singular value decomposition [65] of $C = A^{-1}B$. Then the distance between matrices A and B is[23] (where $|\alpha+\beta|$ is computed modulo 2π in such a way that the results are $\leq \pi$, and where $\lambda \geq 1$)

$$\text{dist} \;=\; |\alpha+\beta| + |\ln\lambda|. \tag{29}$$

It is also possible to combine the two terms by taking the square root of sum of squares instead of adding; numerous other combinations also are possible. This highlights the fact that invariant metrics of nonconstant curvature can be very nonunique. The distinction between $\mathrm{SL}_2\mathbb{R}$ and its universal cover is clarified by considering allowing the angle $\alpha+\beta$ to be an arbitrary real rather than modulo 2π.

Yet another way to look at $\mathrm{SL}_2\mathbb{R}$ is to regard $\mathrm{SL}_2\mathbb{R}/\pm I$ (via a group isomorphism) as the group $SO(2,1)$ of rigid motions (orientation-preserving isometries) of the hyperbolic plane H^2. Then each element of $\mathrm{SL}_2\mathbb{R}$ may be regarded [24] as an element of the tangent bundle (location×direction arrow) of H^2 – for which a left-invariant metric is readily constructed. The universal cover $\widetilde{\mathrm{SL}_2\mathbb{R}}$ then arises by *not* taking the direction-arrow angle modulo 2π.

Finally, here is an explicit 2-parameter family of left-invariant Riemannian metrics:

$$\mathrm{d}s^2 = \frac{\mathrm{d}t^2}{|a+b|} + |a+b|e^{-2t}\mathrm{d}x^2 + (\mathrm{d}y + \sqrt{2b}e^{-t}\mathrm{d}x)^2 \tag{30}$$

where a and b are constants with $b > 0$ and $a + b < 0$.

4. "Nil-geometry," the multiplicative group (often called the "Heisenberg group," because of its connection to the quantum-physically important "Heisenberg algebra" generated by 1, q and $p = \frac{\partial}{\partial q}$) of matrices

$$\begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} \tag{31}$$

under multiplication. A different representation is that the Heisenberg group is the 2-tuples (z,t) where z is complex and t is real, with group operation

$$(z_1,t_1)\circ(z_2,t_2) = (z_1+z_2,\; t_1+t_2+\mathrm{im}(\overline{z_1}z_2))\,; \tag{32}$$

here the identity element is $e = (0,0)$. If we define the "commutator" $[a,b] = aba^{-1}b^{-1}$, then $[a,b]$ is generally nonidentity in the Heisenberg group, i.e. it is a non-abelian group. However, $[[a,b],c] = e$, i.e. double commutators vanish. Thus this group is "nilpotent," hence the name "nil."

Cygan's metric [47]

$$\begin{gathered}\text{dist}\left((z_1,t_1),(z_2,t_2)\right) = \\ \left(|z_1-z_2|^4 + [t_1-t_2+\mathrm{im}(\overline{z_1}z_2)]^2\right)^{1/4}\end{gathered} \tag{33}$$

plainly is invariant under operations applied on the left from this latter group (and indeed, if the exponents $4, 2, 1/4$ were arbitrarily altered, the resulting formula still would be left-invariant).[24] But Cygan's metric is *nonRiemannian.*

A 1-parameter family of Riemannian left-invariant metrics [67] is

$$\mathrm{d}s^2 = \frac{1}{2b}\left(\mathrm{d}x^2 + \mathrm{d}z^2 + (\mathrm{d}y + x\mathrm{d}z)^2\right) \tag{34}$$

where $b > 0$. It is known [173][110] that every Nil-invariant Riemannian manifold has at least one positive and one negative sectional curvature everywhere, and negative scalar curvature $R < 0$???

5. "Sol-geometry," the following group-operation on real 3-tuples (x,y,z):

$$(x,y,z)\circ(X,Y,Z) = (x + e^{-z}X, y + e^{z}Y, z + Z). \tag{35}$$

Note $(0,0,0)$ is the identity element. This is a *sol*vable Lie group. A 1-parameter family of left-invariant metrics is

$$\mathrm{d}s^2 = \frac{2}{b}\left(e^{2z}\mathrm{d}x^2 + e^{-2z}\mathrm{d}y^2 + \mathrm{d}z^2\right) \tag{36}$$

where $b > 0$. Sol geometry also may be thought of as the group of rigid motions of $(1+1)$-dimensional spacetime with metric $\mathrm{d}t^2 - \mathrm{d}x^2$, i.e. the 1D Lorentz transformations, also called[25] the "Poincaré group" [171]. E.Heintze [75] showed that every nonflat Sol manifold has scalar curvature $R < 0$.

All of these manifolds, except for those with H^3 structure, have been completely classified (§4 of [145], [51]).

There are numerous remarks to be made about theorem 9. It is (at least if the TGC is correct) of fundamental and foundational importance in 3-manifold theory. But, although this theorem often is described as "not very difficult," it is discomforting that I have never seen any succinct complete self-contained proof. Proofs rely on a classification of all 3-dimensional Lie groups, which was first performed by Bianchi [19] in a paper written in Italian in 1898 which hardly any modern author has ever read and which relies on extensive case analyses and manual metrical computations[26]. Fortunately, it recently has appeared in English translation. Then additional case analyses, this time group theoretic, discard those Lie groups which lead only to noncompact locally homogeneous manifolds.

The idea behind that proof approach is that any metric symmetric under a Lie group is necessarily "locally homogeneous" in the strong sense that each of its points is equivalent, under some symmetry, to each other. However, it is *not* immediately apparent that it is *necessary* that a locally homogeneous metric (i.e., a metric such that any pair of its points have isometric neighborhoods) be symmetric under any Lie group. To make an analogy in the combinatorial world of graphs rather than the continuum world of manifolds: there are a tremendous number of completely *unsymmetrical* graphs G, each point of which has identical local structure out to any given constant distance. It is for this reason that the above statement of theorem 9 concerns only metrics with transitive automorphism groups. Nevertheless, as we shall see in the next two theorems (and this remark is new, although the ingredients had long been available), Lie-ness *is* necessary, and hence the demand for a transitive automorphism group may be dropped.

[23] Nobody seems to have given an explicit left-invariant metric for $\mathrm{SL}_2\mathbb{R}$ before.
[24] Change the sign of the im to get a right-invariant form of the Cygan metric.
[25] More usually the name "Poincaré group" has denoted the isometries of $(3+1)$-dimensional Minkowski space.
[26] Guido Fubini (1879-1943), a former student of Bianchi's, extended the classification to 4 dimensions in [62].

6.2 Lie groups and curvature homogeneity

A famous 1952 theorem of A.Gleason, D.Montgomery, and L.Zippin, which solved "Hilbert's fifth problem," states that any "topological group" which is "locally connected," "locally compact" and "finite dimensional" in fact necessarily is isomorphic to a finite dimensional Lie group with *analytically-smooth* group-multiplication and group-inverse laws. Montgomery and Zippin's book [118] is highly recommended and carefully gives all fundamental definitions of terms such as "topological space."

Another fundamental theorem [50] whose first version was shown (independently) by G.Birkhoff [21] and S.Kakutani [88] in 1936, states that for any Lie group properly acting on a manifold M, there is a left-invariant Riemannian metric on M. Indeed in view of the Gleason-Montgomery-Zippin theorem, we may demand (if the group acts transitively) that this metric be analytically smooth and Riemannian. In fact, as we have seen there often are *many* such metrics for only one group. For this reason, theorem 9 really only classifies the groups of our metrics, not the metrics themselves. Fortunately, because of the Gleason-Montgomery-Zippin theorem and the transitivity of these groups, this seems entirely adequate for the purpose of specifying their topologies.

Although every Lie group has a left-invariant metric, for many Lie groups there is no *bi*-invariant metric on the group elements. In fact according to the lemma p.296-7 of [110], the *only* bi-invariant metrics[27] of connected Lie groups are those isomorphic to cartesian products of a compact with a commutative group. In particular, in our case Sol, Nil, and $\widetilde{\mathrm{SL}_2\mathbb{R}}$ have no bi-invariant metrics, whereas S^3 (arising from the compact group $SO(3) = SU(2)/\pm I$), E^3 (commutative), and $E^1 \times S^2$ do.

Warning. It has often been stated in the differential geometry and general relativity literature that a Riemannian metric is "homogeneous" if and only if all covariant derivatives $R_{\alpha\beta\mu\nu;\zeta}$ of the Riemann curvature tensor are identically zero. However, at least with *our* definition of "locally homogeneous," which is that any two points have isometric neighborhoods, this is untrue.

One could also consider "curvature homegeneous" manifolds [149] such that any two points A, B have (with an appropriate choice of coordinates systems at A and B) the same Riemann curvature tensor. More generally one could consider the

"k-curvature homogeneous" manifolds: those such that any pair of points A, B have the same Riemann curvature tensor *and* the same covariant derivatives of that tensor, up to and including kth derivatives (where $k \geq 0$ is any fixed integer).

The relationships among these notions for 3-manifolds are summarized in the following

Theorem 10 (3-Metrics with constant principal Ricci curvatures). *A Riemannian 3-manifold M is curvature homogeneous if and only if the 3 eigenvalues of*[28] *the mixed Ricci tensor R_α^β are constant on M. It is 1-curvature homogeneous if and only if it is locally homogeneous. Any locally homogeneous 3-manifold is curvature homogeneous but the reverse implication is untrue. Finally, there exist locally homogeneous 3-manifolds with nonzero $R^\beta_{\alpha;\zeta}$.*

Proof: Obviously, a locally homogeneous 3-manifold, i.e. one in which any two points have isometric neighborhoods, necessarily has Riemann curvature tensor identical at any two points A and B (with appropriate choices of coordinate systems). This therefore also is true of the Ricci tensor $R_{\alpha\beta} = R_{\beta\alpha} = R^\mu{}_{\alpha\mu\beta}$. Indeed, the Maclaurin expansion of the metric tensor $g_{\alpha\beta}$ in the neighborhood of a point P is, in Riemann normal coordinates x^α centered at P,

$$g_{\alpha\beta} = I - \frac{1}{3} R_{\alpha\nu\beta\mu} x^\mu x^\nu - \frac{1}{6} R_{\alpha\nu\beta\mu;\gamma} x^\mu x^\nu x^\gamma + O(|x|^4). \quad (37)$$

Since the metric tensor at P is merely the identity matrix I, index lowering and raising has no effect at P. Hence $R_{\alpha\beta}$ and R_α^β are the same thing at P, so that R_α^β is a *symmetric* 3×3 matrix at P. Hence it has 3 real eigenvalues and 3 orthonormal eigenvectors, and hence by an appropriate rotation and/or reflection of the coordinate system (determined by the eigenvector matrix) we can cause R_α^β at any given point P to be, in fact, a diagonal matrix in some Riemann-normal coordinate system centered there. Hence, if all points P have the same Riemann (and hence Ricci) curvature, it is necessary that all points P have the same 3 Ricci eigenvalues. Conversely, since at each point of a 3-manifold the Riemann curvature tensor is *determined* by the Ricci tensor via

$$R^\alpha{}_{\beta\mu\nu} = \delta^\alpha_\mu R_{\beta\mu} + g_{\beta\mu} R^\alpha_\nu - \delta^\alpha_\mu R_{\beta\nu} - g_{\beta\nu} R^\alpha_\mu + \frac{\delta^\alpha_\mu g_{\beta\nu} - \delta^\alpha_\nu g_{\beta\mu}}{2} R \quad (38)$$

we see that the theorem's "if *and* only if" holds. The fact that 1-curvature homogeneous 3-manifolds are necessarily ∞-curvature homogeneous and indeed locally homogeneous and either S^3, E^3, H^3, $M^2 \times E^1$, or a simply connected complete left-invariant 3-metric of a 3-dimensional Lie group, is theorem B of [147]. Singer's original paper [149] had already shown that k-curvature homogeneous m-manifolds are necessarily locally homogeneous if k is sufficiently large, indeed if $k \geq (m-1)m/2$. The fact that curvature homogeneous 3-manifolds exist, which are not locally homogeneous, was first shown by Takagi [157]. Indeed, work of Kowalski et al. [32] [95] has shown that there exist at least as large an infinity of different curvature homogeneous 3-manifolds (with fixed distinct Ricci eigenvalues) as the number of pairs of univariate analytic functions, whereas the number of locally homogeneous 3-manifolds is of course far smaller. Finally, the (necessarily locally homogeneous) examples we have given above of left-invariant infinitesimal length elements ds^2 for $\widetilde{\mathrm{SL}_2\mathbb{R}}$, Nil, and Sol have (by direct computation) the following mixed Ricci tensors R_α^β:

$$\begin{pmatrix} a & 0 & 0 \\ \sqrt{2}be^{-t}(b-a) & b & 0 \\ 0 & 0 & a \end{pmatrix}, \quad \begin{pmatrix} -b & 0 & 0 \\ 0 & b & 2bx \\ 0 & 0 & -b \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -b \end{pmatrix} \quad (39)$$

[27] A left-invariant metric may be thought of as a symmetric geometric structure whose symmetry group is the Lie group. But if the metric is not bi-invariant, then it is far less enjoyable to regard that structure *as* the Lie group.

[28] The 3-manifolds with constant Ricci eigenvalues have been completely classified [32] [95]. For other works on this topic see [94] [149] [157].

(note in all cases the matrix is triangular, so that the 3 eigenvalues are merely the 3 diagonal entries, which, note, are indeed constants) which have the following as particular nonzero covariant derivatives:

$$R^x_{x;t} = \frac{(a-b)b}{|a+b|}, \quad R^z_{y;x} = b, \quad R^x_{z;x} = -b. \tag{40}$$

Q.E.D.

In view of Sekigawa's theorem B [147], we now make the new point that one may strengthen the statement of theorem 9 to assume nothing about either groups or neighborhoods larger than a single point on the manifold. (Indeed, essentially all the statements in this paper about locally homogeneous 3-manifolds may now be strengthened to apply to 1-curvature homogeneous 3-manifolds.)

Theorem 11 (The Thurston 8 – Stronger statement). *There are exactly 8 kinds of maximal simply connected complete compactifiable[29] 1-curvature homogeneous Riemannian 3-metrics, all of which arise as left-invariant metrics of Lie groups and are listed in theorem 9.*

This statement is strongest possible in the sense that it would become totally false if "0" were substituted for "1."

6.3 Decompositions

Now that we understand the (conjectured) atomic constituents of 3-manifolds, let us now describe the canonical way to cut up the manifold into atoms.

The *connected sum* $M\#N$ of two n-manifolds M, N is the n-manifold got by deleting the interiors of n-balls, one from M and one from N, then gluing the resulting punctured manifolds to each other via a homeomorphism of the S^{n-1} ball surfaces.

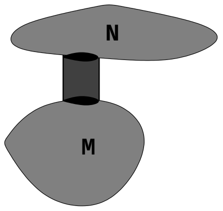

Figure 6.1. Connected sum $M\#N$.

The **Sphere Decomposition theorem**[30] [92][109][70] says that any closed oriented 3-manifold M may be decomposed into a finite connected sum $M = M_1\#M_2\#\cdots\#M_k$ where each M_i is prime. The collection $\{M_i\}$ is *unique* up to permutation of the summands. (A closed 3-manifold M is "prime" if $M = A\#B$ implies either A or B is homeomorphic to S^3. A fundamental theorem of Alexander in 1924 states that any 2-sphere in an S^3, both *smooth*, bounds a ball on each side; Alexander's "horned sphere" shows this is false if the smoothness requirement is dropped.) This sphere decomposition is obtained by choosing an embedded 2-sphere in M which does not bound a 3-ball (if one exists – otherwise the procedure terminates); cutting along it; finally the spherical "holes" in the resulting manifolds are filled in with 3-balls, and we continue on by reducing them recursively[31].Warning: although the #-decomposition of a manifold is unique, the manifold $A\#B$ need not be, because the spheres on A and B can be glued by either an orientation preserving or reversing diffeomorphism.

A 3-manifold M is *irreducible* if every smooth 2-sphere embedded in M bounds a 3-ball in M. Clearly an irreducible 3-manifold is prime. The converse is almost true (lemma 3.13 of [77]) the only prime orientable 3-manifold with a nonseparating S^2 is $S^2 \times S^1$.

The **Jaco-Shalen-Johannson torus decomposition theorem** [83][87] says that if M is a closed oriented *irreducible* 3-manifold, then there is a finite collection of disjoint incompressible 2-tori $T_1, T_2, \ldots, T_k \subset M$ whose removal separates M into a finite collection M_1, M_2, etc. of compact 3-manifolds with toral boundaries, each of which is either atoroidal or Seifert fibered (§6.5). The minimal-cardinality such collection is *unique* up to isotopy. ("Isotopy" here means a homotopy of the 2-tori in which every partly-homotoped version remains embedded.)

To define "incompressible": Let Q be a compact, oriented surface embedded in an irreducible 3-manifold N with $\partial Q \subset \partial N$. (The notation ∂K denotes the *boundary* of K.) Q is *incompressible* if for every closed 2-disc D embedded in N with $D \cap Q = \partial D$, the curve ∂D is contractible in Q to a point. Otherwise Q is *compressible*[32]. For example, let Q be a nonseparating toroidal surface embedded in T^3 and used to cut it (hence we get two bounding surfaces Q_1 and Q_2); these Q_k are incompressible. A compact irreducible 3-manifold is *Haken* if it contains a two-sided incompressible surface of genus $g \geq 1$.

Warning: to get uniqueness, it is necessary to avoid splitting Seifert-fibered pieces (even if they contain incompressible 2-tori) and choose the torus-collection to be minimal. Furthermore, although M determines j and the $M_1, M_2, \ldots, M_j$ uniquely, the converse is untrue: there are an $SL(2,\mathbb{Z})$ of ways to glue back together on each T_i.

6.4 Thurston Geometrization Conjecture

Conjecture 12 (Thurston's geometrization conjecture (TGC)). *After we split a 3-manifold into its connected sum according via the sphere decomposition, and then take the Jaco-Shalen-Johannson torus decompositions of the summands, the resulting components each are diffeomorphic to 3-manifolds having one of the 8 magic geometric structures from theorem 9. This decomposition is unique up to isotopies*

[29] A manifold is "compactificable" if it may be cut into pieces and some of these pieces may be glued together to create a compact boundaryless manifold. For example E^3 is compactifiable to T^3, but no left-invariant 3-metric of the group (which Milnor [110] calls "$E(2,1)$") of rigid motions of the Euclidean plane E^2 is compactifiable – or at least, if it is, then no new compact metrics beyond those listed in theorem 9 result. Similarly [24] at the end of his §2.3 claims that the geometry of $\mathbb{R}^3$ with metric $ds^2 = e^{2az}dx^2 + e^{-2bz}dy^2 + dz^2$ with fixed $a \neq b$, admits no finite volume version.

[30] In fact, Kneser gave an upper bound on the number of pieces required, and Jaco and Tollefsen [84] gave algorithms to find a maximal sphere decomposition.

[31] The sphere decomposition theorem is also true for closed oriented 2-manifolds, and it says that a genus-g manifold is expressible as the connected sum of g tori (except if $g = 0$ when we have S^2).

[32] Two equivalent formulations: Q is incompressible in N iff ① any loop in Q that is contractible in N is contractible in Q, or ② any loop in Q that is uncontractible in Q is uncontractible in N.

of the 2-tori and permutations of the summands in the connected sum decomposition.

Considering known partial results on the TGC[33], an **equivalent conjecture** is that any irreducible non-Seifert atoroidal 3-manifold (which we allow to have boundaries consisting of 2-tori) may be equipped with an H^3 geometric structure.

Note that Thurston's pieces are not necessarily simply-connected, since $S^2 \times S^1$ is an admissible (Seifert-fibered) piece. Thurston proved that compact boundaryless irreducible atoroidal Haken 3-manifolds always admit a hyperbolic structure ([120] p.51). Mostow [121] proved that 3-manifolds admit at most one hyperbolic structure ("Mostow's rigidity theorem"). It is known that hyperbolic manifolds are necessarily atoroidal, irreducible, and not Seifert-fiberable. *The TGC asserts the converse.*

Recently Perelman has claimed a proof of the TGC [131], although the jury is still out on it. An immediate consequence of this would be (as was proven by Thurston) a proof of the notorious **Poincaré Conjecture** (PC) which states that "Every simply connected closed 3-manifold is homeomorphic to S^3." The PC has been proved in every dimension $n \geq 2$ other than 3.

6.5 Seifert fiber spaces

A **Seifert fiber space** is a 3-manifold foliatible into circles. Actually, this was not Seifert's original definition, but it has been shown [56][52] that any 3-manifold foliatable into 1-manifolds is foliatible into circles[34], is Seifert, and that the fibering is always an S^1 bundle over a 2-orbifold X.

Seifert [146] completely classified his spaces. Usually Seifert fiber spaces have an isotopically unique fibering (e.g. if the Euler charcteristic of X is negative) and the cases of nonuniqueness are now totally understood [23][126][127]. It was later also almost entirely understood which Seifert spaces admit foliations into 2-manifolds transverse to the fibers [54][114].

7 Twist

in this section we consider the "no twist postulate" that the universe does not contain a closed geodesic such that parallel translation along it for 1 cycle induces a twist (with the geodesic as axis) other than an integer multiple of 2π radians. In §7.1 we'll examine microscopic-physics reasons for believing this postulate; but there also are reasons to doubt it (which we shall also explain). We shall see that there are many possible different slightly-inequivalent wordings of this postulate; this paper tentatively settles on one (denoted "NTA") for the purposes of concreteness.

§7.3 recounts what is known about "ergodicity" of the "geodesic and frame flows" on manifolds. Finally §7.4 uses that knowledge to find important consequences of the no twist postulate. For example, we determine which 1-curvature homogeneous 3-manifolds are twist-free.

7.1 Physics and twist

Idealize photons as oriented particles, where the orientation is the photon's polarization and is transverse to its direction of motion. Then the polarization of a photon traveling around a twisted closed geodesic would twist each cycle, time-averaging to zero, causing such a photon to cancel itself out. Now really, photons have both particle- and wave-like characteristics; such a photon in a "momentum eigenstate" (i.e. whose momentum is exactly known) would be equidistributed all along the geodesic and hence presumably could not be regarded as *having* a polarization. Under a similar idealization, electrons with spin also could not exist in a momentum-spin eigenstate – *unless* our closed geodesic were untwisted.

But, experimentally, certainly no limit to how accurately one can simultaneously measure the momentum (in any particular fixed direction) and the polarization (or spin) of a photon (or electron) has even been encountered. These quantities are normally thought of as independently specifiable, in principle arbitrarily accurately.

Furthermore, the spin of an electron is normally regarded (for an electron isolated from all external forces) as an unchanging quantum number $+1/2$ or $-1/2$. But if that electron hapened to be traveling round a twisted closed geodesic, that could not be the case.

If it is desired that these properties of flat space quantum mechanics should still work in our (curved-space) universe, then one is led toward the

(Naive) No Twist Assumption (NNTA): The universe 3-manifold does not contain closed geodesics with twist other than an integer multiple of 2π radians.

Unfortunately, there are many possible alternative versions of this assumption, and it is not obvious which (if any) are the right ones. Perhaps *some* twisted closed geodesics should be permitted, provided they are (in some suitable sense – which?) vastly in the minority. Another problem is that this picture of electrons and photons as "point particles moving along a geodesic" is an idealization valid only in the nonquantum $\hbar \to 0$ limit. Since no consensus has yet been reached about what quantum mechanics in curved space should be, the true picture is unclear.

Finally, even if the NNTA and its underlying reasoning is regarded as the only possible alternative, we still must investigate its internal self consistency. What happens to the geodesic twists if the 3-metric is perturbed? The case for the NNTA would be most convincing if the geodesic twists (or at least the allegedly-physically-forbidden nonzero ones) were *topological invariants*. We shall investigate that in the next subsection with the conclusion (theorem 14) that, although geodesic twists in many ways *act like* a topological invariant, in fact they are not one. This seriously undercuts any argument that physics demands the no twist assumption (at least as worded above). But there are two reasons to suspect it does not kill completely:

1. The counterexample theorem 14 provides of just *one*, isolated, twisted closed geodesic, may have little physical relevance.

[33] I.e., it is proved except for the hyperbolic cases.

[34] Despite that, K.Kuperberg [96] showed that there are C^∞ vector fields on S^3 with no periodic orbit; Thurston soon afterwards realized that ① such fields may indeed be taken as real-analytic, and ② they exist not only on S^3 but in fact on every smooth 3-manifold.

2. The whole argument in favor of the NNTA, and indeed the whole basis for wording it in terms of "geodesics" at all, was based on an invalid idealization of "classical point particles." Therefore the counterexample from theorem 14 may be more a sign of the limits of validity of this idealization, than of invalidity of the underlying idea.

So quite likely *some* modified and heavily reworded (but still very similar in spirit) variant of the NNTA is correct, whose statement is currently unknown.

To get a concrete problem to work on, we shall concern ourselves with this particular version:

(Topological-homogenizing) No Twist Assumption (NTA): Any 3-manifold homeomorphic to the universe but maximally homogeneous[35] does not contain closed geodesics with twist other than an integer multiple of 2π radians.

It has several advantages:

1. This statement is a genuine topological invariant – sidestepping objections about "self consistency."
2. This version seems weaker – increasing its probability of being true. We do not exclude a universe merely because it has a twisted closed geodesic; we only exclude it if *every* topologically equivalent homogeneous universe does.
3. The motivation for inserting the words "maximally homogeneous" is partly a feeling that the maximally homogeneous metrics for a topology are the "most beautiful and natural" ones, partly to make analysis easier, and partly because (see §10) of arguments that the universe initially *was* homogeneous.
4. Although the precise wording of the NTA probably is critical for some purposes, that is not the case for the purposes for which it is employed in this paper. Specifically: the theorems we shall obtain in §7.4 are worded so that it *does not matter* whether the NNTA or NTA is employed.

7.2 Twist as a topological invariant

This subsection examines the following

Question. Let M be a smooth 3-manifold and let g be a closed geodesic on M with twist angle θ. If M is distorted via some smooth homotopy to become M', what will happen to g and θ? In particular, will θ remain constant, so that "twists of closed geodesics are topological invariants?" Is it required that $\theta \neq 0$ for it to be an invariant?

Related questions. *How many* closed geodesics g can exist on a 3-manifold M? How many closed geodesics g with *nonzero twist* can exist?

The answers turn out to be surprising[36]. The twists θ of closed geodesics (even if we restrict attention to nonzero twists) are *not* topological invariants of 3-manifolds M. Indeed, we shall describe two homotopic T^3 manifolds, one of which has a twisted geodesic, while the other does not. *But* the set of twist angles behaves very much *like* a topological invariant! For example if M and M' are equivalent under some homotopy and both have (and so do all the intermediate 3-manifolds in the homotopy) *nonnegative* sectional curvatures everywhere, then the set of θ is the same for both M and M'. Seifert [146] argued that the twists of Seifert fibers are invariant under isotopies of the fibration and homotopies of the 3-manifold.

Arguments justifying the answers. Here is an **argument** that θ "must" remain constant. As we shall see later, this argument is flawed, but we also shall see that the gap in it can *sometimes* be patched.

Regard the closed geodesic g as a "loop of string" and parallel translate some direction orthogonal to g along g. The resulting object is a "ribbon." If the twist angle associated with g is a rational number of degrees then the ribbon may be thought of as closing upon itself after it makes some integer number of cycles round g. Now suppose we smoothly distort the 3-manifold in such a way that the consequent distortion of g and of the ribbon also is smooth. Then clearly the ribbon cannot change its number of twists. So we conclude that the twist angle of g, provided it is rational, is invariant under this class (which seems very wide) of distortions of the underlying 3-manifold. Finally, since this works for all *rational* twist angles, by some kind of continuity it must work for all *real* ones. "Q.E.D."

The flaw in this proof was its implicit assumption that g will continue to exist throughout the distortion process. In fact, g can suddenly either split into several closed geodesics, or vanish, as is shown by the following **pimple example.** Let $F(r)$ be a monotonically decreasing real-valued function of $r \geq 0$ whose graph consists of an upper semicircle, joined smoothly to the asymptotic behavior $F(r) \to 0$ as $r \to \infty$. Now consider the d-dimensional surface in $\mathbb{R}^{d+1}$ that is the graph of $F(r)$, where r is the distance to $\vec{0}$ in $\mathbb{R}^d$. If $d = 2$ this surface resembles a "pimple" rising above a flat plane. The equator of the pimple is a closed geodesic g, which suddenly vanishes under a smooth distortion "flattening" the pimple. On the other hand, if the upper hemisphere is raised by height H by some cylindrical joining inserted in place of the equator, then g splits into a continuum infinity of closed geodesics going around that cylinder. Incidentally, when we say a closed geodesic "vanishes," of course a geodesic starting from any point in any direction always exists; what may cease to exist is any requirement that the geodesic close on itself.

The **ellipsoid counterexample** arises by considering distorting the standard round sphere S^2 into a generic ellipsoid. Every geodesic on the sphere is closed and of length 2π. But although the 3 major ellipses of an ellipsoid still are closed geodesics, generic starting points and directions yield geodesics which do not close [93][156][37]

If M everywhere has *negative* sectional curvatures, then neither splitting nor vanishing is possible for closed geodesics.

[35] Constant curvature metrics are more homogeneous than k-curvature homogeneous metrics, which are more homogeneous than j-curvature homogeneous ones if $k > j$.

[36] I am not sure how new these questions or answers are. They are not mentioned in any of the reviews [17][77][91].

[37] A complete understanding of geodesics on ellipsoids was first acquired by Jacobi [85] in 1839; these references redo this in modern language. For the Lusternik-Shnirelman theorem that every manifold diffeomorphic to S^2 has at least 3 simple closed geodesics, see [13]; for the theorem that it has an infinity of geometrically distinct closed geodesics see [14][78].

That is because there is a unique shortest closed geodesic g_c of each possible combinatorial type c, indeed any closed loop of the same type may be shortened to become g_c. (Note also that the total number of closed geodesics on such an M is necessarily countably infinite, since the number of combinatorial types is countable.) In fact, it suffices if M has negative sectional curvatures everywhere *on* g, in all tangent planes to g there (that is, sectional curvatures in other planes do not matter), and that this property continues to hold throughout the distortion process. In other words, for this restricted, but still large, class of manifolds M and geodesics g, our original flawed proof actually is valid.

The reason this is true is Synge's formula[38]

$$\cdot\frac{\mathrm{d}^2\ell}{\mathrm{d}\epsilon^2}\,|_{\epsilon=0} = \int_a^b \left(|\vec{y}'(t)|^2 - K(c'(t),\vec{y}(t))|\vec{y}(t)|^2\right)\mathrm{d}t \qquad (41)$$

for the second variation of the arc length $\ell = \int_a^b |c'(t)|\mathrm{d}t$ of a geodesic curve segment $c(t)|_{t=a}^{t=b}$ of the geodesic when it is perturbed by $\epsilon\vec{y}(t)$. It shows that a closed geodesic is a strict local minimum of arc length, and hence can neither split nor vanish, if the sectional curvature K in the 2-space generated by the geodesic's direction $c'(t)$ and the perturbation direction $\vec{y}(t)$ is always negative *at each point on the geodesic.* (And, of course, the reason a geodesic exists of each possible combinatorial type is by considering shortening a curve of that type.)

We summarize the conclusion of the above argument as a theorem:

Theorem 13 (Twist acts like a 3-manifold invariant). *Let M be a smooth 3-manifold distorted by some smooth homotopy to another: M'. Suppose g is a closed geodesic on M. Suppose that during the homotopy the sectional curvatures of M, at each point on g, in each 2-plane tangent to g, remain negative. Then g, while continuously distorting as the homotopy proceeds, remains a closed geodesic, and its twist angle θ remains constant.*

When we consider the fate of theorem 13 on Thurston's 8 locally homogeneous 3-metrics listed in §6, we find:[39]

1. Hyperbolic 3-manifolds (i.e. based on H^3) have constant negative curvature, so the argument clearly works – their geodesic twists are invariant to all perturbations small enough to keep the sectional curvatures always negative.
2. $H^2 \times E$ has two zero and one negative sectional curvature, so again the argument works: all geodesic twists are invariant under any perturbation which leaves all sectional curvatures nonpositive while keeping at least one negative everywhere.
3. In the case of the locally homogeneous metric EQ 30 for $\widetilde{\mathrm{SL}_2\mathbb{R}}$, the Ricci eigenvalues $\{a, b, a\}$ (see the first matrix of EQ 39 for this metric's mixed Ricci tensor) include two negative eigenvalues a and one positive one b, and with $a + b < 0$. Since the corresponding sectional curvatures in the 3 orthonormal eigendirections include two positive and one negative one, our argument does *not* apply. Furthermore, it follows from corrollary 4.7 of [110] that *every* left-invariant metric for $\widetilde{\mathrm{SL}_2\mathbb{R}}$ has 2 positive and one negative sectional curvatures in the 3 orthonormal Ricci eigendirections (and the scalar curvature R is necessarily negative [75]).
4. For the locally homogeneous metric EQ 34 for Nil, the Ricci eigenvalues $\{-b, b, -b\}$ (see the second matrix of EQ 39) include two negative and one positive eigenvalue. The corresponding sectional curvatures in the 3 orthonormal eigendirections again include two positive and one negative one, so again our argument does *not* apply. (See also [173].)
5. For the locally homogeneous metric EQ 36, for Sol, the Ricci eigenvalues are $\{0, 0, -b\}$; the xz and yz sectional curvatures are everywhere negative, while the xy sectional curvature is positive. Furthermore, it follows from corrollary 4.7 of [110] that *every* left-invariant metric for Sol has 2 positive and one negative sectional curvatures in the 3 orthonormal Ricci eigendirections. Consequently, for any geodesic everywhere moving in a direction close enough to the ∂z direction, our argument *applies*: the twist of any such geodesic will be invariant to all metrical pertubations having a small enough effect on sectional curvatures.
6. S^3 and $S^2 \times E$: argument fails due to positive curvatures.

Even if M everywhere merely has *nonpositive* sectional curvatures, then the negative curvature argument still works, except that g_c now is only unique *up to isometry* and there may be a continuum infinity of closed geodesics of each combinatorial type. This is exactly what happens for compact flat manifolds [174]. But that argument does *not* extend to hold if we merely require nonnegative sectional curvatures everywhere *on g only*, as is shown by a version of the pimple example.

The pimple example with shrinking/expanding cylinder also shows that a continuum infinity of closed geodesics, each with zero twist and each having *zero* sectional curvature all along them, can vanish until only one remains. The sphere-ellipsoid example is an even worse pathology – the instantaneous disappearance of an all-inclusive continuum of closed geodesics – in the case of always positive curvatures.

Furthermore, from the complete classification [174] of flat 3-manifolds we know that closed geodesics on flat 3-manifolds with *nonzero* twist are necessarily isolated, hence under distortions of M cannot vanish and cannot split, and (therefore) the *nonzero* twists of *individual* closed geodesics are homotopy invariants even on flat manifolds, *provided* we only allow homotopies which keep the sectional curvatures everywhere nonnegative along each g under consideration.

Finally, here is an example of a continuum of closed geodesics, each with *nonzero* twist, on a 3-manifold of *constant posi-*

[38] The formula was found and used by J.L.Synge in the decade centered at 1930; some of its history is recounted in [132].

[39] Theorem 1.6 of [110] states that if a connected Lie group has a left-invariant metric with all sectional curvatures nonpositive then it is solvable. From this it immediately follows that the most naive form of our argument, which demands all sectional curvatures nonpositive everywhere on the manifold, could only hope to be applied for locally homogeneous manifolds arising from the following 3 geometries: H^3, E^3 and Sol. The text considers the more sophisticated form of our argument, which only asks for *some* nonpositive sectional curvatures and only at points *along the geodesic*, but reaches essentially the same conclusion.

tive curvature. Consider the "Hopf fibration" of the standard round S^3, which is a fibration of S^3 by great circles. Now do surgery on this manifold to join one end of a segment of one of these geodesics (call it g) to the other end, with a twist angle specifically chosen so that nearby Hopf geodesics (which are linked with and equidistant from g) join to themselves and have (new) length depending only on (and smoothly on) their distance from g. The result should be a 3-manifold chunk topologically equivalent to a solid torus, filled with closed geodesics each of which twists. In particular, we can in this way get a Hopf fibration of 3D elliptic geometry by closed geodesics, each with 180° twist.

These Hopf-fibration-with-surgery examples demonstrate a considerable difference between closed geodesics in spaces of positive and negative sectional curvatures: in the former, continua of twisting closed geodesics can exist, while in the latter, at most a countable number of closed geodesics can exist. However, they do *not* yield an counterexample 3-manifold with a closed geodesic g whose twist θ is not a topological invariant).

Despite all this, a counterexample exists:

Theorem 14 (Twist is not a 3-manifold invariant). *The twist angles of closed geodesics on 3-manifolds are not invariant under homeomorphisms.*

Proof: Imagine flat-torus 3-space[40] (a box with periodic boundary conditions) as filled with a medium of varying refractive index. Can we build an optical device that twists the polarization of light shining around the torus along some closed trajectory? (Note. We are not speaking here of "light" that obeys Maxwell's equations, but rather pseudo-light that follows geodesics, with its "direction of polarization" arising from parallel translation. Although our "optical" terminology thus is technically incorrect, we shall employ it anyway throughout this proof because it is more evocative.)

The answer is yes. Imagine constant refractive index regions separated by planes. As light crosses such a plane it rotates direction. We shall idealize this rotation of direction not as a sharp bent corner but rather as a tiny circular arc, whose effect is to rigidly rotate anything traveling along that arc, by its angle.

Say light initially goes in z direction with polarization in x direction. Use prisms to bend the beam 90° so the light is propagating now in the x direction with polarization in z. Bend 90° so going now in y direction with polarization still z. Bend 90° so going now in z direction with polarization y. The net resulting effect is: incoming light in the z direction with polarization in x is converted to outgoing light still in the z direction but with polarization in y, i.e. there is a 90° rotation of polarization.

By placing such optical devices in our box universe we can alter the light-travel-time metric in such a way that there is now a closed geodesic featuring twist, although previously all closed geodesics had been untwisted. Both the old and the new metric have the same T^3 topology. Q.E.D.

7.3 Ergodicity of geodesic and frame flows

We now lay out the fundamental definitions and results about ergodicity of geodesic and frame flows. This knowledge will be employed next subsection to help get our main results about twist.

A bijective measure-preserving transformation T of some measurable space S is *ergodic* [8] if all T-invariant subsets of S necessarily have either zero or full measure.

It is *mixing* if the measure of $T^N(A)\cap B$ tends (as the number N of iterations tends to ∞) to meas(A)meas(B)/meas(S) for any fixed measurable subsets A,B of S.

The *Bernoulli map* on the unit interval $[0, 1)$ is the following. Let there be a fixed partition of the unit interval into $k \geq 2$ interior-disjoint subintervals of lengths $p_1, \ldots, p_k$. The map $F(x)$ is then piecewise linear with slope $1/p_j$ within subinterval j; finally everything is taken mod 1 to get a selfmap of $[0, 1)$. A transformation is *Bernoulli* if it is isomorphic to some Bernoulli map.

All these things also may be defined for continuous-time flows rather than for iterated transformations (discrete time).

Bernoulli implies the "K" (for Kolmogorov) property implies mixing implies ergodicity, but [8][27][79] none of the the reverse implications hold: for example an irrational rotation T of S^1 is ergodic but not mixing.

Let $n \geq 2$. Physically speaking, it seems almost obvious that the geodesic flow on a compact n-manifold M of constant negative curvature is ergodic and mixing, since the n-volumes and $(n-1)$-surface areas of n-balls of radius r on any manifold of constant negative curvature increase ultimately *exponentially* with r.

The consequence that the present paper needs of ergodicity of the geodesic flow is: that starting from almost anywhere and traveling at unit speed in almost any direction[41] will cause one to come arbitrarily close to any desired point of M, while simultaneously moving in a direction arbitrarily close to any desired direction. In other words, that each generic geodesic densely covers the position×direction space $M \times S^{n-1}$.

On the other hand, since every geodesic on the standard round S^n is a great circle of circumference 2π, ergodicity certainly does *not* hold for compact manifolds of *positive* curvature. Also, consideration of the standard flat torus (unit n-cube with opposite faces identified) shows that the all-directions condition is violated for it, so that ergodicity also does not hold for flat compact manifolds. However, geodesic flow still is *positionally* ergodic on flat compact n-manifolds, i.e. obeys the weaker property that almost any geodesic densely covers M.

These ideas all date back to J.Hadamard and N.I.Lobachevsky in the late 1800s [8][71][11] and have been rigorously proven [123] as well as extended in various ways. For example

Theorem 15 (Anosov [6]). *The geodesic flow on any n-manifold, $n \geq 2$, with everywhere-negative curvature, automatically is ergodic on*[42] $M \times S^{n-1}$.

[40] Many other 3-manifolds could have been used instead, such as S^3 $S^2 \times S^1$.

[41] "Almost any" means, more precisely, a full-measure subset.

[42] Warning: More precisely, we should have said "the tangent space of M," which is not necessarily the same as the product $M \times S^{n-1}$ if $n \geq 3$. That level of explicative precision will be irrelevant for our purposes, and indeed the two concepts are the same if we only think in terms of sets rather than manifolds.

Indeed, it is even Bernoulli [128][137]. The cleanest proof is in the appendix by Misha Brin of [12].

Say a manifold has *κ-nearly constant negative curvature* if all its sectional curvatures are in the interval $[-1, -\kappa)c$ for some constants $\kappa, c > 0$. It is known [34][89][133][172] that the time-1 map of the geodesic flow on any smooth compact oriented boundaryless connected n-manifold M with 0.981-nearly constant negative curvature is "stably ergodic" and "stably K" on $M \times S^{n-1}$ if $n \geq 2$, and that the flow itself is Bernoullian. (The statements that the time-1 map is stably ergodic [or stably K] are slightly stronger than the statements that "the geodesic flow is ergodic [or K].")

So far we have regarded the geodesic flow as a flow on $M \times S^{n-1}$ (position×direction space). One may also consider the *frame flow*, which is a flow on $M \times \mathrm{SO}(n-1)$. This flow is induced by unit-speed movement along geodesics on the n-manifold M, where as we move we consider the rotation, described by a matrix in $\mathrm{SO}(n-1)$, of an orthogonal local coordinate-frame as it is parallel-translated along the geodesic. The reason this is $\mathrm{SO}(n-1)$ and not $\mathrm{SO}(n)$ is that one of the coordinate axes always stays pointing along the geodesic.

Brin and Gromov [30] showed for *odd* $n \neq 7$, the frame flow on a C^3-smooth compact boundaryless connected n-manifold M with negative sectional curvatures is ergodic, K, and Bernoullian. Brin and Karcher [31] showed the same thing for *even* $n \neq 8$ provided the manifold had 0.865-nearly negative curvature. Finally, Burns & Pollicott [33] attacked the remaining cases $n \in \{7, 8\}$, showing that the time-1 frame flow on a C^3 compact oriented boundaryless connected n-manifold M with 0.981-nearly constant negative curvature, for *any* $n \geq 2$, is stably ergodic and stably K, and that the frame flow itself is ergodic, K, and Bernoullian.[43] The best constant κ in the curvature-pinching condition still is unknown and by the preceding results lies somewhere in $[0.25, 0.981]$, with the 0.981 being improvable to 0.865 in even dimensions $n \neq 8$, and with *no* pinching condition being required at all in *odd* dimensions $n \neq 7$.

7.4 Theorems and proofs about twist

The *application* of the NTA and ergodicity to winnowing cosmologies is this:

Theorem 16 (Compact negatively curved manifolds have many closed geodesics that twist). *Let M be a compact 3-manifold with all sectional curvatures always negative. Then there are a countable infinity of closed geodesics τ on M, such that parallel translation one cycle along τ suffers a twist by some angle not an integer multiple of 360°. Indeed, twist angles exist that are arbitrarily close to any desired number ω.*

Proof: First, we argue that a geodesic emanating from almost any starting point P_0 and starting direction $\vec{d}_0$ must return arbitrarily close to itself at some point P and while simultaneously having orientation matching the geodesic's direction $\vec{d}$ at P arbitrarily closely, and simultaneously having twist matching any desired number $\omega \in [0, 2\pi)$ arbitrarily closely. Indeed, no matter how small we choose $\epsilon, \phi, \alpha > 0$, by a volume argument any geodesic after large travel length L must return to within small distance ϵ of P and simultaneously to within a small angular distance θ of $\vec{d}$ if $\epsilon^{n-1}\theta^{n-1}L = \text{const}$; and simultaneously (if $n = 3$) to within α of having twist-angle ω if $\epsilon^2\theta^2\alpha L = \text{const}$.

Now we may, by taking advantage of the exponentially enormous (yet isotropic) sensitivity of the final position of a long geodesic segment to its initial direction and position, slightly perturb that initial direction (and the segment's length L) in order to make our geodesic *exactly* return to P, although the small bend angle θ will still, in general, be nonzero.

We now argue that, upon "shortening" this closed curve to make it become a true closed geodesic, ω does nto change. This is an immediate consequence of theorem 13. In a space of *constant* negative curvature, there is an alternative way to make this argument by using hyperbolic trigonometry [45]. This second way is weaker since it only works if the curvature is constant and it only shows ω changes by at most a small amount, rather than zero. Nevertheless we explain it to give the reader additional insight and confidence.

An isoceles triangle with large legs L and apex angle $180 - \phi$ degrees may be squashed flat without moving anything by more than $\approx \phi$. Consult figure 7.1.

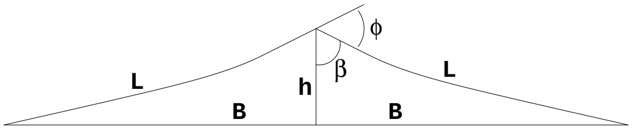

Figure 7.1. A hyperbolic space scenario to be analysed using hyperbolic trigonometry. Regard L as large and ϕ as small.

By the hyperbolic law of sines

$$(\sin\beta)(\sinh L) = \sinh B \tag{42}$$

while by a Napier-Engel rule

$$(\cos\beta)(\tanh L) = \tanh h. \tag{43}$$

Now consider L to be large and ϕ to be small, where $\beta = (\pi - \phi)/2$. Then $\sin\beta \approx 1 - \phi^2/8$ and $\cos\beta \sim \phi/2$ and $\tanh L \to 1$, causing $h \sim \phi/2$ and

$$L - B = L - \operatorname{arcsinh}([\sinh L][\sin\beta]) = \tanh(L)\frac{\phi^2}{8} + O(\phi^4) \sim \frac{\phi^2}{8} \tag{44}$$

[43] The only still-open question is whether the time-1 frame flow need be Bernoullian when $n \in \{7, 8\}$. Also, almost nothing is known about noncompact, but nevertheless finite-volume, 3-manifolds of κ-nearly constant negative curvature. Most experts conjecture that all the compact results mentioned should extend to this finite-volume case, but nobody has been able to prove it despite efforts dating back to the 1960s. If only certain kinds of cusps are permitted, *then* the proofs should extend. Thermodynamics indicates the impossibility of having an "absorbing cusp" that a positive-measure fraction of geodesics flow into, since that would convert a low temperature photon gas to high temperature. Despite the major lack of knowledge about ergodicity in the noncompact finite-volume case, it is still possible [2] to obtain results about the existence of twisted closed geodesics; see the extension of theorem 17 below. Incidentally, the difficulty of this problem is shown by constructions [30][26] of compact smooth boundaryless negatively curved Kahler manifolds, with sectional curvatures -1 and -0.25 only, of even real dimension ≥ 4 (gotten by quotienting complex-hyperbolic spaces, quaternion-hyperbolic spaces, or the octonion projective plane by cocompact lattices) having *non*ergodic frame flows.

both to be small.

This indicates that, if L is large enough and ϕ and ϵ are small enough, we may now shorten our curve to get a genuine closed geodesic with *no* bend angle ϕ, and with nothing moving far enough to destroy validity (for that, something would have to move a distance of order 1).

Finally, due to the small motion and the constant curvature, the twist incurred by our geodesic will necessarily only be altered by $\pm O(\phi)$ and hence will remain nonzero and close to ω. (In fact this twist angle will not change at all.) Q.E.D.

Note that both versions of this argument relied on negative sectional curvatures. The shortening sub-argument based on EQ 41 actually would still work in flat manifolds, but in Euclidean or spherical spaces we would not have had ergodicity[44].

Theorem 17 (The C.C.C.O. 3-universes without twisting closed geodesics). *The only constant curvature compact orientable 3-manifolds in which parallel translation along 1 cycle of any closed geodesic never incurs twist other than an integer multiple of 2π radians, are: T^3 and S^3.*

Proof: The previous theorem showed there are no compact hyperbolic manifolds satisfying the NTA. Now go through the classification of positively curved and flat 3-manifolds of constant curvature in Wolf's book [174]. You will find that the only examples satisfying the no-twist property for all closed geodesics are T^3_∞ and S^3. The easiest way to see this is to consider the abstract group, defined by generators and relations given by Wolf [174] defining each manifold; the key observation, in each case, is the existence of some generator A such that $A^k = 1$ for some minimal integer $k > 1$. The point is that $k \neq 1$, and hence that there is a twist of $2\pi/k$ radians (or an orientation reversal) associated with A.

Alternate argument: (Much of the above proof may be replaced with the following.) Montesinos ([117] ch. 3; see also ch. 4.5-4.6) also discusses many (but not all) of these manifolds from a more intuitive point of view, in a way that makes it very clear they have twist. A different, and full, discussion of the spherical 3-manifolds is in sections 3 and 4 of [63], and it also makes it clear that they all (except S^3) have twist. Conway recently redid the classification (using quaternions) in chapter 4 of his book [42]; the history of previous classifications – which had included some omissions and overcountings! – is described in their §4.5. Q.E.D.

Extension: In the preceding proof, we may replace the role of theorem 17 by a result of Adams et al. [2] that every orientable hyperbolic 3-manifold M either

1. Has a simple twisted closed geodesic, or
2. Arises from a "Fuchsian group" of an orientable hyperbolic 2-manifold. In this case, the volume of M is necessarily infinite, and they still construct a simple closed geodesic, albeit an untwisted one, except in the following cases where M has no simple closed geodesic: when the Fuchsian group is either $\mathbb{Z}$, $\mathbb{Z}\times\mathbb{Z}$, or arises from the hyperbolic thrice-punctured 2-sphere S^2_{-3}.

In other words, theorem 17's assumption that M is *compact* may be replaced by the assumption that it is finite volume and hyperbolic. Also, without assuming M is in any way finite, nor even that it has a finitely generated fundamental group, theorem 17 still holds, albeit at the cost of adjoining T^3_∞, H^3 and the $\mathbb{Z}$, $\mathbb{Z}\times\mathbb{Z}$, and H^3/S^2_{-3} cuspmanifolds to the list of permitted 3-manifolds.[45]

If we now replace theorem 17's assumption of constant curvature with the weaker assumption of local homogeneity, then more manifolds become permissible:

Theorem 18 (The C.O. 3-geometries with twist-free closed geodesics). *Any compact orientable 3-geometry*[46] *in which parallel translation along 1 cycle of any closed geodesic never incurs twist other than an integer multiple of 2π radians, must have one of the following topological types:*

1. T^3,
2. S^3,
3. $S^2 \times S^1$,
4. *$M^2 \times S^1$ where M^2 is a compact hyperbolic 2-manifold (g-holed torus with $g > 1$),*
5. *a torus (T^2) bundle over S^1 where the torus at one end of the S^1 is regarded as the unit square $0 \le x, y \le 1$ with periodic boundary conditions, while the other end that square is regarded as having been distorted into a rhombus (still with periodic boundary conditions) via the linear map with matrix $\left(\begin{smallmatrix} p & r \\ r & q \end{smallmatrix}\right)$ which is mapped into the unit square by taking the coordinates modulo 1, where p, q, r are integers with $|p+q| > 2$ and $pq = r^2+1$. This last metric can be given a Sol geometric structure (see theorem 9).*

Proof: All compact orientable locally homogeneous 3-manifolds (except for the compact hyperbolic manifolds, which were among those handled in a previous theorem) have been completely classified [145][51]. All are either Seifert fiber spaces (i.e. 3-manifolds foliatable into circles S^1) or Sol-manifolds. The underlying geometry is one of the 8 possibilities in theorem 9 in our §6.

Working through the classifications: Page 457 of [145] shows the only compact orientable 3-manifold whose geometry is the same as that of $S^2 \times \mathbb{R}$ and which avoids twist is $S^2 \times S^1$.

[44] But according to M.Berger [17], recent revolutionary work of Joachim Lohkamp has shown that *every* Riemannian n-metric, $n \ge 2$, may be *perturbed* by an arbitrarily small amount in such a way that its geodesic flow becomes ergodic. I.e. presumably "generic" or "bumpy" manifolds have ergodic geodesic flow. If true, this may have some profound implications concerning the validity (or lack thereof) of twist-based arguments in cosmology. But there are 4 caveats: First, this Lohkamp work remains unpublished and incompletely written up. Second, a small perturbation of the metric does not necessarily correspond to a small perturbation of the curvature tensor; indeed Lohkamp requires perturbations of the round sphere severe enough to yield negative curvatures. Unless that deficiency can be overcome, these results will remain irrelevant to physics. Third, it is presently unknown even whether there is *any* smooth positively curved 2-manifold with ergodic geodesic flow. In particular, the geodesic flow on the surface of an ellipsoid is known always to be non-ergodic and indeed features a conserved quantity [93][157]. Indeed, once those references are understood, it becomes easy to see that if an ellipsoid is cut into two hemi-ellipsoids along any one of its 3 major ellipses, which then are glued to the two ends of an appropriate right elliptic cylinder, the resulting convex surface always has non-ergodic geodesic flow, no matter what values are chosen for its 4 defining parameters. Fourth, I do not know if Lohkamp's ideas apply to the frame flow.

[45] Adams et al. [2] also note that S^2_{-3} is the only orientable finite-area hyperbolic 2-manifold with no simple closed geodesic. It is noncompact.

[46] Equivalently, by theorem 11, the word "3-geometry" could be replaced by "1-curvature homogeneous 3-manifold."

Theorem 5.3i of [145] *seems* to show that every compact Sol-geometry has a twisted fiber. If so, that would rule them all out since that fiber could be "shrunk" to get a twisted closed geodesic. However, that impression was incorrect. In fact Dunbar [51] (and the same theorem is mentioned in cite-Bonahon despite the fact that he misleadingly refers to Sol as a "twisted geometry") finds that the compact orientable Sol-geometries are precisely the T^2 bundles over S^1 with gluing map $\binom{p\ r}{s\ q}$ where p, q, r, s are integers with $|p+q| > 2$ and $pq - rs = 1$. This map avoids twist (i.e. is a pure area-preserving *dilation*) if and only if it is a symmetric matrix, i.e. $r = s$. For example: $\binom{2\ 3}{3\ 5}$ and $\binom{-17\ \ 4}{4\ \ -1}$.

Theorems in [145] (e.g. theorem 4.16 handling Nil, where note the proof's remark about how "one cannot have $n = 0$," and theorem 5.3ii handling $\widetilde{\mathrm{SL}_2\mathbb{R}}$) show that $\widetilde{\mathrm{SL}_2\mathbb{R}}$ and Nil (the "twisted geometries"; this time this moniker is justified!) always yield twisted bundles over S^1 and hence also have twisted closed geodesics and hence are forbidden. The reader may be worried at this point due to having acquired a certain skepticism about [145]; if so it is reassuring that there is an

Alternate argument: (Much of the above proof may be replaced with the following.) Instead of relying on classifications of compact orientable 3-geometries, it is also possible to proceed by using Seifert's original (earlier) classification [146] of the orientable Seifert-fiber spaces. As we remarked before [145][24], the latter classification is included in the former since every compact orientable 3-geometry is a Seifert fiber space – except perhaps for Sol-manifolds and compact hyperbolic manifolds, both of which we have already convincingly handled. Seifert showed his spaces are in 1-to-1 correspondence with certain tuples

$$(O, o; P|b; \alpha_1, \beta_1; \alpha_2, \beta_2; \cdots ; ; \alpha_r, \beta_r). \tag{45}$$

(The reader will need to read [146] to understand EQ 45's notation. Seifert would also permit other tuples containing the letters "N" or "n" instead of "O" and "o," but these are non-orientable or contain a non-orientable 2-surface and hence a twisted geodesic, either way being forbidden. Seifert manifolds are also discussed in [24] and [126] and from a different point of view in [117] ch. 4.5 and 4.6.) Here the integer $P \geq 0$ is the genus of the 2-dimensional orbifold of the fibers. If there are no "exceptional fibers" (all of which would lead, upon shrinking, to twisted geodesics) then $r = 0$ so that there are no α and β terms. Finally only $b = 0$ is permitted if there is no twist. Consequently we conclude that the *only* compact orientable Seifert fiber spaces obeying the no-twist postulate are $M^2 \times S^1$ where M^2 is an orientable 2-manifold, which we may (without loss of generality if we are only interested in topological equivalence) take to be of constant curvature of the same sign as $1 - P$.

It is important to note, throughout the above argument, that Seifert fibers always are uncontractible loops (this follows from the structure [56][52] of Seifert spaces as circle bundles over a 2D orbifold), so that shrinking them always yields a closed geodesic. Q.E.D.

For conciseness, let us refer to the Sol manifolds in theorem 18(5) as the "symmetric $T^2|S^1$ bundles."

Summary: Either version of the No-Twist Assumption for closed geodesics implies that our universe must, if it is a compact orientable 3-manifold of constant curvature, be either T^3 or S^3. If we weaken the assumption of constant curvature to 1-curvature homogeneity, then the only additional permitted topologies are $M^2 \times S^1$, where M^2 is any compact orientable 2-manifold of constant curvature, e.g. either M^2 is S^2 or a g-holed torus with $g > 1$ and constant negative curvature, and the symmetric $T^2|S^1$ bundles.

It is presently physically unclear how valid it is to argue that the universe must be twist-free, and if it is valid, it remains unclear which among numerous possible inequivalent phrasings of the No-Twist Assumption should be favored.

8 Implications of charge quantization

The previous paper [151] introduced the idea that charge quantization could be explained by postulating the universe contains a "topologically trapped" primordial magnetic field. Essentially, the idea was that an electron traveling around certain kinds of closed trajectories would "enclose" all of the trapped magnetic flux. If the electron's charge were not an integer multiple of a certain specific charge (corresponding to the requirement that the trapped flux be an integer number of "flux quanta") then slight perturbations of the trajectory would lead to discontinuous physical effects. Assuming such discontinuities are unacceptable logically forces charge quantization.

As was described in [151] and sketched in the present paper's introduction, a universe with two or more kinds of mutually non-isotopic nonseparating surfaces, would, if it initially contained a generic magnetic field, lead to two different, and generically incommensurable, charge quanta, thus ruling out all charges and contradicting reality. This suggests that the universe contains at most *one* isotopy class of nonseparating surface. Furthermore, if the value of the charge quantum is *not* explained by [151]'s cosnmic mechanism but instead purely by microscopic physics, then a universe containing even *one* isotopy class of nonseparating surfaces, would, if it initially contained a generic magnetic field, lead to a different and incommensurable charge quantum, again ruling out charge and contradicting reality. In that case the universe presumably could not contain a nonseparating surface at all.

It would be very good for some future author to go though the TGC's alleged classification of all compact orientable 3-manifold topologies and determine exactly which ones satisfy these conditions. We shall content ourselves by doing this for the (non-conjectural!) classification of the compact orientable 1-curvature homogeneous 3-metrics in theorem 11.

Theorem 19 (1-curvature homogeneous manifolds generically yielding at most one charge quantum). *Let M be a compact 1-curvature homogeneous orientable 3-manifold containing at most 1 isotopy class of nonseparating complete surfaces. Then M is either $S^2 \times S^1$ or has an S^3 geometric structure.*

Proof: We consider the 8 cases in theorem 9 (and 11). If M has $S^2 \times E^1$ geometric structure, then ([24] p.457-458) it is one of 7 possibilities, of which only $S^2 \times S^1$ and $\mathbb{RP}^3 \# \mathbb{RP}^3$ are orientable. The former has only 1 kind of nonseparating surface, but the latter has two (one being an "equator" on each $\mathbb{RP}^3$). It could also be permissible for M to have S^3 geometric structure, since S^3 itself has no nonseparating surface, nor does the Poincaré dodecahedral space (since it is a homology sphere). We have not tried to decide exactly which S^3-based manifold topologies are admissible. Not all of them are, since the "lens spaces" have isotopically non-unique Seifert fibrations [126] and hence non-unique isotopy classes of (transverse) nonseparating surfaces, and hence are forbidden.

If M has $\widetilde{\mathrm{SL}_2\mathbb{R}}$ or Nil structure, then it is a Seifert fiber space ([24], in particular see theorems 4.15ii and 4.16ii for the compact $\widetilde{\mathrm{SL}_2\mathbb{R}}$ and Nil cases, respectively) over a 2D orbifold X. In the case of Nil, X has Euler characteristic $\chi = 0$ (so X is flat) while for $\widetilde{\mathrm{SL}_2\mathbb{R}}$, X has Euler characteristic $\chi < 0$ (so X is multihandled) and hence either way X has a nonseparating curve so that M has at least two nonseparating surfaces. Essentially the same argument would rule out $M^2 \times E^1$, but these also may be ruled out in a more elementary manner: M^2 has at least two handles and hence at least 2 nonseparating curves. We may rule out M being a flat compact orientable manifold (E^3-based) by a case by case examination of the classification [174]. But this also may be ruled out effortlessly by using the fact that M must be a Seifert fiber space over a 2-orbifold X with the fibers in an E^1 direction in the $E^2 \times E^1$ covering space. This immediately leads to more than one kind of nonseparating X. If M is compact and has Sol structure then ([24] theorem 4.17) it has the structure of a 2D bundle over a 1D orbifold S^1, where every possibility for a 2-foliatation leaf necessarily has at least 1 kind of nonseparating curve. That rules out Sol.

If M is hyperbolic (H^3-based) or Euclidean (E^3-based) and compact, then it necessarily has many kinds of uncontractible min-length closed geodesics (cf. theorem 16). For each such closed geodesic G, we may construct a tranverse nonseparating surface as follows. Start with the totally geodesic (but self-intersecting and usually infinite) "plane" surface consisting of all geodesics making a right (or other fixed nonzero) angle to g at a particular point $P \in g$. Color an expanding disc in S centered at P red; whenever a point on S is hit by the expanding red wavefront coming from two different directions simultaneously the wave "freezes" there. Eventually the entire wavefront has frozen and the red portion of S is a polyhedral surface (which may be regarded as a C^0 boundaryless compact 2-manifold) transverse to g at P. This surface is topologically a 2-disc D except that each point $x \in \partial D$ on the boundary of that disc is everywhere glued to itself according to some self-inverse map $x \to f(x) \in \partial D$. If desired, we may now shrink this surface's area to minimum possible to get a smooth, necessarily *not* self-intersecting, isotopic surface necessarily crossing g an odd number of times and hence necessarily nonseparating. (It can be easier to envision the 1-less-dimensionful analogue of this construction on a compact orientable hyperbolic 2-manifold, i.e. a multihandled torus. Such a torus has 4 kinds of uncontractible simple closed curves which come in 2 pairs of mutually transverse kinds, yielding at least 4 kinds of nonseparating curve.) Next, we do the same thing but with a different geodesic g', and different transverse surface S', but now picking S' to *contain* g. We then adopt a similar "red sea" expansion strategy, but this time starting the expansion from the entirety of g and from the shortest segment joining $P' \in g' \cap S'$ to g. The final polyhedral surface will be topologically an annulus A except that again each point $x \in \partial A$ is glued to some other via a self-inverse map. We can also do the analogous construction with the roles of g and g' exchanged. In this way we can obtain two nonseparating surfaces which are manifestly non-isotopic (since they each contain closed geodesics of different isotopy types). Q.E.D.

9 Some exact Einstein-Maxwell solutions

One of the deficiencies of the previous [151] as well as the present paper, is our policy (for the most part) of regarding the universe as a Riemannian 3-manifold rather than a Lorentzian $(3+1)$-manifold in which time is inextricably entangled with space via Einstein's equations of general relativity.

There were two reasons for these decisions to ignore time. First, because at present the foundations of general relativity are not mathematically rigorous (e.g., no general solution existence theorem is known) we had no real choice if we wished to obtain mathematical rigor. Second, a far larger amount is known about the topology of 3-manifolds than about $(3+1)$-manifolds.

But it is not necessary to give up the fight completely. As [151] pointed out, it seems highly plausible that the charge quantization argument there still works even in the presence of general relativity because (1) it is a topological invariant, (2) arguments were made that in a slowly-changing model universe, charge quantization would still work, (3) nonrigorous arguments were made that the presence of small black holes should not affect validity. It was also argued that much of what was said in [150] should still hold in $(3+1)$D because we were there only dealing with homogeneous isotropic universe models.

This section now makes the case completely solid that the charge-quantization mechanism of [151] can work, exactly, under the full $(3+1)$-dimensional Einstein-Maxwell equations. To do so, we shall present several exact solutions of those equations representing suitable model universes with a topologically trapped sourceless magnetic field.

9.1 Some simple electrovac universes

The *uniform-field electrovacuums* (metrics that are exact solutions of the Einstein-Maxwell equations of classical gravity and electromagnetism with cosmical constant Λ) are precisely the following:

$$\mathrm{d}s^2 = \frac{-\mathrm{d}t^2 + \mathrm{d}x^2}{(1 + Jx^2 - Jt^2)^2} + \frac{\mathrm{d}y^2 + \mathrm{d}z^2}{(1 + Ky^2 + Kz^2)^2}. \qquad (46)$$

where J, K are constants such that $\Lambda = 2(K + J)$. These $(3+1)$D metrics may also be regarded as the Cartesian product of 2D and a $(1+1)$D metrics of constant curvature.

Proof: The fact that these are exact solutions corresponding to a uniform magnetic field in the x-direction may be verified by direct substitution. (The mixed Einstein tensor G_α^β is diagonal with entries $G_t^t = G_x^x = K$, $G_y^y = G_z^z = J$.) For the fact that these constitute *all* $(3+1)$D electrovac metrics with constant non-null electromagnetic field tensor (and indeed contain all $(3+1)$-metrics supporting *any* constant non-null 2-form) see [49]. There is also a general theorem of G.S.Hall [72] stating that such a direct product is the *only* way that a $(3+1)$D Lorentzian manifold can have *any* constant symmetric ${}^{0}_{2}$tensor. Q.E.D.

There are **9 = 3 × 3 different cases** of EQ 46, namely those where the 2-metric and the $(1+1)$-metric have constant curvatures $+1$, -1, or 0. We divide these into 3 families; in each family we may choose the constant curvature Riemannian 2-metric, arising from varying y and z while holding x and t fixed, to be either S^2, or any compact hyperbolic metric (e.g. the g-holed torus got by suitable identifications of the sides of a suitably-sized regular $4g$-gon in the hyperbolic plane, where g is any integer with $g \geq 2$), or T^2.

If $J=0$ then this universe is time-independent (i.e., the gravitational effect of the constant magnetic field is exactly compensated by the cosmical constant) and we may choose the x-coordinate to "wrap around" ala S^1. Then the generic result is a homogeneous but anisotropic unchanging spatially-compact universe filled with a uniform magnetic field, in which charge is necessarily quantized. (One exception: in the $\Lambda = K = 0$ flat isotropic case, there is no charge quantization since the magnetic field is zero.)

If $J=1$ then we have the cartesian product of our usual Riemannian 2-space of constant curvature (got by varying y and z while holding t, x fixed) with the "de Sitter line." The latter is a Lorentzian $(1+1)$-space of constant curvature which is isometric to the 2-surface $a^2 + b^2 = 1 + w^2$ in Lorentzian flat 3-space of with infinitesimal metrical line element $\mathrm{d}s^2 = \mathrm{d}a^2 + \mathrm{d}b^2 - \mathrm{d}w^2$. Observe that this surface is spatially compact, i.e., the constant-w curves are circles. So the net result is a homogeneous but anisotropic spatially-compact universe which initially contracts from infinite to some minimum size during the time interval $-\infty < t < 0$, then reexpands during $0 < t < \infty$ to infinite size; it is filled with a spatially uniform (but time-varying) magnetic field. Note: the x-direction in this universe (parallel to the magnetic field, and which is topologically circular) is the one that is exponentially expanding and contracting; in the y and z directions the metric is time-independent. Hence the magnetic field and flux both remain constant. (The energy to supply the magnetic field in the expanded volume in the later universe is supplied by the cosmical constant Λ.) Again, charge in this universe is necessarily quantized, as one may argue by considering (bounded distance) "trips around the universe" taken by an electron moving at sublight speeds[47] with $x =$ constant; the argument is similar to the one in [151] and note that an unboundedly large number of such trips are possible during the infinite life of this universe. Also note that it is possible to choose $\Lambda = 0$ in this kind of model universe.

In all of the above cases the *spatial* topology of our model universe has been $S^1 \times M^2$ where M^2 is a compact Riemannian 2-manifold of constant curvature.

The final family of uniform-field metrics arises if $J = -1$; the Lorentzian $(1+1)$-metric is the "anti de Sitter line," which has periodic time but is noncompact in x. This is just like our previous description of the de Sitter plane, as a 2-surface $a^2 + b^2 = 1 + w^2$ in Lorentzian flat 3-space, except that its 3-metric is now $\mathrm{d}s^2 = \mathrm{d}w^2 - \mathrm{d}a^2 - \mathrm{d}b^2$ so that the angular coordinate in ab planes is now timelike and the w-coordinate spacelike. It is possible to make the time-coordinate noncompact and not periodic by "winding around the circle an infinite number of times" (i.e. using the covering space of the anti de Sitter line), but apparently there is no way to compactify the spatial direction while retaining homogeneity. If so, then there is no way to regard this final family of metrics as homogeneous spatially-compact universes, although we may still regard the magnetic field as being trapped – it is just that that trapping is no longer caused by closed field lines wrapping around an S^1, but rather by them extending to $\pm\infty$ in the x-direction. Charge quantization then would still work, even though this universe is spatially noncompact.

To provide some contrast, let us discuss one more noncompact vacuum-filled model universe in which the charge quantization arguments of [151] do *not* quite work.

Melvin's magnetic universe [106][107][25][161]

$$\mathrm{d}s^2 = (1 + Kr^2)^2(-\mathrm{d}t^2 + \mathrm{d}z^2 + \mathrm{d}r^2) + \frac{r^2\,\mathrm{d}\theta^2}{(1+Kr^2)^2} \qquad (47)$$

features a nonuniform magnetic field of strength proportional to $\sqrt{K}/(1+Kr^2)$, in the z-direction. (This field is source-free and holds together via its own gravity.) We now may (optionally) regard the z-coordinate as being circularly "wrapped," i.e. periodic with any desired period length. This universe is *non*compact (infinite volume) and time-invariant, but if it has finite period length in the z-direction then the magnetic field is topologically trapped. Because the magnetic flux through this universe is *infinite*, the charge quantum would actually be *zero* and hence the arguments of [151] presumably still would allow continuously variable charge.

9.2 Toward greater realism

None of the above model universes are intended to be realistic pictures of our own universe. Our own universe contains matter and radiation at densities far huger than the energy density of any trapped magnetic field; but those model universes contained *vacuum* and magnetic field *only*, and hence were highly anisotropic. They were merely presented to make the case completely solid that the charge quantization mechanism of [151] can work even if we demand exact solution of the full Einstein-Maxwell equations in a curved Lorentzian $(3+1)$-space, whether time-varying or not, and having $\Lambda = 0$ or not.

We now seek more realism. We enquire whether there are exact Einstein-Maxwell solution model universes containing both a uniform magnetic field, *and* some sort of homogeneous isotropic matter (and perhaps a cosmical constant Λ), in arbitrarily variable relative proportions. The answer is yes [86][122].

[47] For example, if we are using as our Riemannian 2-metric, the 2-holed torus got by appropriately gluing the sides of a suitably-sized hyperbolic-plane regular octagon O, then we could consider the electron moving on a regular-octagon trajectory just like O but a tiny bit smaller. By considerations of continuity as in [151] we would conclude that the magnetic flux enclosed by this trajectory had to be an integer number of flux quanta, so that charge had to be quantized.

The impossibility of a *static* such universe is shown by Hall's theorem [72] (since it would necessarily have constant Ricci and Einstein symmetric ${}^0_2$tensors) and indeed the impossibility of one merely containing a *time-invariant* magnetic field is shown by [49]. However, these theorems still permit the possibility of an expanding such model universe containing a uniform magnetic field decreasing with time.[48]

Collins and Hawking [41] claimed that the only Bianchi types [19] of expanding homogenous cosmologies that were capable, in the large time limit, of being asymptotically FRW (i.e. isotropic) were I, V, VII$_o$, VII$_h$, and IX. But to my knowledge nobody has examined magnetism in any Bianchi-type cosmology besides I and V.

Jacobs [86] considered cosmologies of "Bianchi type I," i.e. having a metrical line element of form

$$ds^2 = dt^2 - A(t)^2 dx^2 - B(t)^2 dy^2 - C(t)^2 dz^2, \qquad (48)$$

containing both a uniform magnetic field in the z direction and a perfect fluid obeying the equation of state $P = \gamma\rho$, where P is pressure, ρ is density, and γ is a constant. He found 4 classes of solutions, including 2 classes of axisymmetric solutions (i.e. with $A = B$). One of his axisymmetric universes originates from a point singularity ($A = B \approx C \to 0$ as $t \to 0$), and the other from a "pancake" singularity ($A = B \approx 1 + \alpha t^{1-\gamma}$, $W \approx t$ as $t \to 0$). Note that these metrics can be instantiated in a torus $T^3 \times \mathbb{R}$, i.e. a 3-box with periodic boundary conditions.

Nayak and Bhuyan [122] considered "Bianchi type V," i.e. having a metrical line element of form

$$ds^2 = dt^2 - A(t)^2 dx^2 - e^{-2x}(B(t)^2 dy^2 + C(t)^2 dz^2), \qquad (49)$$

filled with a perfect fluid of unspecified equation of state and a uniform magnetic field. They found that there is exactly one exact solution, which, however, involves two arbitrary real-valued functions of one real variable. This kind of line element seems less interesting because it apparently is uncompactifiable, i.e. cannot represent a spatially compact universe.

Jacobs [86] in 1969 noted that the present-day magnetic field of the universe is $< 10^{-7}$gauss. But it is now known [166] to be $< 3 \times 10^{-15}$Tesla$= 3 \times 10^{-19}$gauss. Consequently, considering the approximately known mass-density of the universe, one may estimate, following Jacobs but updating his numbers, that this magnetic field would have only resulted in an anisotropy in the CMB of order 1 part in 10^{28} and that the initial anisotropy of the universe would have decayed to a small level after 10^{-41} second. In other words, if we live in Jacobs' sort of universe, then its anisotropy would be experimentally undetectable.

In fact, as was discussed in [151], there *does* appear to be some statistically significant anisotropy in the CMB, of order 1 part in 10^5; there is a low-power bidirection pointing roughly towards and away from Virgo. A new possible interpretation of this is that our universe indeed is anisotropic. If so, then our universe cannot be of Jacobs' T^3 kind. However, Jacobs' universe is somewhat atypical because since T^3 is a "naturally" locally isotropic manifold, all his anisotropy arises from the magnetic field – which we know from astronomical observations is (and always was, and always will be) tiny in comparison to matter and radiation densities. In contrast, in a universe of form $S^2 \times S^1$ or $M^2 \times S^1$ (as in EQ 50 below), the metric is "naturally anisotropic," and hence the 10^{-5}-size anisotropy of the CMB would in no way be forbidden.

Both the Jacobs and Nayak-Bhuyan cosmologies may be regarded as permitting cosmical constant Λ because the effect of Λ may be incorporated into the equation of state of the perfect fluid.

We now incorporate perfect fluid into our simple $M^2 \times S^1$ cosmologies, to get more general (and new) exact solutions of form

$$ds^2 = \frac{-dt^2 + A(t)^2 dx^2}{(1 + Jx^2 - Jt^2)^2} + B(t)^2 \frac{dy^2 + dz^2}{(1 + Ky^2 + Kz^2)^2}. \qquad (50)$$

Demand $J = 0$ (because otherwise all expressions become much more complicated). Then computing the mixed Einstein tensor G^β_α, we find that it is diagonal with nonzero terms

$$-B^2 A G^t_t = 2BB'A' + 4KA + B'^2 A, \qquad (51)$$

$$-B^2 G^x_x = 2BB'' + 4K + B'^2 \qquad (52)$$

$$-BAG^y_y = -BAG^z_z = B''A + A''B + A'B'. \qquad (53)$$

Regard x as a periodic coordinate. Then EQ 50 represents a fixed-shape universe of topology $M^2 \times S^1$, where M^2 has constant curvature $K \in \{0, \pm 1\}$. The functions $A(t)$ and $B(t)$ are size parameters.

We can enforce both magnetic flux conservation and fluid isotropy by choosing $A(t)$ and $B(t)$ so that $(G^x_x - G^y_y)B(t)^2$ is time-invariant. $G^y_y = G^z_z$ then is satisfied automatically, so we still have one function worth of freedom left over to use to try to make the fluid's equation of state behave in any desired way. This universe is filled with both an isotropic fluid, and a magnetic field in the x direction.

Summary: We have found exact solutions of the full Einstein-Maxwell $(3 + 1)$D equations representing universes with trapped magnetic fields in which charge quantization would be forced; in the $S^1 \times S^2$ model generically uniquely. All of these model universes are anisotropic.

More realistic magnetic T^3 and $M^2 \times S^1$ universes, also incorporating perfect fluid, were respectively found by Jacobs [86] and by ourselves.

[48] Incidentally, an instructive "paradox" is the following. Any $M^2 \times S^1$ universe with sourcless magnetic field along the S^1s *cannot* have expanding M^2s because that plus flux conservation would imply decreasing magnetic field, leading any observer equipped with a fixed-size wire loop perpendicular to the S^1 fibers, to detect an induced voltage around that loop. Thus there would have to be an induced electric field in the M^2, which by homogeneity would necessarily be uniform – but since it is impossible to "comb the hair" on a surface of genus $g \neq 1$ no such field could exist. "Q.E.D." The resolution of this paradox is that there really is no induced electric field, and if our observer's "wire loop" were in fact made of particles of the (rareifying) perfect fluid it would be expanding just quickly enough to pass constant flux, and hence would have no voltage induced in it. (Voltage *would* arise upon artificially shrinking the loop back down to fixed size but that is irrelevant – electric field is defined by the acceleration of charged gas particles.) Hence the M^2s really can expand contrary to the above "proof."

10 Conclusions

Regard the universe U as a smooth connected boundaryless Riemannian n-manifold. This paper has introduced and analysed three new possible assumptions about the universe, listed in the abstract. A fourth assumption – that the Earth both lies on a closed geodesic (suggested by one possible interpretation of an anomaly in WMAP satellite CMB data) and is not special – was considered in [150].

Which of these assumptions are correct? We do not know, but with more debate and analysis the community should acquire more confidence. We do not wish to leave the reader with the idea that they all are equally likely. My present conjectural beliefs are summarized quantitatively in table 10.1.

A	assumption A	P_A
1	orientable	99%
2	compact	90%
3	constant curvature	20%
4	1-curvature homogeneous	70%
5	0-curvature homogeneous	80%
6	no twisted closed geodesics (NTA)	70%
7	nonzero countable # of closed geodesics through generic point	10%
8	≤ 1 isotopy class nonseparating surfaces	95%
9	global OCSS	5%
10	local OCSS	100%

Figure 10.1. Place your bets. Important possible assumptions A about our universe regarded as connected boundaryless 3-manifold, and my present personal assessment of probability P_A that "the universe's topology is compatible with A being true." Some assumptions subsume others, for example $P_3 \leq P_4 \leq P_5$ since the topologies compatible with assumption 3 are a strict subset of those compatible with assumption 4, in turn subsumed by 5.

We now provide short justifications for these numbers. The fact that the universe seems charge neutral by itself only gives me perhaps 70% confidence the universe is compact, since

1. the universe still might be slightly charged and
2. despite the fact that both approximate and exact neutrality are naively surprising a priori, the former might be less surprising to a more sophisticated viewer.

But the new compactness argument in §3 then reduces my dubiousness by a factor of 3, corresponding to a 70→90% confidence increase.

Why the high confidence the universe has a topology compatible with curvature homogeneity? Postulate that whatever process created the universe would not have favored any particular part of it over any other, so that we would expect some sort of homogeneity. Now in presently accepted physical theories (general relativity) only the Riemann curvature tensor[49] (and not its derivatives) enter, suggesting that the most we should expect is 0-curvature homogeneity. It seems 90% likely that quantizing gravity will also bring in curvature first derivatives, in which case we should expect 1-curvature homogeneity. But requiring *constant* curvature seems a lot more restrictive and therefore less likely a priori.

I am optimistic that the "no twist" postulate, or something very much like it, is correct, despite theorem 14. In particular, the previous paragraph's picture of an initial curvature-homogeneous metric suggests that something like §7.1's wording of the NTA may be the right one, although relying solely on that argument would seem to force $P_6 < P_5$.

The basis for my high confidence that there is *at most one* isotopy class of nonseparating surface in the universe, is that [151] showed that without this assumption and in the presence of an initially generic magnetic field, charges would be forbidden. (I less vigorously support the stronger postulate that there is *exactly* one; this would arise if the cosmic-topological explanation [151] for charge quantization is accepted. On the other hand if charge quantization has some entirely microscopic explanation, then there presumably could be *no* nonseparating surfaces.)

Assumption 7 was prompted by an anomaly in CMB satellite data [158] but even the discoverers of that anomaly are now disfavoring that interpretation of their data [125] which is why I assigned it only 10% probability.

I suspect that postulating the global existence of ≥ 2 commuting vector fields (which in combination with compactness and orientability, would force a torus bundle over S^1, and an additional commuting field would force T^3) would be too strong; that property of flat space quantum mechanics probably will not carry over to curved space because position operators will have no global physical meaning. (See §4-5.) Nevertheless, position operators probably still will have local importance, so that it may still be legitimate to demand the *local* existence of n mutually commuting and orthogonal vector fields.

The most important winnowing effects:

The NTA in combination with the assumptions of $n = 3$, orientability, compactness, and 1-curvature homegeneity would reduce the candidate topologies to: T^3, S^3, $S^2 \times S^1$, $M^2 \times S^1$ with M^2 orientable and hyperbolic, and the symmetric $T^2|S^1$ bundles defined at the end of §7.

Assumption 8 very powerfully winnows topology candidates. For example, the only 1-curvature homogeneous compact orientable candidate 3-metrics meeting the "≤ 1 nonseparating surface" condition are $S^2 \times S^1$ and some undetermined subset of those with S^3 geometric structure. Thus of the six[50] no-twist candidates just mentioned, only two would remain: $S^2 \times S^1$ and S^3, with the latter forbidden by the "exactly 1" version of our postulate.

Assumption 10 yields no candidate-reducing effect whatever (because *every* 3-manifold has an OCSS locally by theorem 4 – hence the "100% confidence" figure for P_{10}!) except for forcing $n \leq 3$.

We now summarize the known assumptions and implications as a chart.

17 possible assumptions:

3: U is 3-dimensional

A: Every point of U lies on at least one, and at most a countable infinity of, closed geodesics.

[49] Actually, the Ricci tensor, but in 3D the Riemann and Ricci curvatures determine each other.

[50] Actually, infinity, since the last two among these 6 types each are infinite families.

C: U has an uncontractible loop and hence has at least one closed geodesic.
B: Universe is compact
E: Electrons exist
F: Universe is spatially flat
G: U has a generic metric with its topology and contains a generic magnetic field
K: U has a constant-curvature metric ([150] argues experimentally observed sky-uniformity implies metric is harmonic, which for 3-metrics is the same as constant-curvature)
H: U is 1-curvature homogeneous, i.e. any two points have the same Riemann curvature tensor and first covariant derivative of that tensor (if appropriate coordinate systems are chosen)
$\mathbf{I}_k$: There are k isotopy classes of nonseparating surface in U
$\mathbf{M_k}$: Milnor-rank k, i.e. k smooth everywhere pairwise commuting linearly independent vector fields exist on U
O: U is orientable (seems to follow from unidirectionality of time, chirality of weak force)
P: n smooth everywhere-orthonormal vector fields exist on U (i.e. U is parallelizable)
R: U is not a rational homology sphere, i.e. U contains a complete nonseparating surface.
S: Have orthogonal curvilinear coordinate system (S_ℓ=locally, S_g=globally).
T: Assume Thurston's geometrization conjecture.
W: Closed geodesics are twist-free.

Summary of important known implications ($\wedge$=and, $\vee$=or)**:**

1. Sky uniformity $\Longrightarrow$ harmonic manifold [150].
2. Harmonic 3-manifold $\Longrightarrow$ Einstein 3-manif. $\Longrightarrow$ K.
3. Isotropic $\Longrightarrow$ K (F.Schur's theorem [68]).
4. P $\Longrightarrow$ O.
5. B $\wedge$ Maxwell equations hold and have solution $\Longrightarrow$ Universe is charge-neutral.
6. F $\wedge$ solution of Einstein's equations for universe filled with homogenous isotropic matter and radiation with positive density exists $\Longrightarrow$ density is exactly critical.
7. B $\wedge$ O $\wedge$ W $\wedge$ 3 $\wedge$ H $\Longrightarrow U \in \{T^3, S^3, S^2 \times S^1, M^2 \times S^1\}$ $\cup$ the "symmetric $T^2|S^1$ bundles" defined immediately before §7.2.
8. F $\wedge$ O $\wedge$ W $\wedge$ 3 $\Longrightarrow U = T^3_\infty$ [150].
9. G $\wedge$ P $\Longrightarrow$ O $\wedge$ (3 $\vee$ the choice of U is restricted) [151].
10. R $\wedge$ G $\wedge$ 3 $\Longrightarrow$ charge quantization [151].
11. R $\wedge$ G $\wedge$ 3 $\wedge$ E $\Longrightarrow I_{\leq 1}$ (and exactly I_1 if we accept Smith's explanation of charge quantization [151]).
12. $M_{\geq 2} \wedge 3 \wedge$ O $\wedge$ B $\Longrightarrow U$ is Torus-Bundle over S^1 [141].
13. U is Torus-Bundle over $S^1 \wedge$ K $\Longrightarrow$ F $\wedge$ Exactly 3 possibilities for U, namely T^3 and 2 twisted variants.
14. $M_{\geq n} \wedge$ O $\Longrightarrow U = T^n$ [141].
15. W $\wedge$ K $\wedge$ B $\Longrightarrow U \in \{T^3, S^3\}$.
16. A $\wedge$ O $\wedge$ W $\Longrightarrow U = T^3_\infty$.
17. B $\wedge$ U is foliated by planes (homeomorphic to $\mathbb{R}^2$) $\Longrightarrow$ $U = T^3$ [140].
18. $S_\ell \wedge$ generic metric $\Longrightarrow n \leq 3$.
19. $S_g \Longrightarrow T^n$.
20. O $\wedge$ B $\wedge$ H $\wedge$ 3 $\wedge$ $I_{\leq 1} \Longrightarrow U$ is either $S^2 \times S^1$ or has S^3 geometric structure.
21. W $\Longrightarrow U$ is not a finite-volume hyperbolic 3-manifold.
22. U has a C^∞ foliation into $(n-1)$-manifolds $\Longleftrightarrow U$ has Euler characteristic 0 [163].

11 Open problems

1. Although the 0-curvature homogeneous 3-metrics are a vastly infinitely larger class than the 1-curvature homogeneous ones classified in §6, it may be that the class of *topologies* admitting a 0-curvature homogeneous compact boundaryless 3-metric is small enough for complete classification. Do it.

2. Decide which among the compact orientable 3-manifold topologies permitted by the TGC (§6) have ≤ 1 isotopy class of nonseparating surface. (In §8 we did this for the "atomic constituents" permitted by the TGC, i.e., the compact orientable 1-curvature homogeneous 3-metrics. The behavior of isotopy classes of nonseparating surfaces under connected sum seems easy to understand but its behavior under the Jaco-Shalen-Johannson torus decomposition is more mysterious.)

3. Which formulation, if any, of §7's No Twist Assumption, is right?

4. Prove or disprove: every smooth 3-torus (T^3) has a globally valid orthogonal curvilinear coordinate system.

12 Acknowledgements

Dennis Sullivan[51] inspired me to begin this paper, and suggested some of the ideas (considering commuting vector fields, spotting some errors, and suggesting much of §7), and was helpful to me throughout. There is a considerable argument that he should have been a coauthor. Igor Rivin[52] also helped.

References

[1] J.Frank Adams: Vector fields on spheres, Annals of Math 75 (1962) 603-632.

[2] Colin Adams, Joel Hass, Peter Scott: Simple Closed Geodesics in Hyperbolic 3-Manifolds, Bull. London Math. Soc. 31 (1999) 81-86.

[3] S. Akbulut & R. Kirby: Mazur manifolds, Michigan Math. J. 26 (1979) 259-284.

[4] S.Akbulut & J.D.McCarthy: Casson's invariant for oriented homology 3-spheres: an exposition, Mathematical Notes 36, Princeton University Press 1990.

[5] Michael T. Anderson: Scalar curvature and Geometrization Conjectures for 3-manifolds, Comparison Geometry, M.S.R.I. Publ. vol. 30 (1997) 49-82. Available from `http://www.math.sunysb.edu/~anderson/papers.html`.

[6] D.V.Anosov: Geodesic flows on closed Riemannian manifolds of negative curvature, Trudy Mat. Inst. Steklov. 90 (1967) 209 pages. This is in Russian. An English translation is available from the Amer. Math. Soc. as Proc. Steklov Institute Math. 90 (1969).

[51] Professor with joint appointment at SUNY Stony Brook and City University of New York (graduate center) math depts.
[52] Professor, Temple Univ. Math. Dept., Philadelpia PA.

[7] M.A.Armstrong: Basic topology, Springer 1997.

[8] V.I. Arnold & A. Avez: Ergodic problems of classical mechanics, W.A. Benjamin Inc., New York & Amsterdam 1968.

[9] J.L. Arraut & M. Craizer: Foliations of M^3 defined by $\mathbb{R}^2$-actions, Ann. Inst. Fourier (Grenoble) 45,4 (1995) 1091-1118.

[10] Martine Babillot: On the mixing property for hyperbolic systems, Israel J. Math. 129 (2002) 61-76.

[11] N.L.Balazs & A.Voros: Chaos on the pseudosphere, Physics Reports 143,3 (1986) 109-240.

[12] Werner Ballmann: Lectures on spaces of nonpositive curvature, Birhauser Basel 1995. With appendix by Misha Brin.

[13] W. Ballmann, G. Thorbergsson, W. Ziller: Existence of closed geodesics on possibly curved manifolds, J. Differential Geom. 18 (1983) 221-252.

[14] Victor Bangert: On the existence of closed geodesics on two-spheres, International J. of Math. 4,1 (1993) 1-10.

[15] G.Bouligand, Une Forme Donnee a la Recherche des Systemes Triples Orthogonaux, C.R.Acad.Sci.Paris, 236 (1953) 2462-2463. F.Backes, Sur les Transformations de Ribaucour des Systemes Triples Orthogonaux, Bull.Acad.Royal Belg., 57, 7 (1971) 768-788. M.Böcher: Über die Reihenentwicklungen der Potentialtheorie. Leipzig, Germany: Teubner, 1894.

[16] Alan F. Beardon: The Geometry of Discrete Groups, Springer (GTM #91) 1983.

[17] Marcel Berger: A panoramic view of Riemannian geometry, Springer 2003.

[18] Luigi Bianchi: Opere, vol. 3. Sistemi Tripli Orthogonali. Edizioni Cremonese, Roma 1955.

[19] Luigi Bianchi: On the three-dimensional spaces which admit a continuous group of motions, Memorie di Matematica e di Fisica della Societa Italiana delle Scienze, Serie Terza, Tomo XI (1898) 267-352. English translation and notes (by R.T.Jantzen) General Relativity & Gravit. 33 (2001) 2157-2253. Further exposes of the same topic are 35 (2003) 487-490 and 491-498. Jantzen has a web page on this: `www34.homepage.villanova.edu/robert.jantzen/bianchi/index.htm`.

[20] R.H.Bing: An alternative proof that 3-manifolds can be triangulated, Annals of Math. 69 (1959) 37-65.

[21] Garrett Birkhoff: A note on topological groups, Compositio Math. 3 (1936) 427-430

[22] Maxime Bôcher: Die Reihenentwickelungender Potentialtheorie, Liepzig 1894.

[23] M.Boileau & J-P.Otal: Scindements de Heegaard et groupe des homeotopies des petites variétés de Seifert, Invent. Math. 106 (1991) 85-107.

[24] Francis Bonahon: Geometric Structures on 3-manifolds, pp.93-164 in Handbook of Geometric Topology (R. Daverman, R. Sher eds.), Elsevier 2002.

[25] W.B. Bonnor: Static magnetic fields in General Relativity, Proc. Physical Soc. London A67 (1954) 225-232.

[26] Armand Borel: Compact Clifford-Klein forms of symmetric spaces, Topology 2 (1963) 11-122. Reprinted vol. 2 of Borel's collected papers.

[27] Richard C. Bradley: On the strong mixing and weak Bernoulli conditions, Z. Wahrscheinlichkeitstheor. Verw. Gebiet. 51,1 (1980) 49-54.

[28] M. Brin: Topological transitivity of one class of dynamical systems and flows of frames on manifolds of negative curvature, Functional Anal. Appl. 9 (1975) 8-16. Original Russian: Funktsional. Anal. i Prilozhen. 9 (1975) 9-19.

[29] Misha Brin: The topology of group extensions of Anosov systems, Math. Notes 18 (1976) 858-864. Original Russian: Mat. Zametki 18 (1975) 453-465.

[30] M.Brin & M.L.Gromov: On the ergodicity of frame flows, Invent. Math. 60 (1980) 1-7.

[31] Michael Brin & H.Karcher: Frame flows on manifolds with pinched negative curvature, Compositio Math. 52,3 (1984) 275-297.

[32] Peter Bueken: Three-dimensional Riemannian manifolds with constant principal Ricci curvatures $\rho_1 = \rho_2 \neq \rho_3$, J.Math'l Physics 37,8 (1996) 4062-4075.

[33] Keith Burns & Mark Pollicott: Stable ergodicity and frame flows, Geometriae Dedicata 98,1 (2003) 189-210.

[34] K.Burns, C.Pugh, A.Wilkinson: Stable ergodicity and Anosov flows, Topology 39 (2000) 149-159.

[35] Riemannian geometry in an orthogonal frame: from original lecture notes by S.P. Finikov of lectures by Elie Cartan at the Sorbonne in 1926-1927. Translated from Russian 1960 edition by Vladislav V. Goldberg; foreword by S. S. Chern; World Scientific, River Edge, NJ 2001.

[36] Élie Cartan: Leçons sur la Geometrie des Espaces de Riemann, Gauthier-Villars Paris. 1st ed. was 1928, second was 1946 reprinted 1951, 1963, 1988. A translation of the 2nd ed. into English by J.Glazebrook, with added notes by R.Hermann, was published as "Geometry of Riemannian spaces" by Math.Sci.Press Brookline MA 1983 (ISBN 0-915692-34-1).

[37] S.Chandrasekhar: Mathematical theory of black holes, Oxford Univ. Press 1983.

[38] G.Chatelet, H.Rosenberg, D.A.Weil: Manifolds which admit $\mathbb{R}^n$ actions, Inst. Hautes Études Sci. Publ. Math. 43 (1974) 245-260; Classification of the topological types of $\mathbb{R}^2$-actions on closed orientable 3-manifolds, Inst. Hautes Études Sci. Publ. Math. 43 (1974) 261-272.

[39] Maurice Cohen: Foliations of 3-manifolds, Amer. Math. Monthly 81 (1974) 462-473.

[40] A.G. Cohen, A. De Rujula, S.L. Glashow: A Matter-Antimatter Universe?, Astrophys.J. 495 (1998) 539-549.

[41] G.W.Collins & S.W.Hawking: Why is the universe isotropic?, Astrophys. J. 180,2 (1973 part 1) 317-334

[42] J.H.Conway & Derek Smith: On quaternions and octonions, A.K.Peters, Jan. 2003.

[43] N.J. Cornish, D.N. Spergel, G.D. Starkman, E. Komatsu: Constraining the Topology of the Universe, Phys.Rev.Lett. 92 (2004) 201302.

[44] Émile Cotton: Sur les variétés à trois dimensions, Ann. Fac. Sci. Toulouse (ser. 2) 1 (Paris 1899) 385-. Based on his dissertation of the same title at Universite de Paris in 1899.

[45] H.S.M.Coxeter: Non-Euclidean geometry, Dover reprint.

[46] H.S.M.Coxeter: Regular polytopes, Dover Publications 3rd reprinted edition 1973.

[47] Jacek Cygan: Wiener's test for Brownian motion on the Heisenberg group, Colloquium Math. 39,2 (1978) 367-373.

[48] J.G.Darboux: Leçons sur les systèmes orthogonaux et les coordonnés curvilignes, Gauthiers-Villars, Paris 1910.

[49] R.Debever & M.Cahen: Champs électromagnétiques constants en relativité générale, Comptes Rendus Acad. Sci. Paris 251 (1960) 1160-1162.

[50] J.J. Duistermaat & J.A.C. Kolk: Lie Groups, Springer, Berlin 2000.

[51] William D. Dunbar: Classification of Solvorbifolds in dimension 3, pp.207-216 in "Braids," Contemp. Math. 78, Amer. Math. Soc. 1988.

[52] R.Edwards, K.Millet, D.Sullivan: Foliations with all leaves compact, Topology 16,1 (1977) 13-32.

[53] James Eells Jr. & R.H.Sampson: Harmonic mappings of Riemanian manifolds, Amer.J.Math. 86 (1964) 109-160.

[54] D.Eisenbud, U.Hirsch, W.Neumann: Transverse foliations of Seifert bundles and self-homeomorphisms of the circle, Commentarii Math. Helvetica 56,4 (1981) 638-660

[55] L.P.Eisenhart: A Treatise on the Differential Geometry of Curves and Surfaces, Dover (New York) 1960 reprint of Ginn (Boston) 1909 original.

[56] D.B.A.Epstein: Periodic flows on 3-manifolds, Annals Math. 95 (1972) 68-82.

[57] J. Escobar & R. Schoen: Conformal metrics with prescribed scalar curvature, Invent. Math. 86,2 (1986) 243-254

[58] M.Ferri & C.Gagliardi: Crystallization moves, Pacific J. Math. 100,1 (1982) 85-103.

[59] A.R.Forsyth: Lectures on the Differential Geometry of Curves and Surfaces, Cambridge University Press 1920.

[60] M.H.Freedman & F.Quinn: Topology of 4-manifolds, Princeton University Press 1990.

[61] Christian Fronsdal: Elementary particles in a curved space, IV: massless particles: Phys. Rev. D 12 (1975) 3819-3830. The previous papers in this series were I: Rev.Mod.Phys. 37 (1965) 221-224; II: Phys.Rev. D 10 (1974) 589-598; III (with R.B.Haugen): Phys.Rev.D 12 (1975) 3810-3818. See [151] for commentary.

[62] Guido Fubini: Sugli spazi a quattro dimensioni che ammettono un gruppo continuo di movimenti, Ann. Mat. pura appli. [3] 9 91904) 33-90; reprinted with his preliminary commentary in his "Opere Scelte," a cura dell'Unione matematica italiana e col contributo del Consiglio nazionale delle ricerche, Roma Edizioni Cremonese, 1957-62.

[63] E.Gausmann and 4 others: Topological lensing in spherical spaces, Classical & Quantum Gravity 18 (2001) 5155-5186.

[64] Raoul-François Gloden: Sur les systèmes triples orthogonaux, Comptes Rendus Acad. Sci. Paris 243 (1956) 1010-1012.

[65] G.H.Golub & C.F.Van Loan: Matrix computations, Johns Hopkins Univ. press 1996.

[66] C.McA.Gordon & J.Luecke: Knots are determined by their complements, J. Amer. Math. Soc. 2 (1989) 371-415.

[67] Michel Goze & Paola Piu: Classification des métriques invariantes à gauche sur le groupe de Heisenberg, Rend. Circ. Mat. Palermo (2) 39,2 (1990) 299-306.

[68] Alfred Gray: A generalization of F.Schur's theorem, J.Math.Soc.Japan 21,3 (1969) 454-457.

[69] T.H.Grönwall: Determination of all triply orthogonal systems containing a family of minimal surfaces, Ann. of Math. (2) 17,2 (1915) 76-100.

[70] Joel Haas: Algorithms for recognizing knots and 3-manifolds, `math.GT/9712269`.

[71] J. Hadamard: Les surfaces á courbures opposées et leurs lignes geodésiques, Journal de Mathematiques pures appl. 5 série IV (1898) 27-73.

[72] G.S.Hall: Covariantly constant tensors and holonomy structure in general relativity, J.Math'l. Phys. 32 (1991) 181-187.

[73] Handbook of geometric topology, (eds. R.J. Daverman, R.B. Sher) Elsevier 2002. See especially the chapters by F.Bonahon and J.G. Ratcliffe.

[74] Paul Harrison: Nature's flawed mirror, Physics World, July 2003.

[75] Ernst Heintze: Riemannsche Solvmanigfaltigkeiten, Geometriae Dedicata 1 (1974) 141-147.

[76] Geoffrey Hemion: The classification of knots and 3-dimensional spaces, Oxford Univ. Press 1992.

[77] John Hempel: 3-Manifolds, Annals of Math. Studies 86, Princeton Univ. Press 1976.

[78] Nancy Hingston: On the growth of the number of closed geodesics on the two-sphere, Internat. Math. Res. Notices 9 (1993) 253-262.

[79] Christopher Hoffman: A K counterexample machine, Trans. Amer. Math. Soc. 351, 10 (1999) 4263-4280.

[80] Gerard 't Hooft: Introduction to General Relativity, Rinton Press, Inc., Princeton NJ 2001..

[81] Birger Iversen: Hyperbolic Geometry, Cambridge UP, 1992. (London Mathematical Society Student Texts #25).

[82] W. Jaco & U. Oertel: An Algorithm to decide if a 3-Manifold Is a Haken Manifold, Topology 23 (1984) 195-209.

[83] W.Jaco & P.Shalen: Seifert fibered spaces in 3-manifolds, Memoirs Amer. Math. Soc. 21 (1979), no. 220 (192 pages).

[84] W. Jaco & J.L. Tollefson: Algorithms for the complete decomposition of a closed 3-manifold, Illinois J. Math. 39 (1995) 358-406.

[85] C.G.J. Jacobi: Note von der geodätischen Linie auf einem Ellipsoid und den verschiedenen Anwendungen einer murkwürdigen analytischen Substitution, Crelle's J. 19 (1839) 309-313.

[86] Kenneth C. Jacobs: Cosmologies of Bianchi type I with a uniform magnetic field, Astrophys. J. 155 (1969) 379-391.

[87] Klaus Johannson: Homotopy equivalence of 3-manifolds with boundaries, Springer (Lecture Notes in Math #761) Berlin 1979.

[88] S.Kakutani: Über die metrization der topologischen Gruppen, proc. Imper. Acad. Japan 12 (1936) 82-84.

[89] A.Katok & A.Kononenko: Cocycle's stability for partially hyperbolic systems, Math. Res. Lett. 3 (1996) 191-210.

[90] R.Kirby & L.Siebenmann: Foundational essays on topological manifold smoothings and triangulations, Princeton Univ. Press (Annals of math. studies #88) 1977.

[91] Wilhelm Klingenberg: Lectures on closed geodesics, Springer-Verlag, Berlin 1978.

[92] H.Kneser: Geschlossene Flächen in dreidimensionalen Mannifaltigkeiten, Jahres Deutsche Math. Verein. 38 (1929) 248-260.

[93] Horst Knörrer: Geodesics on the ellipsoid, Inventiones Mathematicae 59 (1980) 119-143.

[94] O.Kowalski, S.Ž.Ničkević: On Ricci eigenvalues of locally homogeneous Riemannan manifolds, Geometriae Dedicata 62 (1996) 65-72.

[95] Oldřich Kowalski & Vlášek Zdenek: Riemannian 3-manifolds with distinct constant Ricci eigenvalues, Math. Bohemica 124,1 (1999) 45-66. Classification of Riemannian 3-manifolds with distinct constant principal Ricci curvatures, Bulletin of the Belgian Mathematical Society – Simon Stevin 5 (1998) 59-68.

[96] Krystyna Kuperberg: A smooth counterexample to the Seifert conjecture, Annals Math. 140,3 (1994) 723-732.

[97] Elon L. Lima: Commuting vector fields on 2-manifolds, Bull. Amer. Math. Soc. 69 (1963) 366-368.

[98] Elon L. Lima: Commuting vector fields on S^2, Proc. Amer. Math. Soc. 15 (1964) 138-141.

[99] Elon L. Lima: Commuting vector fields on S^3, Ann. of Math. (2) 81 (1965) 70-81.

[100] David Lovelock: Spherically symmetric metrics and field equations in four dimensional space, Il Nuovo Cimento 14B,2 (1973) 260-266.

[101] David Lovelock & Hanno Rund: Tensors, Differential Forms, and Variational Principles, Dover reprint.

[102] Jason Manning: Algorithmic detections and description of hyperbolic structures on 3-manifolds with solvable word problem, Geometry & Topology 6 (2002) 1-26.

[103] A.A. Markov: Insolubility of the Problem of Homeomorphy. An English translation of Markov's original Russian paper in Proceedings of the International Congress of Mathematicians, (1958, published 1960 by Cambridge Univ. Press) 300-306, was made by Afra Zomorodian in 1998 at the University of Illinois at Urbana Champaign starting from a German translation by Rolf Herken. Zomorodian recommends the following additional works: S. I. Adyan & G. S. Makanin: Investigations on algorithmic questions of algebra, Proceedings of the Steklov Institute of Mathematics 3 (1986) 209-219; Yu. V. Matiyasevic: Investigations on some algorithmic problems in algebra and number theory. Proceedings of the Steklov Institute of Mathematics 91986) 227-253; P. S. Novikov: On the algorithmic unsolvability of the word problem in group theory. American Math'l Soc. Translations 9 (1958) 1-122.

[104] S.V. Matveev: A recognition algorithm for the 3-dimensional sphere (after A. Thompson), Mat. Sb. 186 (1995) 69-84; English transl. Math. USSR-Sbornik 186 (1995).

[105] Sergei V. Matveev: Algorithmic topology and classification of 3-manifolds, Springer (algorithms & Computation in Math. #9) 2003.

[106] M.A.Melvin: Pure magnetic and electric geons, Phys.Letters 8,1 (1964) 65-68.

[107] M.A.Melvin: Dynamics of cylindrical electromagnetic geons, Phys. Rev. 139, 1B (1965) 225-243.

[108] Willard Miller, Jr.: Symmetry and separation of variables, Addison-Wesley, Reading, Mass. (Encyclopedia of Math'cs #4) 1977.

[109] J. Milnor: A unique factorization theorem for 3-manifolds, Amer. J. Math 84 (1962) 1-7.

[110] J.Milnor: Curvatures of Left invariant metrics on Lie groups, Adv. Math. 21 (1976) 293-329. Reprinted in his 1994 *Collected Papers in Geometry.*

[111] John Milnor: Towards the Poincaré conjecture and the classification of 3-Manifolds, Notices Amer. Math. Soc. 50,10 (Nov. 2003) 1226-1233.

[112] J.W.Milnor & J.D.Stasheff: Characteristic classes, Princeton Univ. Press 1974.

[113] Charles W. Misner, Kip S. Thorne, John A. Wheeler: gravitation, Freeman 1973.

[114] Shigeaki Miyoshi: On foliated circle bundles over closed orientable 3-manifolds, Commentarii Math. Helvetica 72,3 (1981) 400-410.

[115] Edwin E. Moise: Affine structures in 3-manifolds. V. The triangulation theorem and Hauptvermutung, Annals Math. 56 (1952) 96-114.

[116] E.E. Moise: Geometric topology in dimensions 2 and 3, Springer (GTM #47) 1977.

[117] José M. Montesinos: Classical tessellations and three-manifolds, Springer (Universitext) 1987.

[118] D. Montgomery & L. Zippin: Topological transformation groups, Interscience Publishers, New York 1955.

[119] P.H.Moon & D.E.Spencer: Field theory handbook, Springer (2nd ed.) 1988.

[120] J.W. Morgan & H.Bass (eds.): The Smith conjecture, Academic Press 1983.

[121] G.Mostow: Quasi conformal mappings on n-spaces and the rigidity of hyperbolic space forms, Public. Inst. Hautes Et. Sci. 34 (1968) 53-104.

[122] B.K.Nayak & G.B.Bhuyan: Bianchi type V perfect fluid with source-free electromagnetic fields, General Relat. Gravit. 19,9 (1987)939-948

[123] P.J. Nicholls: The Ergodic Theory of Discrete Groups, Cambridge University Press, 1989.

[124] S.P. Novikov: The topology of foliations (Russian), Trudy Moskov. Mat. Obšč. 14 (1965) 248-278.

[125] A. de Oliveira-Costa, M. Tegmark, M. Zaldarriaga, A. Hamilton: The significance of the largest scale CMB fluctuations in WMAP, Phys.Rev. D69 (2004) 063516.

[126] P.Orlik: Seifert manifolds, Springer (Lecture Notes in Math #291) 1972.

[127] P.Orlik, E.Vogt, H.Zieschang: Zur Topologie gefaseter dreidimensionaler Mannigfaltigkeiten, Topology 6 (1967) 49-64.

[128] Donald S. Ornstein & Benjamin Weiss: Geodesic flows are Bernoullian, Israel J. Math. 14 (1973) 184-198.

[129] Gabriel P. Paternain: Geodesic flows, Birkhauser Boston 1999 (Progress in Math. #180).

[130] W.Pauli: Theory of Relativity, Dover reprint 1981.

[131] Grigory Perelman: The Entropy Formula for the Ricci Flow and Its Geometric Application, `http://arXiv.org/abs/math.DG/0211159`, 11 Nov 2002. Ricci Flow with Surgery on Three-Manifolds, `math.DG/0303109`, 10 Mar 2003. Finite extinction time for solutions of the Ricci flow on certain 3-manifolds, `math.DG/0307245`, 17 July 2003. Perelman promises a fourth manuscript.

[132] Peter Petersen: Aspects of global Riemannian geometry, Bull. Amer. Math. Soc. 36,3 (1999) 297-344.

[133] C. Pugh & M. Shub: Stable ergodicity and julienne quasiconformality, J. Europ. Math. Soc. 2,1 (2000) 1-52.

[134] Michael O. Rabin: Recursive unsolvability of group theoretic problems, Annals of Mathematics 67 (1958) 172-194.

[135] T. Radó: Über den Begriff der Riemannschen Flächen, Acta Litt. Sci. Szeged 2 (1925) 101-121.

[136] A.A.Ranicki (ed.): The Hauptvermutung book, Kluwer (Monographs math. #1) 1996.

[137] M.Ratner: Anosov flows with Gibbs measures are also Bernoullian, Israel J. Math. 17 (1974) 380-391.

[138] Harold Rosenberg: The rank of $S^2 \times S^1$, Amer. J. Math. 87 (1965) 11-24.

[139] Harold Rosenberg: Singularities of $\mathbb{R}^2$ actions, Topology 7 (1968) 143-145.

[140] Harold Rosenberg: Foliations by planes, Topology 7 (1968) 131-138.

[141] H.Rosenberg, R.Roussarie, D.Weil: A classification of closed orientable 3-manifolds of rank two, Ann. of Math. (2) 91 (1970) 449-464.

[142] J. H. Rubinstein: An algorithm to recognize the 3-sphere, Proc. ICM (Zurich, 1994), Vols. 1, 2, Birkhauser, Basel, 1995, pp. 601-611.

[143] Daniel J. Rudolph: Ergodic behaviour of Sullivan's geometric measure on a geometrically finite hyperbolic manifold, Ergodic Theory Dynam. Systems 2, 3-4 (1982) 491-512.

[144] R. Schoen: Conformal Deformation of a Riemannian Metric to Constant Scalar Curvature, J. Diff'l. Geom. 20 (1984).

[145] Peter Scott: The geometries of 3-manifolds, Bull. London Math. Soc. 15,5 (1983) 401-487.

[146] H.Seifert: Topology of 3-dimensional fibered spaces. English translation by W.Heil is pp.369-422 in H.Seifert & W.Threlfall: A textbook of topology, Academic Press 1980; original German was Acta Math. 60 (1933) 147-288.

[147] Kouei Sekigawa: On some 3-dimensional curvature homogeneous spaces, Tensor 31 (1977) 87-97.

[148] Ya. G. Sinaĭ: Topics in ergodic theory, Princeton University Press 1994 (Princeton Math'l Series #44).

[149] I.M.Singer: Infinitesimally homogeneous spaces, Commun. Pure Appl. Math. 13 (1960) 685-697.

[150] Warren D. Smith: On the shape of the universe, #63 at `http://www.math.temple.edu/~wds/homepage/works.html`. Mar 2003.

[151] Warren D. Smith: Charge quantization, the topology of the universe, and the hopeful abolition of monopoles, #55 at `http://www.math.temple.edu/~wds/homepage/works.html`. Jan 2000.

[152] M.Spivak: Comprehensive Introduction to Differential Geometry, 5 vols, Publish or Perish; 3rd edition 1999.

[153] G. Steigman: Observational Tests of Antimatter Cosmologies, Ann. Rev. Astron. & Astrophys. 14 (1976) 339-372.

[154] John Stillwell: Classical topology and combinatorial group theory, Second ed., Springer (GTM #72) 1993.

[155] D.Sullivan: Entropy, Haussdorf measures old and new, and limit sets of geometrically finite Kleinian groups, Acta Mathematica 153 (1984) 259-277.

[156] M.B.Tabanov: New ellipsoidal confocal coordinates and geodesics on ellipsoid, J.Mathematical Sciences 82,6 (1996) 3851-3858.

[157] H.Takagi: On curvature homogeneity of Riemannian manifolds, Tohoku Math. J. 26 (1974) 581-585.

[158] Max Tegmark, A. de Oliveira-Costa, A.J.S.Hamilton: A high resolution foreground-cleaned CMB map from WMAP, Phys.Rev. D68 (2003) 123523.

[159] Abigail Thompson: Algorithmic recognition of 3-manifolds, Bulletin Amer. Math. Soc. 35,1 (Jan. 1998) 57-66.

[160] A.Thompson: Thin position and the recognition problem for S^3, Math. Res. Lett. 1 (1994), 613-630.

[161] Kip S. Thorne: Absolute stability of Melvin's magnetic universe, Phys. Rev. 139, 1B (1965) 244-254.

[162] W.Threlfall & H.Seifert: Topologische Untersuchung der Diskontinuitegungsgruppen des dreidimensionalen sphärischen Raumes, Math. Ann. 104 (1931) 1-70.

[163] W.P. Thurston: Existence of codimension-1 foliations, Ann. Math. 104,2 (1976) 249-268.

[164] W.P. Thurston: Three dimensional manifolds, Kleinian groups, and hyperbolic geometry, Bull. Amer. Math. Soc. 6 (1982) 357-381.

[165] William P. Thurston: Three-Dimensional Geometry and Topology, Princeton Mathematical Series 35, Princeton University Press 1997.

[166] J.P.Vallee: Intergalactic and galactic magnetic fields an updated test, Astrophys.Lett. 23 (1983) 85-94.

[167] C.Weber & H.Seifert: Die beiden Dodekaederräume, Math. Z. 37 (1933) 237-253.

[168] Jeffrey R. Weeks: The shape of space, Second edition, Marcel Dekker, 2002. (Well written popular exposition, includes a rough description of the Thurston geometrization conjecture.)

[169] Steven Weinberg: Photons and gravitons in perturbation theory: Derivation of Maxwell's and Einstein's equations, Phys. Rev. 138, 4B (1965) 988-1002. See §VIII.

[170] Steven Weinberg: Gravitation and cosmology, Wiley 1972.

[171] E.P.Wigner: On unitary representations of the inhomogeneous Lorentz group, Annals of Math. 40,1 (1939) 149-204.

[172] Amie Wilkinson: Stable ergodicity of the time-one map of a geodesic flow, Ergod. Th. and Dynam. Sys. 18,6 (1998) 1545-1588.

[173] J.A.Wolf: Curvature in nilpotent Lie groups, Proc. Amer. Math. Soc. 15 (1964) 271-274.

[174] J.A.Wolf: Spaces of constant curvature, Publish or Perish 1984. Now in 5th ed., 1999.

[175] John W. Wood: Foliations on 3-manifolds, Ann. of Math. (2) 89 (1969) 336-358.

[176] Chien-Shiung Wu, E.Ambler, R.W.Hayward, D.D.Hoppes, R.P.Hudson: Experimental test of parity conservation in beta decay, Phys. Rev. 105 (1957) 1413-1415.

[177] Rugang Ye: Global existence and convergence of Yamabe flow, J. Diff. Geom., 39,1 (1994) 35-50.

[178] Vladimir E. Zakharov: Description of the n-orthogonal curvilinear coordinate systems and Hamiltonian integrable systems of hydrodynamic type, I: Integration of the Lame equations, Duke Math. J. 94,1 (1998) 103-139.