Mean curvature flow

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1090/S0273-0979-2015-01468-0">http://dx.doi.org/10.1090/S0273-0979-2015-01468-0</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Mathematical Society (AMS)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sat Feb 16 03:27:54 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/110410">http://hdl.handle.net/1721.1/110410</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
</tbody>
</table>

Detailed Terms

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
MEAN CURVATURE FLOW

TOBIAS HOLCK COLDING, WILLIAM P. MINICOZZI II, AND ERIK KJÆR PÆDERSEN

Abstract. Mean curvature flow is the negative gradient flow of area, so any hypersurface flows through hypersurfaces in the direction of steepest descent for area and eventually becomes extinct in finite time. Before it becomes extinct, topological changes can occur as it goes through singularities. If the hypersurface is in general or generic position, then we explain what singularities can occur under the flow, what the flow looks like near these singularities, and what this implies for the structure of the singular set. At the end, we will briefly discuss how one may be able to use the flow in low-dimensional topology.

0. Introduction

Imagine that a closed surface in $\mathbb{R}^3$ flows in time to decrease its area as rapidly as possible. Convex points will move inward, while concave points move outward, the speed is slower where the surface is flatter. Independently of whether points move inward or outward, the total area will decrease along the flow and eventually go to zero in finite time. In particular, any closed surface becomes extinct in finite time and, thus, the flow can only be continued smoothly for some finite amount of time before singularities occur.

Mean curvature flow (MCF) is the negative gradient flow for area. This is a nonlinear partial differential equation for the evolving hypersurface that is formally similar to the ordinary heat equation, with some important differences. MCF behaves like the heat equation for a short time with the solution becoming smoother and small-scale variations averaging out. However, after more time, the nonlinearities dominate and the solution becomes singular. To understand the flow, one must understand the singularities it goes through.

MCF has been studied in material science for almost a century to model things such as cell, grain, and bubble growth. In the 1950s, von Neumann studied soap foams whose interface tend to have constant mean curvature, whereas Mullins describes coarsening in metals, in which interfaces are not generally of constant mean curvature. Partly as a consequence, Mullins may have been the first to write down the MCF equation in general. Mullins also found some of the basic self-similar solutions, such as the translating solution now known as the Grim Reaper. MCF and related flows have also been used to model various other physical phenomena as well as being used in image processing.
This paper surveys MCF of hypersurfaces in all dimensions, starting with some of the classical results and then moving to very recent results. Under MCF, any closed hypersurface passes through singularities as it becomes extinct in finite time. There are infinitely many types of singularities that can occur. One of the main themes is that only certain very simple singularities cannot be perturbed away and that these are the most important. We will explain the classification of these “generic” singularities in all dimensions, the structure of the flow near these singularities, and the resulting structure of the singular set itself. Along the way, we will also mention many open problems.

The last few sections are spent in popularizing a result, well known to older surgeons, that connects 4-manifold topology with hypersurfaces in $\mathbb{R}^5$, suggesting the possible use of MCF. Namely, that any closed smooth 4-dimensional manifold homotopy equivalent to $S^4$ can be smoothly embedded as a hypersurface. We do this phrased in modern language, but it is of course only a reformulation of a result due to Kervaire and Milnor [KM] in $\mathbb{R}^5$.

1. Mean curvature flow

Suppose that $M$ is a closed hypersurface in $\mathbb{R}^{n+1}$ and $M_t$ is a variation of $M$. That is, $M_t$ is a one-parameter family of hypersurfaces with $M_0 = M$. If we think of volume as a function on the space of hypersurfaces, then the first variation formula gives the derivative of volume under the variation

$$\frac{d}{dt} \text{Vol}(M_t) = \int_{M_t} \langle \partial_t x, H \mathbf{n} \rangle.$$ 

Here $x$ is the position vector, $\mathbf{n}$ the unit normal, and $H$ the mean curvature scalar given by

$$H = \text{div}_M(\mathbf{n}) = \sum_{i=1}^{n} \langle \nabla_{e_i} \mathbf{n}, e_i \rangle,$$

where $e_i$ is an orthonormal frame for $M$. Equivalently, $H$ is the sum of the principal curvatures of $M$. With this normalization, $H$ is $n/R$ on the round $n$-sphere of radius $R$.

It follows from the first variation formula that the gradient of volume is

$$\nabla \text{Vol} = H \mathbf{n},$$

and the most efficient way to reduce the volume is to choose the variation so that

$$\partial_t x = -\nabla \text{Vol} = -H \mathbf{n}.$$ 

This negative gradient flow for volume is called MCF; see Figure 1. It is formally similar to the heat equation and is smoothing for short time (cf. [EH]). In words, under the MCF, a hypersurface locally moves in the direction where the volume element decreases the fastest. Thus, if $M_t$ flows by MCF, then

$$\frac{d}{dt} \text{Vol}(M_t) = -\langle \nabla \text{Vol}, \nabla \text{Vol} \rangle = - \int_{M_t} H^2.$$ 

The flow contracts a closed hypersurface, eventually leading to its extinction in finite time.

Our chief interest here is what happens before a hypersurface becomes extinct. Is it possible to bring the hypersurface into general position so that one can describe and classify the changes that it goes through? What are the singularities that can
occur during the flow? What does the flow look like near these singularities? What is the structure of the singular set itself? Is it possible to piece together information about the original hypersurface from the changes that it goes through under the flow? In what follows, we will discuss some of the known results addressing these questions.

These are natural questions that one can for ask for many different flows, and advances in one may lead to advances for other flows.

1.1. Curve shortening flow. The simplest case of MCF is when \( n = 1 \) and the hypersurfaces are curves; this is called curve shortening flow. Gage and Hamilton, [GH], showed that curve shortening flow starting from any simple closed convex curve remains simple and convex up to an extinction time where it converges to a point. Moreover, if the flow is rescaled to keep the enclosed area constant,
then the resulting curves converge to a round circle at the extinction time. This is usually summarized in the following way:

**Theorem 1.1 ([GH]).** Under curve shortening flow, every simple closed convex curve in $\mathbb{R}^2$ remains convex and eventually becomes extinct in a “round point”.

A year later, Grayson [G] showed that any simple closed curve eventually becomes convex under the flow; see Figure 2. Thus, by the result of Gage and Hamilton, it becomes extinct in a round point. As we will see, the picture is far more complicated in higher dimensions.

1.2. **Maximum principle.** One of the fundamental tools for studying MCF is the parabolic maximum principle. This has a number of important consequences, including the following key facts (see also Figure 3):

1. If two closed hypersurfaces are disjoint, then they remain disjoint under MCF.
2. If the initial hypersurface is embedded, then it remains embedded under MCF.
3. If a closed hypersurface is convex, then it remains convex under MCF.
4. Likewise, mean convexity (i.e., $H \geq 0$) is preserved under MCF.

It follows from the avoidance property (1) that any closed hypersurface must become extinct under the flow before the extinction of a large sphere containing the initial hypersurface. For shrinking curves, Grayson proved that the singularities are trivial. In higher dimensions, as we will see, the situation is much more complicated.

Grayson showed that his result for curves does not extend to surfaces. In particular, he showed that a dumbbell with a sufficiently long and narrow bar will develop a pinching singularity before extinction. A later proof was given by Angenent [A], using the shrinking donut, that we will discuss shortly, and the avoidance property (1); see Figure 9 where Angenent’s argument is explained. Figures 4–7 show eight snapshots in time of the evolution of a dumbbell.

---

2 Figures 4–7 were created by computer simulation by U. Mayer (see [May]) and are used with permission.
Even though the singularities can be quite complicated, one can define weak solutions of the flow through singularities. There are two main types of weak solutions, each focusing on a different aspect of the flow. The first was Brakke’s MCF of varifolds in \[B\], where the weak solutions evolve to minimize volume. The other approach, called the level set flow, focuses on the avoidance property: a family of sets is a weak solution if it does not violate the avoidance property with any smooth solution. The level set flow was implemented numerically by Osher and Sethian \[OS\] and constructed theoretically by Evans and Spruck \[ES\] and Chen, Giga, and Goto \[CGG\].

Obviously, as long as the flow stays smooth, the evolving hypersurfaces are diffeomorphic. Thus any topological change comes from singularities. However, White used the maximum principle to control the topology of the flow past singularities. For example, a special case of White’s results gives:
Theorem 1.2 ([W4]). Let $M_t$ be a weak MCF in $\mathbb{R}^3$ that starts from a closed surface $M_0$ of genus $g_0$. Then at each time $t > 0$ where $M_t$ is smooth, the genus of $M_t$ is at most $g_0$.

1.3. Shrinkers. Evolution equations often have special solutions, called solitons, that evolve over time by rigid motion or homotheties. The most important ones in MCF are shrinkers which only undergo homothetic changes under the flow. The simplest examples are shrinking round spheres of radius $\sqrt{-2nt}$, where $t < 0$, and shrinking round cylinders $S^k \times \mathbb{R}^{n-k}$ with radius $\sqrt{-2kt}$. More generally, an MCF $M_t$ is a shrinker if

$$M_t = \sqrt{-l} M_{-1}$$

for all $t < 0$.

Here, when $S$ is a subset of $\mathbb{R}^{n+1}$ and $\mu > 0$ is a positive constant, then $\mu S$ is the set $\{\mu s \mid s \in S\}$ where the whole Euclidean space has been scaled by the factor $\mu$.

The shrinkers above become extinct at time $t = 0$ and move by homotheties centered at 0. We could equally well have considered surfaces that under MCF evolve by homothety centered at a different point in space-time.

Angenent [A] constructed a self-similar shrinking donut in $\mathbb{R}^3$, together with similar higher-dimensional examples. Angenent’s example was given by rotating a simple closed curve in the plane around an axis and, thus, had the topology of a torus; see Figure 8. In fact, numerical evidence suggests that, unlike for the case of curves, a complete classification of shrinkers is impossible in higher dimensions as the examples appear to be so plentiful and varied; see for instance Chopp [Ch] and Ilmanen [I2] for numerical examples and the recent rigorously constructed examples by gluing methods by Kapouleas, Kleene, and Möller in [KKM], Möller in [M], and Nguyen in [Nu]; see Figures 10–12.

![Figure 8. Angenent’s shrinking donut from numerical simulations of D. Chopp [Ch]. The vertical z-axis is the axis of rotation and the horizontal r-axis is a line of reflection symmetry.](image)

1.4. The shrinker equation. An easy computation shows that an MCF $M_t$ is a shrinker if and only if $M = M_{-1}$ satisfies the equation

$$H = \frac{\langle x, n \rangle}{2}.$$

That is, $M_t = \sqrt{-l} M_{-1}$ if and only if $M_{-1}$ satisfies $H = \frac{\langle x, n \rangle}{2}$.

---

3This equation differs by a factor of two from Huisken’s definition of a shrinker; this is because Huisken works with the time $-1/2$ slice.
Angenent’s proof for why the neck of the dumbbell pinches before the bells become extinct: Enclose the neck with a small shrinking donut, and place two round spheres inside the bells. By the avoidance property, these four surfaces stay disjoint under the flow. Since the donut becomes extinct before the two spheres, the neck pinches off before the bells become extinct.

We will refer to the time $-1$ slice $M_{-1}$ as a shrinker since this carries all the information of the self-similar flow $M_t = \sqrt{-t} M_{-1}$. The simplest shrinkers are the hyperplane through the origin, the sphere $S_{\sqrt{2}}$, and the cylinders $S_{\sqrt{2k}} \times \mathbb{R}^{n-k}$.

We let $C_k$ be the union of all rotations of $S_{\sqrt{2k}} \times \mathbb{R}^{n-k}$, and then let the space of generalized cylinders $C$ be the union of the $C_k$’s.

The shrinker equation arises variationally in two closely related ways: as minimal surfaces for a conformally changed metric and as critical points for a weighted area functional. We return to the second later, but state the first now:

**Lemma 1.4.** $M$ is a shrinker $\iff$ $M$ is a minimal surface in the metric

$$g_{ij} = e^{-\frac{|x|^2}{2t}} \delta_{ij}.$$  

The proof follows immediately from the first variation. Unfortunately, this metric on $\mathbb{R}^{n+1}$ is not complete (the distance to infinity is finite), and the curvature blows up exponentially.

1.5. **Huisken’s theorem about MCF of convex hypersurfaces.** Huisken [H1] showed that convexity is preserved under MCF and that closed convex hypersurfaces become round:

**Theorem 1.5** ([H1]). Under MCF, every closed convex hypersurface in $\mathbb{R}^{n+1}$ remains convex and eventually becomes extinct in a “round point”.

This is analogous to the result of Gage and Hamilton for convex curves but was proven two years earlier. Huisken’s proof works only for $n > 1$ as he shows that the hypersurfaces become closer to being umbilic and that the limiting shapes are
Figure 10. A shrinker of the type shown to exist by Kapouleas, Kleene, and Møller in [KKM]; see also Nguyen, [Nu], for a similar shrinker. Its existence had been conjectured by Ilmanen in [I2], where this picture is from. (Used with permission.)

Figure 11. A closed numerical example of Chopp, [Ch]. (Used with permission.)

Figure 12. A non-compact numerical example of Chopp, [Ch]. (Used with permission.)

A umbilic. A hypersurface is umbilic if all of the eigenvalues of the second fundamental form are the same; this characterizes the sphere when there are at least two eigenvalues, but is meaningless for curves.

Strict convexity is an open condition, so Theorem 1.5 shows that spherical extinction singularities cannot be eliminated by making an arbitrarily small perturbation
of the surface. We will see later that these are the only compact singularities with this stability property (see also [CM1] and [CM4]).

Using the maximum principle, one can show that various types of convexity are preserved under MCF. We will discuss this in more detail in subsection 2.3.

2. Singularities for MCF

We will now leave convex hypersurfaces and go to the general case. As Grayson’s dumbbell showed, there is no higher-dimensional analog of his theorem for curves. The key for analyzing singularities is a blow-up (or rescaling) analysis, based on two ingredients: monotonicity and rescaling. Rescaling allows one to magnify around a singularity by blowing up the flow to obtain a new flow that models the given singularity. The second ingredient is a monotonicity formula that guarantees that the blow-up or rescaled flow becomes simpler. In fact, we will see next that the limit of the rescaled flows is self-similar.

2.1. Huisken’s monotonicity. Let $\Phi$ be the nonnegative function on $\mathbb{R}^{n+1} \times (-\infty, 0)$ defined by

$$\Phi(x, t) = [-4\pi t]^{-\frac{n+1}{2}} e^{\frac{|x|^2}{4t}}.$$  

The Gaussian function $\Phi$ is a backward heat kernel on $\mathbb{R}^n$ extended to $\mathbb{R}^{n+1}$; in particular, its restriction to any hyperplane through the origin is the backward heat kernel.

G. Huisken proved the following monotonicity formula for MCF [H2], [E1], [E2]:

**Theorem 2.2** ([H2]). If $M_t$ is a solution to the MCF, then $\int_{M_t} \Phi \, dx$ is non-increasing in time. Moreover, the derivative is given by

$$\frac{d}{dt} \int_{M_t} \Phi = -\int_{M_t} \left| Hn + \frac{x}{2t} \right|^2 \Phi \, dx.$$  

A fundamental aspect of this is that Huijsken’s Gaussian volume $\int_{M_t} \Phi$ is constant in time if and only if $M_t$ is a shrinker with $M_t = \sqrt{-t} M_{-1}$.

We have stated the monotonicity for Gaussian areas centered at the origin in space-time. A similar formula holds at other points.

2.2. Tangent flows. If $M_t$ is an MCF, then for all fixed constants $\mu > 0$ one can obtain a new MCF $\tilde{M}_t$ by scaling space by $\mu$ and scaling time by $\mu^2$

$$\tilde{M}_t = \mu M_{\mu^{-2} t}.$$  

The different scaling in time and space comes from that MCF is a parabolic equation where time accounts for one derivative and space for two, just as in the ordinary heat equation. This type of scaling is referred to as parabolic scaling, and it guarantees that the new one-parameter family also flows by MCF. When $\mu$ is large, this magnifies a small neighborhood of the origin in space-time.

If we now take a sequence $\mu_i \to \infty$ and let $M^i_t = \mu_i M_{\mu_i^{-2} t}$, then Huisken’s monotonicity gives uniform Gaussian area bounds on the rescaled sequence. Combining this with Brakke’s weak compactness theorem for MCF, it follows that a subsequence of the $M^i_t$ converges to a limiting flow $M_t^\infty$ (cf., for instance, [W1] and [I2]). Moreover, Huisken’s monotonicity implies that the Gaussian area (centered...
at the origin) is now constant in time, so we conclude that $M^\infty$ is a shrinker. This $M^\infty$ is called a tangent flow at the origin. However, a priori, taking a different subsequence $\mu^i$ might result in a different tangent flow. Whether this can happen is known as the uniqueness of tangent flows question, and it is perhaps the most fundamental question about singularities. We will return to this in Section 5.

For simplicity, we rescaled about the origin in space-time; the same construction can be done at any point of space-time.

2.3. Mean convex flows. A hypersurface is mean convex if the mean curvature is nonnegative. This condition is weaker than convexity since it requires only that the sum of the principle curvatures is nonnegative, rather than that all of them are nonnegative. It follows from the maximum principle that mean convexity is preserved under MCF as long as the flow is smooth. Mean convexity for an MCF is equivalent to that the flow moves strictly inward and this allows one to extend mean convexity to nonsmooth flows.

Mean convex MCF of closed embedded hypersurfaces has a great deal of structure and much is known. In [HS1], [HS2], [W2], and [W3], Huisken and Sinestrari, and White, respectively, classified the tangent flows, showing that the generalized cylinders are the only possibilities. They also showed that all blowups are convex.

Even once one knows that the blowups are cylinders, there is still the possibility of multiplicity. However, White [W3] and Andrews [An] proved that the multiplicity is always one for mean convex flows; cf. [Br], [HaK3]. Haslhofer and Kleiner [HaK1] used the Andrews maximum principle approach to control the normal injectivity radius of mean convex flows and used this to obtain unified proofs of the earlier estimates for mean convex flows. Brendle and Huisken [BrH] and Haslhofer and Kleiner [HaK2] have constructed an MCF with surgery for mean convex surfaces in $\mathbb{R}^3$. Earlier, Huisken and Sinestrari constructed an MCF with surgery for two-convex hypersurfaces in higher dimensions in [HS3].

Finally, using in part this classification and a dimension-reducing argument, White proved a sharp bound on the dimension of the space-time singular set of a mean convex flow (cf. [HaK1, theorem 1.15]):

**Theorem 2.4 ([W2]).** The singular set of a compact mean convex MCF in $\mathbb{R}^{n+1}$ has parabolic Hausdorff dimension at most $n - 1$.

This bound is achieved, for instance, by the marriage ring in $\mathbb{R}^3$; see Figure 17 below. While the bound for the dimension is sharp, it raises the questions of whether the $(n - 1)$-dimensional measure is finite and whether the singular set has more structure. We will return to these questions later.

2.4. Gaussian integrals and the $F$-functionals. We will next define a family of functionals on the space of hypersurfaces given by integrating Gaussian weights with varying centers and scales. For $t_0 > 0$ and $x_0 \in \mathbb{R}^{n+1}$, define $F_{x_0,t_0}$ by

$$F_{x_0,t_0}(M) = (4\pi t_0)^{-n/2} \int_M e^{-\frac{|x-x_0|^2}{4t_0}} \, d\mu.$$ 

We will think of $x_0$ as being the point in space that we focus on and of $\sqrt{t_0}$ as being the scale. By convention, we set $F = F_{0,1}$.  

---

4A general blowup is a limit of rescalings about a sequence of points, whereas a tangent flow is a limit of rescalings about a fixed point.
2.5. **Critical points for the F-functional.** We will say that $M$ is a critical point for $F_{x_0,t_0}$ if it is simultaneously critical with respect to variations in all three parameters, i.e., variations in $M$ and all variations in $x_0$ and $t_0$. Strictly speaking, it is the triplet $(M,x_0,t_0)$ that is a critical point of $F$, but we will refer to $M$ as a critical point of $F_{x_0,t_0}$. The next proposition shows that $M$ is a critical point for $F_{x_0,t_0}$ if and only if it is the time $-t_0$ slice of a self-shrinking solution of MCF that becomes extinct at the point $x_0$ and time 0.

**Proposition 2.5 ([CM1]).** $M$ is a critical point for $F_{x_0,t_0}$ if and only if $M$ is a shrinker becoming extinct at the point $x_0$ in space and at time $t_0$ into the future.

2.6. **F-stable or index 0 critical points.** A closed shrinker is said to be F-stable or just stable if, modulo translations and dilations, the second derivative of the F-functional is nonnegative for all variations at the given shrinker.

There are two equivalent ways of formulating the stability precisely for a closed shrinker. We explain both since each way of thinking about stability has its advantages. The first makes use of the whole family of F-functionals and is the following:

A closed shrinker is said to be F-stable if for every one-parameter family of variations $\Sigma_s$ of $\Sigma$ (with $\Sigma_0 = \Sigma$) there exist variations $x_s$ of $x_0$ and $t_s$ of $t_0$ that make $F'' = (F_{x_s,t_s}(\Sigma_s))'' \geq 0$ at $s = 0$.

The other (obviously equivalent) way of thinking about stability is when we think of a single F-functional and mod out by translations and dilations. This second way will be particularly useful later when we discuss the dynamics of the flow near a closed unstable shrinker.

A closed shrinker is said to be F-stable if for every one-parameter family of variations $\Sigma_s$ of $\Sigma$ (with $\Sigma_0 = \Sigma$) there exist variations $x_s$ of $0$ and $\lambda_s$ of $1$ that make $F'' = (F(\lambda_s \Sigma_s + x_s))'' \geq 0$ at $s = 0$.

**Theorem 2.6 ([CM1]).** In $\mathbb{R}^{n+1}$ the round sphere $S^n$ is the only closed smooth F-stable shrinker.

3. **Generic singularities**

If $M_t$ flows by mean curvature and $t > s$, then Huisken’s monotonicity formula gives

\[(3.1) \quad F_{x_0,t_0}(M_t) \leq F_{x_0,t_0+(t-s)}(M_s).\]

Thus, we see that a fixed $F_{x_0,t_0}$ functional is not monotone under the flow, but the supremum over all of these functionals is monotone. We call this invariant the entropy and denote it by

\[(3.2) \quad \lambda(M) = \sup_{x_0,t_0} F_{x_0,t_0}(M).\]

The entropy has three key properties:

1. $\lambda$ is invariant under dilations, rotations, and translations.
2. $\lambda(M_t)$ is nonincreasing under MCF.
3. If $M$ is a shrinker, then $\lambda(M) = F_{0,1}(M)$. 
It follows from (3) and a result of Stone that 
\[
\lambda(S^n) \text{ is decreasing in } n \text{ and } 
\lambda(S^n) = \sqrt{\frac{2\pi}{e}} \approx 1.5203 > \lambda(S^2) = \frac{4}{\pi} \approx 1.4715 > \lambda(S^3) > \cdots > 1 = \lambda(R^n). 
\]
Moreover, a simple computation shows that 
\[
\lambda(\Sigma \times R) = \lambda(\Sigma). 
\]

A consequence of (1) is, loosely speaking, that the entropy coming from a singularity is independent of the time when it occurs, of the point where it occurs, and even of the scale at which the flow starts to resemble the singularity.

Note also that one way of thinking about (2) is that \( \nabla \text{Vol} \) and \( \nabla \lambda \) point toward the same direction in the sense that \( \langle \nabla \text{Vol}, \nabla \lambda \rangle \geq 0 \). We will use this later.

3.1. How entropy is used. The main point about \( \lambda \) is that it can be used to rule out certain singularities because of the monotonicity of entropy under MCF and its invariance under dilations:

**Corollary 3.4.** If \( M \) is a shrinker that occurs as a tangent flow for \( M_t \) with \( t > 0 \), then 
\[
F_{0,1}(M) = \lambda(M) \leq \lambda(M_0). 
\]

3.2. Classification of entropy stable singularities. The next theorem shows that the only singularities that cannot be perturbed away are the simplest ones, i.e., the generalized cylinders in \( C \).

**Theorem 3.5 ([CM1]).** Suppose that \( M^n \subset R^{n+1} \) is a smooth complete embedded shrinker without boundary and with polynomial volume growth.

1. If \( M \notin C \), then there is a graph \( N \) over \( M \) of a function with arbitrarily small \( C^m \) norm (for any fixed \( m \)) so that \( \lambda(N) < \lambda(M) \).
2. If \( M \) is not \( S^n \) and does not split off a line, then the function in (1) can be taken to have compact support.

In particular, in either case, \( M \) cannot arise as a tangent flow to the MCF starting from \( N \).

Thus, spheres, planes, and cylinders are the only generic shrinkers.

In fact, we have the following stronger result where the shrinker is allowed to have singularities:

**Theorem 3.6 ([CM1]).** Theorem 3.5 holds when \( n \leq 6 \) and \( M \) is smooth off of a singular set with locally finite \( (n-2) \)-dimensional Hausdorff measure.

See [CM2] and [CM3] for more on generic singularities.

3.3. Self-shrinkers with low entropy/Gaussian surface area. It follows from Brakke’s regularity theorem for MCF that \( R^n \) has the least entropy of any shrinker and, in fact, there is a gap to the next lowest. A natural question is:

Can one classify all low entropy shrinkers, and if so what are those?

In [CIMW] it is shown that the round sphere has the least entropy of any closed shrinker.

**Theorem 3.7 ([CIMW]).** Given \( n \), there exists \( \epsilon = \epsilon(n) > 0 \) so that if \( \Sigma \subset R^{n+1} \) is a closed shrinker not equal to the round sphere, then \( \lambda(\Sigma) \geq \lambda(S^n) + \epsilon \). Moreover, if \( \lambda(\Sigma) \leq \min\{\lambda(S^{n-1}), \frac{3}{2}\} \), then \( \Sigma \) is diffeomorphic to \( S^n \).

\(^5\)If \( n > 2 \), then \( \lambda(S^{n-1}) < \frac{3}{2} \) and the minimum is unnecessary.
Theorem 3.7 is suggested by the dynamical approach to MCF of [CM1] and [CM2] that we will discuss in more detail below. The idea is that an MCF starting at a closed $M$ becomes singular, the corresponding shrinker has lower entropy, and, by [CM1], the only shrinkers that cannot be perturbed away are $S^{n-k} \times R^k$ and $\lambda(S^{n-k} \times R^k) \geq \lambda(S^n)$.

The dynamical picture also suggested two closely related conjectures in [CIMW]. The first of these was recently proven by Bernstein and Wang (cf. [KZ]):

Theorem 3.8 ([BW]). Theorem 3.7 holds with $\epsilon = 0$ for any closed hypersurface $M^n$ with $n \leq 6$.

The second conjecture, which remains open, asks whether the theorem also holds for open shrinkers:

Conjecture 3.9 ([CIMW]). Theorem 3.7 holds for any nonflat shrinker $\Sigma^n \subset R^{n+1}$ with $n \leq 6$.

When $n = 1$, Theorem 3.8 follows for curves by combining Grayson’s theorem [G] (cf. [GH]) and the monotonicity of $\lambda$ under curve shortening flow. The conjecture follows for curves from the classification of shrinkers by Abresch and Langer.

One could ask which shrinker has the third least entropy, etc. The entropy of the “Simons cone” over $S^k \times S^k$ in $R^{2k+2}$ is asymptotic to $\sqrt{2}$ as $k \to \infty$, which is also the limit of $\lambda(S^{2k+1})$. Thus, as the dimension increases, the Simons cones have lower entropy than some of the generalized cylinders. For example, the cone over $S^2 \times S^2$ has entropy $\lambda(S^2 \times R^4) < \lambda(S^1 \times R^4)$. In other words, already for $n = 5$, $S^k \times R^{n-k}$ is not a complete list of the lowest entropy shrinkers.

4. Rigidity of cylinders

Generalized cylinders are rigid in a very strong sense. Any other shrinker that is sufficiently close to one of them on a large, but compact set and with a fixed, but arbitrary entropy bound must itself be a cylinder:

Theorem 4.1 ([CIM]). Given $n$, $\lambda_0$ and $C$, there exists $R = R(n, \lambda_0, C)$ so that if $\Sigma^n \subset R^{n+1}$ is a smooth complete embedded shrinker with entropy $\lambda(\Sigma) \leq \lambda_0$ satisfying

\[(\dagger) \ 0 \leq H \ and \ |A| \leq C \ on \ B_R \cap \Sigma,
\]

then $\Sigma \in C$.

We will say that a singular point is cylindrical if at least one tangent flow is a multiplicity one cylinder in $C$. As a corollary of the theorem, we get uniqueness of type of cylindrical tangent flows:

Corollary 4.2 ([CIM]). If a singular point of an MCF is cylindrical, then for every tangent flow there is a multiplicity one cylinder.

This corollary leaves open the possibility that the axis of the cylinder (i.e., the direction of the $R^{n-k}$ factor) might depend on the sequence of rescalings; see Figure 13. Whether this happens is a major problem (known as the uniqueness of tangent flows problem) that we will turn to in the next section.

A tangent flow is the limit of a sequence of rescalings at the singularity, where the convergence is on compact subsets. Thus, it is essential for applications of Theorem
Snapshots of the flow at 3 times near one singular time. The axis of one cylinder could potentially rotate slowly in time.

Figure 13. The essence of uniqueness of tangent flows: Can the flow be close to a cylinder at all times right before the singular time, yet the axis of the cylinder changes as the time gets closer to the singular time?

4.1 like Corollary 4.2 that Theorem 4.1 only requires closeness on a fixed compact set.

The rigidity theorem holds more generally even when the shrinker is not required to be smooth. This is important in applications, including for the proof of Corollary 4.2.

The proof of Theorem 4.1 is by iteration and improvement. Roughly speaking, the theorem assumes that the shrinker is cylindrical on some large scale. The iterative step then shows that it is cylindrical on an even larger scale, but with some loss in the estimates. The improvement step then comes back and says that there was actually no loss if the scale is large enough. Applying these two steps repeatedly gives that the shrinker is roughly cylindrical on all scales, which will easily give the theorem; see [CIM] for details.

Corollary 4.2 suggests the following closely related canonical neighborhood statement:

**Conjecture 4.3.** Let $M_t$ be an MCF flow of smooth closed hypersurfaces in $\mathbb{R}^{n+1}$. If the flow has a cylindrical singularity at time $t_0$ and at the point $x_0 \in \mathbb{R}^{n+1}$, then in an entire space-time neighborhood of $(x_0, t_0)$ the evolving hypersurfaces have positive mean curvature.
For rotationally symmetric hypersurfaces this property of mean convexity near
singularities was shown by ODE techniques by Altschuler, Angenent, and Giga
where they called it the “attracting axis theorem”.

As discussed in subsection 3.3, the sphere has the lowest entropy among closed
shrinkers in each dimension, but there are other shrinkers with entropy below the
cylinders in high dimensions. Thus, Theorem 4.1 is in contrast with Brakke’s regular-
arity theorem that shows that not only is the hyperplane isolated among shrinkers
and has the lowest entropy, but there is a gap to the entropy of all other shrinkers.

4.1. Asymptotic rigidity. One can also ask whether there is a corresponding
rigidity or uniqueness at infinity. Lu Wang proved that there is when the shrinker
is asymptotic to a cone:

**Theorem 4.4 (Wg).** If $\Sigma$ and $\tilde{\Sigma}$ are shrinkers in $\mathbb{R}^{n+1} \setminus B_R$ that have boundary
in $\partial B_R$ and are asymptotic to the same cone, then they coincide.

One of the key ingredients in the proof is a parabolic unique continuation result
of Escauriaza, Seregin, and Sverak [ESS] that was developed to settle a well-known
problem in the regularity theory of the Navier–Stokes equation.

5. Uniqueness of tangent flows

Once one knows that singularities occur, one naturally wonders what the singu-
larities are like. For minimal varieties the first answer, already known to Federer and
Fleming in 1959, is that they weakly resemble cones. For MCF, by the combined
work of Huisken, Ilmanen, and White, singularities weakly resemble shrinkers. Un-
fortunately, the simple proofs leave open the possibility that a minimal variety or
an MCF looked at under a microscope will resemble one blowup, but under higher
magnification, it might (as far as anyone knows) resemble a completely different
blowup. Whether this ever happens is perhaps the most fundamental question
about singularities.

Two of the most prominent early works on uniqueness of tangent cones are Leon
Simon’s hugely influential paper [Si1], where he proves uniqueness for tangent cones
of minimal varieties with smooth cross-section. The other is Allard and Almgren’s
paper [AA], where uniqueness of tangent cones with smooth cross-section is proven
under an additional integrability assumption on the cross-section.

**Theorem 5.1 (CM5).** Let $M_t$ be an MCF in $\mathbb{R}^{n+1}$. At each cylindrical singular
point the tangent flow is unique. That is, any other tangent flow is also a cylinder
with the same $R^k$ factor that points in the same direction.

This theorem solved a major open problem; see, e.g., [W1] p. 534]. Even in
the case of the evolution of mean convex hypersurfaces, where all singularities are
cylindrical, uniqueness of the axis was previously unknown.

Theorem 5.1 is the first general uniqueness theorem for tangent flows to a geo-
metric flow at a noncompact singularity. (In fact, they are also nonintegrable.)
Some special cases of uniqueness of tangent flows for MCF were previously ana-
lyzed assuming either some sort of convexity or that the hypersurface is a surface
of rotation; see [H1], [H2], [HS1], [HS2], [W1], [SS], [AAC], and [GK], [GKS], [GS].

---

6See Brian White [W5] section “Uniqueness of tangent cone” from which part of this discussion
is taken and where one can find more discussion of uniqueness for minimal varieties.
In contrast, uniqueness for blowups at compact singularities is better understood; cf. [AA], [Si1], [H1], [Sc], and [Se].

This noncompactness caused major analytical difficulties, and to address them required entirely new techniques and ideas. This is not so much because of the subtleties of analysis on noncompact domains, though this was an issue, but crucially because the evolving hypersurface cannot be written as an entire graph over the singularity no matter how close one gets to the singularity. Rather, only part of the evolving hypersurface can be written as a graph over a compact piece of the singularity.

5.1. Lojasiewicz inequalities. The main technical tools in [CM5] are two Lojasiewicz-type inequalities for the \( F \) functional on the space of hypersurfaces. Before explaining these in the next subsection, we will review the classical Lojasiewicz inequalities and their role in proving uniqueness for finite dimensional gradient flows.

In real algebraic geometry, the Lojasiewicz inequality, [L], named after Stanislaw Lojasiewicz, gives an upper bound for the distance from a point to the nearest zero of a given real analytic function. Specifically, let \( f : U \to \mathbb{R} \) be a real-analytic function on an open set \( U \) in \( \mathbb{R}^n \), and let \( Z \) be the zero locus of \( f \). Assume that \( Z \) is not empty. Then for any compact set \( K \) in \( U \), there exist \( \alpha \geq 2 \) and a positive constant \( C \) such that, for all \( x \in K \)

\[
\inf_{z \in Z} |x - z|^\alpha \leq C |f(x)|.
\]

Here \( \alpha \) can be large.

Lojasiewicz [L] also proved the following inequality that is often referred to as the gradient inequality. With the same assumptions on \( f \), for every \( p \in U \), there is a possibly smaller neighborhood \( W \) of \( p \) and constants \( \beta \in (0, 1) \) and \( C > 0 \) such that for all \( x \in W \)

\[
|f(x) - f(p)|^\beta \leq C |\nabla_x f|.
\]

Note that this inequality is trivial unless \( p \) is a critical point for \( f \).

An immediate consequence of (5.3) is that every critical point of \( f \) has a neighborhood where every other critical point has the same value; it is easy to construct smooth functions where this is not the case. This consequence of (5.3) for the \( F \) functional near a cylinder is implied by the rigidity result of Corollary 4.2.

Lojasiewicz used his second inequality to show the “Lojasiewicz theorem”:

**Theorem 5.4 ([L]).** If \( f : \mathbb{R}^n \to \mathbb{R} \) is an analytic function, \( x = x(t) : [0, \infty) \to \mathbb{R}^n \) is a curve with \( x'(t) = -\nabla f \), and \( x(t) \) has a limit point \( x_\infty \), then the length of the curve is finite and \( \lim_{t \to \infty} x(t) = x_\infty \). Moreover, \( x_\infty \) is a critical point for \( f \).

**Proof.** To see that \( x(t) \) converges to \( x_\infty \), assume that \( f(x_\infty) = 0 \), and note that if we set \( f(t) = f(x(t)) \), then \( f' = -|\nabla f|^2 \). Moreover, by the second Lojasiewicz inequality, we get that \( f' \leq -f^{2\beta} \) if \( x(t) \) is sufficiently close to \( x_\infty \). (Assume for simplicity that \( x(t) \) stays in a small neighborhood \( x_\infty \) for \( t \) large so that this inequality holds; the general case follows easily.) Then this inequality can be rewritten as

\[
\int_0^\infty f'(t) dt \leq -\int_0^\infty f^{2\beta}(t) dt.
\]

In the end, [CM5] shows that the domain the evolving hypersurface is a graph over is expanding in time and at a definite rate, but this is not all all clear from the outset.
Narrow ledge on steep hillside, becoming more and more narrow and sloping less and less as it spirals infinitely down to the base.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{image}
\caption{There are smooth functions vanishing on an open (compact) set for which the gradient flow lines spiral around the zero locus. The flow lines have infinite length and the Lojasiewicz theorem fails.}
\end{figure}

$(f^{1-2\beta})' \geq (2\beta - 1)$ which integrates to

\begin{equation}
(f(t)) \leq C t^{1/2(2\beta - 1)}.
\end{equation}

We need to show that \eqref{eq:5.5} implies that $\int_0^\infty |\nabla f| \, ds$ is finite. This shows that $x(t)$ converges to $x_\infty$ as $t \to \infty$. To see that $\int_0^\infty |\nabla f| \, ds$ is finite, observe by the Cauchy–Schwarz inequality that

\begin{equation}
\int_0^\infty |\nabla f| \, ds = \int_0^\infty \sqrt{-f'} \, ds \leq \left( -\int_0^\infty f' s^{1+\epsilon} \, ds \right)^{1/2} \left( \int_0^\infty s^{1-\epsilon} \, ds \right)^{1/2}.
\end{equation}

It suffices therefore to show that

\begin{equation}
-\int_0^T f' s^{1+\epsilon} \, ds
\end{equation}

is uniformly bounded. Integrating by parts gives

\begin{equation}
\int_0^T f' s^{1+\epsilon} \, ds = |f s^{1+\epsilon}|_0^T - (1 + \epsilon) \int_0^T f s^\epsilon \, ds.
\end{equation}

If we choose $\epsilon > 0$ sufficiently small depending on $\beta$, then we see that this is bounded independent of $T$ and hence $\int_0^\infty |\nabla f| \, ds$ is finite. \hfill \Box

In contrast, it is easy to construct smooth functions, even on $\mathbb{R}^2$, where the Lojasiewicz theorem fails, i.e., where there are negative gradient flow lines that have more than one limit point (and, thus, also have infinite length); see Figure 14.

The second inequality is used to get the uniqueness, while the first inequality with a sharp enough power is the key ingredient for proving the second in [CM5].

5.2. Lojasiewicz inequalities for noncompact hypersurfaces and MCF.

The uniqueness result relies on two new infinite-dimensional Lojasiewicz type inequalities. Infinite-dimensional Lojasiewicz inequalities were pioneered thirty years ago by Leon Simon. However, unlike previous infinite-dimensional Lojasiewicz inequalities, these new inequalities do not follow from a reduction to the classical
finite-dimensional Lojasiewicz inequalities from the 1960s from algebraic geometry. Rather the inequalities are proven directly, and do not rely on Lojasiewicz’s arguments or results.

Roughly speaking, [CM5] proved the following Lojasiewicz inequalities for the $F$ functional on a general hypersurface $\Sigma$:

$$\text{dist}(\Sigma, C)^2 \leq C |\nabla \Sigma F|,$$

(5.9) $$\left( F(\Sigma) - F(C) \right)^{\frac{2}{3}} \leq C |\nabla \Sigma F|.$$ (5.10)

Equation (5.9) corresponds to Lojasiewicz’s first inequality, whereas (5.10) corresponds to his second inequality. The precise statements of these inequalities are much more complicated than this, but they have the same flavor; see [CM5] and [CM6] for more details.

6. The singular set of MCF with generic singularities

A major theme in PDEs over the last fifty years has been understanding singularities and the set where singularities occur. In the presence of a scale-invariant monotone quantity, blow-up arguments can often be used to bound the dimension of the singular set; see, e.g., [Al], [F]. Unfortunately, these dimension bounds say little about the structure of the set. However, the results of the previous sections lead to a rather complete description of the singular set for MCF with generic singularities:

**Theorem 6.1** ([CM7]). Let $M_t \subset \mathbb{R}^{n+1}$ be an MCF of closed embedded hypersurfaces with only cylindrical singularities. Then the space-time singular set satisfies the following.

- It is contained in finitely many (compact) embedded Lipschitz submanifolds each of dimension at most $(n - 1)$ together with a set of dimension at most $(n - 2)$.
- It is locally the graph of a 2-Hölder function on space (the projection from space-time to space is a finite-to-one covering map).
- The time image of each subset with finite parabolic 2-dimensional Hausdorff measure has measure zero; each such connected subset is contained in a time-slice.

In fact, [CM7] proves considerably more than what is stated in Theorem 6.1; see [CM7] theorem 4.18]. For instance, instead of just proving the first claim of the theorem, the entire stratification of the space-time singular set is Lipschitz of the appropriate dimension. Moreover, this holds without ever discarding any subset of measure zero of any dimension, as is always implicit in any definition of rectifiable. To illustrate the much stronger version, consider the case of evolution of surfaces in $\mathbb{R}^3$. In that case, this gives that the space-time singular set is contained in finitely many (compact) embedded Lipschitz curves with cylinder singularities together with a countable set of spherical singularities. In higher dimensions, the direct generalization of this is proven.

Theorem 6.1 has the following corollaries:

---

In fact, Lipschitz is with respect to the parabolic distance on space-time which is a much stronger assertion than Lipschitz with respect to the Euclidean distance. Note that a function is Lipschitz when the target that has the parabolic metric on $\mathbb{R}$ is equivalent to that which is 2-Hölder for the standard metric on $\mathbb{R}$.
Corollary 6.2 \((\text{CM7})\). Let \(M_t \subset \mathbb{R}^{n+1}\) be an MCF of closed embedded mean convex hypersurfaces or an MCF with only generic singularities, then the conclusion of Theorem \(5.1\) holds.

More can be said in dimensions three and four:

Corollary 6.3 \((\text{CM7})\). If \(M_t\) is as in Theorem \(6.1\) and \(n = 2\) or \(3\), then the evolving hypersurface is completely smooth (i.e., has no singularities) at almost all times. In particular, any connected subset of the space-time singular set is completely contained in a time-slice.

Corollary 6.4 \((\text{CM7})\). For a generic MCF in \(\mathbb{R}^3\) or \(\mathbb{R}^4\) or a flow starting at a closed embedded mean convex hypersurface in \(\mathbb{R}^3\) or \(\mathbb{R}^4\), the conclusion of Corollary \(6.3\) holds.

The conclusions of Corollary \(6.4\) hold in all dimensions if the initial hypersurface is 2- or 3-convex. We have already seen 2-convex hypersurfaces; a hypersurface is \(k\)-convex if the sum of any \(k\) principal curvatures is nonnegative.

A key technical point in \(\text{CM7}\) is to prove a strong parabolic Reifenberg property for MCF with generic singularities. In fact, the space-time singular set is proven to be (parabolically) Reifenberg vanishing. In analysis a subset of Euclidean space is said to be Reifenberg (or Reifenberg flat) if on all sufficiently small scales it is, after rescaling to unit size, close to a \(k\)-dimensional plane; see, e.g., \([\text{Re}], [\text{Si2}], [\text{T}]\). The dimension of the plane is always the same but the plane itself may change from scale.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{koch_curve}
\caption{The Koch curve is close to a line on all scales, yet the line to which it is close changes from scale to scale. It is not rectifiable but admits a H"older parametrization. It also illustrates that uniqueness of blowups is closely related to rectifiability.}
\end{figure}
Many snowflakes, like the Koch snowflake (see Figure 15), are Reifenberg with Hausdorff dimension strictly larger than one. A set is said to be Reifenberg vanishing if the closeness to a $k$-plane goes to zero as the scale goes to zero. It is said to have the strong Reifenberg property if the $k$-dimensional plane depends only on the point but not on the scale. Finally, one sometimes distinguishes between half Reifenberg and full Reifenberg, where half Reifenberg refers to that the set is close to a $k$-dimensional plane, whereas full Reifenberg refers to that and in addition one also has the symmetric property: the plane on the given scale is close to the set.

Using the uniqueness of tangent flows, [CM7] shows that the singular set in space-time is strong (half) Reifenberg vanishing with respect to the parabolic Hausdorff distance. This is done in two steps, showing first that nearby singularities sit inside a parabolic cone (i.e., between two oppositely oriented space-time paraboloids that are tangent to the time-slice through the singularity). In fact, this parabolic cone property holds with vanishing constant. Next, in the complementary region of the parabolic cone in space-time (that is essentially space-like), the parabolic Reifenberg essentially follows from the space Reifenberg that the uniqueness of tangent flows implies.

An immediate consequence, of independent interest, of the parabolic cone property with vanishing constant is that nearby a generic singularity in space-time (“nearby” is with respect to the parabolic distance) all other singularities happen at almost the same time.

These results should be contrasted with a result of Altschuler, Angenent, and Giga [AAG] (cf. [SS]) which shows that in $\mathbb{R}^3$ the evolution of any rotationally symmetric surface obtained by rotating the graph of a function $r = u(x)$, $a < x < b$, around the $x$-axis is smooth except at finitely many singular times, where either a cylindrical or spherical singularity forms; see Figure 16. For more general rotationally symmetric surfaces (even mean convex), the singularities can consist of nontrivial curves. For instance, consider a torus of revolution bounding a region $\Omega$. If the torus is thin enough, it will be mean convex. Since the symmetry is preserved and because the surface always remains in $\Omega$, it can only collapse to a circle; see Figure 17. Thus, at the time of collapse, the singular set is a simple closed curve.

![Figure 16. Finitely many singularities for surfaces of rotation.](image-url)
As a consequence of Theorem 2.4, White showed that a mean convex surface evolving by MCF in $\mathbb{R}^3$ must be smooth at almost all times, and at no time can the singular set be more than 1-dimensional. In fact, White’s general dimension reducing argument [W6], [W7] gives that the singular set of any MCF with only cylindrical singularities has dimension at most $(n-1)$.

These results motivate the following conjecture:

**Conjecture 6.5 ([CM7]).** Let $M_t$ be an MCF of closed embedded hypersurfaces in $\mathbb{R}^{n+1}$ with only cylindrical singularities. Then the space-time singular set has only finitely many components.

If this conjecture was true, then it would follow that in $\mathbb{R}^3$ and $\mathbb{R}^4$ MCF with only generic singularities is smooth except at finitely many times; cf. with the 3-dimensional conjecture at the end of [W1, section 5].

### 7. Isolated singularities

Singularities of an MCF need not be isolated, as can already be seen with shrinking cylinders in $\mathbb{R}^3$ where there is an entire line of singularities at the extinction time or in the marriage ring example where there is a circle of singularities. However, these situations seem unstable as any small change could cause one part of the flow to shrink more quickly, leading to isolated singularities at the first singular time.

**Conjecture 7.1.** Given a closed hypersurface $\Sigma$ and any $\epsilon > 0$, there is a graph $\Sigma'$ with $C^2$ norm at most $\epsilon$ over $\Sigma$ so that MCF starting from $\Sigma'$ has only isolated singularities.

This conjecture is completely open even in the case of mean convex MCF.

The one case where the question is completely understood is for surfaces of revolution. In this case, Altschuler, Angenent, and Giga [AAG] showed that the number of singularities is bounded by the number of critical points of the distance to the axis of revolution. This is exactly what happens for the sphere (where the only critical point is the maximum) and for the dumbbell (where there are two maxima and a single minimum). Since Morse functions are generic and have isolated critical points, it follows that Conjecture 7.1 holds in the simple case of surfaces of revolution.
8. Two conjectures about singularities of MCF

Thus far we have mostly discussed smooth tangent flows (with the exception of Theorem 3.6). However, tangent flows are not always smooth, but we have the following well-known conjecture (see [I1, p. 8]):

**Conjecture 8.1.** Suppose that $M_0 \subset \mathbb{R}^{n+1}$ is a smooth closed embedded hypersurface. A time-slice of any tangent flow of the MCF starting at $M_0$ has a singular set of dimension at most $n - 3$.

Observe, in particular, that Theorem 3.6 classifies entropy stable shrinkers in $\mathbb{R}^{n+1}$ assuming that they have the smoothness of Conjecture 8.1.

In [I1], Ilmanen proved that, in $\mathbb{R}^3$, tangent flows at the first singular time must be smooth, although he left open the possibility of multiplicity. However, he conjectured that the multiplicity must be one.

8.1. **Negative gradient flow near a critical point.** We are interested in the dynamical properties of MCF near a singularity. Specifically, we would like to show that the typical flow line or rather the MCF starting at the typical or generic hypersurface avoids unstable singularities. Before getting to this, it is useful to recall the simple case of gradient flows near a critical point on a finite-dimensional manifold. Suppose therefore that $f : \mathbb{R}^2 \to \mathbb{R}$ is a smooth function with a non-degenerate critical point at 0 (so $\nabla f(0) = 0$, but the Hessian of $f$ at 0 has rank 2).

The behavior of the negative gradient flow $(x', y') = -\nabla f(x, y)$ is determined by the Hessian of $f$ at 0. For instance, if $f(x, y) = \frac{a}{2} x^2 + \frac{b}{2} y^2$ for constants $a$ and $b$, then the negative gradient flow solves the ODEs $x' = -a x$ and $y' = -b y$. Hence, the flow lines are given by $x = e^{-at} x(0)$ and $y = e^{-bt} y(0)$.

The behavior of the flow near a critical point depends on the index of the critical point, as is illustrated by the following examples (see also Figures [18] and [19]):

(Index 0): The function $f(x, y) = x^2 + y^2$ has a minimum at 0. The vector field is $(-2x, -2y)$ and the flow lines are rays into the origin. Thus every flow line limits to 0.

(Index 1): The function $f(x, y) = x^2 - y^2$ has an index 1 critical point at 0. The vector field is $(-2x, 2y)$, and the flow lines are level sets of the function $h(x, y) = xy$. Only points where $y = 0$ are on flow lines that limit to the origin.

(Index 2): The function $f(x, y) = -x^2 - y^2$ has a maximum at 0. The vector field is $(2x, 2y)$ and the flow lines are rays out of the origin. Thus, every flow line limits to $\infty$, and it is impossible to reach 0.

Thus, we see that the critical point 0 is “generic”, or dynamically stable, if and only if it has index 0. When the index is positive, the critical point is not generic and a “random” flow line will miss the critical point.

The stable manifold for a flow near the fixed point is the set of points $x$ so that the flow starting from $x$ is defined for all time, remains near the fixed point, and converges to the fixed point as $t \to \infty$. For instance, in the three examples of indices $0 - 2$, above, the stable manifold is all of $\mathbb{R}^2$, the $x$-axis, and the origin, respectively.
Figure 18. $f(x, y) = x^2 + y^2$ has a minimum at 0. Flow lines: Rays through the origin.

Figure 19. $f(x, y) = x^2 - y^2$ has an index 1 critical point at 0. Flow lines: Level sets of $xy$. Only points where $y = 0$ limit to the origin.

It is also useful to recall what it means for directions to be expanding or contracting under a flow. We will explain this in the example of the negative gradient flow of the function $f(x, y) = \frac{a}{2} x^2 + \frac{b}{2} y^2$. We saw already that the time $t$ flow is $\Psi_t(x, y) = (e^{-at} x, e^{-bt} y)$. In particular, the time 1 flow $\Psi = \Psi_1$ is the diagonal matrix with entries $e^{-a}$ and $e^{-b}$. It follows, that if $a$ is negative, then the $x$ direction
is expanding for the flow, and if $a$ is positive, then the $x$ direction is contracting. Likewise for $b$ and the $y$ direction.

Consider next a slightly more general situation, where $f$ and $g$ are Morse functions. We will assume that the gradients of $f$ and $g$ point toward the same direction meaning that

$$\langle \nabla f, \nabla g \rangle \geq 0.$$ 

We will flow in direction of $-\nabla f$ and would like to claim that the typical flow line avoids the unstable critical points of $g$.

Later the volume function $\text{Vol}$ will play the role of $f$, and the entropy $\lambda$ will play the role of $g$. The assumption that $\nabla \text{Vol}$ and $\nabla \lambda$ point toward the same direction is a consequence of Huisken’s monotonicity formula; see the second key property of the entropy above.

To see this claim, consider an unstable critical point for $g$. Unless it is also a critical point for $f$, then there is nothing to show since there is only one flow line through each point where the gradient does not vanish. We can, therefore, assume that the unstable critical point of $g$ is also a critical point for $f$. It now follows easily from the assumption that $f$ is monotone nonincreasing along the negative gradient flow of $g$, that the given point is also an unstable critical point for $f$ and, hence, the claim follows.

8.2. Flows beginning at a typical hypersurface avoid unstable singularities. We saw above that, at least in finite dimensions, a typical flow line of a negative gradient flow of a function $f$ avoids unstable critical points of a function $g$ when their gradients flow toward the same direction. Applying this to $f = \text{Vol}$ and $g = \lambda$ together with the classification of entropy stable singularities (Theorems 3.5 and 3.6) leads naturally to the following conjecture; we will discuss in the next section some recent work related to this conjecture.

Conjecture 8.2. Suppose that $M^n \subset \mathbb{R}^{n+1}$ is an embedded smooth closed hypersurface where $n \leq 6$. Then there is a graph $N$ over $M$ of a function with arbitrarily small $C^m$ norm (for any fixed $m$) so that for the MCF starting at $N$ all tangent flows are in $C^\alpha$.

9. Dynamics of closed singularities

In this section we will discuss some very recent results related to the previous conjecture. This work shows that, for generic initial data, the MCF never ends up in an unstable closed singularity. The first step of this is to show that, near a closed unstable shrinker, the dynamics of the negative gradient flow of the $F$-functional looks exactly like an infinite-dimensional version of the dynamics of the negative gradient flow of the function $f(x, y) = x^2 - y^2$ near the unstable critical point $(x, y) = (0, 0)$.

\footnote{Note that the only smooth hypersurface that is both a critical point for $\lambda$ (i.e., is a shrinker) and is a critical point for $\text{Vol}$ (i.e., is a minimal surface) is a hyperplane. Namely, any shrinker with $H = 0$ is a cone as $x^2 = 0$, and hence, if smooth, is a hyperplane. We shall not elaborate further on this, but it illustrates that the claim about typical flow lines is more subtle than the above heuristic argument indicates. Namely, otherwise one would have that MCF starting at the typical or generic hypersurface does not become singular as points in space-time with tangent flow that is a hyperplane is a smooth point in space-time by a result of Brakke. This is, however, clearly nonsensible as the flow becomes extinct in finite time and thus must develop a singularity. In addition to this, shrinking spheres indeed are generic singularities.}
We have already seen that MCF is the negative gradient flow of volume. We have seen that singularities are modeled by their blowups, which are shrinkers, and we have explained that the only smooth stable shrinkers, are spheres, planes, and generalized cylinders (i.e., $S^k \times \mathbb{R}^{n-k}$). In particular, the round sphere is the only closed stable singularity for MCF.

Suppose that $M_t$ is a one-parameter family of closed hypersurfaces flowing by MCF. We want to analyze the flow near a singularity in space-time. After translating, we may assume that the singularity occurs at the origin in space-time. If we reparametrize and rescale the flow as follows $t \to M_{-e^{-t}}/\sqrt{e^{-t}}$, then we get a solution to the rescaled MCF equation. The rescaled MCF is the negative gradient flow for the $F$-functional

$$F(\Sigma) = (4\pi)^{-\frac{n}{2}} \int_\Sigma e^{-\frac{|x|^2}{4}},$$

(9.1)

where the gradient is with respect to the weighted inner product on the space of normal variations. The fixed points of the rescaled MCF, or equivalently the critical points of the $F$-functional, are the shrinkers. The dynamics of the MCF near a singularity becomes a question of the dynamics near a fixed point for the rescaled flow. We can therefore treat the rescaled MCF as a special kind of dynamical system that is the gradient flow of a globally defined function and where the fixed points are the singularity models for the original flow.

The paper [CM2] analyzes the behavior of the rescaled flow in a neighborhood of a closed unstable shrinker. Using this analysis, it is shown that generically one never ends up in an unstable closed singularity. A key step is to show that, in a suitable sense, “nearly every” hypersurface in a neighborhood of the unstable shrinkers is wandering or, equivalently, nonrecurrent. In contrast, in a small neighborhood of the round sphere, all closed hypersurfaces are convex and thus all become extinct in a round sphere under MCF by Theorem 1.5. The point in space-time where a closed hypersurface nearby the round sphere becomes extinct may be different from that of the given round sphere. This corresponds to that, under the rescaled MCF, it may leave a neighborhood of the round sphere but does so near a translation of the sphere. Similarly, in a neighborhood of an unstable shrinker, there are closed hypersurfaces that under the rescaled MCF leave the neighborhood of the shrinker but do so in a trivial way, namely, near a translate of the given unstable shrinker. This, of course, leads to no real change. However, a typical closed hypersurface near an unstable shrinker not only leaves a neighborhood of the shrinker, but, when it does, is not close to a rigid motion or dilation of the given shrinker. Thus, we have a genuine improvement, or at least a change, of the singularity. Using that the rescaled MCF is a gradient flow, it was also shown that once the flow leaves a neighborhood of a shrinker it will never return. Together, this not only gives a change, but an actual improvement.

9.1. Dynamics near a closed shrinker. In this subsection, we will explain in what sense the dynamics of the negative gradient flow of the $F$-functional near a closed unstable shrinker looks like an infinite-dimensional version of the dynamics of the negative gradient flow of the function $f(x, y) = x^2 - y^2$ near the index 1 critical point $(x, y) = (0, 0)$.

Let $E$ be the Banach space of $C^{2,\alpha}$ functions on a smooth closed embedded hypersurface $\Sigma \subset \mathbb{R}^{n+1}$ with unit normal $n$. We are identifying $E$ with the space
of $C^{2,\alpha}$ hypersurfaces near $\Sigma$ by mapping a function $u$ to its graph
\begin{equation}
\Sigma_u = \{ p + u(p) \, n(p) \, | \, p \in \Sigma \}.
\end{equation}
If $E_1, E_2$ are subspaces of $E$ with $E_1 \cap E_2 = \{0\}$ and that together span $E$, i.e., so that
\begin{equation}
E = \{ x_1 + x_2 \, | \, x_1 \in E_1, x_2 \in E_2 \},
\end{equation}
then we will say that $E = E_1 \oplus E_2$ is a splitting of $E$.

The essence of the next theorem is that “nearly every” hypersurface in a neighborhood of the given unstable singularity leaves the neighborhood under the rescaled MCF and, when it does, is not near a translate, rotation, or dilation of the given singularity.

**Theorem 9.4** ([CM2]). Suppose that $\Sigma^n \subset \mathbb{R}^{n+1}$ is a smooth closed embedded shrinker, but is not a sphere. There exists an open neighborhood $O = O_{\Sigma}$ of $\Sigma$ and a subset $W$ of $O$ so that the following hold.

- There is a splitting $E = E_1 \oplus E_2$ with $\dim(E_1) > 0$ so that $W$ is contained in the graph $(x, u(x))$ of a continuous mapping $u : E_2 \to E_1$.
- If $\Gamma \in O \setminus W$, then the rescaled MCF starting at $\Gamma$ leaves $O$ and the orbit of $O$ under the group of conformal linear transformations of $\mathbb{R}^{n+1}$.

The space $E_2$ is, loosely speaking, the span of all the contracting directions for the flow together with all the directions tangent to the action of the conformal linear group. It turns out that all the directions tangent to the group action are expanding directions for the flow.

Recall that the (local) stable manifold is the set of points $x$ near the fixed point so that the flow starting from $x$ is defined for all time, remains near the fixed point, and converges to the fixed point as $t \to \infty$. Obviously, Theorem 9.4 implies that the local stable manifold is contained in $W$.

There are several earlier results that analyze rescaled MCF near a closed shrinker, but all of these are for round circles and spheres which are stable under the flow. The earliest are the global results of Gage and Hamilton [GH] and Huisken [H1], mentioned earlier, showing that closed embedded convex hypersurfaces flow to spheres. There is also a stable manifold theorem of Epstein and Weinstein [EW] from the late 1980s for the curve shortening flow that also applies to closed immersed self-shrinking curves, but does not incorporate the group action. In particular, for something to be in Epstein–Weinstein’s stable manifold, then under the rescaled flow it has to limit into the given self-shrinking curve. In other words, for a curve to be in their stable manifold, it is not enough that it limit into a rotation, translation, or dilation of the self-shrinking curve.

**9.2. The heuristics of the local dynamics.** We will very briefly explain the underlying reason for this theorem about the local dynamics near a closed shrinker and why it is an infinite-dimensional and nonlinear version of the simple finite-dimensional examples we discussed earlier.

Suppose $\Sigma$ is a manifold and $h$ is a function on $\Sigma$. Let $w_i$ be an orthonormal basis of the Hilbert space $L^2(\Sigma, e^h \, d\text{Vol})$, where the inner product is given by $\langle v, w \rangle = \int_{\Sigma} v \, w \, e^h \, d\text{Vol}$.

\[\text{Recall that the group of conformal linear transformations of } \mathbb{R}^{n+1} \text{ is generated by the rigid motions and the dilations.}\]
\[ \int_{\Sigma} v \, w \, e^{h} \, d\text{Vol}. \]

For constants \( \mu_i \in \mathbb{R} \), define a function \( f \) on the infinite-dimensional space \( L^2(\Sigma, e^{h} \, d\text{Vol}) \) as follows: If \( w \in L^2(\Sigma, e^{h} \, d\text{Vol}) \), then

\[
(9.5) \quad f(w) = \sum_i \frac{\mu_i}{2} \langle w, w_i \rangle^2.
\]

As in the finite-dimensional case, the negative gradient flow of \( f \) is

\[
(9.6) \quad \Psi_t(w) = e^{-\mu_i t} \langle w_i, w \rangle.
\]

Of particular interest is when \( \Sigma^n \subset \mathbb{R}^{n+1} \) is a shrinker, \( h(x) = -\frac{|x|^2}{4} \), and the basis \( w_i \) are eigenfunctions with eigenvalues \( \mu_i \) of a self-adjoint operator \( L \) of the form

\[
(9.7) \quad L \, w = \Delta \, w + |A|^2 \, w - \frac{1}{2} \langle x, \nabla w \rangle + \frac{1}{2} \, w.
\]

The reason this is of particular interest is because in [CM1] it was shown that the Hessian of the \( F \)-functional is given by

\[
(9.8) \quad \text{Hess}_F(v, w) = - (4 \pi)^{-\frac{3}{2}} \int_{\Sigma} v \, L \, w \, e^{-\frac{|x|^2}{4}}.
\]

For an \( f \) of this form, the negative gradient flow is equal to the heat flow of the linear heat operator \((\partial_t - L)\). Moreover, this linear heat flow is the linearization of the rescaled MCF at the shrinker. It follows that the rescaled MCF near the shrinker is approximated by the negative gradient flow of \( f \). This same fact is also reflected by the fact that if we formally write down the first three terms in the Taylor expansion of \( F \), then we get the value of \( F \) at \( \Sigma \) plus a first-order polynomial which is zero since \( \Sigma \) is a critical point of \( F \) plus a polynomial of degree 2 which is given by the Hessian of \( F \) and is exactly \( f \). This gives a heuristic explanation for the above theorem: the dynamics of the negative gradient flow of the \( F \) functional should be well approximated by the dynamics for its second-order Taylor polynomial.

10. Surgery in Low Dimensions

Surgery theory originated in the seminal paper of M. Kervaire and J. Milnor [KM] in which they classified smooth manifolds homotopy equivalent to a sphere, and developed the basic surgery techniques in the simply connected case. These methods were taken up by W. Browder and S. Novikov [N]. Browder’s point of view was to study the question of existence of a manifold homotopy equivalent to a given space, whereas Novikov’s approach was to investigate whether a given homotopy equivalence is homotopic to a diffeomorphism. D. Sullivan [Su] realized that existence and uniqueness are just two sides of the same question and formulated the theory in terms of a surgery exact sequence. Finally, the theory was vastly generalized to also treat nonsimply connected manifolds by C.T.C. Wall [Wa]. The theory only works in dimensions bigger than 4, but there are nevertheless a few things that work in all dimensions. In the first few sections we give a description of the surgery exact sequence and discuss low-dimensional phenomena. This leads to a description of the Kirby–Siebenmann obstruction to triangulate topological manifolds, it gives a fairly simply proof of topological invariance of Pontrjagin classes, and finally it discusses to what extent surgery works in dimension 4.

The well-known surgery exact sequence of Browder–Novikov–Sullivan–Wall [Wa] breaks down in dimensions below 5. In this section we discuss what remains in low dimensions, and some of the implications this has in higher dimensions.
Surgery deals with existence and uniqueness of manifold structures on a given Poincaré Duality space. We shall not discuss existence, since this is precisely what breaks down in low dimensions, and if a Poincaré Duality space is homotopy equivalent to a manifold, we might as well replace the Poincaré Duality space by a manifold.

The first observation we make is that the terms in the surgery exact sequence are defined in all dimensions.

**Definition 10.1.** Let $M$ be a compact manifold without boundary. An element in the structure set is a homotopy equivalence

$$M_1 \simeq M$$

of manifolds. Two such elements are equivalent if there is a homotopy commutative diagram of manifolds

\[
\begin{array}{ccc}
M_1 & \rightarrow & W \\
\downarrow & & \downarrow \\
M_2 & \rightarrow & M \\
\end{array}
\]

where the inclusions of $M_i \subset W$ are homotopy equivalences and $\partial W$ is the disjoint union of $M_1$ and $M_2$. Such a $W$ is called an $h$-cobordism between $M_1$ and $M_2$.

The structure set is usually denoted $S_h(M)$, and it comes in smooth, PL, and topological versions. It can also be varied by requiring the homotopy equivalences to be simple, in this note however, we shall stick to just homotopy equivalences.

There is a relative version of the structure set where we allow $M$ to have a boundary and require $h : M_1 \rightarrow M$ to be a homeomorphism (PL-homeomorphism, diffeomorphism) on the boundary. This gives rise to the higher structure sets $S(M \times D^n \text{ rel. } \partial)$. In this note we shall not be considering further relativization (e.g., allowing homotopy equivalence on a part of the boundary).

The next term in the surgery exact sequence is the normal invariant. There are many mistakes in the literature concerning basepoints, so to be precise we shall consider the normal invariant to be the set of homotopy classes of based maps from $M_+$ (M with a disjoint basepoint) to $G/O$ in the smooth case, $G/PL$ in the PL case, and $G/Top$ in the topological case. In case $M$ has a boundary, the normal invariant is $[M/\partial M, G/O]_*$, so if $M$ is closed $[M \times D^n/\partial, G/O]_*$ is the same as $[\Sigma^n(M_+), G/O]_*$, which is the reason to prefer $[M_+, G/O]_*$ to the free homotopy classes $[M, G/O]$ even though it is the same.

There is a map $S(M) \rightarrow [M_+, G/O]_*$ and similarly for the Top and PL cases defined as follows: given a homotopy equivalence $M_1 \simeq M$ of smooth manifolds, the uniqueness of the Spivak normal fibration \[S\] produces a commutative diagram

\[
\begin{array}{ccc}
M_1 & \rightarrow & M \\
\downarrow & & \downarrow \\
M & \rightarrow & BG \\
\end{array}
\]
The smooth structures on $M_1$ and $M$ produce lifts to $BO$, and since $BO$ and $BG$ are loop spaces this produces a map to $G/O$, the fiber. In case $M$ has a boundary, these lifts agree on the boundary, so we get a basepoint preserving map $M/\partial M \to G/O$.

The last term in the surgery exact sequence is the Wall group. To indicate this, consider a degree 1 normal map

$$
n_M : M_1 \longrightarrow \xi
$$

i.e., a map sending the fundamental class of $M_1$ to the fundamental class of $M$, covered by a map of bundles from the normal bundle of $M_1$ to some bundle over $M$ (vector bundle, Top-bundle, or PL-bundle). The problem of surgery is to produce a bordism $W$ from $M_1$ to $M_2$ covered by a map of bundles, such that $M_2 \to M$ is a homotopy equivalence. Classically, this is done by making the map $M_1 \to M$ highly connected by surgery, i.e., whenever there is an obstruction

$$
S^i \longrightarrow D^{i+1}
$$

to the map being a homotopy equivalence, $S^i \to M$ is replaced by an embedding $S^i \times D^{n-i} \to M$ (which exists when $i$ is small, the extension of $S^i \to M$ to $S^i \times D^{n-i} \to M$ is assured by the bundle information), and then replace $S^i \times D^{n-i}$ by $D^{i+1} \times S^{n-i-1}$. Actually, we need the bordism covered by bundle maps, but the bordism is obtained by gluing $M \times I$ and $D^{i+1} \times D^{n-i}$ along $S^i \times D^{n-i} \subset M \times 1$ and $S^i \times D^{n-i} \subset \partial(D^{i+1} \times D^{n-i})$. This process breaks down when we reach the middle dimension, and the obstruction is not understood in low dimensions. Using Ranickis algebraic surgery point of view [R1] however, we can immediately pass to algebra, and we then obtain an algebraically defined obstruction in an algebraically defined group, $L^h_n(\mathbb{Z}\pi)$, $\pi = \pi_1(M)$, which is an obstruction in all dimensions, but maybe not the whole obstruction in low dimensions.

There is a map

$$
[M_+ \ast G/O] \ast \to L^h_n(\mathbb{Z}\pi)
$$

defined as follows. Embed the smooth manifold in $\mathbb{R}^k$, $k$ large, let $D(M)$ be the normal disk bundle, and let $S(M)$ be the normal sphere bundle. The Thom map is the map from $S^k$ thought of as $\mathbb{R}^k$ with an extra point at infinity to the Thom space $D(M)/S(M)$ sending everything outside $D(M)$ to the collapsed point. A map $M \to G/O$ gives rise to another bundle $\xi$ over $M$ together with a fiber homotopy equivalence of the corresponding sphere bundles. Consider the composite $S^k \to D(M)/S(M) \to D(\xi)/S(\xi)$. Making this map transverse to the 0-section in $\xi$ produces a manifold $M_1$, and a bundle map from the normal bundle of $M_1$ to $\xi$. The fundamental class of $M_1$ is sent to the fundamental class of $M$ by a Thom isomorphism argument, so we have a surgery problem, and we now pass to algebra to produce an element in $L^h_n(\mathbb{Z}\pi)$.
The bottom part of the surgery exact sequence

\[ S_h(M) \to [M_+, G/O] \to L_n^h(\mathbb{Z}\pi) \]

is now established, and similarly for the case with boundary. The composite is obviously the zero map since starting with a homotopy equivalence there is no obstruction to obtaining a homotopy equivalence. In low dimensions, however, it may not be true that we can perform the surgery to obtain a homotopy equivalent manifold, even though the algebraically defined surgery obstruction vanishes. In dimensions at least 5 the necessary embedding theorems are available so we have exactness in the sense of pointed sets. There is no group structure on the structure set, but in the case when the manifold is of the form \( N \times I \), there is an obvious associative monoid structure on the structure set, a group structure on the other terms, and the maps are indeed homomorphisms. In dimensions at least 5 this monoid structure is easily seen to be a group structure.

The sequence extends to the left as follows: given a closed smooth manifold \( M^n \) with fundamental group \( \pi \) and an element in \( L_{n+1}(\mathbb{Z}\pi) \), one may attempt to produce a bordism

\[ W \to M \times I \]

with \( \partial W \) the disjoint union of \( M \) and \( M_1 \), covered by bundle maps, realizing the given surgery obstruction, with \( M \to M \) the identity and \( M_1 \to M \) a homotopy equivalence. The idea is now to send the given element to this homotopy equivalence thought of as an element in the structure set. This is always possible when the dimension of \( M \) is at least 5. In low dimensions it may or may not be possible, and it is not clear whether we get a well-defined map \( L_{n+1}^h(\mathbb{Z}\pi) \to S_h(M) \) either in low dimensions. So in general we only get a partially defined, maybe not well-defined, action of \( L_{n+1}(\mathbb{Z}\pi) \) on \( S(M) \).

Given an element in \( [\Sigma M_+, G/O]_* \), a transversality argument as above produces a bordism \( W \) and a map \( W \to M \times I \) covered by bundle maps with \( \partial W \) the disjoint union of \( M \) and \( M \) and the map on the boundary the identity of \( M \) to \( M \times 0, M \times 1 \) respectively. We thus get

\[ [\Sigma M_+, G/O] \longrightarrow L_{n+1}^h(\mathbb{Z}\pi) \longrightarrow S(M). \]

The second map is defined on the image of the first map, and the composite is obviously trivial.

It should now be clear how to extend the sequence to the left, and since it will be a manifold with an \( I \)-factor it will be groups and homomorphisms to the left.

The surgery groups of the trivial group were essentially calculated by Kervaire and Milnor \([KM]\) to be \( \mathbb{Z} \) for \( n \equiv 0(4) \), \( \mathbb{Z}/2 \) for \( n \equiv 2(4) \) and 0 otherwise. The surgery obstruction for \( n \equiv 0(4) \) is given as follows. Do surgery below the middle dimension (also possible in dimension 4); the kernel in homology is now an even, symmetric nonsingular matrix over \( \mathbb{Z} \). Such a matrix has index divisable by 8, and the obstruction is the index divided by 8. This can also be calculated as the difference of the indexes of the manifolds divided by 8.

Since every PL-manifold in dimension less than 7 is smoothable, Rohlins theorem states that if \( M^4 \) is a smooth or PL 4-manifold with \( w_1 = w_2 = 0 \), then the index of \( M \) is divisable by 16. Index=16 however, is realizable using the Kummer surface, and using a connected sum, we may realize every multiple of 16.
11. The surgery exact sequence for a disk

Now consider the PL disk. The surgery exact sequence is

\[ \pi_{n+1}(G/PL) \to L_{n+1}(Z) \to S(D^n \text{ rel. } \partial) \to \pi_n(G/PL) \to L_n(Z) \]

when \( n \geq 5 \) consider an element in \( S(D^n \text{ rel. } \partial) \)

\[ \begin{array}{cccc}
S^{n-1} & \xrightarrow{id} & S^{n-1} \\
\downarrow & & \downarrow \\
W & \longrightarrow & D^n 
\end{array} \]

Cutting out a small open disc in the interior of \( W \) produces an \( h \)-cobordism, hence a product, so \( W \) is a disk. Using a cone construction, we get a PL-homeomorphism to \( D^n \) homotopic to the original map relative to the boundary. This latter part is of course what does not work in the smooth case. This shows \( \pi_n(G/PL) \cong L_n(Z) \) for \( n \geq 5 \). For \( n < 5 \), Kervaire and Milnor use that PL/\( O \) is 6-connected and the \( J \)-homomorphism \( O \to G \) is known in low dimensions, showing that \( \pi_n(G/PL) = \begin{cases} 
Z & n \equiv 0(4), \\
Z/2 & n \equiv 2(4), \\
0 & n \equiv 1(2). 
\end{cases} \)

Let us analyze the picture in low dimensions

\[ \pi_5(G/O) \to L_5(Z) \to S(D^4 \text{ rel. } \partial) \to \pi_4(G/O) \to L_4(Z). \]

\( L_5(Z) = \pi_5(G/O) = 0 \) and \( \pi_4(G/O) = L_4(Z) = Z \), but the map by Rohlin’s theorem is multiplication by 2.

**Observation 1.** Surgery theory suggests \( S(D^4 \text{ rel. } \partial) = 0 \).

This is actually true as we shall see (in the equivalent case of \( S^4 \)) in the next section.

Next consider \( n = 3 \),

\[ \pi_4(G/O) \to L_4(Z) \to S(D^3 \text{ rel. } \partial) \to \pi_3(G/O) \to L_3(Z). \]

We have \( \pi_3(G/O) = L_3(Z) = 0, \pi_4(G/O) = L_4(Z) = Z \), but the same argument as before says the map is multiplication by 2, so we get

**Observation 2.** Surgery theory suggests \( S(D^3 \text{ rel. } \partial) = Z/2 \).

This is of course not true by Perelman’s work.

In dimension 2 we have

\[ \pi_3(G/O) \to L_3(Z) \to S(D^2 \text{ rel. } \partial) \to \pi_2(G/O) \to L_2(Z), \]

and the map \( \pi_2(G/O) \to L_2(Z) \) is an isomorphism since the Kervaire invariant element is realized by \( T^2 \to S^2 \) covered by bundles defined by the left invariant framing of \( T^2 \).

**Observation 3.** Surgery theory suggests \( S(D^2 \text{ rel. } \partial) = 0 \).

This is of course correct.

In dimension 1 we have

\[ \pi_2(G/O) \to L_2(Z) \to S(D^1 \text{ rel. } \partial) \to \pi_1(G/O) \to L_1(Z), \]

and the map \( \pi_2(G/O) \to L_2(Z) \) again is an isomorphism for the same reason as above.
Observation 4. Surgery theory suggests $S(D^1 \text{ rel. } \partial) = 0$.

This is of course correct.

Finally, in dimension 0 we have

$$\pi_1(G/O) \to L_1(\mathbb{Z}) \to S(D^0 \text{ rel. } \partial) \to \pi_0(G/O) \to L_0(\mathbb{Z}).$$

Leading to the obvious

Observation 5. Surgery theory suggests $S(D^0) = 0$.

These low-dimensional phenomena do not make sense since surgery does not work in low dimensions, but they do have consequences in higher dimensions and lead to relatively simple proofs of topological invariance of Pontrjagin classes, and the theory of topological manifolds as developed by Kirby and Siebenmann.

12. Dimension 4

Let $M^4$ be a closed smooth 4-dimensional manifold.

**Theorem 12.1.** Assume $[\Sigma(M_+), G/O]_* \to L_5(\mathbb{Z})$ is an epimorphism and $[M_+, G/O]_* \to L_4(\mathbb{Z})$ is a monomorphism. Then $S(M) = 0$.

**Proof.** An element in $S(M)$ is a homotopy equivalence $M_1 \to M$. The composite map to $L_4(\mathbb{Z})$ is always zero so $S(M) \to [M_+, G/O]_*$ must be the zero map. This means there is a normal cobordism $W$ with boundary the disjoint union of $M_1$ and $M$, and a degree 1 normal map $W \to M \times I$. This normal map has a surgery obstruction $\sigma \in L_5(\mathbb{Z})$. Now choose an element in $[\Sigma(M_+), G/O]_*$ which hits $-\sigma$. This produces a normal cobordism $W_1 \to M \times I$ which is the identity on the boundary. Gluing $W$ and $W_1$ together, the surgery obstruction is 0, but we are in dimension 5, so we may perform the surgery to obtain an $h$-cobordism between $M_1$ and $M$ and a homotopy equivalence to $M \times I$. $\square$

**Corollary 12.2.** Assume $M$ is a closed 4-dimensional manifold homotopy equivalent to $S^4$. Then $M$ is $h$-cobordant to $S^4$.

**Proof.** $L_5(\mathbb{Z}) = 0$, so the first condition in the theorem above is obviously satisfied. $G/O$ is connected and simply connected, so $[S^4_+, G/O]_* = \pi_4(G/O) = \mathbb{Z}$, and the map to $L_4(\mathbb{Z}) = \mathbb{Z}$ is multiplication by 2 by Rohlin's theorem as above. This means $S(S^4) = 0$, hence $M$ is $h$-cobordant to $S^4$. $\square$

**Theorem 12.3.** If $M$ is a closed 4-dimensional manifold homotopy equivalent to $S^4$, then $M$ be smoothly embedded in $\mathbb{R}^5$.

**Proof.** We want to show that $M$ can be smoothly embedded in $\mathbb{R}^5$. By Corollary 12.2 above there is an $h$-cobordism $W$ between $M$ and $S^4$. Now $W \times S^1$ is a 6-dimensional $h$-cobordism and $\text{Wh}(\mathbb{Z}) = 0$, so $M \times S^1$ is diffeomorphic to $S^4 \times S^1$. Hence, the universal covers $M \times \mathbb{R}$ and $S^4 \times \mathbb{R}$ are diffeomorphic, but $S^4 \times \mathbb{R}$ is diffeomorphic to $\mathbb{R}^5 \setminus 0$. $\square$
13. Bounded surgery

The material in this section is taken from [FP].

Definition 13.1. Let $X$ be a metric space, let $M_1$ and $M$ be topological spaces, and let $p : M_1 \to M$ be a proper map. Then a map $f : M_1 \to M$ is said to be a bounded homotopy equivalence if there is a map $g : M \to M_1$ and homotopies $G$ from $g \circ f$ to the identity and $H$ from $f \circ g$ to the identity such that $p(H(z \times I))$ and $p(f(K(y \times I)))$ have uniformly bounded diameter.

Remark 13.2. This concept of course is only interesting when $M_1$ and $M$ are non-compact.

In [FP] a bounded surgery theory was developed classifying manifolds up to bounded homotopy equivalence for a large class of metric spaces $X$, e.g., $X = \mathbb{R}^n$, and a surgery exact sequence was established in dimensions $\geq 5$. The normal invariant term is the same as in the compact case, but the $L$-group term was defined using certain additive categories that give an algebraic criterion for bounded homotopy equivalence similarly to the Whitehead theorem in the classical case. This additive category is denoted $C_X(\mathbb{Z})$ in the simply connected case. We thus have an exact sequence

$$\cdots \to [\Sigma(M_+), G/O]_* \to L_{n+1}(C_X(\mathbb{Z})) \to S_b(M \to X) \to [M_+, G/O]_* \to L_n(C_X(\mathbb{Z})).$$

An especially interesting case is $X = \mathbb{R}^n$ and $M = D^k \times \mathbb{R}^n$ with $p$ projection on the second factor. In this case our observations above lead to

Theorem 13.3. When $n + k \geq 5$, the bounded structure set

$$S(D^k \times \mathbb{R}^n \to \mathbb{R}^n \text{ rel. } \partial)$$

is 0 when $k \neq 3$, and $\mathbb{Z}/2$ when $n = 3$.

Proof. Crossing with $\mathbb{R}^n$ is obviously an isomorphism on the normal invariant term, since this term is homotopy theoretic and $\mathbb{R}^n$ is contractible. One may prove algebraically [R2] that crossing with $\mathbb{R}^n$ induces an isomorphism $L_k(\mathbb{Z}) \to L_{n+k}(C_{\mathbb{R}^n}(\mathbb{Z}))$, but then the observations above calculate the maps in the bounded surgery sequence which is exact when $n + k \geq 5$. \qed

To study the differences between PL and Top manifolds, one considers $\pi_k(\widetilde{\text{Top}}(n), \widetilde{\text{PL}}(n))$. An element is given by a homeomorphism $\Delta(k) \times \mathbb{R}^n \to \Delta(k) \times \mathbb{R}^n$, where $\Delta(k)$ is a standard $k$-simplex, restricting to a PL-homeomorphism $\sigma \times \mathbb{R}^n \to \sigma \times \mathbb{R}^n$ for each face $\sigma$ of $\Delta(k)$. Such an element is 0 if there is an isotopy, fixing the boundary, to a PL-homeomorphism. If all such elements were 0, a relatively simple argument due to Kirby would produce a PL-structure on a topological manifold. In this direction we get

Theorem 13.4. When $n + k \geq 5$ there is a monomorphism

$$\pi_k(\widetilde{\text{Top}}(n), \widetilde{\text{PL}}(n)) \to S_b(D^k \times \mathbb{R}^n \to \mathbb{R}^n).$$

Since we have calculated the target, this means that $\pi_k(\widetilde{\text{Top}}, \widetilde{\text{PL}})$ is 0 for $k \neq 3$, and for $k = 3$ it is either 0 or $\mathbb{Z}/2$. Notice this implies topological invariance of Pontrjagin classes since the stable maps $BO \to B\text{PL}$ and $B\text{PL} \to B\text{Top}$ are then both rational homotopy equivalences.
Proof. To define the map, consider a homeomorphism \( \Delta(k) \times \mathbb{R}^n \to \Delta(k) \times \mathbb{R}^n \) which restricts to a PL-homeomorphism on each simplex in the boundary crossed with \( \mathbb{R}^k \). Hence to a PL-homeomorphism on the boundary. Identifying \( D^k \) and \( \Delta(k) \), a homeomorphism is obviously a bounded homotopy equivalence, bounded by 0, so it defines an element in the structure set (which is a group in this case). To see this map is monic, we change the metric on \( D^k \times \mathbb{R}^n \) such that \( D^k \times \{x\} \) has the standard metric on \( D^k \) multiplied by \(|x| + 1\). Let us denote this metric space by \( X \). We now have a proper Lipschitz map \( X \to \mathbb{R}^n \), and it turns out that it induces an isomorphism of structure sets \( S(D^k \times \mathbb{R}^n \to X) \to S(D^k \times \mathbb{R}^n \to \mathbb{R}^n) \). This is a five lemma argument. On the normal invariant term the metric space plays no role, and on the \( L \)-group term an algebraic argument provides the necessary argument; see [FP]. Now consider an element \( h \) going to 0 in \( S(D^k \times \mathbb{R}^n \to \mathbb{R}^n) \). The above-mentioned map obviously factors through \( S(D^k \times \mathbb{R}^n \to X) \) since a homeomorphism is a bounded homotopy equivalence no matter what metric space it is measured in. This means there is a bounded isotopy relative to the boundary of \( h \) to a PL-homeomorphism \( g \), which is a uniformly bounded distance from the identity when measured in \( X \). Now consider \( h \circ g^{-1} : D^k \times \mathbb{R}^n \to D^k \times \mathbb{R}^n \). This is the identity on the boundary. Thinking of \( D^k \times \mathbb{R}^n \) as an open subset of \( D^k \times D^n \) in a radial way, one sees that this homeomorphism may be completed to a homeomorphism of \( D^k \times D^n \) by the identity because boundedness measured in \( X \) translates to smallness near the boundary of \( D^n \). An Alexander isotopy now provides an isotopy relative to the boundary of \( h \circ g^{-1} \) to the identity, so \( h \) is isotopic to \( g \).

It can be proved that \( \pi_3(\widetilde{\text{Top}}(n), \widetilde{\text{PL}}(n)) = \mathbb{Z}/2 \) when \( n+k \geq 6 \). In other words the above map is onto, but this requires a proof that a certain bounded homotopy equivalence is bounded homotopic to a homeomorphism using Quinn's end theorem, and is probably not easier than Kirby and Siebenmann's original argument.

Acknowledgments

The authors are grateful to Dave Gabai for his interest and numerous stimulating discussions.

About the authors

Tobias Colding and William Minicozzi are professors at the Massachusetts Institute of Technology. Their research is in analysis and geometry. Erik Pedersen is professor and chair at the University of Copenhagen, working on problems in topology and related fields.

References


Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307

E-mail address: colding@math.mit.edu

Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307

E-mail address: minicozz@math.mit.edu

Department of Mathematics, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark

E-mail address: erik@math.ku.dk