

## Introduction

In this book we present a complete and detailed proof of

**The Poincaré Conjecture: every closed, smooth, simply connected 3-manifold is diffeomorphic<sup>1</sup> to  $S^3$ .**

This conjecture was formulated by Henri Poincaré [58] in 1904 and has remained open until the recent work of Perelman. The arguments we give here are a detailed version of those that appear in Perelman’s three preprints [53, 55, 54]. Perelman’s arguments rest on a foundation built by Richard Hamilton with his study of the Ricci flow equation for Riemannian metrics. Indeed, Hamilton believed that Ricci flow could be used to establish the Poincaré Conjecture and more general topological classification results in dimension 3, and laid out a program to accomplish this. The difficulty was to deal with singularities in the Ricci flow. Perelman’s breakthrough was to understand the qualitative nature of the singularities sufficiently to allow him to prove the Poincaré Conjecture (and Theorem 0.1 below which implies the Poincaré Conjecture). For a detailed history of the Poincaré Conjecture, see Milnor’s survey article [50].

A class of examples closely related to the 3-sphere are *the 3-dimensional spherical space-forms*, i.e., the quotients of  $S^3$  by free, linear actions of finite subgroups of the orthogonal group  $O(4)$ . There is a generalization of the Poincaré Conjecture, called the **3-dimensional spherical space-form conjecture**, which conjectures that any closed 3-manifold with finite fundamental group is diffeomorphic to a 3-dimensional spherical space-form. Clearly, a special case of the 3-dimensional spherical space-form conjecture is the Poincaré Conjecture.

As indicated in Remark 1.4 of [54], the arguments we present here not only prove the Poincaré Conjecture, they prove the 3-dimensional space-form conjecture. In fact, the purpose of this book is to prove the following more general theorem.

---

<sup>1</sup>Every topological 3-manifold admits a differentiable structure and every homeomorphism between smooth 3-manifolds can be approximated by a diffeomorphism. Thus, classification results about topological 3-manifolds up to homeomorphism and about smooth 3-manifolds up to diffeomorphism are equivalent. In this book ‘manifold’ means ‘smooth manifold.’

**THEOREM 0.1.** *Let  $M$  be a closed, connected 3-manifold and suppose that the fundamental group of  $M$  is a free product of finite groups and infinite cyclic groups. Then  $M$  is diffeomorphic to a connected sum of spherical space-forms, copies of  $S^2 \times S^1$ , and copies of the unique (up to diffeomorphism) non-orientable 2-sphere bundle over  $S^1$ .*

This immediately implies an affirmative resolution of the Poincaré Conjecture and of the 3-dimensional spherical space-form conjecture.

**COROLLARY 0.2.** *(a) A closed, simply connected 3-manifold is diffeomorphic to  $S^3$ . (b) A closed 3-manifold with finite fundamental group is diffeomorphic to a 3-dimensional spherical space-form.*

Before launching into a more detailed description of the contents of this book, one remark on the style of the exposition is in order. Because of the importance and visibility of the results discussed here, and because of the number of incorrect claims of proofs of these results in the past, we felt that it behooved us to work out and present the arguments in great detail. Our goal was to make the arguments clear and convincing and also to make them more easily accessible to a wider audience. As a result, experts may find some of the points are overly elaborated.

## 1. Overview of Perelman's argument

In dimensions less than or equal to 3, any Riemannian metric of constant Ricci curvature has constant sectional curvature. Classical results in Riemannian geometry show that the universal cover of a closed manifold of constant positive curvature is diffeomorphic to the sphere and that the fundamental group is identified with a finite subgroup of the orthogonal group acting linearly and freely on the universal cover. Thus, one can approach the Poincaré Conjecture and the more general 3-dimensional spherical space-form problem by asking the following question. Making the appropriate fundamental group assumptions on 3-manifold  $M$ , how does one establish the existence of a metric of constant Ricci curvature on  $M$ ? The essential ingredient in producing such a metric is the Ricci flow equation introduced by Richard Hamilton in [29]:

$$\frac{\partial g(t)}{\partial t} = -2\text{Ric}(g(t)),$$

where  $\text{Ric}(g(t))$  is the Ricci curvature of the metric  $g(t)$ . The fixed points (up to rescaling) of this equation are the Riemannian metrics of constant Ricci curvature. For a general introduction to the subject of the Ricci flow see Hamilton's survey paper [34], the book by Chow-Knopf [13], or the book by Chow, Lu, and Ni [14]. The Ricci flow equation is a (weakly) parabolic partial differential flow equation for Riemannian metrics on a smooth manifold. Following Hamilton, one defines a Ricci flow to be a family of

Riemannian metrics  $g(t)$  on a fixed smooth manifold, parameterized by  $t$  in some interval, satisfying this equation. One considers  $t$  as time and studies the equation as an initial value problem: Beginning with any Riemannian manifold  $(M, g_0)$  find a Ricci flow with  $(M, g_0)$  as initial metric; that is to say find a one-parameter family  $(M, g(t))$  of Riemannian manifolds with  $g(0) = g_0$  satisfying the Ricci flow equation. This equation is valid in all dimensions but we concentrate here on dimension 3. In a sentence, the method of proof is to begin with any Riemannian metric on the given smooth 3-manifold and flow it using the Ricci flow equation to obtain the constant curvature metric for which one is searching. There are two examples where things work in exactly this way, both due to Hamilton. (i) If the initial metric has positive Ricci curvature, Hamilton proved over twenty years ago, [29], that under the Ricci flow the manifold shrinks to a point in finite time, that is to say, there is a finite-time singularity, and, as we approach the singular time, the diameter of the manifold tends to zero and the curvature blows up at every point. Hamilton went on to show that, in this case, rescaling by a time-dependent function so that the diameter is constant produces a one-parameter family of metrics converging smoothly to a metric of constant positive curvature. (ii) At the other extreme, in [36] Hamilton showed that if the Ricci flow exists for all time and if there is an appropriate curvature bound together with another geometric bound, then as  $t \rightarrow \infty$ , after rescaling to have a fixed diameter, the metric converges to a metric of constant negative curvature.

The results in the general case are much more complicated to formulate and much more difficult to establish. While Hamilton established that the Ricci flow equation has short-term existence properties, i.e., one can define  $g(t)$  for  $t$  in some interval  $[0, T)$  where  $T$  depends on the initial metric, it turns out that if the topology of the manifold is sufficiently complicated, say it is a non-trivial connected sum, then no matter what the initial metric is one must encounter finite-time singularities, forced by the topology. More seriously, even if the manifold has simple topology, beginning with an arbitrary metric one expects to (and cannot rule out the possibility that one will) encounter finite-time singularities in the Ricci flow. These singularities, unlike in the case of positive Ricci curvature, occur along proper subsets of the manifold, not the entire manifold. Thus, to derive the topological consequences stated above, it is not sufficient in general to stop the analysis the first time a singularity arises in the Ricci flow. One is led to study a more general evolution process called *Ricci flow with surgery*, first introduced by Hamilton in the context of four-manifolds, [35]. This evolution process is still parameterized by an interval in time, so that for each  $t$  in the interval of definition there is a compact Riemannian 3-manifold  $M_t$ . But there is a discrete set of times at which the manifolds and metrics undergo topological and metric discontinuities (surgeries). In each of the complementary

intervals to the singular times, the evolution is the usual Ricci flow, though, because of the surgeries, the topological type of the manifold  $M_t$  changes as  $t$  moves from one complementary interval to the next. From an analytic point of view, the surgeries at the discontinuity times are introduced in order to ‘cut away’ a neighborhood of the singularities as they develop and insert by hand, in place of the ‘cut away’ regions, geometrically nice regions. This allows one to continue the Ricci flow (or more precisely, restart the Ricci flow with the new metric constructed at the discontinuity time). Of course, the surgery process also changes the topology. To be able to say anything useful topologically about such a process, one needs results about Ricci flow, and one also needs to control both the topology and the geometry of the surgery process at the singular times. For example, it is crucial for the topological applications that we do surgery along 2-spheres rather than surfaces of higher genus. Surgery along 2-spheres produces the connected sum decomposition, which is well-understood topologically, while, for example, Dehn surgeries along tori can completely destroy the topology, changing any 3-manifold into any other.

The change in topology turns out to be completely understandable and amazingly, the surgery processes produce exactly the topological operations needed to cut the manifold into pieces on which the Ricci flow can produce the metrics sufficiently controlled so that the topology can be recognized.

The bulk of this book (Chapters 1-17 and the Appendix) concerns the establishment of the following long-time existence result for Ricci flow with surgery.

**THEOREM 0.3.** *Let  $(M, g_0)$  be a closed Riemannian 3-manifold. Suppose that there is no embedded, locally separating  $\mathbb{R}P^2$  contained<sup>2</sup> in  $M$ . Then there is a Ricci flow with surgery defined for all  $t \in [0, \infty)$  with initial metric  $(M, g_0)$ . The set of discontinuity times for this Ricci flow with surgery is a discrete subset of  $[0, \infty)$ . The topological change in the 3-manifold as one crosses a surgery time is a connected sum decomposition together with removal of connected components, each of which is diffeomorphic to one of  $S^2 \times S^1$ ,  $\mathbb{R}P^3 \# \mathbb{R}P^3$ , the non-orientable 2-sphere bundle over  $S^1$ , or a manifold admitting a metric of constant positive curvature.*

While Theorem 0.3 is central for all applications of Ricci flow to the topology of three-dimensional manifolds, the argument for the 3-manifolds described in Theorem 0.1 is simplified, and avoids all references to the nature of the flow as time goes to infinity, because of the following finite-time extinction result.

---

<sup>2</sup>I.e., no embedded  $\mathbb{R}P^2$  in  $M$  with trivial normal bundle. Clearly, all orientable manifolds satisfy this condition.

**THEOREM 0.4.** *Let  $M$  be a closed 3-manifold whose fundamental group is a free product of finite groups and infinite cyclic groups<sup>3</sup>. Let  $g_0$  be any Riemannian metric on  $M$ . Then  $M$  admits no locally separating  $\mathbb{R}P^2$ , so that there is a Ricci flow with surgery defined for all positive time with  $(M, g_0)$  as initial metric as described in Theorem 0.3. This Ricci flow with surgery becomes extinct after some time  $T < \infty$ , in the sense that the manifolds  $M_t$  are empty for all  $t \geq T$ .*

This result is established in Chapter 18 following the argument given by Perelman in [54], see also [15].

We immediately deduce Theorem 0.1 from Theorems 0.3 and 0.4 as follows: Let  $M$  be a 3-manifold satisfying the hypothesis of Theorem 0.1. Then there is a finite sequence  $M = M_0, M_1, \dots, M_k = \emptyset$  such that for each  $i$ ,  $1 \leq i \leq k$ ,  $M_i$  is obtained from  $M_{i-1}$  by a connected sum decomposition or  $M_i$  is obtained from  $M_{i-1}$  by removing a component diffeomorphic to one of  $S^2 \times S^1$ ,  $\mathbb{R}P^3 \# \mathbb{R}P^3$ , a non-orientable 2-sphere bundle over  $S^1$ , or a 3-dimensional spherical space-form. Clearly, it follows by downward induction on  $i$  that each connected component of  $M_i$  is diffeomorphic to a connected sum of 3-dimensional spherical space-forms, copies of  $S^2 \times S^1$ , and copies of the non-orientable 2-sphere bundle over  $S^1$ . In particular,  $M = M_0$  has this form. Since  $M$  is connected by hypothesis, this proves the theorem. In fact, this argument proves the following:

**COROLLARY 0.5.** *Let  $(M_0, g_0)$  be a connected Riemannian manifold with no locally separating  $\mathbb{R}P^2$ . Let  $(\mathcal{M}, G)$  be a Ricci flow with surgery defined for  $0 \leq t < \infty$  with  $(M_0, g_0)$  as initial manifold. Then the following four conditions are equivalent:*

- (1)  *$(\mathcal{M}, G)$  becomes extinct after a finite time, i.e.,  $M_T = \emptyset$  for all  $T$  sufficiently large,*
- (2)  *$M_0$  is diffeomorphic to a connected sum of three-dimensional spherical space-forms and  $S^2$ -bundles over  $S^1$ ,*
- (3) *the fundamental group of  $M_0$  is a free product of finite groups and infinite cyclic groups,*
- (4) *no prime<sup>4</sup> factor of  $M_0$  is acyclic, i.e., every prime factor of  $M_0$  has either non-trivial  $\pi_2$  or non-trivial  $\pi_3$ .*

**PROOF.** Repeated application of Theorem 0.3 shows that (1) implies (2). The implication (2) implies (3) is immediate from van Kampen's theorem.

<sup>3</sup>In [54] Perelman states the result for 3-manifolds without prime factors that are acyclic. It is a standard exercise in 3-manifold topology to show that Perelman's condition is equivalent to the group theory hypothesis stated here; see Corollary 0.5.

<sup>4</sup>A three-manifold  $P$  is prime if every separating 2-sphere in  $P$  bounds a three-ball in  $P$ . Equivalently,  $P$  is prime if it admits no non-trivial connected sum decomposition. Every closed three-manifold decomposes as a connected sum of prime factors with the decomposition being unique up to diffeomorphism of the factors and the order of the factors.

The fact that (3) implies (1) is Theorem 0.4. This shows that (1), (2) and (3) are all equivalent. Since three-dimensional spherical space-forms and  $S^2$ -bundles over  $S^1$  are easily seen to be prime, (2) implies (4). Thus, it remains only to see that (4) implies (3). We consider a manifold  $M$  satisfying (4), a prime factor  $P$  of  $M$ , and universal covering  $\tilde{P}$  of  $P$ . First suppose that  $\pi_2(P) = \pi_2(\tilde{P})$  is trivial. Then, by hypothesis  $\pi_3(P) = \pi_3(\tilde{P})$  is non-trivial. By the Hurewicz theorem this means that  $H_3(\tilde{P})$  is non-trivial, and hence that  $\tilde{P}$  is a compact, simply connected three-manifold. It follows that  $\pi_1(P)$  is finite. Now suppose that  $\pi_2(P)$  is non-trivial. Then  $P$  is not diffeomorphic to  $\mathbb{R}P^3$ . Since  $P$  is prime and contains no locally separating  $\mathbb{R}P^2$ , it follows that  $P$  contains no embedded  $\mathbb{R}P^2$ . Then by the sphere theorem there is an embedded 2-sphere in  $P$  that is homotopically non-trivial. Since  $P$  is prime, this sphere cannot separate, so cutting  $P$  open along it produces a connected manifold  $P_0$  with two boundary 2-spheres. Since  $P_0$  is prime, it follows that  $P_0$  is diffeomorphic to  $S^2 \times I$  and hence  $P$  is diffeomorphic to a 2-sphere bundle over the circle.  $\square$

REMARK 0.6. (i) The use of the sphere theorem is unnecessary in the above argument for what we actually prove is that if every prime factor of  $M$  has non-trivial  $\pi_2$  or non-trivial  $\pi_3$ , then the Ricci flow with surgery with  $(M, g_0)$  as initial metric becomes extinct after a finite time. In fact, the sphere theorem for closed 3-manifolds follows from the results here.

(ii) If the initial manifold is simpler then all the time-slices are simpler: If  $(\mathcal{M}, G)$  is a Ricci flow with surgery whose initial manifold is prime, then every time-slice is a disjoint union of connected components, all but at most one being diffeomorphic to a 3-sphere and if there is one not diffeomorphic to a 3-sphere, then it is diffeomorphic to the initial manifold. If the initial manifold is a simply connected manifold  $M_0$ , then every component of every time-slice  $M_T$  must be simply connected, and thus *a posteriori* every time-slice is a disjoint union of manifolds diffeomorphic to the 3-sphere. Similarly, if the initial manifold has finite fundamental group, then every connected component of every time-slice is either simply connected or has the same fundamental group as the initial manifold.

(iii) The conclusion of this result is a natural generalization of Hamilton's conclusion in analyzing the Ricci flow on manifolds of positive Ricci curvature in [29]. Namely, under appropriate hypotheses, during the evolution process of Ricci flow with surgery the manifold breaks into components each of which disappears in finite time. As a component disappears at some finite time, the metric on that component is well enough controlled to show that the disappearing component admits a non-flat, homogeneous Riemannian metric of non-negative sectional curvature, i.e., a metric locally isometric to either a round  $S^3$  or to a product of a round  $S^2$  with the usual metric on  $\mathbb{R}$ . The existence of such a metric on a component immediately gives the

topological conclusion of Theorem 0.1 for that component, i.e., that it is diffeomorphic to a 3-dimensional spherical space-form, to  $S^2 \times S^1$  to a non-orientable 2-sphere bundle over  $S^1$ , or to  $\mathbb{R}P^3 \# \mathbb{R}P^3$ . The biggest difference between these two results is that Hamilton's hypothesis is geometric (positive Ricci curvature) whereas Perelman's is homotopy theoretic (information about the fundamental group).

(iv) It is also worth pointing out that it follows from Corollary 0.5 that the manifolds that satisfy the four equivalent conditions in that corollary are exactly the closed, connected, 3-manifolds that admit a Riemannian metric of positive scalar curvature, cf, [62] and [26].

One can use Ricci flow in a more general study of 3-manifolds than the one we carry out here. There is a conjecture due to Thurston, see [69], known as Thurston's Geometrization Conjecture or simply as the Geometrization Conjecture for 3-manifolds. It conjectures that every 3-manifold without locally separating  $\mathbb{R}P^2$ 's (in particular every orientable 3-manifold) is a connected sum of prime 3-manifolds each of which admits a decomposition along incompressible<sup>5</sup> tori into pieces that admit locally homogeneous geometries of finite volume. Modulo questions about cofinite-volume lattices in  $SL_2(\mathbb{C})$ , proving this conjecture leads to a complete classification of 3-manifolds without locally separating  $\mathbb{R}P^2$ 's, and in particular to a complete classification of all orientable 3-manifolds. (See Peter Scott's survey article [63].) By passing to the orientation double cover and working equivariantly, these results can be extended to all 3-manifolds.

Perelman in [55] has stated results which imply a positive resolution of Thurston's Geometrization conjecture. Perelman's proposed proof of Thurston's Geometrization Conjecture relies in an essential way on Theorem 0.3, namely the existence of Ricci flow with surgery for all positive time. But it also involves a further analysis of the limits of these Ricci flows as time goes to infinity. This further analysis involves analytic arguments which are exposed in Sections 6 and 7 of Perelman's second paper ([55]), following earlier work of Hamilton ([36]) in a simpler case of bounded curvature. They also involve a result (Theorem 7.4 from [55]) from the theory of manifolds with curvature locally bounded below that are collapsed, related to results of Shioya-Yamaguchi [67]. The Shioya-Yamaguchi results in turn rely on an earlier, unpublished work of Perelman proving the so-called 'Stability Theorem.' Recently, Kapovich, [43] has put a preprint on the archive giving a proof of the stability result. We have been examining another approach, one suggested by Perelman in [55], avoiding the stability theorem, cf, [44] and [51]. It is our view that the collapsing results needed for the Geometrization Conjecture are in place, but that before a definitive statement that the Geometrization Conjecture has been resolved can be made these

---

<sup>5</sup>I.e., embedded by a map that is injective on  $\pi_1$ .

arguments must be subjected to the same close scrutiny that the arguments proving the Poincaré Conjecture have received. This process is underway.

In this book we do not attempt to explicate any of the results beyond Theorem 0.3 described in the previous paragraph that are needed for the Geometrization Conjecture. Rather, we content ourselves with presenting a proof of Theorem 0.1 above which, as we have indicated, concerns initial Riemannian manifolds for which the Ricci flow with surgery becomes extinct after finite time. We are currently preparing a detailed proof, along the lines suggested by Perelman, of the further results that will complete the proof of the Geometrization Conjecture.

As should be clear from the above overview, Perelman's argument did not arise in a vacuum. Firstly, it resides in a context provided by the general theory of Riemannian manifolds. In particular, various notions of convergence of sequences of manifolds play a crucial role. The most important is geometric convergence (smooth convergence on compact subsets). Even more importantly, Perelman's argument resides in the context of the theory of the Ricci flow equation, introduced by Richard Hamilton and extensively studied by him and others. Perelman makes use of almost every previously established result for 3-dimensional Ricci flows. One exception is Hamilton's proposed classification results for 3-dimensional singularities. These are replaced by Perelman's strong qualitative description of singularity development for Ricci flows on compact 3-manifolds.

The first five chapters of the book review the necessary background material from these two subjects. Chapters 6 through 11 then explain Perelman's advances. In Chapter 12 we introduce the standard solution, which is the manifold constructed by hand that one 'glues in' in doing surgery. Chapters 13 through 17 describe in great detail the surgery process and prove the main analytic and topological estimates that are needed to show that one can continue the process for all positive time. At the end of Chapter 17 we have established Theorem 0.3. Chapter 18 and 19 discuss the finite-time extinction result. Lastly, there is an appendix on some topological results that were needed in the surgery analysis in Chapters 13-17.

## 2. Background material from Riemannian geometry

**2.1. Volume and injectivity radius.** One important general concept that is used throughout is the notion of a manifold being non-collapsed at a point. Suppose that we have a point  $x$  in a complete Riemannian  $n$ -manifold. Then we say that the manifold is  $\kappa$ -non-collapsed at  $x$  provided that the following holds: For any  $r$  such that the norm of the Riemann curvature tensor,  $|\text{Rm}|$ , is  $\leq r^{-2}$  at all points of the metric ball,  $B(x, r)$ , of radius  $r$  centered at  $x$ , we have  $\text{Vol } B(x, r) \geq \kappa r^n$ . There is a relationship between this notion and the injectivity radius of  $M$  at  $x$ . Namely, if  $|\text{Rm}| \leq r^{-2}$  on  $B(x, r)$  and if  $B(x, r)$  is  $\kappa$ -non-collapsed then the injectivity radius of  $M$

at  $x$  is greater than or equal to a positive constant that depends only on  $r$  and  $\kappa$ . The advantage of working with the volume non-collapsing condition is that, unlike for the injectivity radius, there is a simple equation for the evolution of volume under Ricci flow.

Another important general result is the Bishop-Gromov volume comparison result that says that if the Ricci curvature of a complete Riemannian  $n$ -manifold  $M$  is bounded below by a constant  $(n-1)K$ , then for any  $x \in M$  the ratio of the volume of  $B(x, r)$  to the volume of the ball of radius  $r$  in the space of constant curvature  $K$  is a non-increasing function whose limit as  $r \rightarrow 0$  is 1.

All of these basic facts from Riemannian geometry are reviewed in the first chapter.

**2.2. Manifolds of non-negative curvature.** For reasons that should be clear from the above description and in any event will become much clearer shortly, manifolds of non-negative curvature play an extremely important role in the analysis of Ricci flows with surgery. We need several general results about them. The first is the soul theorem for manifolds of non-negative sectional curvature. A *soul* is a compact, totally geodesic submanifold. The entire manifold is diffeomorphic to the total space of a vector bundle over any of its souls. If a non-compact  $n$ -manifold has positive sectional curvature, then any soul for it is a point, and in particular, the manifold is diffeomorphic to Euclidean space. In addition, the distance function  $f$  from a soul has the property that, for every  $t > 0$ , the pre-image  $f^{-1}(t)$  is homeomorphic to an  $(n-1)$ -sphere and the pre-image under this distance function of any non-degenerate interval  $I \subset \mathbb{R}^+$  is homeomorphic to  $S^{n-1} \times I$ .

Another important result is the splitting theorem, which says that, if a complete manifold of non-negative sectional curvature has a geodesic line (an isometric copy of  $\mathbb{R}$ ) that is distance minimizing between every pair of its points, then that manifold is a metric product of a manifold of one lower dimension and  $\mathbb{R}$ . In particular, if a complete  $n$ -manifold of non-negative sectional curvature has two ends, then it is a metric product  $N^{n-1} \times \mathbb{R}$  where  $N^{n-1}$  is a compact manifold.

Also, we need some of the elementary comparison results from Toponogov theory. These compare ordinary triangles in the Euclidean plane with triangles in a manifold of non-negative sectional curvature whose sides are minimizing geodesics in that manifold.

**2.3. Canonical neighborhoods.** Much of the analysis of the geometry of Ricci flows revolves around the notion of canonical neighborhoods. Fix some  $\epsilon > 0$  sufficiently small. There are two types of non-compact canonical neighborhoods:  $\epsilon$ -necks and  $\epsilon$ -caps. An  $\epsilon$ -neck in a Riemannian 3-manifold  $(M, g)$  centered at a point  $x \in M$  is a submanifold  $N \subset M$  and

a diffeomorphism  $\psi: S^2 \times (-\epsilon^{-1}, \epsilon^{-1}) \rightarrow N$  such that  $x \in \psi(S^2 \times \{0\})$  and such that the pullback of the rescaled metric,  $\psi^*(R(x)g)$ , is within  $\epsilon$  in the  $C^{[1/\epsilon]}$ -topology of the product of the round metric of scalar curvature 1 on  $S^2$  with the usual metric on the interval  $(-\epsilon^{-1}, \epsilon^{-1})$ . (Throughout,  $R(x)$  denotes the scalar curvature of  $(M, g)$  at the point  $x$ .) An  $\epsilon$ -cap is a non-compact submanifold  $\mathcal{C} \subset M$  with the property that a neighborhood  $N$  of infinity in  $\mathcal{C}$  is an  $\epsilon$ -neck, such that every point of  $N$  is the center of an  $\epsilon$ -neck in  $M$ , and such that the *core*,  $\mathcal{C} \setminus \overline{N}$ , of the  $\epsilon$ -cap is diffeomorphic to either a 3-ball or a punctured  $\mathbb{R}P^3$ . It will also be important to consider  $\epsilon$ -caps that, after rescaling to make  $R(x) = 1$  for some point  $x$  in the cap, have bounded geometry (bounded diameter, bounded ratio of the curvatures at any two points, and bounded volume). If  $C$  represents the bound for these quantities, then we call the cap a  $(C, \epsilon)$ -cap. See FIG. 1. An  $\epsilon$ -tube in  $M$  is a submanifold of  $M$  diffeomorphic to  $S^2 \times (0, 1)$  which is a union of  $\epsilon$ -necks and with the property that each point of the  $\epsilon$ -tube is the center of an  $\epsilon$ -neck in  $M$ .

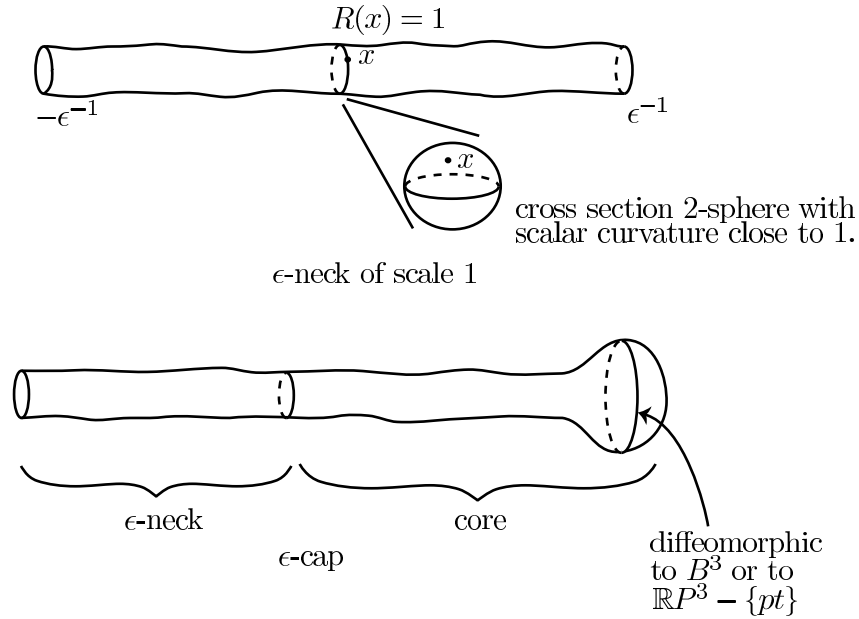


FIGURE 1. Canonical neighborhoods.

There are two other types of canonical neighborhoods in 3-manifolds – (i) a  $C$ -component and (ii) an  $\epsilon$ -round component. The  $C$ -component is a compact, connected Riemannian manifold of positive sectional curvature diffeomorphic to either  $S^3$  or  $\mathbb{R}P^3$  with the property that rescaling the metric by  $R(x)$  for any  $x$  in the component produces a Riemannian manifold whose diameter is at most  $C$ , whose sectional curvature at any point and in any 2-plane direction is between  $C^{-1}$  and  $C$ , and whose volume is between  $C^{-1}$

and  $C$ . An  $\epsilon$ -round component is a component on which the metric rescaled by  $R(x)$  for any  $x$  in the component is within  $\epsilon$  in the  $C^{[1/\epsilon]}$ -topology of a round metric of scalar curvature 1.

As we shall see, the singularities at time  $T$  of a 3-dimensional Ricci flow are contained in subsets that are unions of canonical neighborhoods with respect to the metrics at nearby, earlier times  $t' < T$ . Thus, we need to understand the topology of manifolds that are unions of  $\epsilon$ -tubes and  $\epsilon$ -caps. The fundamental observation is that, provided that  $\epsilon$  is sufficiently small, when two  $\epsilon$ -necks intersect (in more than a small neighborhood of the boundaries) their product structures almost line up, so that the two  $\epsilon$ -necks can be glued together to form a manifold fibered by  $S^2$ 's. Using this idea we show that, for  $\epsilon > 0$  sufficiently small, if a connected manifold is a union of  $\epsilon$ -tubes and  $\epsilon$ -caps, then it is diffeomorphic to  $\mathbb{R}^3$ ,  $S^2 \times \mathbb{R}$ ,  $S^3$ ,  $S^2 \times S^1$ ,  $\mathbb{R}P^3 \# \mathbb{R}P^3$ , the total space of a line bundle over  $\mathbb{R}P^2$ , or the non-orientable 2-sphere bundle over  $S^1$ . This topological result is proved in the appendix at the end of the book. **We shall fix  $\epsilon > 0$  sufficiently small so that these results hold.**

There is one result relating canonical neighborhoods and manifolds of positive curvature of which we make repeated use: Any complete 3-manifold of positive curvature does not admit  $\epsilon$ -necks of arbitrarily high curvature. In particular, if  $M$  is a complete Riemannian 3-manifold with the property that every point of scalar curvature greater than  $r_0^{-2}$  has a canonical neighborhood, then  $M$  has bounded curvature. This turns out to be of central importance and is used repeatedly.

All of these basic facts about Riemannian manifolds of non-negative curvature are recalled in the second chapter.

### 3. Background material from Ricci flow

Hamilton [29] introduced the Ricci flow equation,

$$\frac{\partial g(t)}{\partial t} = -2\text{Ric}(g(t)).$$

This is an evolution equation for a one-parameter family of Riemannian metrics  $g(t)$  on a smooth manifold  $M$ . The Ricci flow equation is weakly parabolic and is strictly parabolic modulo the ‘gauge group’, which is the group of diffeomorphisms of the underlying smooth manifold. One should view this equation as a non-linear, tensor version of the heat equation. From it, one can derive the evolution equation for the Riemannian metric tensor, the Ricci tensor, and the scalar curvature function. These are all parabolic equations. For example, the evolution equation for scalar curvature  $R(x, t)$  is

$$(0.1) \quad \frac{\partial R}{\partial t}(x, t) = \Delta R(x, t) + 2|\text{Ric}(x, t)|^2,$$

illustrating the similarity with the heat equation. (Here  $\Delta$  is the Laplacian with non-positive spectrum.)

**3.1. First results.** Of course, the first results we need are uniqueness and short-time existence for solutions to the Ricci flow equation for compact manifolds. These results were proved by Hamilton ([29]) using the Nash-Moser inverse function theorem, ([28]). These results are standard for strictly parabolic equations. By now there is a fairly standard method for working ‘modulo’ the gauge group (the group of diffeomorphisms) and hence arriving at a strictly parabolic situation where the classical existence, uniqueness and smoothness results apply. The method for the Ricci flow equation goes under the name of ‘DeTurck’s trick.’

There is also a result that allows us to patch together local solutions  $(U, g(t))$ ,  $a \leq t \leq b$ , and  $(U, h(t))$ ,  $b \leq t \leq c$ , to form a smooth solution defined on the interval  $a \leq t \leq c$  provided that  $g(b) = h(b)$ .

Given a Ricci flow  $(M, g(t))$  we can always translate, replacing  $t$  by  $t + t_0$  for some fixed  $t_0$ , to produce a new Ricci flow. We can also rescale by any positive constant  $Q$  by setting  $h(t) = Qg(Q^{-1}t)$  to produce a new Ricci flow.

**3.2. Gradient shrinking solitons.** Suppose that  $(M, g)$  is a complete Riemannian manifold, and suppose that there is a constant  $\lambda > 0$  with the property that

$$\text{Ric}(g) = \lambda g.$$

In this case, it is easy to see that there is a Ricci flow given by

$$g(t) = (1 - 2\lambda t)g.$$

In particular, all the metrics in this flow differ by a constant factor depending on time and the metric is a decreasing function of time. These are called *shrinking solitons*. Examples are compact manifolds of constant positive Ricci curvature.

There is a closely related, but more general, class of examples: the *gradient shrinking solitons*. Suppose that  $(M, g)$  is a complete Riemannian manifold, and suppose that there is a constant  $\lambda > 0$  and a function  $f: M \rightarrow \mathbb{R}$  satisfying

$$\text{Ric}(g) = \lambda g - \text{Hess}^g f.$$

In this case, there is a Ricci flow which is a shrinking family after we pull back by the one-parameter family of diffeomorphisms generated by the time-dependent vector field  $\frac{1}{1-2\lambda t} \nabla_g f$ . An example of a gradient shrinking soliton is the manifold  $S^2 \times \mathbb{R}$  with the family of metrics being the product of the shrinking family of round metrics on  $S^2$  and the constant family of standard metrics on  $\mathbb{R}$ . The function  $f$  is  $s^2/4$  where  $s$  is the Euclidean parameter on  $\mathbb{R}$ .

**3.3. Controlling higher derivatives of curvature.** Now let us discuss the smoothness results for geometric limits. The general result along these lines is Shi's theorem, see [65, 66]. Again, this is a standard type of result for parabolic equations. Of course, the situation here is complicated somewhat by the existence of the gauge group. Roughly, Shi's theorem says the following. Let us denote by  $B(x, t_0, r)$  the metric ball in  $(M, g(t_0))$  centered at  $x$  and of radius  $r$ . If we can control the norm of the Riemann curvature tensor on a backward neighborhood of the form  $B(x, t_0, r) \times [0, t_0]$ , then for each  $k > 0$  we can control the  $k^{\text{th}}$  covariant derivative of the curvature on  $B(x, t_0, r/2^k) \times [0, t_0]$  by a constant over  $t^{k/2}$ . This result has many important consequences in our study because it tells us that geometric limits are smooth limits. Maybe the first result to highlight is the fact (established earlier by Hamilton) that if  $(M, g(t))$  is a Ricci flow defined on  $0 \leq t < T < \infty$ , and if the Riemann curvature is uniformly bounded for the entire flow, then the Ricci flow extends past time  $T$ .

In the third chapter this material is reviewed and, where necessary, slight variants of results and arguments in the literature are presented.

**3.4. Generalized Ricci flows.** Because we cannot restrict our attention to Ricci flows, but rather must consider more general objects, Ricci flows with surgery, it is important to establish the basic analytic results and estimates in a context more general than that of Ricci flow. We choose to do this in the context of generalized Ricci flows.

A generalized 3-dimensional Ricci flow consists of a smooth manifold  $\mathcal{M}$  of dimension 4 (the space-time) together with a time function  $\mathbf{t}: \mathcal{M} \rightarrow \mathbb{R}$  and a smooth vector field  $\chi$ . These are required to satisfy:

- (1) Each  $x \in \mathcal{M}$  has a neighborhood of the form  $U \times J$ , where  $U$  is an open subset in  $\mathbb{R}^3$  and  $J \subset \mathbb{R}$  is an interval, in which  $\mathbf{t}$  is the projection onto  $J$  and  $\chi$  is the unit vector field tangent to the one-dimensional foliation  $\{u\} \times J$  pointing in the direction of increasing  $\mathbf{t}$ . We call  $\mathbf{t}^{-1}(t)$  the  $t$  time-slice. It is a smooth 3-manifold.
- (2) The image  $\mathbf{t}(\mathcal{M})$  is a connected interval  $I$  in  $\mathbb{R}$ , possibly infinite. The boundary of  $\mathcal{M}$  is the pre-image under  $\mathbf{t}$  of the boundary of  $I$ .
- (3) The level sets  $\mathbf{t}^{-1}(t)$  form a codimension-one foliation of  $\mathcal{M}$ , called the horizontal foliation, with the boundary components of  $\mathcal{M}$  being leaves.
- (4) There is a metric  $G$  on the horizontal distribution, i.e., the distribution tangent to the level sets of  $\mathbf{t}$ . This metric induces a Riemannian metric on each  $t$  time-slice, varying smoothly as we vary the time-slice. We define the curvature of  $G$  at a point  $x \in \mathcal{M}$  to be the curvature of the Riemannian metric induced by  $G$  on the time-slice  $M_t$  at  $x$ .

- (5) Because of the first property the integral curves of  $\chi$  preserve the horizontal foliation and hence the horizontal distribution. Thus, we can take the Lie derivative of  $G$  along  $\chi$ . The Ricci flow equation is then

$$\mathcal{L}_\chi(G) = -2\text{Ric}(G).$$

Locally in space-time the horizontal metric is simply a smoothly varying family of Riemannian metrics on a fixed smooth manifold and the evolution equation is the ordinary Ricci flow equation. This means that the usual formulas for the evolution of the curvatures as well as much of the analytic analysis of Ricci flows still hold in this generalized context. In the end, a Ricci flow with surgery is a more singular type of space-time, but it will have an open dense subset which is a generalized Ricci flow, and all the analytic estimates take place in this open subset.

The notion of canonical neighborhoods make sense in the context of generalized Ricci flows. There is also the notion of a strong  $\epsilon$ -neck. Consider an embedding  $\psi: (S^2 \times (-\epsilon^{-1}, \epsilon^{-1})) \times (-1, 0]$  into space-time such that the time function pulls back to the projection onto  $(-1, 0]$  and the vector field  $\chi$  pulls back to  $\partial/\partial t$ . If there is such an embedding into an appropriately shifted and rescaled version of the original generalized Ricci flow so that the pull-back of the rescaled horizontal metric is within  $\epsilon$  in the  $C^{[1/\epsilon]}$ -topology of the product of the shrinking family of round  $S^2$ 's with the Euclidean metric on  $(-\epsilon^{-1}, \epsilon^{-1})$ , then we say that  $\psi$  is a strong  $\epsilon$ -neck in the generalized Ricci flow.

**3.5. The maximum principle.** The Ricci flow equation satisfies various forms of the maximum principle. The fourth chapter explains this principle, which is due to Hamilton (see Section 4 of [34]), and derives many of its consequences, which are also due to Hamilton (cf. [36]). This principle and its consequences are at the core of all the detailed results about the nature of the flow. We illustrate the idea by considering the case of the scalar curvature. A standard scalar maximum principle argument applied to Equation (0.1) proves that the minimum of the scalar curvature is a non-decreasing function of time. In addition, it shows that if the minimum of scalar curvature at time 0 is positive then we have

$$R_{\min}(t) \geq R_{\min}(0) \left( \frac{1}{1 - \frac{2t}{n} R_{\min}(0)} \right),$$

and thus the equation develops a singularity at or before time  $n/(2R_{\min}(0))$ .

While the above result about the scalar curvature is important and is used repeatedly, the most significant uses of the maximum principle involve the tensor version, established by Hamilton, which applies for example to the Ricci tensor and the full curvature tensor. These have given the most

significant understanding of the Ricci flows, and they form the core of the arguments that Perelman uses in his application of Ricci flow to 3-dimensional topology. Here are the main results established by Hamilton:

- (1) For 3-dimensional flows, if the Ricci curvature is positive, then the family of metrics becomes singular at finite time and as the family becomes singular, the metric becomes closer and closer to round; see [29].
- (2) For 3-dimensional flows, as the scalar curvature goes to  $+\infty$  the ratio of the absolute value of any negative eigenvalue of the Riemann curvature to the largest positive eigenvalue goes to zero; see [36]. This condition is called *pinched toward positive curvature*.
- (3) Motivated by a Harnack inequality for the heat equation established by Li-Yau [48], Hamilton established a Harnack inequality for the curvature tensor under the Ricci flow for complete manifolds  $(M, g(t))$  with bounded, non-negative curvature operator; see [32]. In the applications to three dimensions, we shall need the following consequence for the scalar curvature: Suppose that  $(M, g(t))$  is a Ricci flow defined for all  $t \in [T_0, T_1]$  of complete manifolds of non-negative curvature operator with bounded curvature. Then

$$\frac{\partial R}{\partial t}(x, t) + \frac{R(x, t)}{t - T_0} \geq 0.$$

In particular, if  $(M, g(t))$  is an ancient solution (i.e., defined for all  $t \leq 0$ ) of bounded, non-negative curvature, then  $\partial R(x, t)/\partial t \geq 0$ .

- (4) If a complete 3-dimensional Ricci flow  $(M, g(t))$ ,  $0 \leq t \leq T$ , has non-negative curvature, if  $g(0)$  is not flat, and if there is at least one point  $(x, T)$  such that the Riemann curvature tensor of  $g(T)$  has a flat direction in  $\wedge^2 TM_x$ , then  $M$  has a cover  $\widetilde{M}$  so that for each  $t > 0$  the Riemannian manifold  $(\widetilde{M}, g(t))$  splits as a Riemannian product of a surface of positive curvature and a Euclidean line. Furthermore, the flow on the cover  $\widetilde{M}$  is the product of a 2-dimensional flow and the trivial one-dimensional Ricci flow on the line; see Sections 8 and 9 of [30].
- (5) In particular, there is no Ricci flow  $(U, g(t))$  with non-negative curvature tensor defined for  $0 \leq t \leq T$  with  $T > 0$ , such that  $(U, g(T))$  is isometric to an open subset in a non-flat, 3-dimensional metric cone.

**3.6. Geometric limits.** In the fifth chapter we discuss geometric limits of Riemannian manifolds and of Ricci flows. Let us review the history of these ideas. The first results about geometric limits of Riemannian manifolds go back to Cheeger in his thesis in 1967; see [6]. Here Cheeger obtained topological results. In [25] Gromov proposed that geometric limits should exist in the Lipschitz topology and suggested a result along these

lines, which also was known to Cheeger. In [23], Greene-Wu gave a rigorous proof of the compactness theorem suggested by Gromov and also enhanced the convergence to be  $C^{1,\alpha}$ -convergence by using harmonic coordinates; see also [56]. Assuming that all the derivatives of curvature are bounded, one can apply elliptic theory to the expression of curvature in harmonic coordinates and deduce  $C^\infty$ -convergence. These ideas lead to various types of compactness results that go under the name Cheeger-Gromov compactness for Riemannian manifolds. Hamilton in [33] extended these results to Ricci flows. We shall use the compactness results for both Riemannian manifolds and for Ricci flows. In a different direction, geometric limits were extended to the non-smooth context by Gromov in [25] where he introduced a weaker topology, called the Gromov-Hausdorff topology and proved a compactness theorem.

Recall that a sequence of based Riemannian manifolds  $(M_n, g_n, x_n)$  is said to *converge geometrically* to a based, complete Riemannian manifold  $(M_\infty, g_\infty, x_\infty)$  if there is a sequence of open subsets  $U_n \subset M_\infty$  with compact closures, with  $x_\infty \in U_1 \subset \overline{U}_1 \subset U_2 \subset \overline{U}_2 \subset U_3 \subset \dots$  with  $\cup_n U_n = M_\infty$ , and embeddings  $\varphi_n: U_n \rightarrow M_n$  sending  $x_\infty$  to  $x_n$  so that the pullback metrics,  $\varphi_n^* g_n$ , converge uniformly on compact subsets of  $M_\infty$  in the  $C^\infty$ -topology to  $g_\infty$ . Notice that the topological type of the limit can be different from the topological type of the manifolds in the sequence. There is a similar notion of geometric convergence for a sequence of based Ricci flows.

Certainly, one of the most important consequences of Shi's results, cited above, is that, in concert with Cheeger-Gromov compactness, it allows us to form smooth geometric limits of sequences of based Ricci flows. We have the following result of Hamilton's; see [33]:

**THEOREM 0.7.** *Let  $(M_n, g_n(t), (x_n, 0))$  be a sequence of based Ricci flows defined for  $t \in (-T, 0]$  with the  $(M_n, g_n(t))$  being complete. Suppose that:*

- (1) *There is  $r > 0$  and  $\kappa > 0$  such that for every  $n$  the metric ball  $B(x_n, 0, r) \subset (M_n, g_n(0))$  is  $\kappa$ -non-collapsed.*
- (2) *For each  $A < \infty$  there is  $C = C(A) < \infty$  such that the Riemann curvature on  $B(x_n, 0, A) \times (-T, 0]$  is bounded by  $C$ .*

*Then after passing to a subsequence there is a geometric limit which is a based Ricci flow  $(M_\infty, g_\infty(t), (x_\infty, 0))$  defined for  $t \in (-T, 0]$ .*

To emphasize, the two conditions that we must check in order to extract a geometric limit of a subsequence based at points at time zero are: (i) uniform non-collapsing at the base point in the time zero metric, and (ii) for each  $A < \infty$  uniformly bounded curvature for the restriction of the flow to the metric balls of radius  $A$  centered at the base points.

Most steps in Perelman's argument require invoking this result in order to form limits of appropriate sequences of Ricci flows, often rescaled to make the scalar curvatures at the base point equal to 1. If, before rescaling, the

scalar curvature at the base points goes to infinity as we move through the sequence, then the resulting limit of the rescaled flows has non-negative sectional curvature. This is a consequence of the fact that the sectional curvatures of the manifolds in the sequence are uniformly pinched toward positive. It is for exactly this reason that non-negative curvature plays such an important role in the study of singularity development in 3-dimensional Ricci flows.

#### 4. Perelman's advances

So far we have been discussing the results that were known before Perelman's work. They concern almost exclusively Ricci flow (though Hamilton in [35] had introduced the notion of surgery and proved that surgery can be performed preserving the condition that the curvature is pinched toward positive, as in (2) above). Perelman extended in two essential ways the analysis of Ricci flow – one involves the introduction of a new analytic functional, *the reduced length*, which is the tool by which he establishes the needed non-collapsing results, and the other is a delicate combination of geometric limit ideas and consequences of the maximum principle together with the non-collapsing results in order to establish bounded curvature at bounded distance results. These are used to prove in an inductive way the existence of canonical neighborhoods, which is a crucial ingredient in proving that it is possible to do surgery iteratively, creating a flow defined for all positive time.

While it is easiest to formulate and consider these techniques in the case of Ricci flow, in the end one needs them in the more general context of Ricci flow with surgery since we inductively repeat the surgery process, and in order to know at each step that we can perform surgery we need to apply these results to the previously constructed Ricci flow with surgery. We have chosen to present these new ideas only once – in the context of generalized Ricci flows – so that we can derive the needed consequences in all the relevant contexts from this one source.

**4.1. The reduced length function.** In Chapter 6 we come to the first of Perelman's major contributions. Let us first describe it in the context of an ordinary 3-dimensional Ricci flow, but viewing the Ricci flow as a horizontal metric on a space-time which is the manifold  $M \times I$ , where  $I$  is the interval of definition of the flow. Suppose that  $I = [0, T)$  and fix  $(x, t) \in M \times (0, T)$ . We consider paths  $\gamma(\tau)$ ,  $0 \leq \tau \leq \bar{\tau}$ , in space-time with the property that for every  $\tau \leq \bar{\tau}$  we have  $\gamma(\tau) \in M \times \{t - \tau\}$  and  $\gamma(0) = x$ . These paths are said to be *parameterized by backward time*. See FIG. 2. The  $\mathcal{L}$ -length of such a path is given by

$$\mathcal{L}(\gamma) = \int_0^{\bar{\tau}} \sqrt{\tau} (R(\gamma(\tau)) + |\gamma'(\tau)|^2) d\tau,$$

where the derivative on  $\gamma$  refers to the spatial derivative. There is also the closely related *reduced length*

$$\ell(\gamma) = \frac{\mathcal{L}(\gamma)}{2\sqrt{\bar{\tau}}}.$$

There is a theory for the functional  $\mathcal{L}$  analogous to the theory for the usual energy function<sup>6</sup>. In particular, there is the notion of an  $\mathcal{L}$ -geodesic, and the reduced length as a function on space-time  $\ell_{(x,t)}: M \times [0, t) \rightarrow \mathbb{R}$ . One establishes a crucial monotonicity for this reduced length along  $\mathcal{L}$ -geodesics. Then one defines the reduced volume

$$\tilde{V}_{(x,t)}(U \times \{\bar{t}\}) = \int_{U \times \{\bar{t}\}} \bar{\tau}^{-3/2} e^{-\ell_{(x,t)}(q, \bar{\tau})} d\text{vol}_g(q),$$

where, as before  $\bar{\tau} = t - \bar{t}$ . Because of the monotonicity of  $\ell_{(x,t)}$  along  $\mathcal{L}$ -geodesics, the reduced volume is also non-increasing under the flow (forward in  $\bar{\tau}$  and hence backward in time) of open subsets along  $\mathcal{L}$ -geodesics. This is the fundamental tool which is used to establish non-collapsing results which in turn are essential in proving the existence of geometric limits.

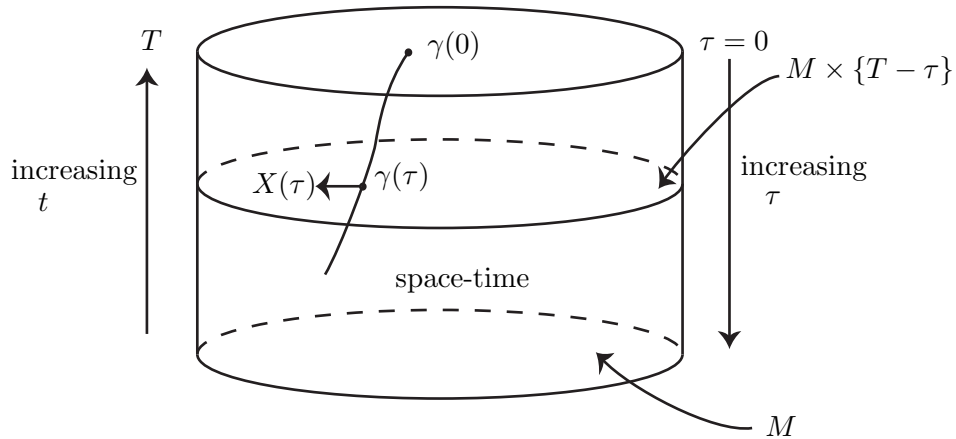


FIGURE 2. Curves in space-time parameterized by  $\tau$ .

The definitions and the analysis of the reduced length function and the reduced volume as well as the monotonicity results are valid in the context of the generalized Ricci flow. The only twist to be aware of is that in the more general context one cannot always extend  $\mathcal{L}$ -geodesics; they may run ‘off the edge’ of space-time. Thus, the reduced length function and reduced volume cannot be defined globally, but only on appropriate open subsets of a time-slice (those reachable by minimizing  $\mathcal{L}$ -geodesics). But as long as

<sup>6</sup>Even though this functional is called a length, the presence of the  $|\gamma'(\tau)|^2$  in the integrand means that it behaves more like the usual energy functional for paths in a Riemannian manifold.

one can flow an open set  $U$  of a time-slice along minimizing  $\mathcal{L}$ -geodesics in the direction of decreasing  $\bar{\tau}$ , the reduced volumes of the resulting family of open sets form a monotone non-increasing function of  $\bar{\tau}$ . This turns out to be sufficient to extend the non-collapsing results to Ricci flow with surgery, provided that we are careful in how we choose the parameters that go into the definition of the surgery process.

**4.2. Application to non-collapsing results.** As we indicated in the previous paragraph, one of the main applications of the reduced length function is to prove non-collapsing results for 3-dimensional Ricci flows with surgery. In order to make this argument work, one takes a weaker notion of  $\kappa$ -non-collapsed by making a stronger curvature bound assumption: one considers points  $(x, t)$  and constants  $r$  with the property that  $|\text{Rm}| \leq r^{-2}$  on  $P(x, t, r, -r^2) = B(x, t, r) \times (t - r^2, t]$ . The  $\kappa$ -non-collapsing condition applies to these balls and says that  $\text{Vol}(B(x, t, r)) \geq \kappa r^3$ . The basic idea in proving non-collapsing is to use the fact that, as we flow forward in time via minimizing  $\mathcal{L}$ -geodesics, the reduced volume is a non-decreasing function. Hence, a lower bound of the reduced volume of an open set at an earlier time implies the same lower bound for the corresponding open subset at a later time. This is contrasted with direct computations (related to the heat kernel in  $\mathbb{R}^3$ ) that say if the manifold is highly collapsed near  $(x, t)$  (i.e., satisfies the curvature bound above but is not  $\kappa$ -non-collapsed for some small  $\kappa$ ) then the reduced volume  $\tilde{V}_{(x,t)}$  is small at times close to  $t$ . Thus, to show that the manifold is non-collapsed at  $(x, t)$  we need only find an open subset at an earlier time that is reachable by minimizing  $\mathcal{L}$ -geodesics and that has a reduced volume bounded away from zero.

One case where it is easy to do this is when we have a Ricci flow of compact manifolds or of complete manifolds of non-negative curvature. Hence, these manifolds are non-collapsed at all points with a non-collapsing constant that depends only on the geometry of the initial metric of the Ricci flow. Non-collapsing results are crucial and are used repeatedly in dealing with Ricci flows with surgery in Chapters 10 – 17, for these give one of the two conditions required in order to take geometric limits.

**4.3. Application to ancient  $\kappa$ -non-collapsed solutions.** There is another important application of the length function, which is to the study of non-collapsed, ancient solutions in dimension 3. In the case that the generalized Ricci flow is an ordinary Ricci flow either on a compact manifold or on a complete manifold (with bounded curvatures) one can say much more about the reduced length function and the reduced volume. Fix a point  $(x_0, t_0)$  in space-time. First of all, one shows that every point  $(x, t)$  with  $t < t_0$  is reachable by a minimizing  $\mathcal{L}$ -geodesic and thus that the reduced length is defined as a function on all points of space at all times  $t < t_0$ . It turns out to be a locally Lipschitz function in both space and time and

hence its gradient and its time derivative exist as  $L^2$ -functions and satisfy important differential inequalities in the weak sense.

These results apply to a class of Ricci flows called  $\kappa$ -solutions, where  $\kappa$  is a positive constant. By definition a  $\kappa$ -solution is a Ricci flow defined for all  $t \in (-\infty, 0]$ , each time-slice is a non-flat, complete 3-manifold of non-negative, bounded curvature and each time-slice is  $\kappa$ -non-collapsed. The differential inequalities for the reduced length from any point  $(x, 0)$  imply that, for any  $t < 0$ , the minimum value of  $\ell_{(x,0)}(y, t)$  for all  $y \in M$  is at most  $3/2$ . Furthermore, again using the differential inequalities for the reduced length function, one shows that for any sequence  $t_n \rightarrow -\infty$ , and any points  $(y_n, t_n)$  at which the reduced length function is bounded above by  $3/2$ , there is a subsequence of based Riemannian manifolds,  $(M, \frac{1}{|t_n|}g(t_n), y_n)$ , with a geometric limit, and this limit is a gradient shrinking soliton. This gradient shrinking soliton is called an *asymptotic soliton* for the original  $\kappa$ -solution, see FIG. 3.

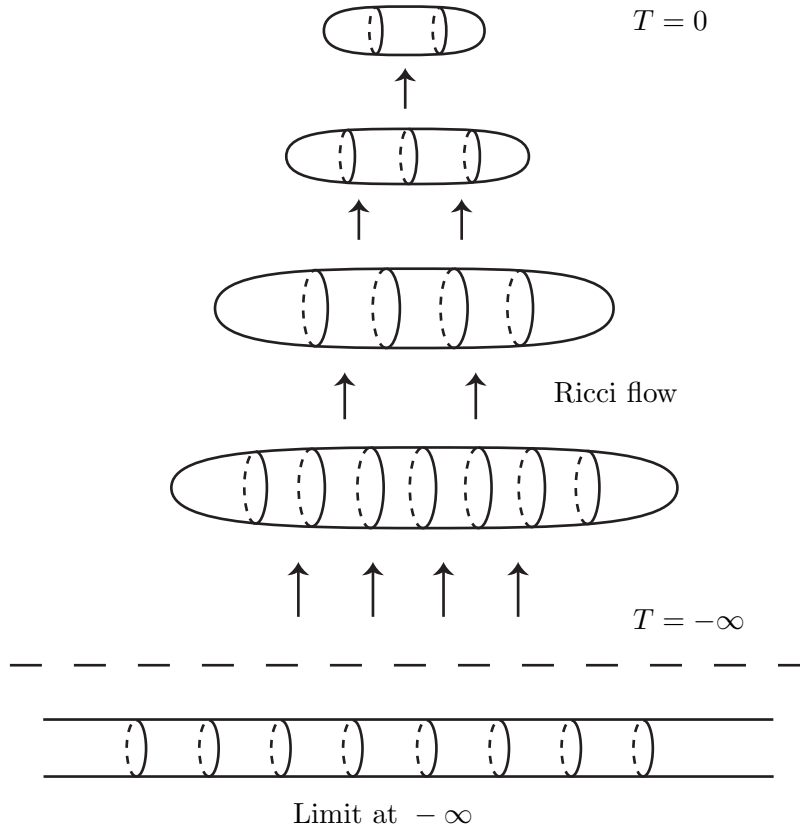


FIGURE 3. The asymptotic soliton.

The point is that there are only two types of 3-dimensional gradient shrinking solitons – (i) those finitely covered by a family of shrinking round

3-spheres and (ii) those finitely covered by a family of shrinking round cylinders  $S^2 \times \mathbb{R}$ . If a  $\kappa$ -solution has a gradient shrinking soliton of the first type then it is in fact isomorphic to its gradient shrinking soliton. More interesting is the case when the  $\kappa$ -solution has a gradient shrinking soliton which is of the second type. If the  $\kappa$ -solution does not have strictly positive curvature, then it is isomorphic to its gradient shrinking soliton. Furthermore, there is a constant  $C_1 < \infty$  depending on  $\epsilon$  (which remember is taken sufficiently small) such that a  $\kappa$ -solution of strictly positive curvature either is a  $C_1$ -component, or is a union of cores of  $(C_1, \epsilon)$ -caps and points that are the center points of  $\epsilon$ -necks.

In order to prove the above results (for example the uniformity of  $C_1$  as above over all  $\kappa$ -solutions) one needs the following result:

**THEOREM 0.8.** *The space of based  $\kappa$ -solutions, based at points with scalar curvature zero in the zero time-slice is compact.*

This result does not generalize to ancient solutions that are not non-collapsed because, in order to prove compactness, one has to take limits of subsequences, and in doing this the non-collapsing hypothesis is essential. See Hamilton's work [34] for more on general ancient solutions (i.e., those that are not necessarily non-collapsed).

Since  $\epsilon > 0$  is sufficiently small so that all the results from the appendix about manifolds covered by  $\epsilon$ -necks and  $\epsilon$ -caps hold, the above results about gradient shrinking solitons lead to a rough qualitative description of all  $\kappa$ -solutions. There are those which do not have strictly positive curvature. These are gradient shrinking solitons, either an evolving family of round 2-spheres times  $\mathbb{R}$  or the quotient of this family by an involution. Non-compact  $\kappa$ -solutions of strictly positive curvature are diffeomorphic to  $\mathbb{R}^3$  and are the union of an  $\epsilon$ -tube and a core of a  $(C_1, \epsilon)$ -cap. The compact ones of strictly positive curvature are of two types. The first type are positive, constant curvature shrinking solitons. Solutions of the second type are diffeomorphic to either  $S^3$  or  $\mathbb{R}P^3$ . Each time-slice of a  $\kappa$ -solution of the second type either is of uniformly bounded geometry (curvature, diameter, and volume) when rescaled so that the scalar curvature at a point is 1, or admits an  $\epsilon$ -tube whose complement is either a disjoint union of the cores of two  $(C_1, \epsilon)$ -caps.

This gives a rough qualitative understanding of  $\kappa$ -solutions. Either they are round, or they are finitely covered by the product of a round surface and a line, or they are a union of  $\epsilon$ -tubes and cores of  $(C_1, \epsilon)$ -caps, or they are diffeomorphic to  $S^3$  or  $\mathbb{R}P^3$  and have bounded geometry (again after rescaling so that there is a point of scalar curvature 1). This is the source of canonical neighborhoods for Ricci flows: the point is that this qualitative result remains true for any point  $x$  in a Ricci flow that has an appropriate size neighborhood within  $\epsilon$  in the  $C^{[1/\epsilon]}$ -topology of a neighborhood

in a  $\kappa$ -solution. For example, if we have a sequence of based generalized flows  $(\mathcal{M}_n, G_n, x_n)$  converging to a based  $\kappa$ -solution, then for all  $n$  sufficiently large  $x$  will have a canonical neighborhood, one that is either an  $\epsilon$ -neck centered at that point, a  $(C_1, \epsilon)$ -cap whose core contains the point, a  $C_1$ -component, or an  $\epsilon$ -round component.

**4.4. Bounded curvature at bounded distance.** Perelman's other major breakthrough is his result establishing bounded curvature at bounded distance for blow-up limits of generalized Ricci flows. As we have alluded to several times, many steps in the argument require taking (smooth) geometric limits of a sequence of based generalized flows about points of curvature tending to infinity. To study such a sequence we rescale each term in the sequence so that its curvature at the base point becomes 1. Nevertheless, in taking such limits we face the problem that even though the curvature at the point we are focusing on (the points we take as base points) was originally large and has been rescaled to be 1, there may be other points in the same time-slice of much larger curvature, which, even after the rescalings, can tend to infinity. If these points are at uniformly bounded (rescaled) distance from the base points, then they would preclude the existence of a smooth geometric limit of the based, rescaled flows. In his arguments, Hamilton avoided this problem by always focusing on points of maximal curvature (or almost maximal curvature). That method will not work in this case. The way to deal with this possible problem is to show that a generalized Ricci flow satisfying appropriate conditions satisfies the following. For each  $A < \infty$  there are constants  $Q_0 = Q_0(A) < \infty$  and  $Q(A) < \infty$  such that any point  $x$  in such a generalized flow for which the scalar curvature  $R(x) \geq Q_0$  and for any  $y$  in the same time-slice as  $x$  with  $d(x, y) < AR(x)^{-1/2}$  satisfies  $R(y)/R(x) < Q(A)$ . As we shall see, this and the non-collapsing result are the fundamental tools that allow Perelman to study neighborhoods of points of sufficiently large curvature by taking smooth limits of rescaled flows, so essential in studying the prolongation of Ricci flows with surgery.

The basic idea in proving this result is to assume the contrary and take an incomplete geometric limit of the rescaled flows based at the counterexample points. The existence of points at bounded distance with unbounded, rescaled curvature means that there is a point at infinity at finite distance from the base point where the curvature blows up. A neighborhood of this point at infinity is cone-like in a manifold of non-negative curvature. This contradicts Hamilton's maximum principle result (5) in Section 3.5) that the result of a Ricci flow of manifolds of non-negative curvature is never an open subset of a cone. (We know that any 'blow-up limit' like this has non-negative curvature because of the curvature pinching result.) This contradiction establishes the result.

## 5. The standard solution and the surgery process

Now we are ready to discuss 3-dimensional Ricci flows with surgery.

**5.1. The standard solution.** In preparing the way for defining the surgery process, we must construct a metric on the 3-ball that we shall glue in when we perform surgery. This we do in Chapter 12. We fix a non-negatively curved, rotationally symmetric metric on  $\mathbb{R}^3$  that is isometric near infinity to  $S^2 \times [0, \infty)$  where the metric on  $S^2$  is the round metric of scalar curvature 1, and outside this region has positive sectional curvature, see FIG. 4. Any such metric will suffice for the gluing process, and we fix one and call it the *standard metric*. It is important to understand Ricci flow with the standard metric as initial metric. Because of the special nature of this metric (the rotational symmetry and the asymptotic nature at infinity), it is fairly elementary to show that there is a unique solution of bounded curvature on each time-slice to the Ricci flow equation with the standard metric as the initial metric; this flow is defined for  $0 \leq t < 1$ ; and for any  $T < 1$  outside of a compact subset  $X(T)$  the restriction of the flow to  $[0, T]$  is close to the evolving round cylinder. Using the length function, one shows that the Ricci flow is non-collapsed, and that the bounded curvature and bounded distance result applies to it. This allows one to prove that every point  $(x, t)$  in this flow has one of the following types of neighborhoods:

- (1)  $(x, t)$  is contained in the core of a  $(C_2, \epsilon)$ -cap, where  $C_2 < \infty$  is a given universal constant depending only on  $\epsilon$ .
- (2)  $(x, t)$  is the center of a strong  $\epsilon$ -neck.
- (3)  $(x, t)$  is the center of an evolving  $\epsilon$ -neck whose initial slice is at time zero.

These form the second source of models for canonical neighborhoods in a Ricci flow with surgery. Thus, we shall set  $C = C(\epsilon) = \max(C_1(\epsilon), C_2(\epsilon))$  and we shall find  $(C, \epsilon)$ -canonical neighborhoods in Ricci flows with surgery.

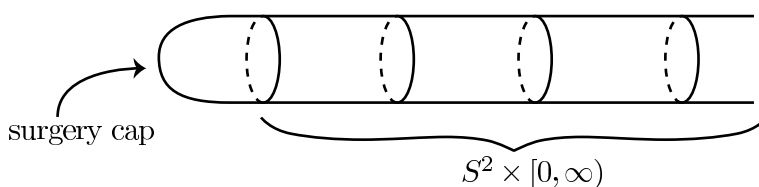


FIGURE 4. The standard metric.

**5.2. Ricci flows with surgery.** Now it is time to introduce the notion of a Ricci flow with surgery. To do this we formulate an appropriate notion of 4-dimensional space-time that allows for the surgery operations. We define *space-time* to be a 4-dimensional Hausdorff singular space with a time function  $\mathbf{t}$  with the property that each time-slice is a compact, smooth

3-manifold, but level sets at different times are not necessarily diffeomorphic. Generically space-time is a smooth 4-manifold, but there are *exposed regions* at a discrete set of times. Near a point in the exposed region space-time is a 4-manifold with boundary. The singular points of space-time are the boundaries of the exposed regions. Near these, space-time is modeled on the product of  $\mathbb{R}^2$  with the square  $(-1, 1) \times (-1, 1)$ , the latter having a topology in which the open sets are, in addition to the usual open sets, open subsets of  $(0, 1) \times [0, 1)$ , see FIG. 5. There is a natural notion of smooth functions on space-time. These are smooth in the usual sense on the open subset of non-singular points. Near the singular points, and in the local coordinates described above, they are required to be pull-backs from smooth functions on  $\mathbb{R}^2 \times (-1, 1) \times (-1, 1)$  under the natural map. Space-time is equipped with a smooth vector field  $\chi$  with  $\chi(\mathbf{t}) = 1$ .

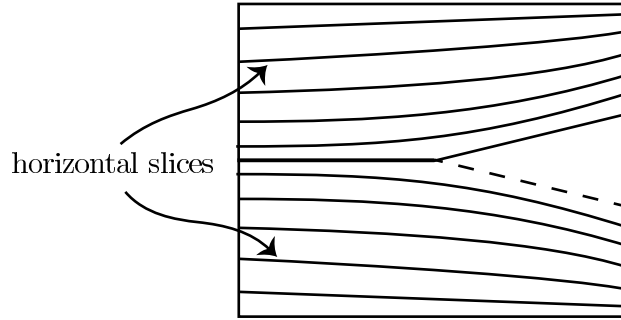


FIGURE 5. Model for singularities in space-time.

A Ricci flow with surgery is a smooth horizontal metric  $G$  on a space-time with the property that the restriction of  $G$ ,  $\mathbf{t}$  and  $\chi$  to the open subset of smooth points forms a generalized Ricci flow. We call this the *associated generalized Ricci flow* for the Ricci flow with surgery.

**5.3. The inductive conditions necessary for doing surgery.** With all this preliminary work out of the way, we are ready to show that one can construct Ricci flow with surgery which is precisely controlled both topologically and metrically. This result is proved inductively, one interval of time after another, and it is important to keep track of various properties as we go along to ensure that we can continue to do surgery. Here we discuss the conditions we verify at each step.

Fix  $\epsilon > 0$  sufficiently small and let  $C = \max(C_1, C_2) < \infty$ , where  $C_1$  is the constant associated to  $\epsilon$  for  $\kappa$ -solutions and  $C_2$  is the constant associated to  $\epsilon$  for the standard solution. We say that a point  $x$  in a generalized Ricci flow has a  $(C, \epsilon)$ -canonical neighborhood if one of the following holds:

- (1)  $x$  is contained in a  $C$ -component of its time-slice.

- (2)  $x$  is contained in a connected component of its time-slice that is within  $\epsilon$  of round in the  $C^{[1/\epsilon]}$ -topology.
- (3)  $x$  is contained in the core of a  $(C, \epsilon)$ -cap.
- (4)  $x$  is the center of a strong  $\epsilon$ -neck.

We shall study Ricci flows with surgery defined for  $0 \leq t < T < \infty$  whose associated generalized Ricci flows satisfy the following properties:

- (1) The initial metric is normalized, meaning that for the metric at time zero the norm of the Riemann curvature is bounded above by 1 and the volume of any ball of radius 1 is at least half the volume of the unit ball in Euclidean space.
- (2) The curvature of the flow is pinched toward positive.
- (3) There is  $\kappa > 0$  so that the associated generalized Ricci flow is  $\kappa$ -non-collapsed on scales at most  $\epsilon$ , in the sense that we require only that balls of radius  $r \leq \epsilon$  be  $\kappa$ -non-collapsed.
- (4) There is  $r_0 > 0$  such that any point of space-time at which the scalar curvature is  $\geq r_0^{-2}$  has a  $(C, \epsilon)$ -canonical neighborhood.

The main result is that, having a Ricci flow with surgery defined on some time interval satisfying these conditions, it is possible to extend it to a longer time interval in such a way that it still satisfies the same conditions, possibly allowing the constants  $\kappa$  and  $r_0$  defining these conditions to get closer to zero, but keeping them bounded away from 0 on each compact time interval. We repeat this construction inductively. It is easy to see that there is a bound on the number of surgeries in each compact time interval. Thus, in the end, we create a Ricci flow with surgery defined for all positive time. As far as we know, it may be the case that in the entire flow, defined for all positive time, there are infinitely many surgeries.

**5.4. Surgery.** Let us describe how we extend a Ricci flow with surgery satisfying all the conditions listed above and becoming singular at time  $T < \infty$ . Fix  $T^- < T$  so that there are no surgery times in the interval  $[T^-, T)$ . Then we can use the Ricci flow to identify all the time-slices  $M_t$  for  $t \in [T^-, T)$ , and hence view this part of the Ricci flow with surgery as an ordinary Ricci flow. Because of the canonical neighborhood assumption, there is an open subset  $\Omega \subset M_{T^-}$  on which the curvature stays bounded as  $t \rightarrow T$ . Hence, by Shi's results, there is a limiting metric at time  $T$  on  $\Omega$ . Furthermore, the scalar curvature is a proper function, bounded below, from  $\Omega$  to  $\mathbb{R}$ , and each end of  $\Omega$  is an  $\epsilon$ -tube where the cross-sectional area of the 2-spheres goes to zero as we go to the end of the tube. We call such tubes  $\epsilon$ -horns. We are interested in  $\epsilon$ -horns whose boundary is contained in the part of  $\Omega$  where the scalar curvature is bounded above by some fixed finite constant  $\rho^{-2}$ . We call this region  $\Omega_\rho$ . Using the bounded curvature at bounded distance result, and using the non-collapsing hypothesis, one shows that given any  $\delta > 0$  there is  $h = h(\delta, \rho, r_0)$  such that for any  $\epsilon$ -horn

$\mathcal{H}$  whose boundary lies in  $\Omega_\rho$  and for any  $x \in \mathcal{H}$  with  $R(x) \geq h^{-2}$ , the point  $x$  is the center of a strong  $\delta$ -neck.

Now we are ready to describe the surgery procedure. It depends on our choice of standard solution on  $\mathbb{R}^3$  and on a choice of  $\delta > 0$  sufficiently small. For each  $\epsilon$ -horn in  $\Omega$  whose boundary is contained in  $\Omega_\rho$ , fix a point of curvature  $(h(\delta, \rho, r_0))^{-2}$  and fix a strong  $\delta$ -neck centered at this point. Then we cut the  $\epsilon$ -horn open along the central 2-sphere  $S$  of this neck and remove the end of the  $\epsilon$ -horn that is cut off by  $S$ . Then we glue in a ball of a fixed radius around the tip from the standard solution, after scaling the metric on this ball by  $(h(\delta, \rho, r_0))^2$ . To glue these two metrics together we must use a partition of unity near the 2-spheres that are matched. There is also a delicate point that we first bend in the metrics slightly so as to achieve positive curvature near where we are gluing. This is an idea due to Hamilton, and it is needed in order to show that the condition of curvature pinching toward positive is preserved. In addition, we remove all components of  $\Omega$  that do not contain any points of  $\Omega_\rho$ .

This operation produces a new compact 3-manifold. One continues the Ricci flow with surgery by letting this Riemannian manifold at time  $T$  evolve under the Ricci flow. See FIG. 6.

**5.5. Topological effect of surgery.** Looking at the situation just before the surgery time, we see a finite number of disjoint submanifolds, each diffeomorphic to either  $S^2 \times I$  or the 3-ball, where the curvature is large. In addition there may be entire components of where the scalar curvature is large. The effect of 2-sphere surgery is to do a finite number of ordinary topological surgeries along 2-spheres in the  $S^2 \times I$ . This simply effects a partial connected-sum decomposition and may introduce new components diffeomorphic to  $S^3$ . We also remove entire components, but these are covered by  $\epsilon$ -necks and  $\epsilon$ -caps so that they have standard topology (each one is diffeomorphic to  $S^3$ ,  $\mathbb{R}P^3$ ,  $\mathbb{R}P^3 \# \mathbb{R}P^3$ ,  $S^2 \times S^1$ , or the non-orientable 2-sphere bundle over  $S^1$ ). Also, we remove  $C$ -components and  $\epsilon$ -round components (each of these is either diffeomorphic to  $S^3$  or  $\mathbb{R}P^3$  or admits a metric of constant positive curvature). Thus, the topological effect of surgery is to do a finite number of ordinary 2-sphere topological surgeries and to remove a finite number of topologically standard components.

## 6. Extending Ricci flows with surgery

We consider Ricci flows with surgery that are defined on the time interval  $0 \leq t < T$ , with  $T < \infty$ , and that satisfy four conditions. These conditions are: (i) normalized initial metric, (ii) curvature pinched toward positive, (iii) all points of scalar curvature  $\geq r^{-2}$  have canonical neighborhoods, and (iv) the flow is  $\kappa$ -non-collapsed on scales  $\leq \epsilon$ . The crux of the argument is to show that it is possible to extend to such a Ricci flow with surgery to a Ricci flow with surgery defined for all  $t \in [0, T')$  for some  $T' > T$ , keeping

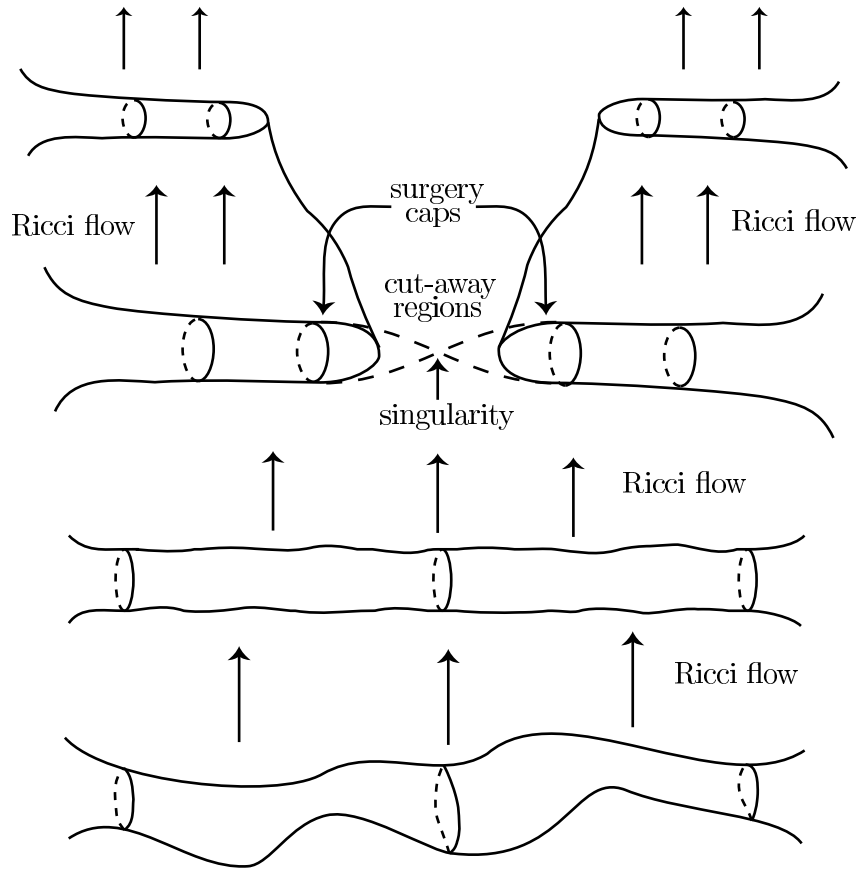


FIGURE 6. Surgery.

these four conditions satisfied (possibly with different constants  $r' < r$  and  $\kappa' < \kappa$ ). In order to do this we need to choose the surgery parameter  $\delta > 0$  sufficiently small. There is also the issue of whether the surgery times can accumulate.

Of course, the initial metric does not change as we extend surgery so that the condition that the normalized initial metric is clearly preserved as we extend surgery. As we have already remarked, Hamilton had proved earlier that one can do surgery in such a way as to preserve the condition that the curvature is pinched toward positive. The other two conditions require more work, and, as we indicated above, the constants may decay to zero as we extend the Ricci flow with surgery.

If we have all the conditions for the Ricci flow with surgery up to time  $T$ , then the analysis of the open subset on which the curvature remains bounded holds, and given  $\delta > 0$  sufficiently small, we do surgery on the central  $S^2$  of a strong  $\delta$ -neck in each  $\epsilon$ -horn meeting  $\Omega_\rho$ . In addition we remove entirely all components that do not contain points of  $\Omega_\rho$ . We then

glue in the cap from the standard solution. This gives us a new compact 3-manifold and we restart the flow from this manifold.

The  $\kappa$ -non-collapsed result is extended to the new part of the Ricci flow with surgery using the fact that it holds at times previous to  $T$ . To establish this extension one uses  $\mathcal{L}$ -geodesics in the associated generalized Ricci flow and reduced volume as indicated before. In order to get this argument to work, one must require  $\delta > 0$  to be sufficiently small; how small is determined by  $r_0$ .

The other thing that we must establish is the existence of canonical neighborhoods for all points of sufficiently large scalar curvature. Here the argument is by contradiction. We consider all Ricci flows with surgery that satisfy all four conditions on  $[0, T)$  and we suppose that we can find a sequence of such containing points (automatically at times  $T' > T$ ) of arbitrarily large curvature where there are not canonical neighborhoods. In fact, we take the points at the first such time violating this condition. We base our flows at these points. Now we consider rescaled versions of the generalized flows so that the curvature at these base points is rescaled to 1. We are in a position to apply the bounded curvature and bounded distance results to this sequence, and of course the  $\kappa$ -non-collapsing results which have already been established. There are two possibilities. The first is that the rescaled sequence converges to an ancient solution. This ancient solution has non-negative curvature by the pinching hypothesis. General results about 3-manifolds of non-negative curvature imply that it also has bounded curvature. It is  $\kappa$ -non-collapsed. Thus, in this case the limit is a  $\kappa$ -solution. This produces the required canonical neighborhoods for the base points of the tail of the sequence modeled on the canonical neighborhoods of points in a  $\kappa$ -solution. This contradicts the assumption that none of these points has a canonical neighborhood.

The other possibility is that one can take a partial smooth limit but that this limit does not extend all the way back to  $-\infty$ . The only way this can happen is if there are surgery caps that prevent extending the limit back to  $-\infty$ . This means that the base points in our sequence are all within a fixed distance and time (after the rescaling) of a surgery region. But in this case results from the nature of the standard solution show that if we have taken  $\delta > 0$  sufficiently small, then the base points have canonical neighborhoods modeled on the canonical neighborhoods in the standard solution, again contradicting our assumption that none of the base points has a canonical neighborhood. In order to show that our base points have neighborhoods near those of the standard solution, one appeals to a geometric limit argument as  $\delta \rightarrow 0$ . This argument uses the uniqueness of the Ricci flow for the standard solution. (Actually, Bruce Kleiner pointed out to us that one only needs a compactness result for the space of all Ricci flows with the standard metric as initial metric, not uniqueness, and the compactness result can be

proved by the same arguments that prove the compactness of the space of  $\kappa$ -solutions.)

Interestingly enough, in order to establish the uniqueness of the Ricci flow for the standard solution, as well as to prove that this flow is defined for time  $[0, 1)$  and to prove that at infinity it is asymptotic to an evolving cylinder, requires the same results – non-collapsing and the bounded curvature at bounded distance that we invoked above. For this reason, we order the material described here as follows. First, we introduce generalized Ricci flows, and then introduce the length function in this context and establish the basic monotonicity results. Then we have a chapter on stronger results for the length function in the case of complete manifolds with bounded curvature. At this point we are in a position to prove the needed results about the Ricci flow from the standard solution. Then we are ready to define the surgery process and prove the inductive non-collapsing results and the existence of canonical neighborhoods.

In this way, one establishes the existence of canonical neighborhoods. Hence, one can continue to do surgery, producing a Ricci flow with surgery defined for all positive time. Since these arguments are inductive, it turns out that the constants in the non-collapsing and in the canonical neighborhood statements decay in a predetermined rate as time goes to infinity.

Lastly, there is the issue of ruling out the possibility that the surgery times accumulate. The idea here is very simple: Under Ricci flow during an elapsed time  $T$ , volume increases at most by a multiplicative factor which is a fixed exponential of the time  $T$ . Under each surgery there is a removal of at least a fixed positive amount of volume depending on the surgery scale  $h$ , which in turns depends on  $\delta$  and  $r_0$ . Since both  $\delta$  and  $r_0$  are bounded away from zero on each finite interval, there can be at most finitely many surgeries in each finite interval. This argument allows for the possibility, noted in Section 5.3, that in the entire flow all the way to infinity there are infinitely many surgeries. It is still unknown whether that possibility ever happens.

This completes our outline of the proof of Theorem 0.3.

## 7. Finite-time extinction

The last topic we discuss is the proof of the finite-time extinction for Ricci flows with initial metrics satisfying the hypothesis of Theorem 0.4.

As we present it, the finite extinction result has two steps. The first step is to show that there is  $T < \infty$  (depending on the initial metric) such that for all  $t \geq T$ , all connected components of the  $t$ -time-slice  $M_t$  have trivial  $\pi_2$ . First, an easy topological argument shows that only finitely many of the 2-sphere surgeries in a Ricci flow with surgery can be along homotopically non-trivial 2-spheres. Thus, after some time  $T_0$  all 2-sphere surgeries are along homotopically trivial 2-spheres. Such a surgery does not affect  $\pi_2$ .

Thus, after time  $T_0$ , the only way that  $\pi_2$  can change is by removal of components with non-trivial  $\pi_2$ . (An examination of the topological types of components that are removed shows that there are only two types of such components with non-trivial  $\pi_2$ : 2-sphere bundles over  $S^1$  and  $\mathbb{R}P^3 \# \mathbb{R}P^3$ .) We suppose that at every  $t \geq T_0$  there is a component of  $M_t$  with non-trivial  $\pi_2$ . Then we can find a connected open subset  $\mathcal{X}$  of  $\mathfrak{t}^{-1}([T_0, \infty))$  with the property that for each  $t \geq T_0$  the intersection  $\mathcal{X}(t) = \mathcal{X} \cap M_t$  is a component of  $M_t$  with non-trivial  $\pi_2$ . We define a function  $W_2: [T_0, \infty) \rightarrow \mathbb{R}$  associated with such an  $\mathcal{X}$ . The value  $W_2(t)$  is the minimal area of all homotopically non-trivial 2-spheres mapping into  $\mathcal{X}(t)$ . This minimal area  $W_2(t)$  is realized by a harmonic map of  $S^2$  into  $\mathcal{X}(t)$ . The function  $W_2$  varies continuously under Ricci flow and at a surgery is lower semi-continuous. Furthermore, using an idea that goes back to Hamilton (who applied it to minimal disks) one shows that the forward difference quotient of the minimal area satisfies

$$\frac{dW_2(t)}{dt} \leq -4\pi + \frac{3}{(4t+1)}W_2(t).$$

(Here, the explicit form of the bound for the forward difference quotient depends on the way we have chosen to normalize initial metric and also on Hamilton's curvature pinching result.)

But any function  $W_2(t)$  with these properties and defined for all  $t > T_0$ , becomes negative at some finite  $T_1$  (depending on the initial value). This is absurd since  $W_2(t)$  is the minimum of positive quantities. This contradiction shows that such a path of components with non-trivial  $\pi_2$  cannot exist for all  $t \geq T_0$ . In fact, it even gives a computable upper bound on how long such a component  $\mathcal{X}$ , with every time-slice having non-trivial  $\pi_2$ , can exist in terms of the minimal area of a homotopically non-trivial 2-sphere mapping into  $\mathcal{X}(T_0)$ . It follows that there is  $T < \infty$  with the property that every component of  $M_T$  has trivial  $\pi_2$ . This condition then persists for all  $t \geq T$ .

Three remarks are in order. This argument showing that eventually every component of the time-slice  $t$  has trivial  $\pi_2$  is not necessary for the topological application (Theorem 0.4), or indeed, for any other topological application. The reason is the sphere theorem (see [39]), which says that if  $\pi_2(M)$  is non-trivial then either  $M$  is diffeomorphic to an  $S^2$  bundle over  $S^1$  or  $M$  has a non-trivial connected sum decomposition. Thus, we can establish results for all 3-manifolds if we can establish them for 3-manifolds with  $\pi_2 = 0$ . Secondly, the reason for giving this argument is that it is pleasing to see Ricci flow with surgery implementing the connected sum decomposition required for geometrization of 3-manifolds. Also, this argument is a simpler version of the one that we use to deal with components with non-trivial  $\pi_3$ . Lastly, these results on Ricci flow do not use the sphere theorem so that establishing the cutting into pieces with trivial  $\pi_2$  allows us to give

a different proof of this result (though admittedly one using much deeper ideas).

Let us now fix  $T < \infty$  such that for all  $t \geq T$  all the time-slices  $M_t$  have trivial  $\pi_2$ . There is a simple topological consequence of this and our assumption on the initial manifold. If  $M$  is a compact 3-manifold whose fundamental group is either a non-trivial free product or an infinite cyclic group, then  $M$  admits a homotopically non-trivial embedded 2-sphere. Since we began with a manifold  $M_0$  whose fundamental group is a free product of finite groups and infinite cyclic groups, it follows that for  $t \geq T$  every component of  $M_t$  has finite fundamental group. Fix  $t \geq T$ . Then each component of  $M_t$  has a finite cover that is simply connected, and thus, by an elementary argument in algebraic topology, each component of  $M_t$  has non-trivial  $\pi_3$ . The second step in the finite-time extinction argument is to use a non-trivial element in this group analogously to the way we used homotopically non-trivial 2-spheres to show that eventually the manifolds have trivial  $\pi_2$ .

There are two approaches to this second step: the first is due to Perelman in [54] and the other due to Colding-Minicozzi in [15]. In their approach Colding-Minicozzi associate to a non-trivial element in  $\pi_3(M)$  a non-trivial element in  $\pi_1(\text{Maps}(S^2, M))$ . This element is represented by a one-parameter family of 2-spheres (starting and ending at the constant map) representing a non-trivial element  $\xi \in \pi_3(M_0)$ . They define the *width* of this homotopy class by  $W(\xi, t)$  by associating to each representative the maximal energy of the 2-spheres in the family and then minimizing over all representatives of the homotopy class. Using results of Jost [42], they show that this function satisfies the same forward difference inequality that  $W_2$  satisfies (and has the same continuity property under Ricci flow and the same semi-continuity under surgery). Since  $W(\xi, t)$  is always  $\geq 0$  if it is defined, this forward difference quotient inequality implies that the manifolds  $M_t$  must eventually become empty.

While this approach seemed completely natural to us, and while we believe that it works, we found the technical details daunting<sup>7</sup> (because one is forced to consider critical points of index 1 of the energy functional rather than minima). For this reason we chose to follow Perelman's approach. He represents a non-trivial element in  $\pi_3(M)$  as a non-trivial element in  $\xi \in \pi_2(\Lambda M, *)$  where  $\Lambda M$  is the free loop space of  $M$ . He then associates to a family  $\Gamma: S^2 \rightarrow \Lambda M$  of homotopically trivial loops an invariant  $W(\Gamma)$  which is the maximum of the areas of minimal spanning disks for the loops  $\Gamma(c)$  as  $c$  ranges over  $S^2$ . The invariant of a non-trivial homotopy class  $\xi$  is then the infimum over all representatives  $\Gamma$  for  $\xi$  of  $W(\Gamma)$ . As before, this function is continuous under Ricci flow and is lower semi-continuous under

---

<sup>7</sup>Colding and Minicozzi tell us they plan to give an expanded version of their argument with a more detailed proof.

surgery (unless the surgery removes the component in question). It also satisfies a forward difference quotient

$$\frac{dW(\xi)}{dt} \leq -2\pi + \frac{3}{4t+1}W(\xi).$$

The reason for the term  $-2\pi$  instead of  $-4\pi$  which occurs in the other cases is that we are working with minimal 2-disks instead of minimal 2-spheres. Once this forward difference quotient estimate (and the continuity) have been established the argument finishes in the same way as the other cases: a function  $W$  with the properties we have just established cannot be non-negative for all positive time. This means the component in question, and indeed all components at later time derived from it, must disappear in finite time. Hence, under the hypothesis on the fundamental group in Theorem 0.4 the entire manifold must disappear at finite time.

Because this approach uses only minima for the energy or area functional, one does not have to deal with higher index critical points. But one is forced to face other difficulties though – namely boundary issues. Here, one must prescribe the deformation of the family of boundary curves before computing the forward difference quotient of the energy. The obvious choice is the curve-shrinking flow (see [2]). Unfortunately, this flow can only be defined when the curve in question is immersed and even in this case the curve-shrinking flow can develop singularities even if the Ricci flow does not. Following Perelman, or indeed [2], one uses the device of taking the product with a small circle and using loops, called *ramps*, that go around that circle once. In this context the curve-shrinking flow remains regular as long as the Ricci flow does. One then projects this flow to a flow of families of 2-spheres in the free loop space of the time-slices of the original Ricci flow. Taking the length of the circle sufficiently small yields the boundary deformation needed to establish the forward difference quotient result. This requires a compactness result which holds under local total curvature bounds. This compactness result holds outside a subset of time-interval of small total measure, which is sufficient for the argument. At the very end of the argument we need an elementary but complicated result on annuli, which we could not find in the literature. For more details on these points see Chapter 18.

## 8. Acknowledgements

In sorting out Perelman’s arguments we have had the aid of many people. First of all, during his tour of the United States in the Spring of 2003, Perelman spent enormous amounts of time explaining his ideas and techniques both in public lectures and in private sessions to us and to many others. Since his return to Russia, he has also freely responded to our questions by e-mail. Richard Hamilton has also given unstintingly of his time, explaining to us his results and his ideas about Perelman’s results. We have benefitted

tremendously from the work of Bruce Kleiner and John Lott. They produced a lengthy set of notes filling in the details in Perelman's arguments [45]. We have referred to those notes countless times as we came to grips with Perelman's ideas. In late August and early September of 2004, Kleiner, Lott and the two of us participated in a workshop at Princeton University, supported by the Clay Mathematics Institute, going through Perelman's second paper (the one on Ricci flow with surgery) in detail. This workshop played a significant role in convincing us that Perelman's arguments were complete and correct and also in convincing us to write this book. We thank all the participants of this workshop and especially Guo-Feng Wei, Peng Lu, Yu Ding, and X.-C. Rong who, together with Kleiner and Lott, made significant contributions to the workshop. Before, during, and after this workshop, we have benefitted from private conversations too numerous to count with Bruce Kleiner and John Lott.

Several of the analytic points were worked out in detail by others, and we have freely adapted their work. Rugang Ye wrote a set of notes proving the Lipschitz properties of the length function and proving the fact that the cut locus for the reduced length function is of measure zero. Closely related to this, Ye gave detailed arguments proving the requisite limiting results for the length function required to establish the existence of a gradient shrinking soliton. None of these points was directly addressed by Perelman, and it seemed to us that they needed careful explanation. Also, the proof of the uniqueness of the standard solution was established jointly by the second author and Peng Lu, and the proof of the refined version of Shi's theorem where one has control on a certain number of derivatives at time zero was also shown to us by Peng Lu. We also benefitted from Ben Chow's expertise and vast knowledge of the theory of Ricci flow especially during the ClayMath/MSRI Summer School on Ricci Flow in 2005. We had several very helpful conversations with Tom Mrowka and also with Robert Bryant, especially about annuli of small area.

The second author gave courses at Princeton University and ran seminars on this material at MIT and Princeton. Natasa Sesum and Xiao-Dong Wang, see [64], wrote notes for the seminars at MIT, and Edward Fan and Alex Subotic took notes for the seminars and courses at Princeton, and they produced preliminary manuscripts. We have borrowed freely from these manuscripts, and it is a pleasure to thank each of them for their efforts in the early stages of this project.

The authors thank all the referees for their time and effort which helped us to improve the presentation. In particular, we wish to thank Cliff Taubes who carefully read the entire manuscript and gave us innumerable comments and suggestions for clarifications and improvements, and Arthur Greenspoon who carefully copy-edited the book. We also thank Terry Tao for much

helpful feedback which improved the exposition. We also thank Colin Rourke for his comments on the finite-time extinction argument.

We thank Lori Lejeune for drawing the figures and John Etnyre for technical help in inserting these figures into the manuscript.

Lastly, during this work we were both generously supported by the Clay Mathematical Institute, and it is a pleasure to thank the Clay Mathematics Institute for its support and to thank its staff, especially Vida Salahi, for their help in preparing this manuscript. We also thank the National Science Foundation for its continued support and the second author thanks the Jim Simons Foundation for its support during the period that he was a faculty member at MIT.

### 9. List of related papers

For the readers' convenience we gather here references to all the closely related articles.

First and foremost are Perelman's three preprints, [53], [55], and [54]. The first of these introduces the main techniques in the case of Ricci flow, the second discusses the extension of these techniques to Ricci flow with surgery, and the last gives the short-cut to the Poincaré Conjecture and the 3-dimensional spherical space-form conjecture, avoiding the study of the limits as time goes to infinity and collapsing space arguments. There are the detailed notes by Bruce Kleiner and John Lott, [45], which greatly expand and clarify Perelman's arguments from the first two preprints. There is also a note on Perelman's second paper by Yu Ding [17]. There is the article by Colding-Minicozzi [15], which gives their alternate approach to the material in Perelman's third preprint. Collapsing space arguments needed for the full geometrization conjecture are discussed in Shioya-Yamaguchi [67]. Lastly, after we had submitted a preliminary version of this manuscript for refereeing, H.-D. Cao and X.-P. Zhu published an article on the Poincaré Conjecture and Thurston's Geometrization Conjecture; see [5].