Geometrization of 3-Manifolds via the Ricci Flow

Michael T. Anderson

Introduction

The classification of closed surfaces is a milestone in the development of topology, so much so that it is now taught to most mathematics undergraduates as an introduction to topology. Since the solution of the uniformization problem for surfaces by Poincaré and Koebe, this topological classification is now best understood in terms of the geometrization of 2-manifolds: every closed surface Σ admits a metric of constant Gauss curvature $+1$, 0, or -1 and so is uniformized by one of the standard space-form geometries $\mathbb{S}^2$, $\mathbb{R}^2$, $\mathbb{H}^2$. Hence any surface Σ is a quotient of either the 2-sphere, the Euclidean plane, or the hyperbolic disc by a discrete group Γ acting freely and isometrically.

The classification of higher-dimensional manifolds is of course much more difficult. In fact, due to the complexity of the fundamental group, a complete classification as in the case of surfaces is not possible in dimensions ≥ 4. In dimension 3 this argument does not apply, and the full classification of 3-manifolds has long been a dream of topologists. As a very special case, this problem includes the Poincaré Conjecture.

In this article we report on remarkable recent work of Grisha Perelman [15]-[17], which may well have solved the classification problem for 3-manifolds (in a natural sense). Perelman's work is currently under intense investigation and scrutiny by many groups around the world. At this time, much of his work has been validated by experts in the area. Although at the moment it is still too soon to declare a definitive solution to the problem, Perelman's ideas are highly original and of deep insight. Morever, his results are already being used by others in research on related topics. These circumstances serve to justify the writing of an article at this time, which otherwise might be considered premature.

The work of Perelman builds on prior work of Thurston and Hamilton. In the next two sections we discuss the Thurston picture of 3-manifolds and the Ricci flow introduced and analyzed by Hamilton. For additional background, in particular on the Poincaré Conjecture, see Milnor's *Notices* survey [14] and references therein. For much more detailed commentary and discussion on Perelman's work, see [13].

The Geometrization Conjecture

While the Poincaré Conjecture has existed for about one hundred years, the remarkable insights of Thurston in the late 1970s led to the realistic possibility of understanding and classifying all closed 3-manifolds in a manner similar to the classification of surfaces via the uniformization theorem.

To explain this, we first need to consider what are the corresponding geometries in 3-dimensions. In terms of Riemannian geometry, a *geometric structure* on a manifold M is a complete, locally homogeneous Riemannian metric g. Thus, M may be described as the quotient $\Gamma \setminus G/H$, where G is the isometry group of the universal cover $(\widetilde{M}, g)$ and Γ, H are discrete and compact subgroups of the Lie group G respectively. Thurston showed that there are eight such simply connected geometries G/H in dimension 3 which admit compact quotients.[1]

As in two dimensions, the most important geometries are those of constant curvature: hyperbolic

Michael T. Anderson is professor of mathematics at the State University of New York, Stony Brook. His email address is anderson@math.sunysb.edu.

Partially supported by NSF Grant DMS 0305865.

[1] *The Thurston classification is essentially a special case of the much older Bianchi classification of homogeneous space-time metrics arising in general relativity; cf. [3] for further remarks on the dictionary relating these classifications.*

geometry $\mathbb{H}^3$ of curvature -1, Euclidean geometry $\mathbb{R}^3$ of curvature 0, and spherical geometry $\mathbb{S}^3$ of curvature $+1$. The remaining five geometries are products or twisted products with the 2-dimensional geometries. Trivial S^1 bundles over a surface of genus $g > 1$ have $\mathbb{H}^2 \times \mathbb{R}$ geometry, while nontrivial bundles have $\widetilde{SL(2,\mathbb{R})}$ geometry; nontrivial S^1 bundles over T^2 have Nil geometry, while nontrivial T^2 bundles over S^1 have Sol geometry (or Nil or $\mathbb{R}^3$ geometry); finally, S^1 bundles over S^2 have $\mathbb{S}^2 \times \mathbb{R}$ (or $\mathbb{S}^3$) geometry. For example, any *Seifert fibered* 3-manifold, a 3-manifold admitting a locally free S^1 action, has such a geometric structure.

Geometric 3-manifolds, that is 3-manifolds admitting a geometric structure, are the building blocks of more complicated 3-manifolds. For simplicity, we assume throughout the article that all manifolds M are orientable. The building blocks are then assembled along 2-spheres S^2, via connected sum, and along tori T^2. As a simple example of such an assembly, let $\{M_i\}$ be a finite collection of Seifert fibered 3-manifolds over surfaces Σ_i with nonempty boundary, so that ∂M_i consists of tori. These tori may then be glued together pairwise by diffeomorphisms to obtain a closed 3-manifold or a 3-manifold with toral boundary. A 3-manifold assembled in this way is called a *graph manifold*. (One assigns a vertex to each Seifert fibered space and an edge to each torus connecting two such Seifert spaces). A torus bundle over S^1 is a graph manifold, since it is the union of two Seifert fibered spaces over $S^1 \times I$. Graph manifolds were introduced, and their structure completely analyzed, by Waldhausen.

Conversely, let M be an arbitrary closed 3-manifold, as above always orientable. One then decomposes or splits it into pieces according to the structure of the simplest surfaces embedded in M, namely spheres and tori. Topologically, this is accomplished by the following classical results in 3-manifold topology.

Sphere (or Prime) Decomposition (Kneser, Milnor)

Let M be a closed 3-manifold. Then M admits a finite connected sum decomposition

$$M = (K_1\#\ldots\#K_p)\#(L_1\#\ldots\#L_q)\#(\#_1^r S^2 \times S^1). \quad (1)$$

The K and L factors here are closed *irreducible* 3-manifolds; i.e. every embedded 2-sphere S^2 bounds a 3-ball. The K factors have infinite fundamental group and are aspherical 3-manifolds ($K(\pi,1)$'s), while the L factors have finite fundamental group and have universal cover a homotopy 3-sphere. Since $M\#S^3 = M$, we assume no L factor is S^3 unless $M = L = S^3$. The factors in (1) are then unique up to permutation and are obtained from M by performing surgery on a collection of essential, i.e. topologically nontrivial, 2-spheres in M (replacing regions $S^2 \times I$ by two copies of B^3); see Figure 1 for a schematic representation.

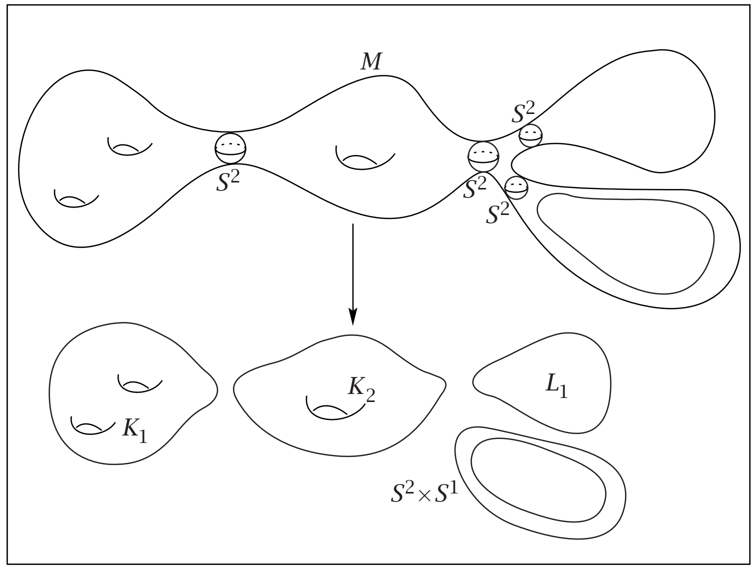

Figure 1. Sphere decomposition.

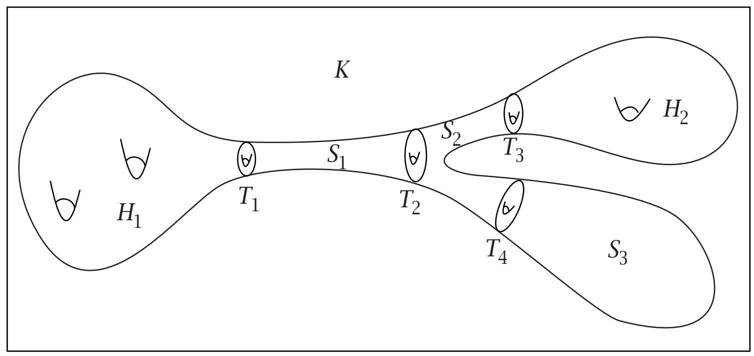

Figure 2. Torus decomposition, (S_i Siefert fibered, H_j torus-irreducible).

The K factors in (1) may also contain topologically essential tori. A torus T^2 embedded in M is called *incompressible* if the inclusion map induces an injection on π_1. A 3-manifold N is called torus-irreducible if every embedded incompressible torus may be deformed to a torus in ∂N. Hence, if $\partial N = \emptyset$, then N has no incompressible tori.

Torus Decomposition (Jaco-Shalen, Johannsen)

Let M be a closed, irreducible 3-manifold. Then there is a finite collection, possibly empty, of disjoint incompressible tori in M that separate M into a finite collection of compact 3-manifolds (with toral boundary), each of which is torus-irreducible or Seifert fibered.

A coarser, but essentially equivalent, decomposition is given by tori separating M into torus-irreducible and graph manifold components; see Figure 2.

With the simple exceptions of $S^2 \times S^1$ and its oriented $\mathbb{Z}_2$ quotient $S^2 \times_{\mathbb{Z}_2} S^1 \simeq \mathbb{RP}^3\#\mathbb{RP}^3$, essential 2-spheres are obstructions to the existence of a geometric structure on a 3-manifold. The same is true for essential tori, unless M happens to be a Seifert fibered or a Sol 3-manifold. Thus, the sphere and torus decompositions divide M topologically into pieces where these known obstructions are removed.

Geometrization Conjecture *(Thurston). Let M be a closed, oriented 3-manifold. Then each component of the sphere and torus decomposition admits a geometric structure.*

The geometrization conjecture gives a complete and effective classification of all closed 3-manifolds, closely resembling in many respects the classification of surfaces. More precisely, it reduces the classification to that of geometric 3-manifolds. The classification of geometric 3-manifolds is rather simple and completely understood, except for the case of hyperbolic 3-manifolds, which remains an active area of research.

As an illustration of the power of the Thurston Conjecture, let us see how it implies the Poincaré Conjecture. If M is a simply connected 3-manifold, then the sphere decomposition (1) implies that M must be an L factor. The geometrization conjecture implies that L is geometric, and so $L = S^3/\Gamma$. Hence, $M = L = S^3$.

Thurston's formulation and work on the geometrization conjecture revolutionized the field of 3-manifold topology; see [18], [19] and further references therein. He recognized that in the class of all (irreducible) 3-manifolds, hyperbolic 3-manifolds are overwhelmingly the most prevalent, as is the case with surfaces, and developed a vast array of new ideas and methods to understand the structure of 3-manifolds. Thurston and a number of other researchers proved the geometrization conjecture in several important cases, the most celebrated being the *Haken manifold theorem*: if M is an irreducible Haken 3-manifold, i.e. M contains an incompressible surface of genus ≥ 1, then the geometrization conjecture is true for M.

An important ingredient in the Thurston approach is the deformation and degeneration of hyperbolic structures on noncompact manifolds (or the deformation of singular hyperbolic structures on compact manifolds). The eight geometric structures are rigid in that there are no geometries which interpolate continuously between them. Hence, on a composite 3-manifold M, the geometric structure on each piece must degenerate in passing from one piece to the next; there is no single structure or metric giving the geometrization of all of M. For example, in Figure 2 the H pieces may be hyperbolic 3-manifolds separated by tori from Seifert fibered pieces S. Although this splitting is topologically well defined, the geometries do not match in the glueing region, and metrically there is no natural region in which to perform the glueing.

Independently and around the same time as Thurston, Gromov [6], [7] also studied the deformation and degeneration of more general Riemannian metrics with merely bounded curvature in place of constant curvature. The idea is that one can control the behavior of a metric, or of a family of metrics, given a uniform bound on the Riemann curvature tensor $Riem$ of the metric.[2] This leads to the important Gromov compactness theorem, the structure theory of almost flat manifolds, and the theory of collapsing Riemannian manifolds, worked out in detail with Cheeger and Fukaya.

One version of these results is especially relevant for our purposes. Let (M, g) be a closed Riemannian manifold, normalized to unit volume, and suppose

$$|Riem| \leq \Lambda, \tag{2}$$

for some arbitrary constant $\Lambda < \infty$. The metric g provides a natural decomposition of M into thick and thin parts, $M = M^\nu \cup M_\nu$, where

$$\begin{aligned} M^\nu &= \{x \in M : volB_x(1) \geq \nu\}, \\ M_\nu &= \{x \in M : volB_x(1) < \nu\}; \end{aligned} \tag{3}$$

here $B_x(1)$ is the geodesic ball about x of radius 1 and $\nu > 0$ is an arbitrary but fixed small number. Now consider the class of all Riemannian n-manifolds of unit volume satisfying (2), and consider the corresponding decompositions (3). Then the geometry and topology of M^ν is a priori controlled. For any given $\nu > 0$, there are only a finite number, (depending on Λ and ν), of possible topological types for M^ν. Moreover, the space of metrics on M^ν is compact in a natural sense; any sequence has a subsequence converging in the $C^{1,\alpha}$ topology, $\alpha < 1$ (modulo diffeomorphisms). For ν sufficiently small, the complementary thin part M_ν admits an F-structure in the sense of Cheeger-Gromov; in dimension 3 this just means that M_ν is a graph manifold with toral (or empty) boundary. In particular, the topology of M_ν is strongly restricted. A metric on M_ν is highly collapsed in the sense that the circles in the Seifert fibered pieces of M_ν and the tori glueing these pieces together have very small diameter, depending on ν; see Figure 3 for a schematic picture. Moreover, for any fixed $\nu > 0$, the distance between M^ν and the arbitrarily thin part $M_{\nu'}$ becomes arbitrarily large as $\nu'/\nu \to 0$.

We point out that similar results hold locally and for complete noncompact manifolds; thus the unit volume normalization above is not essential.

---

[2] *The curvature tensor is a complicated (3,1) tensor expressed in terms of the second derivatives of the metric; in a local geodesic normal coordinate system at a given point, the components of Riem are given by* $R^l_{ijk} = -\frac{1}{2}(\partial_i\partial_k g_{jl} + \partial_j\partial_l g_{ik} - \partial_j\partial_k g_{il} - \partial_i\partial_l g_{jk})$.

The Thurston approach to geometrization has made a great deal of progress on the "hyperbolic part" of the conjecture. In comparison with this, relatively little progress has been made on the "positive curvature part" of the conjecture, for example the Poincaré Conjecture. It is worth pointing out that among the eight geometries, the constant curvature geometries $\mathbb{H}^3$ and $\mathbb{S}^3$ are by far the most important to understand (in terms of characterizing which manifolds are geometric); the other (mixed) geometries are much simpler in comparison.

From the point of view of Riemannian geometry, the Thurston conjecture essentially asserts the existence of a "best possible" metric on an arbitrary closed 3-manifold. In the case that M is not itself geometric, one must allow the optimal metric to have degenerate regions. The discussion and figures above suggest that the degeneration should be via the pinching off of 2-spheres (sphere decomposition) and collapse of graph manifolds along circles and tori (torus decomposition).

The Ricci Flow

One method to find a best metric on a manifold is to find a natural evolution equation, described by a vector field on the space of metrics, and try to prove that the flow lines exist for all time and converge to a geometric limit. In case a flow line does not converge, the corresponding metrics degenerate, and one then needs to relate the degeneration with the topology of M.

There is essentially only one simple and natural vector field (or more precisely family of vector fields) on the space of metrics. It is given by

$$\frac{d}{dt}g(t) = -2Ric_{g(t)} + \lambda(t) \cdot g(t). \tag{4}$$

Here Ric is the Ricci curvature, given in local coordinates by $R_{ij} = (Ric)_{ij} = \sum_k R^k_{ikj}$, so that Ric is a trace of the Riemann curvature. The constant 2 is just for convenience and could be changed by rescaling the time parameter; $\lambda(t)$ is a constant depending on time t. The Ricci flow, introduced by Hamilton [11], is obtained by setting $\lambda = 0$, i.e.

$$\frac{d}{dt}g(t) = -2Ric_{g(t)}. \tag{5}$$

The reason (4) is the only natural flow equation is essentially the same as that leading to the Einstein field equations in general relativity. The Ricci curvature is a symmetric bilinear form, as is the metric. Besides multiples of the metric itself, it is the only such form depending on at most the second derivatives of the metric, and invariant under coordinate changes, i.e. a $(2, 0)$ tensor formed from the metric. By rescaling the metric and time variable t, one may transform (5) into (4). For example, rescaling the Ricci flow (5) so that the volume of $(M, g(t))$ is preserved leads to the flow equation (4) with $\lambda = 2 ⨍ R$, twice the mean value of the scalar curvature R.

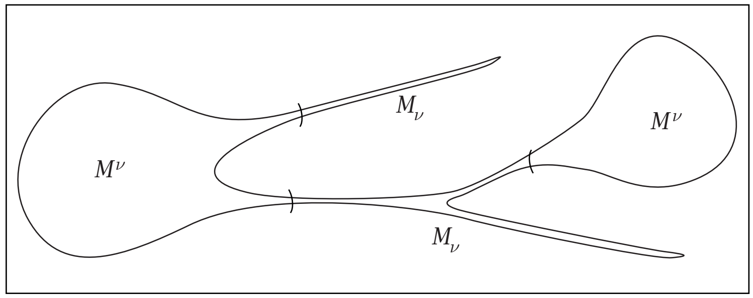

Figure 3. Thick-thin decomposition.

In a suitable local coordinate system, equation (5) has a very natural form. Thus, at time t, choose local harmonic coordinates so that the coordinate functions are locally defined harmonic functions in the metric $g(t)$. Then (5) takes the form

$$\frac{d}{dt}g_{ij} = \Delta g_{ij} + Q_{ij}(g, \partial g), \tag{6}$$

where Δ is the Laplace-Beltrami operator on functions with respect to the metric $g = g(t)$ and Q is a lower-order term quadratic in g and its first-order partial derivatives. This is a nonlinear heat-type equation for g_{ij}. From the analysis of such PDE, one obtains existence and uniqueness of solutions to the Ricci flow on some time interval, starting at any smooth initial metric. This is the reason for the minus sign in (5); a plus sign leads to a backwards heat-type equation, which has no solutions in general.

Here are a few simple examples of explicit solutions to the Ricci flow. If the initial metric $g(0)$ is of constant Ricci curvature, $Ric = a \cdot g$, then the evolution $g(t)$ is just a rescaling of $g(0)$: $g(t) = (1 - 2at)g(0)$. Note that if $a > 0$, then the flow contracts the metric, while if $a < 0$, the flow expands the metric, uniformly in all directions. Hence, if one rescales $g(t)$ to have constant volume, the resulting curve is constant. The stationary points of the volume-normalized Ricci flow are exactly the class of Einstein metrics, i.e. metrics of constant Ricci curvature. In dimension 3, Einstein metrics are of constant curvature and so give the $\mathbb{H}^3$, $\mathbb{R}^3$ and $\mathbb{S}^3$ geometries.

More generally, if $Ric(x, t) > 0$, then the flow contracts the metric $g(t)$ near x, to the future, while if $Ric(x, t) < 0$, then the flow expands $g(t)$ near x. At a general point, there will be directions of positive and negative Ricci curvature along which the metric will locally contract or expand.

Suppose $g(0)$ is a product metric on $S^1 \times \Sigma$, where Σ is a surface with constant curvature metric. Then $g(t)$ remains a product metric, where the length of the S^1 factor stays constant while the surface factor expands or contracts according to the sign of its curvature.

Finally, the Ricci flow commutes with the action of the diffeomorphism group and so preserves all isometries of an initial metric. Thus, geometric 3-manifolds remain geometric. For the "nonpositive" mixed geometries $\mathbb{H}^2 \times \mathbb{R}$, $\widetilde{SL(2,R)}$, Nil, Sol, the volume-normalized Ricci flow contracts the S^1 or T^2 fibers and expands the base surface factor, while for the positive mixed geometry $\mathbb{S}^2 \times \mathbb{R}$, the volume-normalized flow contracts the S^2 and expands in the $\mathbb{R}$-factor.

Now consider the Ricci flow equation (5) in general. From its form it is clear that the flow $g(t)$ will continue to exist if and only if the Ricci curvature remains bounded. This suggests one should consider evolution equations for the curvature, induced by the flow for the metric. The simplest of these is the evolution equation for the scalar curvature $R = tr_g Ric = g^{ij}R_{ij}$:

$$\frac{d}{dt}R = \Delta R + 2|Ric|^2. \tag{7}$$

Evaluating (7) at a point realizing the minimum R_{min} of R on M gives the important fact that R_{min} is monotone nondecreasing along the flow. In particular, the Ricci flow preserves positive scalar curvature (in all dimensions). Moreover, if $R_{min}(0) > 0$, then the same argument gives $\frac{d}{dt}R_{min} \geq \frac{2}{n}R_{min}^2$, $n = dimM$, by the Cauchy-Schwarz inequality $|Ric|^2 \geq \frac{1}{n}R^2$. A simple integration then implies

$$t \leq \frac{n}{2R_{min}(0)}. \tag{8}$$

Thus, the Ricci flow exists only up to a maximal time $T \leq n/2R_{min}(0)$ when $R_{min}(0) > 0$. In contrast, in regions where the Ricci curvature stays negative definite, the flow exists for infinite time.

The evolution of the Ricci curvature has the same general form as (7):

$$\frac{d}{dt}R_{ij} = \Delta R_{ij} + \widetilde{Q}_{ij}. \tag{9}$$

The expression for $\widetilde{Q}$ is much more complicated than the Ricci curvature term in (7) but involves only quadratic expressions in the curvature. However, $\widetilde{Q}$ involves the full Riemann curvature tensor $Riem$ of g and not just the Ricci curvature (as (7) involves Ricci curvature and not just scalar curvature). An elementary but important feature of dimension 3 is that the full Riemann curvature $Riem$ is determined algebraically by the Ricci curvature. This implies that, in general, Ricci flow has a much better chance of "working" in dimension 3. For example, an analysis of $\widetilde{Q}$ shows that the Ricci flow preserves positive Ricci curvature in dimension 3: if $Ric_{g(0)} > 0$, then $Ric_{g(t)} > 0$, for $t > 0$. This is not the case in higher dimensions. On the other hand, in any dimension > 2, the Ricci flow does not preserve negative Ricci curvature, nor does it preserve a general lower bound $Ric \geq -\lambda$, for $\lambda > 0$. For the remainder of the paper, we assume then that $dimM = 3$.

In the Gromov compactness result and thick/thin decomposition (3), the hypothesis of a bound on $|Riem|$ can now also be replaced by a bound on $|Ric|$ (since we are in dimension 3). Further, on time intervals $[0, t]$ where $|Ric|$ is bounded, the metrics $g(t)$ are all quasi-isometric to each other: $cg(0) \leq g(t) \leq Cg(0)$ as bilinear forms, where c, C depend on t. Hence, the arbitrarily thin region M_ν, $\nu << 1$, can only arise, under bounds on $|Ric|$, in arbitrarily large times.

The discussion above shows that the Ricci flow is very natural and has many interesting properties. One can see some relations emerging with the Thurston picture for 3-manifolds. However, the first real indication that the flow is an important new tool in attacking geometric problems is the following result of Hamilton:

• *Space-form Theorem* [8]. If $g(0)$ is a metric of positive Ricci curvature on a 3-manifold M, then the volume-normalized Ricci flow exists for all time and converges to the round metric on S^3/Γ, where Γ is a finite subgroup of $SO(4)$ acting freely on S^3.

Thus the Ricci flow "geometrizes" 3-manifolds of positive Ricci curvature. Since this groundbreaking result, it has been an open question whether it can be generalized to initial metrics with positive scalar curvature.

Although the evolution of the curvature along the Ricci flow is very complicated for general initial metrics, a detailed analysis of (9) leads to the following important results:

• *Curvature pinching estimate* [10], [12]. Let $g(t)$ be a solution to the Ricci flow on a closed 3-manifold M. Then there is a nonincreasing function $\phi : (-\infty, \infty) \to \mathbb{R}$, tending to 0 at ∞, and a constant C, depending only on $g(0)$, such that

$$Riem(x,t) \geq -C - \phi(R(x,t)) \cdot |R(x,t)|. \tag{10}$$

This statement means that all the sectional curvatures R_{ijji} of $g(t)$, where e_i is any orthonormal basis at (x, t), are bounded below by the right side of (10).

This estimate does not imply a lower bound on $Riem(x, t)$ uniform in time. However, when combined with the fact that the scalar curvature $R(x, t)$ *is* uniformly bounded below, it implies that $|Riem|(x, t) >> 1$ only where $R(x, t) >> 1$. Hence, to control the size $|Riem|$ of the full curvature, it suffices to obtain just an upper bound on the scalar curvature R. This is remarkable, since the scalar curvature is a much weaker invariant of the metric than the full curvature. Moreover, at points where the curvature is sufficiently large, (10) shows that $Riem(x, t)/R(x, t) \geq -\delta$, for δ small. Thus, if one scales the metric to make $R(x,t) = 1$, then $Riem(x,t) \geq -\delta$. In such a scale, the metric then has almost nonnegative curvature near (x,t).

• *Harnack estimate* [9]. Let $(N, g(t))$ be a solution to the Ricci flow with bounded and nonnegative curvature $Riem \geq 0$, and suppose $g(t)$ is a complete Riemannian metric on N. Then for $0 < t_1 \leq t_2$,

$$(11) \quad R(x_2, t_2) \geq \frac{t_1}{t_2} \exp\left(-\frac{d_{t_1}^2(x_1, x_2)}{2(t_2 - t_1)}\right) R(x_1, t_1),$$

where d_{t_1} is the distance function on (M, g_{t_1}). This estimate allows one to relate or control the geometry of the solution at different space-time points.

An estimate analogous to (11) in general, i.e. without the assumption $Riem \geq 0$, has been one of the major obstacles to further progress in the Ricci flow.

The analysis above shows that the Ricci flow tends to favor positive curvature. The flow tends to evolve to make the curvature more positive, and the strongest results have been proved in the case of positive curvature, somewhat in contrast to the Thurston approach.

Singularity Formation

The deeper analysis of the Ricci flow is concerned with the singularities that arise in finite time. As (8) already shows, the Ricci flow will not exist for an arbitrarily long time in general. In the case of initial metrics with positive Ricci curvature, this is resolved by rescaling the Ricci flow to constant volume. Hamilton's space-form theorem shows that the volume-normalized flow exists for all time and converges smoothly to a round metric. However, the situation is necessarily much more complicated outside the class of positive Ricci curvature metrics. Consider for instance initial metrics of positive scalar curvature. Any manifold which is a connected sum of S^3/Γ and $S^2 \times S^1$ factors has metrics of positive scalar curvature (compare with the sphere decomposition (1)). Hence, for obvious topological reasons, the volume-normalized Ricci flow could not converge nicely to a round metric; even the renormalized flow must develop singularities.

Singularities occur frequently in numerous classes of nonlinear PDEs and have been extensively studied for many decades. Especially in geometric contexts, the usual method to understand the structure of singularities is to rescale or renormalize the solution on a sequence converging to the singularity to make the solution bounded and try to pass to a limit of the renormalization. Such a limit solution serves as a model for the singularity, and one hopes (or expects) that the singularity models have special features making them much simpler than an arbitrary solution of the equation.

A singularity can form for the Ricci flow only where the curvature becomes unbounded. Suppose then that one has $\lambda_i^2 = |Riem|(x_i, t_i) \to \infty$, on a sequence of points $x_i \in M$, and times $t_i \to T < \infty$. It is then natural to consider the rescaled metrics and times

$$(12) \qquad \bar{g}_i(\bar{t}_i) = \lambda_i^2 g(t), \quad \bar{t}_i = \lambda_i^2(t - t_i).$$

The metrics $\bar{g}_i$ are also solutions of the Ricci flow and have bounded curvature at $(x_i, 0)$. For suitable choices of x_i and t_i, the curvature will be bounded near x_i, and for nearby times to the past, $\bar{t}_i \leq 0$; for example, one might choose points where the curvature is maximal on $(M, g(t))$, $0 \leq t \leq t_i$.

The rescaling (12) expands all distances by the factor λ_i and time by the factor λ_i^2. Thus, in effect one is studying very small regions, of spatial size on the order of $r_i = \lambda_i^{-1}$ about (x_i, t_i), and "using a microscope" to examine the small-scale features in this region on a scale of size about 1. Implicit in this analysis is a change of coordinates near x_i, i.e. use of local diffeomorphisms in conjunction with the metric rescaling.

A local version of the Gromov compactness theorem will then allow one to pass to a limit solution of the Ricci flow, at least locally defined in space and time, provided that the local volumes of the rescalings are bounded below; more explicitly, one needs $x_i \in M^\nu(\bar{g}_i(\bar{t}_i))$, for some fixed $\nu > 0$; see (3). In terms of the original unscaled flow, this means that the metric $g(t)$ should not be locally collapsed, on the scale of its curvature, i.e.

$$volB_{x_i}(r_i, t_i) \geq \nu r_i^3.$$

A maximal connected limit $(N, \bar{g}(\bar{t}), x)$ containing the base point $x = \lim x_i$ is then called a *singularity model*. Observe that the topology of the limit N may well be distinct from the original manifold M, most of which may have been blown off to infinity in the rescaling.

To describe the potential usefulness of this process, suppose one does have local noncollapse on the scale of the curvature and that we have chosen points of maximal curvature in space and time $0 \leq t \leq t_i$. One then obtains, at least in a subsequence, a limit solution to the Ricci flow $(N, \bar{g}(\bar{t}), x)$, based at x, defined at least for times $(-\infty, 0]$; moreover, $\bar{g}(\bar{t})$ is a complete Riemannian metric on N. These are called ancient solutions of the Ricci flow in Hamilton's terminology. The estimates in (10) and (11) can now be used to show that such singularity models do in fact have important features making them much simpler than general solutions of the Ricci flow. As discussed following (10), the pinching estimate implies that the limit has nonnegative curvature. Moreover, the topology of complete manifolds N of nonnegative curvature is completely understood in dimension 3. If N is noncompact, then N is diffeomorphic to $\mathbb{R}^3$, $S^2 \times \mathbb{R}$, or a quotient of these spaces. If N is compact, then a slightly stronger form of Hamilton's theorem above implies N is diffeomorphic to S^3/Γ, $S^2 \times S^1$ or $S^2 \times_{\mathbb{Z}_2} S^1$. Moreover, the Harnack estimate (11) holds on the limit.

These general features of singularity models are certainly encouraging. Nevertheless, there are many problems to overcome to obtain any real benefit from this picture.

I. One needs to prove noncollapse at the scale of the curvature to obtain a singularity model.

II. In general, the curvature may blow up at many different rates or scales, and it is not nearly sufficient to understand just the structure of the singularity models at points of (space-time) maximal curvature. Somewhat analogous phenomena (usually called bubbling) arise in many other geometrical variational problems, for instance harmonic maps, Yang-Mills fields, Einstein metrics, and others. (In such elliptic contexts, these problems of multiple scales have been effectively resolved.)

III. Even if one can solve the two previous issues, this leaves the main issue. One needs to relate the structure of the singularities with the topology of the underlying manifold.

The study of the formation of singularities in the Ricci flow was initiated by Hamilton in [10]; cf. also [4] for a recent survey. Although there has been further technical progress over the last decade, the essential problems on the existence and structure of singularity models and their relation with topology remained unresolved until the appearance of Perelman's work last year.

Perelman's Work

Perelman's recent work [15]-[17] (together with a less crucial paper still to appear) implies a complete solution of the Geometrization Conjecture. This is accomplished by introducing numerous highly original geometric ideas and techniques to understand the Ricci flow. In particular, Perelman's work completely resolves issues I–III above. We proceed by describing, necessarily very briefly, some of the highlights.

I. Noncollapse

Consider the Einstein-Hilbert action

$$\mathcal{R}(g) = \int_M R(g) dV_g \tag{13}$$

as a functional on the space of Riemannian metrics $\mathbb{M}$ on a manifold M. Critical points of $\mathcal{R}$ are Ricci-flat metrics ($Ric = 0$). The action may be adjusted, for instance by adding a cosmological constant -2Λ, to give an action whose critical points are Einstein metrics of constant Ricci curvature.[3] It is natural to try to relate the Ricci flow with $\mathcal{R}$; for instance, is the Ricci flow the gradient flow of $\mathcal{R}$ (with respect to a natural L^2 metric on the space $\mathbb{M}$)? However, while rather close to being true, it has long been recognized that this is not the case. In fact, the gradient flow of $\mathcal{R}$ does not even exist, since it implies a backwards heat-type equation for the scalar curvature R (similar to (7) but with a minus sign before Δ).

Consider now the following functional enhancing $\mathcal{R}$:

$$\mathcal{F}(g, f) = \int_M (R + |\nabla f|^2) e^{-f} dV_g. \tag{14}$$

This is a functional on the larger space $\mathbb{M} \times C^\infty(M, \mathbb{R})$, or equivalently a family of functionals on $\mathbb{M}$, parametrized by $C^\infty(M, \mathbb{R})$.[4] Fix any smooth measure dm on M and define the Perelman coupling by requiring that (g, f) satisfy

$$e^{-f} dV_g = dm. \tag{15}$$

The resulting functional

$$\mathcal{F}^m(g, f) = \int_M (R + |\nabla f|^2) dm \tag{16}$$

becomes a functional on $\mathbb{M}$. At first sight this may appear much more complicated than (13); however, for any $g \in \mathbb{M}$ there exists a large class of functions f (or measures dm) such that the L^2 gradient flow of $\mathcal{F}^m$ exists at g and is given simply by

$$\frac{d\tilde{g}}{dt} = -2(Ric_{\tilde{g}} + D^2 f), \tag{17}$$

where $D^2 f$ is the Hessian of f with respect to $\tilde{g}$. The evolution equation (17) for $\tilde{g}$ is just the Ricci flow (5) modified by an infinitesimal diffeomorphism: $D^2 f = (d/dt)(\phi_t^* \tilde{g})$, where $(d/dt)\phi_t = \nabla f$. Thus, the gradient flow of $\mathcal{F}^m$ is the Ricci flow, up to diffeomorphisms. (Different choices of dm correspond to different choices of diffeomorphism.) In particular, the functional $\mathcal{F}^m$ increases along the Ricci flow.

What can one do with this more complicated functional? It turns out that, given any initial metric $g(0)$ and $t > 0$, the function f (and hence the measure dm) can be freely specified at $g(t)$, where $g(t)$ evolves by the Ricci flow (5). Perelman then uses

[3] *The action (13) leads to the vacuum Einstein field equations in general relativity for Lorentz metrics on a 4-manifold. The term $\lambda(t)$ in (4) is of course analogous to the cosmological constant.*

[4] *The functional (14) arises in string theory as the low-energy effective action [5, §6]; the function or scalar field f is called the dilaton. It is interesting to note in this context that the gravitational field and the dilaton field arise* simultaneously *from the low-energy quantization of the string world sheet (σ-model) [5, p. 837].*

this freedom to probe the geometry of $g(t)$ with suitable choices of f. For instance, he shows by a very simple study of the form of $\mathcal{F}^m$ that the collapse or noncollapse of the metric $g(t)$ near a point $x \in M$ can be detected from the size of $\mathcal{F}^m(g(t))$ by choosing e^{-f} to be an approximation to a delta function centered at x. The more collapsed $g(t)$ is near x, the more negative the value of $\mathcal{F}^m(g(t))$. The collapse of the metric $g(t)$ on any scale in finite time is then ruled out by combining this with the fact that the functional $\mathcal{F}^m$ is increasing along the Ricci flow. In fact, this argument is carried out with respect to a somewhat more complicated *scale-invariant* functional than $\mathcal{F}$; motivated by certain analogies in statistical physics, Perelman calls this the entropy functional.

II. Singularity Models

A second highlight of [15] is essentially a classification of all complete singularity models $(N, g(t))$ that arise in finite time. Complete here means the metric $g(0)$ is a complete Riemannian metric on N; we also drop the overbar from the notation from now on. If N is smooth and compact, then it follows from Hamilton's space-form theorem that N is diffeomorphic to S^3/Γ, $S^2 \times S^1$ or $S^2 \times_{\mathbb{Z}_2} S^1$. In the more important and difficult case where N is complete and noncompact, Perelman proves that the geometry of N near infinity is as simple and natural as possible. At time 0 and at points x with $r(x) = dist(x, x_0) \gg 1$, for a fixed base point x_0, a large neighborhood of x in the scale where $R(x) = 1$ is ε-close to a large neighborhood in the standard round product metric on $S^2 \times \mathbb{R}$. Here ε may be made arbitrarily small by choosing $r(x)$ sufficiently large. Such a region is called an ε-neck. Thus the geometry near infinity in N is that of a union of ε-necks, where the slowly varying radius of S^2 may either be uniformly bounded or diverge to infinity, but only at a rate much smaller than $r(x)$. Moreover, this structure also holds on a long time interval to the past of 0, so that on such regions the solution is close to the (backwards) evolving Ricci flow on $S^2 \times \mathbb{R}$. Topologically, N is diffeomorphic to $\mathbb{R}^3$ or (N, g) is isometric to $S^2 \times \mathbb{R}$.

Perelman shows that this structural result for the singularity models themselves also holds for the solution $g(t)$ very near any singularity time T. Thus, at *any* base point (x, t) where the curvature is sufficiently large, the rescaling as in (12) of the space-time by the curvature is smoothly close, on large compact domains, to corresponding large domains in a complete singularity model. The "ideal" complete singularity models do actually describe the geometry and topology near any singularity. Consequently, one has a detailed understanding of the small-scale geometry and topology everywhere on $(M, g(t))$, for t near T. In particular, this basically proves a general version of the Harnack inequality (11).

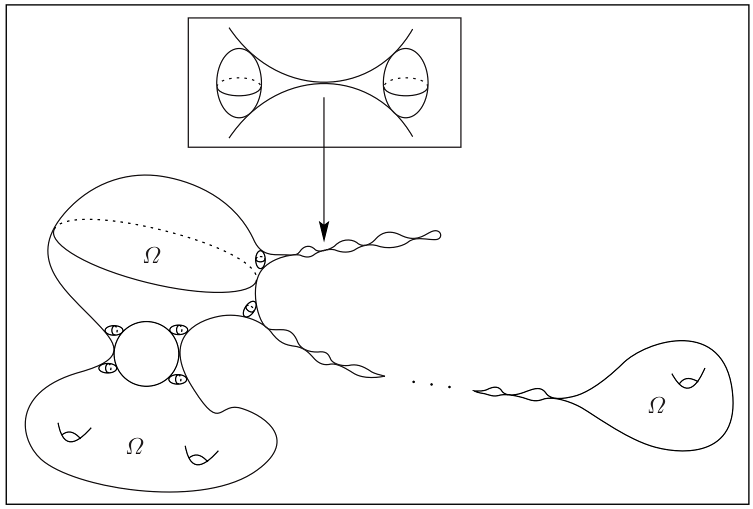

Figure 4. Horns on singular limit.

These results are of course rather technical, and the proofs are not simple. However, they are not exceptionally difficult and mainly rely on new insights and tools to understand the Ricci flow. A key idea is the use of the noncollapse result above on all relevant scales.

III. Relation with Topology

The basic point now is the appearance of 2-spheres S^2 near the singularities. Recall from (1) that one first needs to perform the sphere decomposition on M before it can be geometrized. There is no geometry corresponding to the sphere decomposition.[5] While the sphere decomposition is the simplest operation to carry out topologically, geometrically and analytically it is by far the hardest to understand. How does one detect 2-spheres in M on which to perform surgery from the geometry of a metric? We now see that such 2-spheres, embedded in the ε-necks above, arise naturally near the singularities of the Ricci flow.

The idea then is to surger the 3-manifold M along the 2-spheres just before the first singularity time T. Figure 4 gives a schematic picture of the partially singular metric $g(T)$ on M. The metric $g(T)$ is smooth on a maximal domain $\Omega \subset M$, where the curvature is locally bounded but is singular, i.e. ill-defined, on the complement where the curvature blows up as $t \to T$.

Suppose first that $\Omega = \emptyset$, so that the curvature of $g(t)$ blows up everywhere on M as $t \to T$. One says that the solution to the Ricci flow becomes extinct at time T. Note that $R(x, t) \gg 1$ for all $x \in M$ and t near T (by the pinching estimate (10)). Given the understanding of the singularity models above,

[5] *One might think that the $S^2 \times \mathbb{R}$ geometry corresponds to sphere decomposition, but this is not really correct; at best, this can be made sense of only in an idealized or limiting context.*

it is not difficult to see that M is then diffeomorphic to S^3/Γ, $S^2 \times S^1$, or $S^2 \times_{\mathbb{Z}_2} S^1$. In this situation we are done, since M is then geometric.

If $\Omega \neq \emptyset$, then the main point is that small neighborhoods of the boundary $\partial\Omega$ consist of horns. A horn is a metric on $S^2 \times [0, \delta]$ where the S^2 factor is approximately round of radius $\rho(r)$, with $\rho(r)$ small and $\rho(r)/r \to 0$ as $r \to 0$. Thus, a horn is a union of ε-necks assembled on smaller and smaller scales. The boxed figure in Figure 4 represents a partially singular metric on the smooth manifold $S^2 \times I$, consisting of a pair of horns joined by a degenerate metric. At time T there may be infinitely many components of Ω, of arbitrarily small size, containing such horns. However, all but finitely many of these components are doubled horns, each topologically again of the form $S^2 \times I$.

In quantitative terms, there is a small constant $\rho_0 > 0$ such that if Ω contains no horns with sphere $S^2 \times \{\delta\}$ of radius $\geq \rho_0$, then, as above when $\Omega = \emptyset$, M is diffeomorphic to S^3/Γ, $S^2 \times S^1$, or $S^2 \times_{\mathbb{Z}_2} S^1$, and we are done. If there are horns containing a sphere $S^2 \times \delta$ of a definite size ρ_0 in Ω, one then performs a surgery on each such horn by truncating it along the S^2 of radius ρ_0 and glueing in a smooth 3-ball, giving then a disjoint collection of 3-manifolds.

Having now disconnected M by surgery on 2-spheres into a finite number of components, one then continues with the Ricci flow separately on each component. A conceptually simple, but technically hard, argument based on the decrease of volume associated with each surgery shows that the surgery times are locally finite: on any finite time interval there are only many finitely times at which singularities form.

As a concrete example, suppose the initial metric $g(0)$ on M has positive scalar curvature. Then the estimate (8) shows that Ricci flow completely terminates, i.e. becomes extinct, in finite time. Hence only finitely may surgeries are applied to M during the Ricci flow and it follows from the work above that M is diffeomorphic to a finite connected sum of S^3/Γ and $S^2 \times S^1$ factors.

The upshot of this procedure is that if one successively throws away or ignores such components which become extinct in finite times (and which have already been identified topologically), the Ricci flow with surgery then exists for infinite time $[0, \infty)$. What then does the geometry of the remaining components $\{\hat{M}_i\}$ of M look like at a sufficiently large time T_0? Here the thick-thin decomposition of Gromov-Thurston appears. Fix any $\{\hat{M}\} \in \{\hat{M}_i\}$ and consider the rescaled metric $\hat{g}(t) = t^{-1}g(t)$, for $t = T_0$; it is easy to see from the Ricci flow equation that $\mathrm{vol}(\hat{M}, \hat{g}(t))$ is uniformly bounded. For ν sufficiently small, Perelman proves that there is sufficient control on the ν-thick part $\hat{M}^\nu$, as defined in (3), to see that $\hat{M}^\nu$ is diffeomorphic to a complete hyperbolic 3-manifold H (with finitely many components). The smooth Ricci flow exists on $\hat{M}^\nu$ for infinite time, and the rescalings $t^{-1}g(t)$ converge to the hyperbolic metric of curvature $-\frac{1}{4}$ as $t \to \infty$. (Since the Ricci flow exists for all time, it is reasonable to expect that the volume-normalized flow converges to an Einstein metric, necessarily a hyperbolic metric in our situation.) While there is less control on the ν-thin part $\hat{M}_\nu$, there is enough to conclude that $\hat{M}_\nu$ is diffeomorphic to a graph manifold G (with finitely many components). Although there may still be infinitely many surgeries required to continue the Ricci flow for all time, all further surgeries take place in $\hat{M}_\nu = G$.[6]

Thus, the original 3-manifold M has been decomposed (at large finite time) topologically as

$$M = (K_1\#\ldots\#K_p)\#(\#_1^q S^3/\Gamma_i)\#((\#_1^r S^2 \times S^1). \tag{18}$$

Perelman has recently shown [17] that the S^3/Γ and $S^2 \times S^1$ factors necessarily become extinct in bounded time (with bound depending on the initial metric), so that only the K factors exist after a sufficiently long time. (This result is not needed, however, for the geometrization conjecture.)

Moreover, each $K = K_i$ decomposes via the thick/thin decomposition as a union

$$K = H \cup G, \tag{19}$$

where H is a complete hyperbolic manifold of finite volume (possibly disconnected) and G is a graph manifold (possibly disconnected). The union of H and G is along a collection of embedded tori. Perelman uses the proofs in [11] or [1], [2] to conclude that each such torus is incompressible in K.

This process gives then both the sphere and torus decomposition of the manifold M. Although it is not asserted that the Ricci flow detects the further decomposition of G into Seifert fibered components, this is comparatively elementary from a topological standpoint. The torus-irreducible components of K have been identified as hyperbolic manifolds.

This completes our brief survey of the geometrization conjecture. Perelman's work has created a great deal of excitement in the mathematical research community, as well as in the scientifically interested public at large. While at the moment further evaluation of the details of his work are still being carried out, the beauty and depth of these new contributions are clear.

I am very grateful to Bruce Kleiner, John Lott, and Jack Milnor for their many suggestions and comments, leading to significant improvements in the paper.

[6] *It is not asserted that the bound (2) holds on $\hat{M}_\nu$ for all t large, for some $\Lambda < \infty$.*

References

[1] M. ANDERSON, Scalar curvature and geometrization conjectures for 3-manifolds, in *Comparison Geometry* (Berkeley 1993–94), MSRI Publications, vol. 30, 1997, pp. 49–82.

[2] ____ , Scalar curvature and the existence of geometric structures on 3-manifolds. I, *J. Reine Angew. Math.* **553** (2002), 125–182.

[3] L. ANDERSSON, The global existence problem in general relativity, preprint, 1999, gr-qc/9911032; to appear in *50 Years of the Cauchy Problem in General Relativity*, P. T. Chruściel and H. Friedrich, eds.

[4] H.-D. CAO and B. CHOW, Recent developments on the Ricci flow, *Bull. Amer. Math. Soc. (N.S.)* **36** (1999), 59–74.

[5] E. D'HOKER, String theory, in *Quantum Fields and Strings: A Course for Mathematicians*, *Vol. 2*, Amer. Math. Soc., Providence, RI, 1999.

[6] M. GROMOV, *Structures Métriques pour les Variétés Riemanniennes*, Cedic/Fernand Nathan, Paris, 1981.

[7] ____ , Volume and bounded cohomology, *Inst. Hautes Études Sci. Publ. Math.* **50** (1982), 5–100.

[8] R. HAMILTON, Three manifolds of positive Ricci curvature, *J. Differential Geom.* **17** (1982), 255–306.

[9] ____ , The Harnack estimate for the Ricci flow, *J. Differential Geom.* **37** (1993), 225–243.

[10] ____ , Formation of singularities in the Ricci flow, *Surveys in Differential Geometry, Vol. 2*, International Press, 1995, pp. 7–136.

[11] ____ , Non-singular solutions of the Ricci flow on three-manifolds, *Comm. Anal. Geom.* **7** (1999), 695–729.

[12] T. IVEY, Ricci solitons on compact three-manifolds, *Differential Geom. Appl.* **3** (1993), 301–307.

[13] B. KLEINER and J. LOTT, Ricci flow website: `http://www.math.lsa.umich.edu/research/ricciflow/perelman.html`.

[14] J. MILNOR, Towards the Poincaré Conjecture and the classification of 3-manifolds, preprint, 2003; *Notices Amer. Math. Soc.* **50** (2003), 1226–1233.

[15] G. PERELMAN, The entropy formula for the Ricci flow and its geometric applications, preprint, 2002, math.DG/0211159.

[16] ____ , Ricci flow with surgery on three-manifolds, preprint, 2003, math.DG/0303109.

[17] ____ , Finite extinction time for the solutions to the Ricci flow on certain three-manifolds, preprint, 2003, math.DG/0307245.

[18] W. THURSTON, *The Geometry and Topology of Three-Manifolds*, preprint, 1978, Princeton University, available online at `http://www.msri.org/publications/books/gt3m`; and *Three-Dimensional Geometry and Topology, Vol. 1*, Princeton University Press, 1997.

[19] ____ , Three dimensional manifolds, Kleinian groups and hyperbolic geometry, *Bull. Amer. Math. Soc. (N. S.)* **6** (1982), 357–381.