The Inverse Mean Curvature Flow and the Riemannian Penrose Inequality

Gerhard Huisken[1] and Tom Ilmanen[2]

In this paper we develop the theory of weak solutions for the inverse mean curvature flow of hypersurfaces in a Riemannian manifold, and apply it to prove the Riemannian Penrose inequality for a connected horizon, to wit: the total mass of an asymptotically flat 3-manifold of nonnegative scalar curvature is bounded below in terms of the area of each smooth, compact, connected, "outermost" minimal surface in the 3-manifold. A minimal surface is called *outermost* if it is not separated from infinity by any other compact minimal surface. The result was announced in [HI1].

Let M be a smooth Riemannian manifold of dimension $n \geq 2$ with metric $g = (g_{ij})$. A classical solution of the inverse mean curvature flow is a smooth family $x : N \times [0, T] \to M$ of hypersurfaces $N_t := x(N, t)$ satisfying the parabolic evolution equation

$$(\star) \qquad \frac{\partial x}{\partial t} = \frac{\nu}{H}, \qquad x \in N_t, \qquad 0 \leq t \leq T,$$

where H, assumed to be positive, is the mean curvature of N_t at the point x, ν is the outward unit normal, and $\partial x/\partial t$ denotes the normal velocity field along the surface N_t.

Without special geometric assumptions (see [Ge,U,Ho]), the mean curvature may tend to zero at some points and singularities develop. In §1 we introduce a level-set formulation of $(\star)$, where the evolving surfaces are given as level-sets of a scalar function u via

$$N_t = \partial\{x : u(x) < t\},$$

and $(\star)$ is replaced by the degenerate elliptic equation

$$(\star\star) \qquad \operatorname{div}_M\left(\frac{\nabla u}{|\nabla u|}\right) = |\nabla u|,$$

This document was prepared using $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX.

[1]Mathematisches Institut, Auf der Morgenstelle 10, D-72076 Tübingen, Germany. Partially supported by Sonderforschungsbereich 382, Tübingen.

[2]Department of Mathematics, Northwestern University, Evanston IL 60201, USA, and Max-Planck Institut für Mathematik, Inselstrasse 22-26, D-04103 Leipzig, Germany. Partially supported by an NSF Postdoctoral Fellowship, NSF Grants DMS-9626405 and DMS-9708261, and a Sloan Foundation Fellowship.

where the left hand side describes the mean curvature of the level-sets and the right hand side yields the inverse speed. This formulation in divergence form admits locally Lipschitz continuous solutions and is inspired by the work of Evans-Spruck [ES] and Chen-Giga-Goto [CGG] on the mean curvature flow.

Using a minimization principle similar to those of Lichnewsky-Temam [LT], Luckhaus [Lu], Zhou and Hardt-Zhou [Z,HZ], and Visintin [V], together with elliptic regularization, in §2-3 we prove existence and uniqueness of a Lipschitz continuous weak solution of ($\star\star$) having level-sets of bounded nonnegative weak mean curvature. The existence result (Theorem 3.1) assumes only mild conditions on the underlying manifold M, with no restrictions on dimension. The minimization principle used in the definition of weak solutions of ($\star\star$) allows the evolving surfaces to jump instantaneously over a positive 3-volume (momentary "fattening"), which is desirable for our main application. See Pasch [Pa] for numerical computation of this phenomenon.

The minimization property is also essential for the uniqueness, compactness and regularity properties of the solution. Furthermore, it implies connectedness of the evolving surfaces (§4), which is essential in our argument.

The Riemannian Penrose Inequality. In §5-§8 we employ the inverse mean curvature flow in asymptotically flat 3-manifolds to prove the Riemannian Penrose Inequality as stated below.

An end of a Riemannian 3-manifold (M, g) is called *asymptotically flat* if it is realized by an open set that is diffeomorphic to the complement of a compact set K in $\mathbf{R}^3$, and the metric tensor g of M satisfies

$$|g_{ij} - \delta_{ij}| \leq \frac{C}{|x|}, \qquad |g_{ij,k}| \leq \frac{C}{|x|^2}, \tag{0.1}$$

as $|x| \to \infty$. The derivatives are taken with respect to the Euclidean metric $\delta = (\delta_{ij})$ on $\mathbf{R}^3 \setminus K$. In addition, we require the Ricci curvature of M to satisfy

$$Rc \geq -\frac{Cg}{|x|^2}. \tag{0.2}$$

Following Arnowitt, Deser and Misner [ADM] the ADM mass, or total energy, of the end is defined by a flux integral through the sphere at infinity,

$$m := \lim_{r\to\infty} \frac{1}{16\pi} \int_{\partial B_r^\delta(0)} (g_{ii,j} - g_{ij,i}) n^j \, d\mu_\delta. \tag{0.3}$$

This is a geometric invariant of the given end, despite being expressed in coordinates. It is finite precisely when the the scalar curvature R of g satisfies

$$\int_M |R| < \infty.$$

See Bartnik [B] and Chruściel [Ch3] for these facts, detailed below in Lemma 7.3, and see [AM,CK,F,Wa1,Y] for further discussion of asymptotic flatness.

The Riemannian case of the *Positive Mass Theorem*, first proven by Schoen and Yau [SY3], states that an aymptotically flat 3-manifold of nonnegative scalar curvature (possibly with a compact, minimal boundary) has

$$m \geq 0,$$

with equality only in the case of Euclidean space. A number of other proofs and approaches to this theorem have appeared in the physics and mathematics literature: these include Schoen-Yau [SY2,SY9-SY11] by the method of minimal surfaces and conformal changes; Geroch [G] and Jang [J1-J3] introducing the inverse mean curvature flow; Jang [J4], Kijowski [K1,K2], Jezierski-Kijowski [JK1,JK2], Chruściel [Ch1,Ch2,Ch4,Ch6], and Jezierski [Je1-Je3] using the p-Laplacian; Witten [W], Parker-Taubes [PT], Reula [R], Choquet-Bruhat [C], Reula-Tod [RT], and Chruściel [Ch5] using harmonic spinors; Penrose-Sorkin-Woolgar [PSW] using a single complete null geodesic; Lohkamp [Lo] for a recent Riemannian geometry approach; and others. See also the survey article by Lee-Parker [LP]. Such analytic methods have many additional consequences for the geometry and topology of positive scalar curvature manifolds; see [SY1,SY5,SY6] and [Lc,GL1-GL3].

In 1973, motivated by a physical argument (see below), Penrose [P2] conjectured that

$$16\pi m^2 \geq |N|,$$

where $|N|$ is the area of a minimal surface in M. Partial or related results have been obtained by Gibbons [G1,G4] in the case of thin shells using quermassintegrals, Herzlich [He] via the Dirac operator, Bartnik [Ba6] for metrics with quasi-spherical foliations, Jezierski [Je1,Je3] in certain cases using the 5-Laplacian, and Bray [B1] using isoperimetric regions. For further information about the role of isoperimetric regions in positive mass, see [CY,HY].

The following theorem is the main result of this paper.

Main Theorem (Riemannian Penrose Inequality). *Let M be a complete, connected 3-manifold. Suppose that*

(i) M has nonnegative scalar curvature,

(ii) M is asymptotically flat satisfying (0.1) *and* (0.2), *with ADM mass m,*

(iii) The boundary of M is compact and consists of minimal surfaces, and M contains no other compact minimal surfaces.

Then

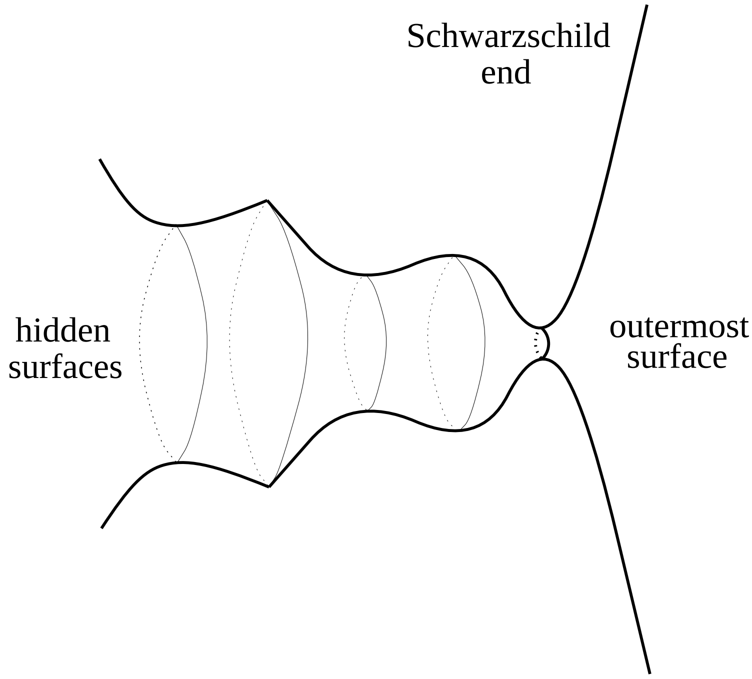
(0.4)
$$m \geq \sqrt{\frac{|N|}{16\pi}},$$

where $|N|$ is the area of any connected component of ∂M. Equality holds if and only if M is isometric to one-half of the spatial Schwarzschild manifold.

The *spatial Schwarzschild manifold* is the manifold $\mathbf{R}^3 \setminus \{0\}$ equipped with the metric $g := (1 + m/2|x|)^4\delta$. Note that it possesses an inversive isometry fixing the sphere $\partial B_{m/2}(0)$, which is an area-minimizing sphere of area $16\pi m^2$. The manifold M to which the theorem applies is then $\mathbf{R}^3 \setminus B_{m/2}(0)$. It is the standard spacelike slice of the exterior region of the Schwarzschild solution of the Einstein vacuum equations, describing a static, rotationally symmetric black hole.

The second clause of condition (iii) seems restrictive at first. Suppose, however, that M is asymptotically flat and has compact, minimal boundary. Then (iii) is satisfied by the infinite portion of M that stands outside the union of all compact minimal surfaces in M (Lemma 4.1). Accordingly, a connected, asymptotically flat 3-manifold that satisfies conditions (i)-(iii) is called an *exterior region*. An exterior region has the topology of $\mathbf{R}^3$ minus a finite number of balls, and its boundary consists of area-minimizing 2-spheres. These facts, well known in the literature, are collected in section §4.

Condition (iii) ensures that ∂M is not shielded from spacelike infinity by a smaller minimal surface further out. It is easy to construct a rotationally symmetric example of positive scalar curvature in which the inequality holds for the "outermost" minimal surface, but fails for arbitrarily large minimal surfaces hidden beyond the outermost one. (See figure.) For an example of this shielding effect with several boundary components, see Gibbons [G3].



Large hidden surfaces.

Our result allows ∂M to be disconnected, but only applies to each *connected* component of ∂M. This limitation seems to be fundamental, due to the use of the Gauss-Bonnet formula in deriving the Geroch Monotonicity formula.[3]

Physical Interpretation. The manifold (M, g) arises as a spacelike hypersurface in an asymptotically flat Lorentzian manifold $L^{3,1}$ modelling a relativistic isolated gravitating system governed by Einstein's equations. In this setting, the result can be interpreted as an optimal lower bound for the total energy of the system measured at spacelike infinity in terms of the size of the largest black hole contained inside. The positive scalar curvature of M corresponds to the physical hypothesis of nonnegative energy density on L, expressed by the so-called *weak energy condition* ([HE,Wa1]), provided that M is a maximal hypersurface in L (see [Ba1]). If M is totally geodesic in L (at least along N), then the minimal surface N corresponds to a so-called *closed trapped surface* that presages the formation of a space-

[3]Using a novel and surprising technique, Bray [B2] has recently proven the most general form of the Riemannian Penrose Inequality, namely $16\pi m^2 \geq |\partial M|$ with no connectedness requirement on ∂M. This is a considerable strengthening. He employs a quasistatic flow that couples a moving surface with an evolving conformal factor; it bears a family resemblance to the Hele-Shaw flow.

time singularity and ultimately a black hole [HP,HE,Wa1]. The shielding effect corresponds to causal separation.

Penrose originally conjectured (0.4) as a consequence of a physical argument. He assumed that M could be taken as the momentarily static (i.e. totally geodesic) initial slice of a spacetime L with a well-defined, asymptotically flat lightlike infinity existing for all time, and invoked various physical hypotheses and principles including the strong or the dominant energy condition, the Hawking-Penrose Singularity Theorems, the Weak Cosmic Censorship Hypothesis, the Hawking Area Monotonicity Theorem, long-term settling to a stationary black hole, and the No-Hair Theorem (classification of stationary black holes), which he collectively referred to as the "establishment viewpoint" of black hole formation. From these he deduced a chain of inequalities leading to the purely Riemannian assertion (0.4) for the 3-manifold M. For details of this beautiful argument, see [P2,P3,JW], as well as relevant discussion in [H1,G,H3,G2,G3,To,Wa1].

Of these ingredients, the most controversial is the famous (Weak) Cosmic Censorship Hypothesis [P1,P3,Wa2], which asserts that in a generic asymptotically flat spacetime, any singularity is surrounded by a well-defined event horizon, so that it cannot be seen from lightlike infinity. In particular, the singularity cannot have any causal effects on the outside world. Penrose viewed (0.4) as a Riemannian test of Cosmic Censorship. Thus our result rules out a class of potential counterexamples to this important conjecture.

For further information on the physical interpretation, see the announcement [HI1]. See also the informative review article of Penrose [P3].

Outline of Proof. We now sketch our proof of the Penrose Inequality. (A more detailed overview appears in the announcement [HI1].) Hawking [H2] introduced the Hawking quasi-local mass of a 2-surface, defined by

$$m_H(N) := \frac{|N|^{1/2}}{(16\pi)^{3/2}} \left(16\pi - \int_N H^2 \right),$$

and observed that it approaches the ADM mass for large coordinate spheres. Geroch [G] introduced the inverse mean curvature flow and discovered that the Hawking mass of the evolving surface is monotone nondecreasing, provided that the surface is connected and the scalar curvature of M is nonnegative. This gave an argument in favor of positive mass. Jang-Wald [JW] realized that if there were a classical solution of the flow starting at the inner boundary and approaching large coordinate spheres as $t \to \infty$, the Geroch Monotonicity result would imply the Penrose Inequality.

The heart of our proof (§5) is to justify the Geroch Monotonicity Formula (5.8) for our weak solutions, even in the presence of jumps. In this step, it is crucial that the evolving surface start and remain connected, as was established in §4. In §6, we extend the Monotonicity Formula to each connected component of a multi-component boundary by exploiting the ability of the evolving surface to jump over regions in 3-space.

Section 7 establishes that the Hawking mass of the evolving surface is eventually bounded by the ADM mass, leading to the proof of the Main Theorem in §8.

The positive integrand on the right-hand side of the Geroch Monotonicity Formula can serve as a kind of "mass density" on M. It is noncanonical in that it depends on the evolving family of surfaces. Remarkably, it vanishes precisely on standard expanding spheres in flat space and Schwarzschild – thus the rigidity statement of the Main Theorem. We revisit the Schwarzschild example in §8 to show that the imputed mass distribution – in particular, whether it resides in the black hole or is distributed throughout the field surrounding it – depends on where you start the flow.

Though noncanonical, the "mass density" can often be estimated in terms of the local geometry of M, leading to explicit lower bounds for the ADM mass. We exploit this idea in §9 to study the quasi-local mass (or gravitational capacity) proposed by Bartnik [Ba3,Ba4,Ba5]. We prove two conjectures of Bartnik: first, the Bartnik capacity is positive on any nonflat fragment of space, and second, the ADM mass is the supremum of the Bartnik capacities of all compact subsets. We end with conjectures that refine those of Bartnik [Ba3,Ba5] about the convergence and existence of 3-manifolds that minimize the ADM mass.

In a future paper we will prove smoothness of the flow outside some finite region, convergence of the Hawking quasi-local mass to the ADM mass, and asymptotic convergence to the center of mass in the sense of Huisken-Yau [HY], under suitable asymptotic regularity for (M, g).

The Full Penrose Conjecture. Inequality (0.4) is subsumed under the space-time form of the Penrose Conjecture, in which the assumption on the extrinisic curvature of M within L is dropped. This allows M to carry nonzero momentum and angular momentum, and would generalize the space-time version of the positive mass theorem (see Schoen-Yau [SY4,SY7,SY8]). Let $E_{ij} := R^L_{ij} - R^L g^L_{ij}/2$ be the Einstein tensor of L. We say that L satisfies the *dominant energy condition* if $E(m,n) \geq 0$ for all future timelike vectors m, n.

Conjecture. *Let N be an "outermost" spacelike 2-surface with null mean curvature vector in an asymptotically flat Lorentzian manifold L satisfying the dominant energy condition. Then*

$$16\pi m^2 \geq |N|, \tag{0.5}$$

where m is the mass of L.

The mass of L is defined as $\sqrt{E^2 - P_1^2 - P_2^2 - P_3^2}$, where (E, P^1, P^2, P^3) is the ADM energy-momentum 4-vector of L (see [Wa1, p.293]). For the definition of asymptotically flat for a spacetime L, see [AM,CK,F,Wa1,Y].

A major puzzle is the correct formulation of "outermost". The careful 4-dimensional definition and analysis of a *trapping horizon* in Hayward [Ha] may provide clues. An obvious possibility is to require the existence of an asymptotically flat, spacelike slice extending from N to infinity and containing no compact surface (besides the boundary) with nonspacelike mean curvature vector. Of course, such a slice may be useful in proving the inequality. See Jang [J2], Schoen-Yau [SY8] for interesting and relevant calculations.

In this connection, it is worth noting that the Hawking mass is also monotone under the codimension-two inverse mean curvature flow

$$\frac{\partial x}{\partial t} = -\frac{\vec{H}}{|H|^2}, \qquad x \in N_t, \quad t \geq 0.$$

of a spacelike 2-surface with spacelike mean curvature vector in a Lorentz manifold satisfying the dominant energy condition, as can be verified by calculation. Unfortunately, this 2x2 system of evolution equations is backward-forward parabolic.

There are sharper versions of the Penrose Inequality for charged black holes [J3] and rotating black holes [P3, p.663]. The former can now be rigourously established using our Geroch Monotonicity Formula (5.24), as was pointed out to us by G. Gibbons, while the latter would require understanding space-time issues like those in the above conjecture.

Another important case treated by Penrose's original argument involves an infinitely thin shell of matter collapsing at the speed of light, surrounding a region of flat Minkowski space. Gibbons [G1,G4] reduces this to the Minkowski-like inequality

$$\int_N |\vec{H}| \geq 4\sqrt{\pi}|N|^{1/2}$$

for a spacelike 2-surface in $\mathbf{R}^{3,1}$ with inner-pointing, spacelike mean curvature vector $\vec{H}$, which he then establishes using recent work of Trudinger [Tr] on quermassintegrals. Accordingly, the Penrose Inequality has often been called the Isoperimetric Inequality for black holes.

Acknowledgements. We would like to thank B. Andrews, R. Bartnik, G. Belletini, P. Chruściel, M. Friedman, R. Geroch, G. Gibbons, R. Hamilton, M. Herzlich, J. Isenberg, P. Li, R. Schoen, R. M. Wald, S.-T. Yau, and R. Ye for stimulating and useful conversations. The paper started at the CIME meeting in Cetraro and the PDG meeting in Oberwolfach, in summer 1996.

§1 Variational Formulation of Inverse Mean Curvature Flow

Because the inverse mean curvature flow does not remain smooth, we give a weak formulation based on minimizing a functional. In general, the surfaces satisfying the weak formulation are $C^{1,\alpha}$, but jump discontinuously across a set of positive measure when a certain minimizing criterion ceases to be met. Nevertheless, the area of the surface evolves continuously, except possibly at $t = 0$.

Notation. Let N be a smooth hypersurface in the Riemannian manifold M with metric g. We write $h = (h_{ij})$ for the induced metric on N, ∇ for the connection on M, D for the induced connection on N, and Δ for the Laplace-Beltrami operator on N.

Suppose $N = \partial E$ where E is an open set in M, and let $\nu = (\nu^i)$ be the outward unit normal to E along N. Define the second fundamental form $A = (A^N_{ij})$ of N by

$$A(e_1, e_2) := \langle \nabla_{e_1}\nu, e_2 \rangle, \qquad e_1, e_2 \in T_xN, \quad x \in N,$$

mean curvature scalar $H := h^{ij}A_{ij}$, and mean curvature vector $\vec{H} := -H\nu$. For a standard ball, $H > 0$ and $\vec{H}$ points inward.

Elementary Observations. A simple example of inverse mean curvature flow is the expanding sphere $\partial B_{R(t)}$ in $\mathbf{R}^n$, where

$$R(t) := e^{t/(n-1)}.$$

Gerhardt [Ge] proved that any compact, starshaped initial surface remains star-shaped and smooth under the flow, and becomes an expanding round sphere as $t \to \infty$. See Gerhardt [Ge] or Urbas [U] for similar results for more general expanding flows. Some noncompact self-similar examples are given in [HI2].

Note that $(\star)$ is invariant under the transformation

$$x \mapsto \lambda x, \qquad t \mapsto t,$$

so t is unitless. This suggests that we will not get much regularity in time, and that any singularities encountered will have a purely spatial character.

Let $(N_t)_{0 \leq t < T}$ be a smooth family of hypersurfaces satisfying $(\star)$. Write v for the outward normal speed. By the first variation formula (see (1.11)), the area element evolves in the normal direction by

$$\frac{\partial}{\partial t} d\mu_t = Hv\, d\mu_t = d\mu_t. \tag{1.1}$$

Here $(\partial/\partial t)d\mu_t$ is the Lie derivative of the form $d\mu_t$ along a bundle of trajectories orthogonal to N_t. Therefore the area satisfies

$$\frac{d}{dt}|N_t| = |N_t|, \qquad |N_t| = e^t |N_0|, \qquad t \geq 0.$$

Next we derive the evolution of the mean curvature. For a general surface moving with normal speed v, by taking the trace of the Riccati equation, we get

$$\frac{\partial H}{\partial t} = \Delta(-v) - |A|^2 v - Rc(\nu,\nu)v, \tag{1.2}$$

which yields for $(\star)$,

$$\begin{aligned} \frac{\partial H}{\partial t} &= -\Delta\left(\frac{1}{H}\right) - \frac{|A|^2}{H} - \frac{Rc(\nu,\nu)}{H} \\ &= \frac{\Delta H}{H^2} - \frac{2|DH|^2}{H^3} - \frac{|A|^2}{H} - \frac{Rc(\nu,\nu)}{H}, \end{aligned} \tag{1.3}$$

where Rc is the Ricci curvature of M. This equation is cause for optimism, because in view of the fact that $|A|^2 \geq H^2/(n-1)$, the parabolic maximum principle yields the curvature bound

$$\max_{N_t} H^2 \leq \max_{N_0} H^2 + C, \tag{1.4}$$

as long as the Ricci curvature is bounded below and the flow remains smooth.

This estimate constrasts sharply with ordinary mean curvature flow (see [Hu]), where the $|A|^2$ term has a positive sign, unleashing a menagerie of singularities [I2]. Together with the variational formulation to be given below, the negative $|A|^2$ is the key to the regularity theory. Unfortunately, the same term has a tendency to cause $H \to 0$.

By a computation similar to [Hu], we find the following evolution equation for the norm of the second fundamental form

$$\frac{d}{dt}|A|^2 = \frac{1}{H^2}\bigg(\Delta|A|^2 - 2|DA|^2 + 4DH \cdot \frac{A}{H} \cdot DH + 2|A|^4 - 4H \operatorname{tr} A^3 \\ + A * A * Rm + A * \nabla Rm\bigg),$$

where $*$ represents a linear combination of contractions. The $|A|^4/H^2$ term suggests that the second fundamental form blows up, especially where $H \searrow 0$.

Torus Counterexample. We give an example to establish that smoothness cannot be preserved. Suppose that N_0 is a thin torus in $\mathbf{R}^3$, obtained as the boundary of an ε-neighborhood of a large round circle. Then $H > 0$ initially, and the flow exists for a short time. Now (1.4) yields a uniform lower bound for the speed, so the torus steadily fattens up, and if this were to continue, H would become negative in the donut hole. This shows that N_t cannot remain smooth, and suggests that the surface must inexorably change topology, as indeed does occur in our formulation.

Level-Set Description. Because it is desirable that H remain positive (or at least nonnegative), we impose a *unidirectional ansatz* that compels this. This is accomplished via a level-set formulation, inspired by Evans-Spruck [ES], Chen-Giga-Goto [CGG]. Assume that the flow is given by the level-sets of a function $u : M \to \mathbf{R}$ via

$$E_t := \{x : u(x) < t\}, \qquad N_t := \partial E_t.$$

Wherever u is smooth with $\nabla u \neq 0$, equation $(\star)$ is equivalent to

$$(\star\star) \qquad \operatorname{div}_M \left(\frac{\nabla u}{|\nabla u|} \right) = |\nabla u|,$$

where the left side gives the mean curvature of $\{u = t\}$ and the right side gives the inverse speed.

Weak Formulation. It appears that $(\star)$ is not a gradient flow, nor is $(\star\star)$ an Euler-Lagrange equation. Yet in the context of geometric measure theory, the mean curvature bound (1.4) calls out for a minimization principle. Freezing the $|\nabla u|$ term on the right-hand side, consider equation $(\star\star)$ as the Euler-Lagrange equation of the functional

$$J_u(v) = J_u^K(v) := \int_K |\nabla v| + v|\nabla u| \, dx.$$

The idea of freezing the time derivative and minimizing an elliptic functional is seen in [LT,Lu,HZ,V], where it is the the key to existence, compactness, and regularity.

Definition. Let u be a locally Lipschitz function on the open set Ω. Then u is a *weak solution (subsolution, supersolution respectively) of* $(\star\star)$ *on* Ω provided

$$(1.5) \qquad J_u^K(u) \leq J_u^K(v)$$

for every locally Lipschitz function v ($v \leq u$, $v \geq u$ respectively) such that $\{v \neq u\} \subset\subset \Omega$, where the integration is performed over any compact set K containing $\{u \neq v\}$. (It does not matter which such set we use, so we will usually drop the K.)

By virtue of the identity

$$J_u(\min(v, w)) + J_u(\max(v, w)) = J_u(v) + J_u(w),$$

whenever $\{v \neq w\}$ is precompact, applied with $w = u$, we see that u is a weak solution if and only if u is simultaneously a weak supersolution and a weak subsolution.

Equivalent Formulation. Let us particularize the above definition to individual level sets. If F is a set of locally finite perimeter, let $\partial^* F$ denote its reduced boundary. If a set X is supposed to be $(n-1)$-dimensional, write $|X|$ for the $(n-1)$-dimensional Hausdorff measure of X. Write $E \Delta F := (E \setminus F) \cup (F \setminus E)$ for the symmetric difference of sets E and F.

Given $K \subseteq M$ and u locally Lipschitz, define the functional

$$J_u(F) = J_u^K(F) := |\partial^* F \cap K| - \int_{F\cap K} |\nabla u|,$$

for a set F of locally finite perimeter. We say that E *minimizes* J_u *in a set* A *(on the outside, inside respectively)* if

$$J_u^K(E) \leq J_u^K(F),$$

for each F such that $F\Delta E \subset\subset A$ (with $F \supseteq E$, $F \subseteq E$ respectively), and any compact set K containing $F\Delta E$. Again, the choice of such K does not matter, and we will generally drop K in the sequel.

In general, we have

$$\mathcal{H}^{n-1}\lfloor\partial^*(E\cup F) + \mathcal{H}^{n-1}\lfloor\partial^*(E\cap F) \leq \mathcal{H}^{n-1}\lfloor\partial^* E + \mathcal{H}^{n-1}\lfloor\partial^* F,$$

as Radon measures, so

$$J_u(E\cup F) + J_u(E\cap F) \leq J_u(E) + J_u(F), \tag{1.6}$$

whenever $E\Delta F$ is precompact. (The inequality reflects possible cancellation of oppositely oriented pieces.) Therefore E minimizes J_u (in A) if and only if E minimizes J_u on the inside and on the outside (in A).

1.1 Lemma. *Let u be a locally Lipschitz function in the open set Ω. Then u is a weak solution (subsolution, supersolution respectively) of $(\star)$ in Ω if and only if for each t, $E_t := \{u < t\}$ minimizes J_u in Ω (on the outside, inside respectively).*

The reader may look ahead to Example 1.5 to gain some insight into the effect of this minimization principle.

Proof of Lemma 1.1. 1. Let v be a locally Lipschitz function such that $\{v \neq u\} \subset\subset \Omega$ and K a compact set containing $\{v \neq u\}$. Set $E_t := \{u < t\}$, $F_t := \{v < t\}$ and note that $F_t\Delta E_t \subseteq K$ for every t. Select $a < b$ such that $a < u, v < b$ on K, and compute by the co-area formula (see [S,p.70]),

$$\begin{aligned} J_u^K(v) &= \int_K |\nabla v| + v|\nabla u| \\ &= \int_a^b |\partial^* F_t \cap K| - \int_K \int_a^b \chi_{\{v(x)<t\}} |\nabla u| + b\int_K |\nabla u| \\ &= \int_a^b J_u^K(F_t) + b\int_K |\nabla u|. \end{aligned} \tag{1.7}$$

If each E_t minimizes J_u in Ω, this formula shows that u minimizes (1.5) in Ω, and is thereby a weak solution of ($\star\star$). The same argument treats weak supersolutions and subsolutions separately. This proves one direction of the Lemma.

2. Now suppose that u is a supersolution of ($\star\star$). Fix t_0 and F such that

$$F \subseteq E_{t_0}, \qquad E_{t_0} \setminus F \subset\subset \Omega.$$

We aim to show that $J_u(E_{t_0}) \leq J_u(F)$. Since J_u is lower semicontinuous, we may assume that

$$J_u(F) \leq J_u(G), \tag{1.8}$$

for all G with $G \Delta E_{t_0} \subseteq F \Delta E_{t_0}$. Now define the nested family

$$F_t := \begin{cases} F \cap E_t, & t \leq t_0, \\ E_t, & t > t_0. \end{cases}$$

By (1.8), $J_u(F) \leq J_u(E_t \cup F)$, so by (1.6),

$$J_u(F_t) \leq J_u(E_t) \qquad \text{for all } t.$$

Define v so that $F_t = \{v < t\}$, that is,

$$v := \begin{cases} t_0 & \text{on } E_{t_0} \setminus F \\ u & \text{elsewhere.} \end{cases}$$

Note that $v \in BV_{loc} \cap L^\infty_{loc}$ and $\{v \neq u\} \subset\subset \Omega$, so $J_u(v)$ makes sense. Approximating v by smooth functions with $|\nabla v_i| \rightharpoonup |\nabla v|$ as measures, we find that $J_u(u) \leq J_u(v)$. Furthermore, (1.7) is valid for v. So by (1.7), we conclude

$$\int_a^b J_u(E_t)\, dt \leq \int_a^b J_u(F_t)\, dt.$$

Together with the above, this shows that $J_u(E_t) = J_u(F_t)$ for a.e. t. By (1.6), this shows $J_u(E_t \cup F) \leq J_u(F)$ for a.e. $t \leq t_0$. Passing $t \nearrow t_0$, we obtain by lower semicontinuity

$$J_u(E_{t_0}) \leq J_u(F).$$

We have proven: if u is a supersolution of ($\star\star$), then for each t_0, E_{t_0} minimizes J_u on the inside.

3. Next assume that u is a subsolution of $(\star\star)$. Similarly to the above, we prove: for each t, $\{u \le t\}$ minimizes J_u on the outside. Choose $t_i \nearrow t_0$, and note that $\{u \le t_i\}$ converges to E_{t_0} locally in L^1. Using lower semicontinuity of J_u and a standard replacement argument, it follows that E_{t_0} minimizes J_u on the outside. $\square$

Initial Value Problem. We will usually combine the above definition with an initial condition consisting of an open set E_0 with a boundary that is at least C^1. We say that u is a *weak solution of* $(\star\star)$ *with initial condition* E_0 if

$$u \in C^{0,1}_{loc}(M),\ E_0 = \{u < 0\},\ \text{and } u \text{ satisfies (1.5) in } M \setminus E_0. \tag{††}$$

This imposes the Dirichlet boundary condition $u = 0$ on ∂E_0. It is reasonable to expect Lipschitz continuity up to the boundary in view of (1.4), provided that the mean curvature of ∂E_0 is bounded at least in a weak sense. For particularly bad initial conditions, however, this requirement should be weakened (see [HI2]).

Let E_t be a nested family of open sets in M, closed under ascending union. Define u by the characterization $E_t = \{u < t\}$. We say that $(E_t)_{t>0}$ is a *weak solution of* $(\star)$ *with initial condition* E_0 if

$$u \in C^{0,1}_{loc}(M) \text{ and } E_t \text{ minimizes } J_u \text{ in } M \setminus E_0 \text{ for each } t > 0. \tag{†}$$

Observe that (†) allows slightly more competitors than (††), since it permits $F \Delta E_t$ to touch ∂E_0. Nevertheless, using Lemma 1.1 and approximating up to the boundary, it is quite straightforward to show

1.2 Lemma. (††) *is equivalent to* (†).

At this point, the crucial question arises to whether the equation holds at the initial time. By approximating $s \searrow t$, we see that (††) and (†) are equivalent to

$$u \in C^{0,1}_{loc}(M) \text{ and } \{u \le t\} \text{ minimizes } J_u \text{ in } \Omega \setminus E_0 \text{ for each } t \ge 0. \tag{1.9}$$

However, we cannot deduce that E_0 minimizes J_u in $M \setminus E_0$, since in general $\{u = 0\}$ may have positive measure. In fact, the equation can be satisfied at $t = 0$ only if E_0 is assumed to be in the correct class to begin with! We will take up this essential point in Lemma 1.4 below.

Regularity. Our variational formulation was partly inspired by Visintin [V], who noticed that an appropriate minimization principle can force H to be bounded. As a result, the regularity theory is essentially spatial in character, and succumbs to methods of geometric measure theory. The price of keeping H bounded is temporal *irregularity* in the form of jumps.

Let f be a bounded measurable function on a domain Ω with smooth metric g. Suppose E contains an open set U and minimizes the functional

$$|\partial^* F| + \int_F f$$

with respect to competitors F such that

$$F \supseteq U, \qquad F \Delta E \subset\subset \Omega.$$

We say that E minimizes area plus bulk energy f in Ω, respecting the obstacle U.

The following regularity theorem is obtained by [Mi,E] in the C^1 case; [Ma,A,Ln,T1] in the $C^{1,\alpha}$ case, $0 < \alpha < 1$, and [BK,Ln,SWZ] in the $C^{1,1}$ case. In the references for (ii), the case of a nonflat background metric g is not treated. However, carefully reworking the monograph [T2] suffices to establish the estimate with the given dependencies.

1.3 Regularity Theorem. *Let $n < 8$.*

(i) If ∂U is C^1 then ∂E is a C^1 submanifold of Ω.

(ii) If ∂U is $C^{1,\alpha}$, $0 < \alpha \leq 1/2$, then ∂E is a $C^{1,\alpha}$ submanifold of Ω. The $C^{1,\alpha}$ estimates depend only on the distance to $\partial\Omega$, $\operatorname{ess\,sup}|f|$, $C^{1,\alpha}$ bounds for ∂U, and C^1 bounds (including positive lower bounds) for the metric g.

(iii) If ∂U is C^2 and $f = 0$ (the case of pure area minimization with obstacle), then ∂E is $C^{1,1}$, and C^∞ where it does not contact the obstacle U.

Let u solve (††). For the duration of the paper, set

$$E_t := \{u < t\}, \qquad E_t^+ := \operatorname{int}\{u \leq t\}, \qquad N_t := \partial E_t, \qquad N_t^+ := \partial E_t^+.$$

For $n < 8$, the Regularity Theorem, Lemma 1.2, and (1.9) imply that $N_t^+ = \partial\{u > t\}$ and that N_t and N_t^+ possess locally uniform $C^{1,\alpha}$ estimates depending only on the local Lipschitz bounds for u.[4] In particular, for all $t > 0$,

$$N_s \to N_t \quad \text{as } s \nearrow t, \qquad N_s \to N_t^+ \quad \text{as } s \searrow t, \tag{1.10}$$

[4]M. Heidusch [Hd] has recently improved this to $C^{1,1}$ regularity.

in the sense of local $C^{1,\beta}$ convergence, $0 \le \beta < \alpha$. If ∂E_0 is $C^{1,\alpha}$, the estimates and convergence also hold as $s \searrow 0$.

When $n \geq 8$, the Regularity Theorem and the convergence (1.10) remain true away from a closed singular set Z of dimension at most $n-8$ and disjoint from $\bar{U}$.

Weak Mean Curvature. In order to give a PDE interpretation of our solutions, we will define the mean curvature of a surface that is not quite smooth. Let N be a hypersurface in M, X a compactly supported vectorfield defined on M, and $(\Phi^s)_{-\varepsilon<s<\varepsilon}$ the flow of diffeomorphisms generated by X, $\Phi^0 = id_M$. The *first variation formula* for area [S, p.80] states

$$\frac{d}{ds}\bigg|_{s=0} |\Phi^s(N) \cap W| = \int_{N\cap W} \operatorname{div}_N X \, d\mu = \int_{N\cap W} H\nu \cdot X \, d\mu, \tag{1.11}$$

for any precompact open set W containing the support of X. Here $d\mu$ is the surface measure of N, and $\operatorname{div}_N X(x) := \sum_i \nabla_{e_i} X(x) \cdot e_i$, where $e_1, \ldots e_{n-1}$ is any orthonormal basis of $T_x N$. When N is smooth, the first inequality is proven by differentiating under the integral sign, and the second by integration by parts.

When N is C^1, or C^1 with a small singular set and locally finite Hausdorff measure, we use the second equality as a definition. A locally integrable function H on N is called the *mean curvature* provided it satisfies (1.11) for every X in $C_c^\infty(TM)$.

For minimizers of (1.5), by the co-area formula and the dominated convergence theorem,

$$\begin{aligned}
0 &= \frac{d}{ds}\bigg|_{s=0} J_u(u \circ \Phi^s) \\
&= \frac{d}{ds}\bigg|_{s=0} \left(\int_{-\infty}^{\infty} \int_{N_t \cap W} |\det d\Phi^s(x)| \, d\mathcal{H}^{n-1}(x) \, dt + \int_W u(\Phi^s(x)) |\nabla u(x)| \, dx \right) \\
&= \int_{-\infty}^{\infty} \int_{N_t \cap W} \operatorname{div}_{N_t} X \, d\mathcal{H}^{n-1} \, dt + \int_W \nabla u \cdot X |\nabla u| \, dx.
\end{aligned}$$

Thus the linear functional of X given by the first term is bounded by $C \int_M |X| \, dx$, and therefore can be represented by a vectorfield $H\nu$ on M, where H is some bounded measurable function and ν is a measurable unit vectorfield on M. By the co-area formula, we obtain

$$0 = \int_{-\infty}^{\infty} \int_{N_t} (H\nu + \nabla u) \cdot X \, d\mathcal{H}^{n-1} \, dt, \qquad X \in C_c^1(TM).$$

By Lebesgue differentiation and comparison with the above, this shows that for a.e. t and a.e. $x \in N_t$, ν is the unit normal, H is the weak mean curvature, and

$$H = |\nabla u| \qquad \text{a.e. } x \in N_t, \qquad \text{a.e. } t. \tag{1.12}$$

For later use, we mention the following convergence fact. If N^i is a sequence of C^1 hypersurfaces, $N^i \to N$ locally in C^1, and

$$\sup_i \operatorname*{ess\,sup}_{N^i} |H_{N^i}| < \infty,$$

it follows from (1.11) and the Riesz Representation Theorem that H_N exists weakly as a locally L^1 function, with the weak convergence

$$\int_{N^i} H_{N^i} \nu_{N^i} \cdot X \to \int_N H_N \nu_N \cdot X, \qquad X \in C^0_c(TM), \tag{1.13}$$

and lower semicontinuity

$$\operatorname*{ess\,sup}_N |H_N| \le \liminf_{i\to\infty} \operatorname*{ess\,sup}_{N_i} |H_{N_i}|, \qquad \int_N \phi |H_N|^2 \le \liminf_{i\to\infty} \int_{N_i} \phi |H_{N_i}|^2, \tag{1.14}$$

for any $\phi \in C^0_c(M)$.

Minimizing Hulls. To explain the jumps, we introduce some terminology. Let Ω be an open set. We call E a *minimizing hull (in Ω)* if E minimizes area on the outside in Ω, that is, if

$$|\partial^* E \cap K| \le |\partial^* F \cap K|,$$

for any F containing E such that $F \setminus E \subset\subset \Omega$, and any compact set K containing $F \setminus E$. We say that E is a *strictly minimizing hull (in Ω)* if equality implies that $F \cap \Omega = E \cap \Omega$ a.e. (See [BT] for a related definition.)

It is easily checked that the Lebesgue points of a minimizing hull form an open set in Ω. In general we prefer to work with this unique, open, representatative.

The intersection of a countable collection of minimizing hulls is a minimizing hull, and the same goes for strictly minimizing. Now let E be any measurable set. Define $E' = E'_\Omega$ to be the intersection of (the Lebesgue points of) all the strictly minimizing hulls in Ω that contain E. Working modulo sets of measure zero, this may be realized by a countable intersection, so E' itself is a strictly minimizing hull, and open. We call E' the *strictly minimizing hull of E (in Ω).* Note that $E'' = E'$.

If $E \subset\subset M$ and M grows at infinity (for example, if M is asymptotically flat), then E', taken in M, is precompact as well. In this case E' satisfies the "shrink-wrap" obstacle problem

$$|\partial E'| \leq |\partial F| \qquad \text{whenever } E \subseteq F \subset\subset M.$$

Sometimes there are several distinct solutions, but they are all contained in E' a.e.

If ∂E is C^2, the $C^{1,1}$ regularity given in Theorem 1.3(iii) applies to E'. For the weak mean curvature, we have

$$(1.15) \qquad H_{\partial E'} = 0 \quad \text{on } \partial E' \setminus \partial E, \qquad\qquad H_{\partial E'} = H_{\partial E} \geq 0 \quad \mathcal{H}^{n-1} \text{ a.e. on } \partial E' \cap \partial E.$$

In this language, our variational definition of the flow has the following important consequence. M need not be complete.

1.4 Minimizing Hull Property. *Suppose that u satisfies (††) and M has no compact components. Then*

(i) For $t > 0$, E_t is a minimizing hull in M,

(ii) For $t \geq 0$, E_t^+ is a strictly minimizing hull in M,

(iii) For $t \geq 0$, $E_t' = E_t^+$, provided that E_t^+ is precompact,

(iv) For $t > 0$, $|\partial E_t| = |\partial E_t^+|$, provided that E_t^+ is precompact. This extends to $t = 0$ precisely if E_0 is a minimizing hull.

Note that the function u can take an arbitrary constant value on each compact component of M that does not meet E_0; the resulting nonuniqueness would falsify (ii) and (iii).

Proof. (i) By (†),

$$(1.16) \qquad |\partial E_t \cap K| + \int_{F \setminus E_t} |\nabla u| \leq |\partial^* F \cap K|,$$

for $t > 0$, any F with $F \Delta E_t \subset\subset M \setminus E_t$, and K containing $F \Delta E_t$. That is, E_t is a minimizing hull, $t > 0$.

(ii) By (1.9) and the fact that $|\nabla u| = 0$ a.e. on $\{u = t\}$,

$$(1.17) \qquad |\partial E_t^+ \cap K| + \int_{F \setminus E_t^+} |\nabla u| \leq |\partial^* F \cap K|,$$

for $t \geq 0$, any F with $F\Delta E_t^+ \subset\subset M \setminus E_t$, and suitable K, proving that E_t^+ is a minimizing hull, $t \geq 0$. To prove strictly minimizing, suppose F contains E_t^+ and

$$|\partial^* F \cap K| = |\partial E_t^+ \cap K|.$$

Then by (1.17), $\nabla u = 0$ a.e. on $F \setminus E_t^+$. And clearly, F itself is a minimizing hull, so by a measure zero modification, we may assume F is open. Then u is constant on each connected component of $F \setminus \bar{E}_t^+$. Since M has no compact components and F has no wasted perimeter, no such component can have closure disjoint from $\bar{E}_t^+$, so $u = t$ on $F \setminus E_t^+$, so $F \subseteq E_t^+$. This proves that E_t^+ is a strictly minimizing hull.

(iii) It follows that $E_t' \subseteq E_t^+$. If E_t^+ is precompact, then E_t' must coincide with E_t^+, for otherwise $|\partial E_t'| < |\partial E_t^+|$, contradicting (1.17).

(iv) If E_t^+ is precompact, then so is E_t, and by cross substitution in (1.16) and (1.17), we get $|\partial E_t| \leq |\partial E_t^+|$ and $|\partial E_t^+| \leq |\partial E_t|$, proving (iv) for $t > 0$, and for $t = 0$ if E_0 happens to be a minimizing hull itself. □

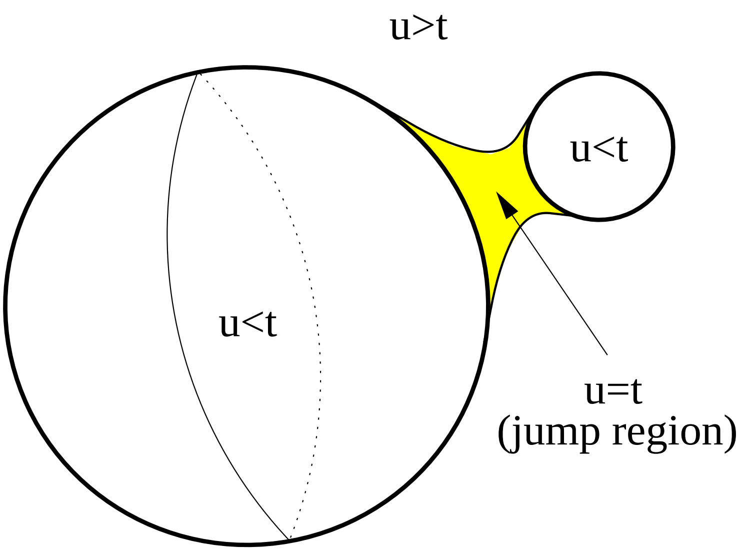
Returning to the question of the initial condition, we see by Lemma 1.4(iv) that E_t minimizes J_u in $M \setminus E_0$ for all $t \geq 0$ if and only if (†) holds and E_0 is a minimizing hull. This gives a stronger attainment of the initial condition than (†) alone. In this case, one might say that $(E_t)_{0 \leq t < \infty}$ is a weak solution of ($\star$). However, this terminology differs so subtly from our previous locutions that we will avoid it.

As a consequence of Lemma 1.4, the effect of the minimization principle (1.5) can be described heuristically as follows. The statement is not precise since, in principle, there may be an infinite number of jump times.

(1) As long as E_t remains a minimizing hull, it flows by the usual inverse mean curvature flow.

(2) When this condition is violated, E_t jumps to E_t' and continues.

In particular, the mean curvature is nonnegative on the flowing surfaces after time zero, even if the curvature of ∂E_0 is mixed.



Two Spheres Joining.

1.5 Two Spheres Example. In the following example, the surface does indeed jump. Suppose E_0 is the union of two balls in R^3 at some distance from one another. Each ball expands until the first instant $t = t_{jump}$ when the two spheres can be enclosed by a connected surface of equal area. At this instant, a catenoidal bridge forms between the two balls and the flow continues.

Note that $t_{jump} < t_{touch}$, the time when the two spheres would make contact classically. If we flow up to time t_{touch} without jumping, then u minimizes (1.5) on $\{u < t_{touch}\}$, but not on larger sets. So, not only does the flow meet the obstacle of its own history, but if it waits too long to jump, it cannot continue at all.

G. Bellettini originally showed us this example, observing that any choice of jump time (before the spheres touch) yields a viscosity solution of ($\star\star$), so viscosity solutions are nonunique. The variational principle J_u serves to select the jump time.

A more extreme form of jumping is

(1.18) if u weakly solves ($\star\star$), then so does $\min(u, t)$ for every constant t.

That is, the surfaces would like to jump instantly to infinity in order to exploit the negative bulk energy term in $J_u(F)$. To prove this, recall that for $s \le t$ and K containing $E_s \Delta F$,

$$|\partial E_s \cap K| - \int_{E_s \cap K} |\nabla u| \le |\partial^* F \cap K| - \int_{F \cap K} |\nabla u|.$$

If u is replaced by $\min(u,t)$, the left hand side only increases, while the right-hand side is unchanged. This shows that $\min(u,t)$ is a weak solution, by Lemma 1.1.

1.6 Exponential Growth Lemma. *Let $(E_t)_{t>0}$ solve (†) with initial condition E_0. As long as E_t remains precompact, we have the following.*

(i) $e^{-t}|\partial E_t|$ is constant for $t>0$.

(ii) If E_0 is a minimizing hull, then $|\partial E_t| = e^t|\partial E_0|$.

Proof. It follows from the minimization property (†) that $J_u(E_t)$ is independent of t, $t>0$. By the co-area formula, this implies that $|\partial E_t| - \int_0^t |\partial E_s|\,ds$ is constant for $t>0$, which yields (i). A similar exponential growth applies to $|\partial E_t^+|$, $t \geq 0$. If E_0 is a minimizing hull, then Lemma 1.4(iv) yields (ii). □

§2 Further Properties of Inverse Mean Curvature Flow

In this section we derive compactness and uniqueness theorems for weak solutions of (1.5), and state the relationship between classical and weak solutions.

2.1 Compactness Theorem. *Let u_i be a sequence of solutions of (1.5) on open sets Ω_i in M such that*

$$u_i \to u, \qquad \Omega_i \to \Omega,$$

locally uniformly, and for each $K \subset\subset \Omega$,

$$\sup_K |\nabla u_i| \leq C(K),$$

for large i. Then u is a solution of (1.5) on Ω.

Remarks. 1. For each t, the Regularity Theorem 1.3(ii) implies that any subsequence of N_t^i has a further subsequence that converges in $C^{1,\alpha}$ to some hypersurface $\tilde{N}$, away from the singular set Z for u. Whenever there is no jump, $\tilde{N} = N_t = N_t^+$, so for the full sequence,

$$N_t^i \to N_t \qquad \text{locally in } C^{1,\alpha} \text{ in } \Omega \setminus Z, \tag{2.1}$$

for any t not belonging to the countable set of jump times.

2. The Lemma and Remark 1 still hold if we allow u_i to solve (1.5) with respect to a metric g_i that converges to g in C^1_{loc}.

Proof. 1. Let v be a locally Lipschitz function such that $\{v \neq u\} \subset\subset \Omega$. We must prove that $J_u(u) \leq J_u(v)$. First we assume $v < u + 1$.

Let $\phi \in C^1_c(\Omega)$ be a cutoff function such that $\phi = 1$ on $\{v \neq u\}$. Then

$$v_i := \phi v + (1 - \phi)u_i$$

is a valid comparison function for u_i. So by (1.5), for sufficiently large i,

$$\begin{aligned}\int_U |\nabla u_i| + u_i|\nabla u_i| &\leq \int_U |\nabla v_i| + v_i|\nabla u_i| \\ &= \int_U |\phi\nabla v + (1-\phi)\nabla u_i + \nabla\phi(v - u_i)| + (\phi v + (1-\phi)u_i)|\nabla u_i|\end{aligned}$$

for appropriate U, so

$$\int_U \phi|\nabla u_i|(1 + u_i - v) \leq \int_U \phi|\nabla v| + |\nabla\phi(v - u_i)|.$$

The last term converges to zero. Since $1 + u_i - v$ is eventually positive and converges uniformly, it follows by lower semicontinuity that

$$\int_U \phi|\nabla u|(1 + u - v) \leq \int_U \phi|\nabla v|,$$

so u satisfies (1.5) for all $v < u + 1$.

2. Next assume that u satisfies (1.5) for all $w \leq u + k$, and prove that it does for each $v \leq u + 2k$. Define

$$v_1 := \min(v, u + k), \qquad v_2 := \max(v - k, u).$$

Then v_1, $v_2 \leq u + k$, so inserting v_1 into $J_u(u) \leq J_u(v)$, we get

$$\int |\nabla u| + u|\nabla u| \leq \int_{v \leq u+k} |\nabla v| + v|\nabla u| + \int_{v > u+k} |\nabla u| + (u + k)|\nabla u|.$$

Inserting v_2 into the same inequality, we get

$$\int |\nabla u| + u|\nabla u| \leq \int_{v \leq u+k} |\nabla u| + u|\nabla u| + \int_{v > u+k} |\nabla v| + (v - k)|\nabla u|.$$

Adding these two inequalities and cancelling the extra $J_u(u)$, we get $J_v(u) \le J_u(v)$, as required. $\square$

2.2 Uniqueness Theorem. *Assume that M has no compact component; M need not be complete.*

(i) If u and v solve (1.5) *on an open set Ω in M, and $\{v > u\} \subset\subset \Omega$, then $v \le u$ on Ω.*

(ii) If $(E_t)_{t>0}$ and $(F_t)_{t>0}$ solve (†) *in a manifold M and the initial conditions satisfy $E_0 \subseteq F_0$, then $E_t \subseteq F_t$ as long as E_t is precompact in M.*

(iii) In particular, for a given E_0, there exists at most one solution $(E_t)_{t>0}$ of (†) *such that each E_t is precompact.*

Without the compactness assumption, $\min(u, t)$ gives an example of nonuniqueness. Another example occurs in the metric

$$g := a^2(t)dr^2 + e^{t/2}g_{S^2},$$

where g_{S^2} is the standard metric. The evolution is $N_t = \{t - c\} \times S^2$ quite independent of the choice of a. For $a(t) := 1 + 1/t$, $t > 0$, the metric is asymptotic to a cylinder at one end and $\mathbf{R}^3$ at the other, and a noncompact E_t rushes in from infinity at any time $t = c$. This gives nonuniqueness with $E_0 = \emptyset$.

Proof. (i) First assume that u is a *strict* weak supersolution of ($\star\star$) in the sense that for any Lipschitz function $w \ge u$ with $\{w \ne u\} \subset\subset \Omega$, we have

$$\int |\nabla u| + u|\nabla u| + \varepsilon \int (w - u)|\nabla u| \le \int |\nabla w| + w|\nabla u|, \tag{2.2}$$

where $\varepsilon > 0$. Replace w by $u + (v - u)_+$ to obtain

$$\int_{v>u} |\nabla u| + u|\nabla u| + \varepsilon \int_{v>u} (v - u)|\nabla u| \le \int_{v>u} |\nabla v| + v|\nabla u|,$$

and replace v by $v - (v - u)_+$ and u by v in (1.5) to obtain

$$\int_{v>u} |\nabla v| + v|\nabla v| \le \int_{v>u} |\nabla u| + u|\nabla v|. \tag{2.3}$$

Adding these, we get

$$\int_{v>u} (v - u)(|\nabla v| - |\nabla u|) + \varepsilon \int_{v>u} (v - u)|\nabla u| \le 0. \tag{2.4}$$

We expect the first integral to be much smaller than the second one near the point of contact, because $|\nabla u|$ nearly cancels $|\nabla v|$. To bring this out, we need something like an upper bound for the positive quantity $\int (v-u)|\nabla u|$, so we employ the minimizing property of u once again. We replace v by $u+(v-s-u)_+$, $s \geq 0$, in (1.5) and integrate over s to obtain

$$\int_0^\infty \int_{v-s>u} |\nabla u| + u|\nabla u| \, dx \, ds \leq \int_0^\infty \int_{v-s>u} |\nabla v| + (v-s)|\nabla u| \, dx \, ds.$$

(Since the functional is nonlinear, there can be no action at a distance; the gap must be calibrated.) Switching the order of integration, we have

$$\int_M |\nabla u| \int_{-\infty}^\infty \chi_{\{s>0\}}\chi_{\{v-u>s\}}(1+u-v+s)\, ds \, dx \leq \int_M |\nabla v| \int_{-\infty}^\infty \chi_{\{s>0\}}\chi_{\{v-u>s\}} \, ds \, dx$$

which yields

$$\int_{v>u} -\frac{(v-u)^2}{2}|\nabla u| \leq \int_{v>u} (v-u)(|\nabla v| - |\nabla u|).$$

Inserting this into (2.4) yields

$$\int_{v>u} -\frac{(v-u)^2}{2}|\nabla u| + \varepsilon(v-u)|\nabla u| \leq 0.$$

Assuming $v \leq u+\varepsilon$, the above inequality implies that $|\nabla u| = 0$ a.e. on $\{v>u\}$, and then (2.3) implies that $|\nabla v| = 0$ a.e. on $\{v>u\}$. Then u and v are constant on each component of $\{v>u\}$. Since $\{v>u\}$ is precompact in Ω and Ω has no compact component, we conclude that $v \leq u+\varepsilon$ implies $v \leq u$.

For general v, by subtracting a constant we can arrange that $0 < \sup(v-u) \leq \varepsilon$, contradicting what has just been proven. This establishes the result for any u that satisfies (2.2).

Now let u be an arbitrary weak supersolution of ($\star\star$). We may assume that $u>0$. Observe that for any $\varepsilon > 0$, $u^\varepsilon := u/(1-\varepsilon)$ satisfies (2.2) and $\{v > u^\varepsilon\}$ is precompact. The above discussion shows that $v \leq u^\varepsilon$, so $v \leq u$.

(ii) Recall by (1.18) that $v^t := \min(v,t)$ solves (1.5) on $M \setminus \bar{F}_0$. Let $W := E_t \setminus \bar{F}_0$, a precompact open set. Since $E_0 \subseteq F_0$, we have $v^t < u+\delta$ near ∂W for each $\delta > 0$, so $\{v^t > u+\delta\} \subset\subset W$, so (i) implies $v^t \leq u+\delta$ on W, so $v^t \leq u$ on W. Since $u<t$ on W, $v \leq u$ on W. Thus $E_t \subseteq F_t$.

(iii) is immediate from (ii). $\square$

The next proposition shows that smooth flows satisfy the weak formulation in the domain they foliate.

2.3 Smooth Flow Lemma. *Let $(N_t)_{c\leq t<d}$ be a smooth family of surfaces of positive mean curvature that solves $(\star)$ classically. Let $u = t$ on N_t, $u < c$ in the region bounded by N_c, and $E_t := \{u < t\}$. Then for $c \leq t < d$, E_t minimizes J_u in $E_d \setminus E_c$.*

Proof. The exterior normal, defined by $\nu_u := \nabla u/|\nabla u|$, is a smooth unit vector field on Ω with $\operatorname{div} \nu_u = H_{N_t} = |\nabla u| > 0$. Using ν_u as a calibration, by the divergence theorem

$$|\partial E_t| - \int_{E_t} |\nabla u| = \int_{\partial E} \nu_{\partial E} \cdot \nu_u - \int_E |\nabla u| = \int_{\partial^* F} \nu_{\partial^* F} \cdot \nu_u - \int_F |\nabla u| \leq |\partial^* F| - \int_F |\nabla u|,$$

for any finite perimeter set F differing compactly from E_t. $\square$

The next proposition says that the weak evolution of a smooth, positively curved, strictly minimizing hull is smooth for a short time.

2.4 Smooth Start Lemma. *Let E_0 be a precompact open set in M such that ∂E_0 is smooth with $H > 0$ and $E_0 = E_0'$. Then any weak solution $(E_t)_{0<t<\infty}$ of $(\dagger)$ with initial condition E_0 coincides with the unique smooth, classical solution for a short time, provided that E_t remains precompact for a short time.*

Proof. Since $(\star)$ is parabolic when $H > 0$, a classical solution F_t exists for a short time. By Lemma 2.3, this gives rise to a smooth solution v of $(\dagger\dagger)$ defined in an open neighborhood W of E_0. By 1.4(iii) and the hypotheses, $E_0^+ = E_0' = E_0$. Since E_t is precompact, E_t converges in Hausdorff distance to E_0^+ as $t \searrow 0$, which shows that E_t is precompact in W for a short time. Then by Theorem 2.2(ii), $E_t = F_t$ for a short time. $\square$

Remark. It can be shown that the weak solution remains smooth until the first moment when either $E_t \neq E_t'$, $H \searrow 0$, or $|A| \nearrow \infty$. Whether the latter two possibilities occur is unknown.

§3 Elliptic Regularization and Existence

In this section we prove existence of solutions of the initial value problem $(\dagger\dagger)$ by means of elliptic regularization, an approximation scheme. We also get a useful local estimate of H and $|\nabla u|$ that is independent of the size of u.

Recall that u is *proper* if each set $\{s \le u \le t\}$ is compact. Write $H_+ := \max(0, H_{\partial E_0})$.

3.1 Weak Existence Theorem. *Let M be a complete, connected Riemannian n-manifold without boundary. Suppose there exists a proper, locally Lipschitz, weak subsolution of* (††) *with a precompact initial condition.*

Then for any nonempty, precompact, smooth open set E_0 in M, there exists a proper, locally Lipschitz solution u of (††) *with initial condition E_0, which is unique on $M \setminus E_0$. Furthermore, the gradient of u satisfies the estimate*

$$|\nabla u(x)| \le \sup_{\partial E_0 \cap B_r(x)} H_+ + \frac{C(n)}{r}, \qquad a.e.\ x \in M \setminus E_0, \tag{3.1}$$

for each $0 < r \le \sigma(x)$, where $\sigma(x) > 0$ is defined in Definition 3.3.

Remarks. 1. The subsolutions required by the Theorem will exist whenever M has some mild growth at infinity. In particular, if M is asymptotically flat, or more generally, asymptotically conic, then M will possess a subsolution of the form

$$v = C \log |x|$$

in the asymptotic region.

2. As a consequence of (3.1), we can drop the local Lipschitz bounds in the Compactness Theorem in favor of adding a constant.

Let v be the given subsolution at infinity, $F_L := \{v < L\}$. We may assume that $E_0 \subseteq F_0$. The domain $\Omega_L := F_L \setminus \bar{E}_0$ is precompact.

To prove the theorem, we employ the following approximate equation, known as *elliptic regularization* (see [I1]).

$$(\star)_\varepsilon \qquad \begin{cases} E^\varepsilon u^\varepsilon := \operatorname{div}\left(\dfrac{\nabla u^\varepsilon}{\sqrt{|\nabla u^\varepsilon|^2 + \varepsilon^2}}\right) - \sqrt{|\nabla u^\varepsilon|^2 + \varepsilon^2} = 0 & \text{in } \Omega_L, \\ u^\varepsilon = 0 & \text{on } \partial E_0, \\ u^\varepsilon = L - 2 & \text{on } \partial F_L. \end{cases}$$

The relation between ε and L will be revealed below.

This equation has a geometric interpretation: it states that the downward translating graph

$$N_t^\varepsilon := \operatorname{graph}\left(\frac{u^\varepsilon(x)}{\varepsilon} - \frac{t}{\varepsilon}\right), \qquad -\infty < t < \infty,$$

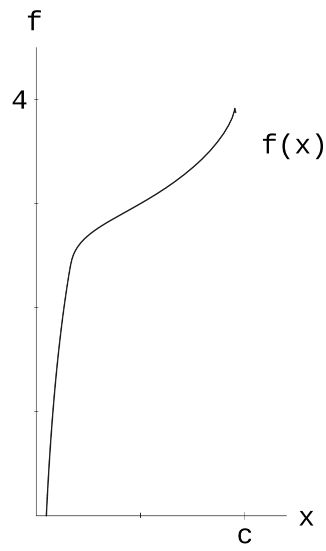
solves the inverse mean curvature flow $(\star)$ in the manifold $M \times \mathbf{R}$. To see this, define

$$U^\varepsilon(x,z) := u^\varepsilon(x) - \varepsilon z, \qquad (x,z) \in \Omega_L \times \mathbf{R}, \tag{3.2}$$

so $N_t^\varepsilon = \{U^\varepsilon = t\}$. Then assuming smoothness, one checks that U^ε satisfies $(\star\star)$ on $\Omega_L \times \mathbf{R}$ if and only if u^ε satisfies $(\star)_\varepsilon$ on Ω_L.

3.2 Example. Note that the natural units of ε are $1/x$, since t is unitless. So by scaling $\varepsilon \to 0$, we are actually scaling $x \to \infty$. This suggests that our more difficult problems will be in the large rather than in the small.

In fact, for positive ε, solutions of $(\star)_\varepsilon$ do not exist globally, but only on a ball $B_{c/\varepsilon}$. The Maple-computed figure exhibits a rotationally symmetric solution $u(x) = f(|x|)$ on $\mathbf{R}^3$ with $\varepsilon = 1$. Near $x = 0$, $u \sim (n-1)\log|x|$. At $|x| = c \approx 1.98$, u ceases to exist; its first and second derivatives diverge to ∞.



Downward Translating Solution.

A solution u^ε can be obtained from u by

$$u^\varepsilon(x) := u(\varepsilon x), \qquad 0 < |x| < c/\varepsilon.$$

(Thus L cannot exceed c/ε in the flat case.) As $\varepsilon \to 0$, then modulo vertical translations, u^ε converges to the solution

$$u(x) := (n-1)\log|x|, \qquad x \in \mathbf{R}^n \setminus \{0\},$$

of $(\star\star)$, corresponding to an expanding sphere.

Estimate of H. We will derive a local estimate for H for smooth solutions of $(\star)$. The $-|A|^2/H$ term in (1.3) is so strong that the estimate is independent of the oscillation of u. The estimate improves on (1.4).

3.3 Definition. For any $x \in M$, define $\sigma(x) \in (0, \infty]$ to be the supremum of radii r such that $B_r(x) \subset\subset M$,

$$Rc \geq -\frac{1}{100nr^2} \qquad \text{in } B_r(x),$$

and there exists a C^2 function p on $B_r(x)$ such that

$$p(x) = 0, \qquad p \geq d_x^2 \qquad \text{in } B_r(x),$$

yet

$$|\nabla p| \leq 3d_x \quad \text{and} \quad \nabla^2 p \leq 3g \qquad \text{on } B_r(x),$$

where d_x is the distance to x. In flat space, $p(y) := |y - x|^2$ does the job with $r = \infty$. More generally, a sufficient condition is the existence of a diffeomorphism of $B_r(x)$ with a smooth ball in $\mathbf{R}^n$ such that

$$|g - \delta| \leq \frac{1}{100} \quad \text{and} \quad |g_{ij,k}| \leq \frac{1}{100r} \qquad \text{in } B_r(x).$$

Then $p(y) := 100|y - x|^2/99$ will do the trick, where $|\cdot|$ represents distance in coordinates.

Now we will obtain an estimate of H using the maximum principle. Fix x and $0 < r < \sigma(x)$, and write B_r for $B_r(x)$. Define the speed function $\psi = 1/H$ and transform (1.3) to

$$\begin{aligned} \frac{\partial \psi}{\partial t} &= \psi^2(\Delta\psi + |A|^2\psi + Rc(\nu,\nu)\psi) \\ &\geq \psi^2\Delta\psi + \frac{\psi}{n-1} - \frac{\psi^3}{100nr^2}, \end{aligned} \tag{3.3}$$

on $N_t \cap B_r$. We seek a function $\phi = \phi(y)$ that vanishes on ∂B_r and is a subsolution of (3.3) along $N_t \cap B_r$. For this purpose, it suffices that

$$\frac{\partial \phi}{\partial t} \leq \phi^2 \Delta\phi + \frac{\phi}{2n}, \tag{3.4}$$

provided we assume $\phi \leq r$.

We have the following relations to the ambient derivatives of ϕ,

$$\frac{\partial \phi}{\partial t} = \psi \nu \cdot \nabla \phi, \qquad \Delta \phi = S : \nabla^2 \phi - \frac{1}{\psi} \nu \cdot \nabla \phi,$$

where $\partial\phi/\partial t$ is the derivative following a particle on N_t, S is the orthogonal projection from $T_x M$ to $T_x N_t$, and ν is the exterior normal of E_t.

If we assume that $\phi < \psi$ initially on $N_t \cap B_r$, then at the first point of contact, we have $\phi = \psi$. Therefore, for the purpose of producing a subsolution, we may replace ψ by ϕ in the ambient relations and substitute the modified relations into (3.4), to obtain the following sufficient condition for $\phi = \phi(y)$ to be a subsolution of (3.3),

$$0 \le \phi^2 S : \nabla^2 \phi - 2\phi \nu \cdot \nabla \phi + \frac{\phi}{2n}, \tag{3.5}$$

where this is required to hold for every $x \in B_r$ and every hyperplane S and unit vector ν in $T_x M$. Define

$$\phi(y) := \frac{A}{r}(r^2 - p(y))_+,$$

where p is as defined above. Then $\phi = 0$ on ∂B_r, and $\phi \le r$ provided $A \le 1$. By the definition of p, $|\nabla p| \le 3r$ and $S : \nabla^2 p \le 3(n-1)$, so by calculation, ϕ satisfies (3.5) provided $3nA^2 + 6A \le 1/2n$, for which $A \le 1/6n(n+1)$ suffices. This shows that the function

$$h(x) := \frac{Br}{(r^2 - p(y))_+}$$

is a smooth supersolution on $N_t \cap B_r$ of the evolution equation (1.3) for H wherever h is finite, provided $B \ge 6n(n+1)$.

Fix t, assume $x \in N_t$, and define the parabolic boundary of the flow to be

$$P_r = P_r(x,t) := (B_r \cap N_0) \times \{0\} \cup (\cup_{0 \le s \le t} (B_r \cap \partial N_s) \times \{s\})$$

and

$$H_r = H_r(x,t) := \sup_{(y,s) \in P_r} H(y,s),$$

where we allow N_s to have a smooth boundary ∂N_s. Set

$$B = \max(rH_r, 6n(n+1)),$$

which ensures that h is a supersolution and $h \geq H$ on P_r. By the maximum principle, $h \geq H$ everywhere on $N_s \cap B_r$, $0 \leq s \leq t$, so in particular

$$H(x,t) \leq h(x,t) = \frac{B}{r} = \max\left(H_r, \frac{6n(n+1)}{r}\right).$$

We have proven the following lemma.

Interior Estimate of H. *Let $(N_t)_{0\leq s\leq t}$ solve $(\star)$ smoothly in M, where N_t may have boundary. Then for each $x \in N_t$ and each $r < \sigma(x)$, we have*

$$H(x,t) \leq \max\left(H_r, \frac{C(n)}{r}\right), \tag{3.6}$$

where H_r is the maximum of H on P_r, the parabolic boundary of the intersection of the flow with $B_r(x)$, $0 \leq s \leq t$.

To solve equation $(\star)_\varepsilon$, we will estimate solutions of the following family of equations,

$$(\star)_{\varepsilon,\tau} \qquad \begin{cases} E^\varepsilon u^{\varepsilon,\tau} := \operatorname{div}\left(\dfrac{\nabla u^{\varepsilon,\tau}}{\sqrt{|\nabla u^{\varepsilon,\tau}|^2+\varepsilon^2}}\right) - \sqrt{|\nabla u^{\varepsilon,\tau}|^2+\varepsilon^2} = 0 & \text{in } \Omega_L, \\ u = 0 & \text{on } \partial E_0, \\ u = \tau & \text{on } \partial F_L, \end{cases}$$

for $0 \leq \tau \leq L-2$.

3.4 Lemma. *Suppose the subsolution v provided in Theorem 3.1 is smooth, with $\nabla v \neq 0$. Then for every $L > 0$, there is $\varepsilon(L) > 0$ such that for $0 < \varepsilon \leq \varepsilon(L)$ and $0 \leq \tau \leq L-2$, a smooth solution of $(\star)_{\varepsilon,\tau}$ on $\bar{\Omega}_L$ satisfies the following estimates*

$$u^{\varepsilon,\tau} \geq -\varepsilon \quad \text{in } \bar{\Omega}_L, \qquad u^{\varepsilon,\tau} \geq v + \tau - L \quad \text{in } \bar{F}_L \setminus F_0. \tag{3.7}$$

$$|\nabla u^{\varepsilon,\tau}| \leq H_+ + \varepsilon \quad \text{on } \partial E_0, \qquad |\nabla u^{\varepsilon,\tau}| \leq C(L) \quad \text{on } \partial F_L, \tag{3.8}$$

$$|\nabla u^{\varepsilon,\tau}(x)| \leq \max_{\partial\Omega_L \cap B_r(x)} |\nabla u^{\varepsilon,\tau}| + \varepsilon + \frac{C(n)}{r}, \qquad x \in \bar{\Omega}_L, \tag{3.9}$$

$$|u^{\varepsilon,\tau}|_{C^{2,\alpha}(\bar{\Omega}_L)} \leq C(\varepsilon, L), \tag{3.10}$$

for any r with $0 < r \leq \sigma(x)$. Here $H_+ = \max(0, H_{\partial E_0})$.

Note that (3.8) and (3.9), with $r = \sigma(x)$, yield uniform $C^{0,1}$ estimates (independent of ε and L) on each compact set, for all sufficiently large L.

Proof of Lemma 3.4. 1. We will use supersolutions and subsolutions, (3.6), and standard theorems. Write $u = u^{\varepsilon,\tau}$.

The particular size of Ω_L on which u exists is determined by the availability of subsolutions, as can be seen in Example 3.2. Subsolutions are delicate and require conditions on the asymptotic region, since the surfaces really would like to jump instantly to infinity.

First we construct a subsolution that bridges from E_0 to where v starts. It is a perturbation of zero, and hence allows for unrestricted jumps in the compact part of the manifold. Define $G_0 = E_0$, $G_s := \{x : \operatorname{dist}(x, E_0) < s\}$. Select s_L so that G_{s_L} contains F_L. (This is possible since M is connected and E_0 is nonempty.) Let Σ be the cut locus of E_0 in M. On $M \setminus E_0 \setminus \Sigma$, the distance function is smooth, each point is connected to E_0 by a unique length-minimizing geodesic γ, and ∂G_s foliates a neighborhood of γ. Differentiating along γ, we have by (1.2),

$$\frac{\partial H}{\partial s} = -|A|^2 - Rc(\nu,\nu) \le C_1 \qquad \text{on } \partial G_s \setminus \Sigma, \qquad 0 \le s \le s_L,$$

where $C_1 = C_1(L)$, yielding

$$H_{\partial G_s} \le \max_{\partial E_0} H_+ + C_1 s \le C_2 \qquad \text{on } \partial G_s \setminus \Sigma, \qquad 0 \le s \le s_L,$$

where $C_2 = C_2(L)$. Now consider the prospective subsolution

$$v_1(x) := f(s) = f(\operatorname{dist}(x, G)), \qquad x \in \bar{G}_{s_L} \setminus E_0,$$

with $f' < 0$. Then

$$\left(g^{ij} - \nu^i \nu^j\right) \nabla^2_{ij} v_1 = f' H_{\partial G_s} \ge C_2 f',$$

and hence

$$\begin{aligned}\sqrt{(f')^2 + \varepsilon^2}\, E^\varepsilon v_1 &= \left(g^{ij} - \frac{(f')^2 \nu^i \nu^j}{(f')^2 + \varepsilon^2}\right) \nabla^2_{ij} v_1 - (f')^2 - \varepsilon^2 \\ &\ge C_2 f' + \frac{\varepsilon^2 f''}{(f')^2 + \varepsilon^2} - (f')^2 - \varepsilon^2,\end{aligned}$$

on the smooth part, $G_{s_L} \setminus E_0 \setminus \Sigma$. Therefore, $E^\varepsilon_2 v_1 \ge 0$ on this set, provided

$$((f')^2 + \varepsilon^2)((f')^2 + \varepsilon^2 - f' C_2) \le \varepsilon^2 f''.$$

Set

$$f(s) := \frac{\varepsilon}{A}(-1 + e^{-As}), \qquad 0 \le s \le s_L.$$

Then we have $\varepsilon^2 \le |f'| \le \varepsilon$, provided that we impose $\varepsilon \le \varepsilon(A, L) := e^{-As_L}$. Choosing $A = A(L) := 4 + 2C_2$, we have

$$((f')^2 + \varepsilon^2)((f')^2 + \varepsilon^2 - f'C_2) \le 2\varepsilon^2(2\varepsilon^2 + C_2|f'|) \le 2\varepsilon^2(2 + C_2)|f'| = \varepsilon^2 f'',$$

as required. This shows that for sufficiently small ε, the function

$$v_1(x) = \frac{\varepsilon}{4 + 2C_2}(-1 + e^{-(4+2C_2)s}), \qquad s = \operatorname{dist}(x, \partial G_0),$$

is a smooth subsolution for E^ε on $G_{s_L} \setminus E_0 \setminus \Sigma$.

We claim that v_1 is a *viscosity* subsolution of E^ε on all of $G_{s_L} \setminus \bar{E}_0$ (see [CIL]). Suppose ϕ is a smooth function tangent to v_1 from above at a point x. Since $f' < 0$, $\{\phi = v_1(x)\}$ is locally a smooth hypersurface tangent to G_s at x from outside, from which it follows by solving the Riccati equation backward along γ that $x \notin \Sigma$. Therefore, we have $E^\varepsilon\phi(x) \ge 0$, which proves the claim.

Since $u \ge v_1$ on the boundary, it follows by the maximum principle for viscosity solutions that

$$u \ge v_1 \ge -\varepsilon \quad \text{in } \bar{\Omega}_L, \qquad \frac{\partial u}{\partial \nu} \ge -\varepsilon \quad \text{on } \partial E_0. \tag{3.11}$$

2. Next, consider the function

$$v_2 := \frac{L-1}{L} v + \tau - (L - 1).$$

Clearly $E^0 v_2 > 0$ on $\bar{F}_L \setminus F_0$. Since the domain is compact, for all sufficiently small ε we obtain $E^\varepsilon v_2 > 0$. Note that

$$u \ge -\varepsilon \ge v_2 \quad \text{on } \partial F_0, \qquad u = \tau = v_2 \quad \text{on } \partial F_L,$$

since $0 \le \tau \le L - 2$. Then by the maximum principle,

$$u \ge v_2 \ge v + \tau - L \quad \text{in } \bar{F}_L \setminus F_0, \qquad \frac{\partial u}{\partial \nu} \ge -C(L) \quad \text{on } \partial F_L, \tag{3.12}$$

Since any constant is a supersolution of $(\star)_{\varepsilon,\tau}$, we obtain

$$u \le \tau \quad \text{in } \bar{\Omega}_L, \qquad \frac{\partial u}{\partial \nu} \le 0 \quad \text{on } \partial F_L. \tag{3.13}$$

3. Next we construct a supersolution along ∂E_0. Choose a smooth function v_3 vanishing on ∂E_0, such that

$$H_+ < \frac{\partial v_3}{\partial \nu} \le H_+ + \varepsilon \qquad \text{along } \partial E_0.$$

This implies that for sufficiently small $\delta > 0$, $|\nabla v_3| > 0$ and $E^0 v_3 < 0$ in the neighborhood $U := \{0 \le v_3 \le \delta\}$. Now define the sped-up function

$$v_4 := \frac{v_3}{1 - v_3/\delta}, \qquad x \in U.$$

Then $E^0 v_4 < 0$ as well, and $v_4 \to \infty$ on $\partial U \setminus \partial E_0$. Then for sufficiently small ε, depending on L, $E^\varepsilon v_4 < 0$ on the set $V := \{0 \le v_4 \le L\}$. Since $u \le L - 2$ by (3.13), we have $u \le v_4$ on ∂V. Then by the maximum principle, we obtain $u \le v_4$ on V, and therefore

$$\frac{\partial u}{\partial \nu} \le \frac{\partial v_4}{\partial \nu} = \frac{\partial v_3}{\partial \nu} \le H_+ + \varepsilon \qquad \text{on } \partial E_0, \tag{3.14}$$

for sufficiently small ε. Collecting together (3.11), (3.12), (3.13), and (3.14), we have proven (3.7) and (3.8).

4. Let $N_t^{\varepsilon,\tau}$ denote the level-set $\{U = t\}$ of the function $U(x,z) := u^{\varepsilon,\tau}(x) - \varepsilon z$, $-\infty < t < \infty$. Equation $(\star)_{\varepsilon,\tau}$ asserts

$$H_{N_t^{\varepsilon,\tau}} = \sqrt{|\nabla u|^2 + \varepsilon^2},$$

which says as pointed out above that $N_t^{\varepsilon,\tau}$ solves the inverse mean curvature flow in $\Omega_L \times \mathbf{R}$. Let $\tilde{B} := B_r^{n+1}(x,z)$ be an $(n+1)$-dimensional ball in $M \times \mathbf{R}$. Since the parabolic boundary of the flow $N_t^{\varepsilon,\tau}$ is nothing more than various translates of $\partial\Omega_L$ and $|\nabla u|$ is independent of z, we may apply (3.6) to $N_t^{\varepsilon,\tau}$ in $\tilde{B}$ to yield

$$\begin{aligned}\sqrt{|\nabla u|^2 + \varepsilon^2} &\le \sup_t \max_{\partial N_t^{\varepsilon,\tau} \cap \tilde{B}} \sqrt{|\nabla u|^2 + \varepsilon^2} + \frac{C(n)}{r} \\ &\le \max_{\partial\Omega_L \cap B_r(x)} |\nabla u| + \varepsilon + \frac{C(n)}{r},\end{aligned}$$

for $x \in M \setminus E_0$ and any r with $0 < r \le \sigma_M(x) = \sigma_{M\times\mathbf{R}}(x, u(x)/\varepsilon)$. This is (3.9).

Then (3.8) and (3.9) yield the estimate $|u|_{C^{0,1}(\bar{\Omega}_L)} \le C(L)$ for all sufficiently small ε. Following the method of [GT, Thm. 13.2], the Nash-Moser-De Giorgi estimates yield $|u|_{C^{1,\alpha}(\bar{\Omega}_L)} \le C(\varepsilon, L)$, where $\alpha = \alpha(\Omega_L)$. The Schauder estimates [GT] complete the proof of (3.10). □

3.5 Approximate Existence Lemma. *Under the hypotheses of Lemma 3.4, a smooth solution of* $(\star)_\varepsilon$ *exists.*

Proof. 1. We use the method of continuity applied to $(\star)_{\varepsilon,\tau}$, $0 \le \tau \le L-2$. First let us prove that there is a solution for $\tau = 0$ and small enough ε. Set $u = \varepsilon w$ and rewrite $(\star)_{\varepsilon,\tau}$ as

$$F^\varepsilon(w) := \operatorname{div}\left(\frac{\nabla w}{\sqrt{|\nabla w|^2+1}}\right) - \varepsilon\sqrt{|\nabla w|^2+1} = 0,$$

with $w = 0$ on $\partial\Omega_L$. Clearly the map

$$F\,:\,C^{2,\alpha}_0(\bar\Omega_L)\times\mathbf{R}\to C^\alpha(\bar\Omega_L),$$

defined by $F(w,\varepsilon) := F^\varepsilon(w)$ is C^1, and possesses the solution $F^0(0) = 0$. The linearization of F^0 at $w = 0$ is given by

$$\mathcal{D}F^0|_0 = \Delta\,:\,C^{2,\alpha}_0(\bar\Omega_L)\to C^\alpha(\bar\Omega_L),$$

the ordinary Laplace-Beltrami operator, an isomorphism. Then by the Implicit Function Theorem there is a solution of $F^\varepsilon(w) = 0$ for sufficiently small ε, and hence of $(\star)_{\varepsilon,\tau}$ with $\tau = 0$.

2. Next we fix ε and vary τ. Let I be the set of τ such that $(\star)_{\varepsilon,\tau}$ possesses a solution. Now I contains 0 by Step 1, and $I\cap[0,L-2]$ is closed by estimate (3.10) and the Arzela-Ascoli theorem.

Let us prove that I is open. Let π be the boundary values map $u\mapsto u|\partial\Omega$, and define $G^\tau(u) := (E^\varepsilon(u), \pi(u) - \tau\chi_{\partial F_L})$, so that $(\star)_{\varepsilon,\tau}$ is equivalent to $G^\tau(u) = (0,0)$. Clearly the map

$$F\,:\,C^{2,\alpha}(\bar\Omega_L)\times\mathbf{R}\to C^\alpha(\bar\Omega_L)\times C^{2,\alpha}(\partial\Omega_L),$$

defined by $F(u,\tau) = G^\tau(u)$, is C^1. The linearization of G^τ at a solution u is given by

$$\mathcal{D}G^\tau|_u = \begin{pmatrix}\mathcal{D}E^\varepsilon|_u\\ \pi\end{pmatrix}\,:\,C^{2,\alpha}(\bar\Omega_L)\to C^\alpha(\bar\Omega_L)\times C^{2,\alpha}(\partial\Omega_L).$$

Now E^ε has the form

$$E^\varepsilon(u) = \nabla_i A^i(\nabla u) + B(\nabla u),$$

which is independent of u, so it follows by the maximum principle that the linearization

$$\mathcal{D}^\varepsilon E|_u(v) = \nabla_i(A^i_{p_j}(\nabla u)\nabla_j v) + B_{p_j}(\nabla u)\nabla_j v$$

possesses only the zero solution. Then using existence theory and Schauder estimates up to the boundary, $\mathcal{D}G^\tau|_u$ is seen to be an isomorphism.

Then by the Implicit Function Theorem, the set of τ for which $G^\tau(u) = (0,0)$ is solvable (namely I) is open. Therefore $L-2 \in I$, which proves existence of u^ε in $C^{2,\alpha}$. Smoothness follows by Schauder estimates. □

Now we are prepared to prove the existence of the exact solution u. Because the downward translating graphs are an exact solution of $(\star)$ one dimension higher, we may use the Compactness Theorem as it stands to pass these sets to limits, obtaining a family of cylinders in $M \times \mathbf{R}$, which we will slice by M to obtain a family of surfaces weakly solving $(\star)$.

Proof of Theorem 3.1. 1. First assume that v is smooth with nonvanishing gradient. By Lemma 3.5, for each $L > 0$ there exists a smooth solution $u^\varepsilon = u^{\varepsilon, L-2}$ of $(\star)_\varepsilon$ on Ω_L, where $\varepsilon = \varepsilon(L) \to 0$ as $L \to \infty$. Combining (3.8) and (3.9), we find that

$$|\nabla u^\varepsilon(x)| \leq \max_{\partial E_0 \cap B_r(x)} H_+ + 2\varepsilon + \frac{C(n)}{r},$$

on each compact subset of $M \setminus E_0$ and sufficiently large L. By the Arzela-Ascoli theorem, there exists $L_i \to \infty$, $\varepsilon_i \to 0$, a subsequence u_i, and a locally Lipschitz function u such that

$$u_i \to u,$$

locally uniformly on $M \setminus E_0$, and u satisfies estimate (3.1) in the limit. By (3.7) with $\tau = L-2$,

$$u \geq 0 \quad \text{in } M \setminus E_0, \qquad u \to \infty \quad \text{as } x \to \infty.$$

Let $U_i(x,z) = u_i(x) - \varepsilon_i z$ as in (3.2), and set $U(x,z) := u(x)$. Then

$$U_i \to U,$$

locally uniformly on $(M \setminus E_0) \times \mathbf{R}$ with local Lipschitz bounds. The sets

$$N_t^i := \{U_i = t\} = \operatorname{graph}\left(\frac{u_i}{\varepsilon_i} - \frac{t}{\varepsilon_i}\right), \qquad -\infty < t < \infty,$$

smoothly solve $(\star)$ with $H > 0$ in the region $\Omega_{L_i} \times \mathbf{R}$, so Lemma 2.3 implies that U_i satisfies the variational formulation (1.5) on $\Omega_{L_i} \times \mathbf{R}$. Then the Compactness Theorem 2.1 implies that U satisfies (1.5) on $(M \setminus \bar{E}_0) \times \mathbf{R}$.

Let us check that u satisfies (1.5) on $M \setminus \bar{E}_0$. Let v be a locally Lipschitz function such that $\{v \neq u\} \subset\subset M \setminus \bar{E}_0$. Let $\phi(z)$ be a cutoff function with $|\phi_z| \leq 1$, $\phi = 1$ on $[0, S]$ and $\phi = 0$ on $\mathbf{R} \setminus (-1, S+1)$. Inserting $V(x,z) := \phi(z)v(x)$ into $J_U(U) \leq J_U(V)$, we obtain

$$\int_{K\times[-1,S+1]} |\nabla u| + u|\nabla u|\,dx\,dz \leq \int_{K\times[-1,S+1]} \phi|\nabla v| + v|\phi_z| + \phi v|\nabla u|\,dx\,dz,$$

where K contains $\{u \neq v\}$. Dividing by S and passing $S \to \infty$ proves that u satisfies (1.5). Finally, extend u negatively to E_0 so that $E_0 = \{u < 0\}$. This completes the proof of existence of the initial value problem (††) in the event that v is smooth with $\nabla v \neq 0$.

2. In the general case, fix $L > 0$ and select an open set U_L with smooth boundary such that $F_L \subset\subset U_L \subset\subset M$. Modify the metric on U_L near ∂U_L to obtain a complete metric g_L on U_L such that $g_L = g$ on F_L, $g_L \geq g$ on all of U_L, and near ∂U_L, g_L is isometric to a Riemannian cone of the form

$$g_L = Cs^2 g_{\partial U_L} + ds^2 \qquad \text{in } \partial U_L \times [C, \infty).$$

For some $\alpha > 0$, $\alpha \log s$ is a smooth subsolution of (1.5) on $\partial U_L \times [C, \infty)$. By step 1, there exists a proper solution u_L of (††) in (M_L, g_L) with initial condition E_0.

By the reasoning of (1.18), $\min(v, L)$ is a weak subsolution of ($\star\star$) in U_L with respect to g. Since $g = g_L$ where $v < L$ and otherwise $g_L \geq g$, it follows that $\min(v, L)$ remains a weak subsolution when considered with respect to the metric g_L. Then by the Comparison Theorem 2.2(ii), we have $u_L \geq v$ on F_L. Passing $L \to \infty$ and taking a convergent subsequence via (3.1), the Compactness Theorem 2.1 yields a solution u defined everywhere on M, with $u \geq v$. Theorem 2.2(iii) implies that u is unique. The gradient bound (3.1) is preserved. ☐

Remarks. 1. By passing *all* of the surfaces N_t^i to limits, we can show that at a jump, the interior U of $\{u = t\} \times \mathbf{R}$ is foliated by surfaces that minimize area subject to an obstacle condition consisting of the vertical walls ∂E_t and ∂E_t^+. Each such surface is either a vertical cylinder or a smooth graph over an open subset of U. This construction yields a high-speed movie of the jump from N_t to N_t^+.

2. We mention an alternate weak formulation of the inverse mean curvature flow. We call the locally Lipschitz function u a solution if there exists a measurable vector field ν

such that

$$(3.15)\qquad \begin{gathered} |\nu| \leq 1, \qquad \nabla u \cdot \nu = |\nabla u| \text{ a.e.,} \\ \int_\Omega \nabla\phi \cdot \nu + \phi|\nabla u| = 0 \qquad \text{for all } \phi \in C^1_c(\Omega). \end{gathered}$$

The vector field ν extends $\nabla u/|\nabla u|$ as a calibration across the gulfs. In fact, ν is the projection to TM of the normal vector of the foliation constructed above. The following properties hold.

(i) Formulation (3.15) implies the variational formulation (1.5).

(ii) Compactness holds for (3.15), and consequently the existence procedure does too.

(iii) Uniqueness then implies that (3.15) is equivalent to (1.5) under the hypotheses of Theorem 3.1.

§4 Topological Consequences

We review the existence and topology of exterior regions, and prove that in an exterior region, a surface evolving weakly by inverse mean curvature flow remains connected. This sets the stage for bounding the Euler characteristic in §5.

Exterior and Trapped Regions. Let us recall some definitions and well-known results; see [G2,Hm,MSY,MY,SSY,Ga]. Let M be a complete 3-manifold with asymptotically flat ends. We allow M to have a smooth, compact boundary consisting of minimal surfaces (that is, surfaces of vanishing mean curvature).

Let K_1 be the closure of the union of the images of all smooth, compact, immersed minimal surfaces in M. Since the region near infinity is foliated by spheres of positive mean curvature, K_1 is compact.

The *trapped region* K is defined to be the union of K_1 together with the bounded components of $M \setminus K_1$. The set K is compact as well. By minimization, there exists a smooth, embedded, stable surface separating any two ends of M, so $M \setminus K$ contains exactly one connected component corresponding to each end of M. The following lemma tells us about the smoothness and topology of these components.

4.1 Lemma. *(i) The topological boundary of the trapped set K consists of smooth, embedded minimal 2-spheres. The metric completion, M', of any connected component of $M \setminus K$*

is an "exterior region", that is, M' is connected and asymptotically flat, has a compact, minimal boundary, and contains no other compact minimal surfaces (even immersed).

(ii) A 3-dimensional exterior region M' is diffeomorphic to $\mathbf{R}^3$ minus a finite number of open 3-balls with disjoint closures. The boundary of M' minimizes area in its homology class.

Note that there is no curvature hypothesis. We take the metric completion rather than the closure because some component of K might be a nonseparating surface. The Lemma embodies various well-known theorems in the literature [MSY,MY,Ga], but we sketch the proof for completeness.

Proof. (i) First we show that K can be generated by a tamer collection of surfaces. Let C be the perimeter of some bounded open set containing K, and let $\mathcal{A}$ be the collection of smooth, compact, embedded minimal hypersurfaces in M such that

(a) $|N| \leq C$, and

(b) the double cover of N that orients its normal bundle is a stable minimal immersion. (If N is transversely orientable, this says that N itself is stable.)

Let P be one of the surfaces used in defining K_1, that is, an arbitrary smooth, connected, compact, immersed minimal surface in M. Suppose $P \notin \mathcal{A}$. Let $\hat{M}$ be the metric completion of $M \setminus P$, and let $\hat{P}$ be the union of the (one or two) boundary components that correspond to P. Let N minimize area among smooth surfaces in $\hat{M}$ that separate $\hat{P}$ from infinity. Note that $|N| \leq C$.

Since $P \notin \mathcal{A}$, either P has a segment of transverse double points, $|P| > C$, or P fails the stability (b). It follows that N cannot contain any component of $\hat{P}$. By the strong maximum principle, N lies in $M \setminus P$. It follows that $N \in \mathcal{A}$.

Therefore, either $P \in \mathcal{A}$, or P is separated from infinity by a disjoint surface $N \in \mathcal{A}$. This shows we may restrict to the class $\mathcal{A}$ in defining K.

Now ∂K is contained in the closure of the union of $\mathcal{A}$. But the class $\mathcal{A}$ is compact in the C^2 topology (see [SSY]), so in fact, each point of ∂K lies in some N in $\mathcal{A}$.

Let N, N_1 be distinct, connected surfaces in $\mathcal{A}$ that meet ∂K. We claim that they are disjoint. For otherwise, by the strong maximum principle, they cross, and using $N \cup N_1$ as a barrier for an area minimization problem, we find that $N \cup N_1$ is contained in $\operatorname{int} K$, a contradiction.

Next the surfaces in $\mathcal{A}$ that meet ∂K are actually isolated from one another. Suppose we have connected surfaces N, $N_i \in \mathcal{A}$ such that each N_i is disjoint from N, but the C^2 limit surface touches N. Then by the strong maximum principle, the limit surface equals N, so for sufficiently large i, N_i is either a single or double cover over N under the nearest point projection. In either case, the region between N_i and N is part of the interior of K, so that eventually N_i does not meet ∂K. This shows that a surface $N \in \mathcal{A}$ that meets ∂K is isolated from the other surfaces in $\mathcal{A}$ that meet ∂K.

It follows that ∂K consists of a finite union of disjoint, connected, stable minimal hypersurfaces. Each surface is either a component of K, or bounds K on one side. This proves the smoothness of ∂K. The spherical topology comes from (ii). The claims about M' are immediate.

(ii) Let M' be an exterior region. If M' is not simply connected or has more than one end, then its universal cover must have several ends, which as above, are separated by some smooth, compact minimal surface. Clearly this surface is not contained entirely in the boundary, so its projection to M' would violate the exterior region hypothesis. Therefore M' is simply connected, with one end.

By the Loop Theorem [He,MY] and simple connectedness, $\partial M'$ consists of 2-spheres. By simple connectedness, M' contains no one-sided $\mathbf{RP}^2$, so by [MSY], any fake 3-cell in M' would be surrounded by a least area sphere, which would contradict the exterior region hypothesis. The manifold obtained by filling in the boundary spheres with balls and compactifying at infinity is therefore a simply connected 3-manifold without fake 3-cell, hence is S^3. This implies that M' has the topology claimed.

To see that $\partial M'$ is area-minimizing: using large spheres as barriers, there is some area-minimizing compact surface N homologous to $\partial M'$, but since M' is an exterior region, N must lie in $\partial M'$, and therefore $N = \partial M'$. $\square$

Remark. Statement (i) can alternately be proven by flowing a large celestial 2-sphere forever by mean curvature. The mean curvature remains positive, so the flow is unidirectional, and the maximum principle keeps the spheres from the different ends from running into each other. The flowing surface can split into further spheres which eventually stabilize at a finite union of stable, smooth 2-spheres. The area swept out is free of minimal surfaces. This argument can be made rigorous by recent techniques of White [Wh].

The following lemma implies that a connected surface evolving in an exterior region remains connected. We are grateful to P. Li for clarifying the role of topology in part (ii) of the lemma.

4.2 Connectedness Lemma.

(i) A solution u of (1.5) *has no strict local maxima or minima.*

(ii) Suppose M is connected and simply connected with no boundary and a single, asymptotically flat end, and $(E_t)_{t>0}$ is a solution of (†) *with initial condition E_0. If ∂E_0 is connected, then N_t remains connected as long as it stays compact.*

Proof. (i) If u possesses a strict local maximum or minimum, then there is an "island" $E \subset\subset \Omega$, that is, a connected, precompact component of $\{u > t\}$ or of $\{u < t\}$ for some t. Define the Lipschitz function v by $v = u$ on $\Omega \setminus E$, $v = t$ on E. Then (1.5) yields

$$\int_E |\nabla u| + u|\nabla u| \leq \int_E t|\nabla u|.$$

In the case of $\{u > t\}$, this immediately implies that $u = t$ on E, a contradiction. In the case of $\{u < t\}$, we select t to ensure that $u > t-1$ on E, and again obtain a contradiction.

(ii) Let $\Omega := M \setminus \bar{E}_0$. Let $t > 0$. By (1.10), each N_t is approximated in C^1 by earlier surfaces N_{s_i} at which there are no jumps. Therefore we may assume that $N_t = \{u = t\}$.

Let $W := \{u < t\}$. Because u is proper, $W \cap \Omega$ is bounded. But by (i), no component of W can be compactly contained in Ω. Therefore, each component of W meets $\partial\Omega = \partial E_0$. Since ∂E_0 is connected, it follows that W is connected.

Let $X := \{u > t\}$. Since u is proper, X contains a neighborhood of infinity, which is connected since M has only one end. On the other hand, by (i), X has no precompact component. Therefore X is connected.

If N_t has more than one component, then by the connectedness of X and W, there is a closed loop γ that starts near ∂E_0, crosses N_t via the first component, and returns via the second component. By intersection theory, γ cannot be homotoped to zero, a contradiction to the simple connectedness. □

§5 Geroch Monotonicity

The Geroch Monotonicity Formula states that the Hawking quasi-local mass of a connected surface in a manifold of nonnegative scalar curvature is monotone during the inverse mean curvature flow. It was used by Geroch [G] and Jang-Wald [JW] to argue for the Penrose Inequality in case the flow remains smooth forever. In this section, we establish the formula for the weak flow. An advantage of the elliptic regularization approximation is that it preserves most of the geometry, so a computation at the ε-level is feasible, which may then be passed to limits.

Monotonicity Calculation. Let us first demonstrate the monotonicity in the smooth case. Recall that the Hawking mass is defined by $m_H(N) := |N|^{1/2}(16\pi - \int_N H^2)/(16\pi)^{3/2}$. We find by (1.1) and (1.3),

$$\frac{d}{dt}\int_{N_t} H^2 = \int_{N_t} -2H\Delta\left(\frac{1}{H}\right) - 2|A|^2 - 2Rc(\nu,\nu) + H^2.$$

The Gauss equation implies

$$K = K_{12} + \lambda_1\lambda_2 = \frac{R}{2} - Rc(\nu,\nu) + \frac{1}{2}\left(H^2 - |A|^2\right)$$

where K is the Gauss curvature of N, K_{12} is the sectional curvature of M in the direction of T_xM, λ_1, λ_2 are the principal curvatures of N in M, and R is the scalar curvature of M. Eliminating Rc in favor of R, we find

$$\begin{aligned}\frac{d}{dt}\int_{N_t} H^2 &= \int_{N_t} -2\frac{|DH|^2}{H^2} - |A|^2 - R + 2K \\ &= 4\pi\chi(N_t) + \int_{N_t} -2\frac{|DH|^2}{H^2} - \frac{1}{2}H^2 - \frac{1}{2}(\lambda_1-\lambda_2)^2 - R \\ &\le \frac{1}{2}\left(16\pi - \int_{N_t} H^2\right)\end{aligned}$$

provided that N_t is *connected* and $R \ge 0$. (A similar calculation is used in studying positive scalar curvature via stable minimal surfaces, including the Schoen-Yau proof of the positive mass theorem.)

It follows that the quantity

$$e^{t/2}(16\pi - \int_{N_t} H^2)$$

is nondecreasing. Since $|N_t|^{1/2}$ equals $e^{t/2}$ up to a factor, this shows that $m_H(N_t)$ is nondecreasing.

The estimate of $\int_{N_t} H^2$ is quite interesting in itself because it is independent of the detailed geometry of M except the scalar curvature.

Heuristically, when the surface jumps, we should obtain by Lemma 1.4(iv) and (1.15),

$$\int_{\partial E_t^+} H^2 \leq \int_{\partial E_t} H^2, \qquad |\partial E_t^+| = |\partial E_t|,$$

at least for $t > 0$, which implies that the monotonicity is preserved even at a jump. This reasoning is valid at $t = 0$ as well, provided E_0 is a minimizing hull. This is how the equation divines that there are no other minimal surfaces in M. In the remainder of this section, we substantiate this discussion.

Monotonicity Calculation, ε-Version. We will now reproduce the Geroch formula in a weak setting. Assume for now that N_0 is smooth and M possesses a smooth subsolution at infinity. Let $u^\varepsilon := u^{\varepsilon, L-2}$ be an approximator defined over Ω_L as in the proof of Theorem 3.1. Recall that $N_t^\varepsilon := \mathrm{graph}(u^\varepsilon/\varepsilon - t/\varepsilon)$, $-\infty < t < \infty$, is a smooth inverse mean curvature flow. Let ν be the downward unit normal to N_t^ε. Note $H = -\vec{H} \cdot \nu > 0$.

Select a cutoff function $\phi \in C_c^2(\mathbf{R})$ such that $\phi \geq 0$, $\mathrm{spt}\, \phi \subseteq [1, 5]$, and $\int \phi(z)\, dz = 1$. Fix an arbitrary $T > 0$ and require $L \geq T + 7$, $\varepsilon \leq 1$, so that ∂N_t^ε is disjoint from $M \times \mathrm{spt}\, \phi$ for $0 \leq t \leq T$. The boundary term disappears and we calculate

$$\begin{aligned}
\frac{d}{dt} \int_{N_t^\varepsilon} \phi H^2 &= \int_{N_t^\varepsilon} 2\phi H \frac{\partial H}{\partial t} + H^2 \nabla\phi \cdot \frac{\nu}{H} + \phi H^2 \\
&= \int_{N_t^\varepsilon} \phi \left(-2H\Delta\left(\frac{1}{H}\right) - 2|A|^2 - 2Rc(\nu,\nu) \right) + \nabla\phi \cdot \nu H + \phi H^2 \\
&= \int_{N_t^\varepsilon} \phi \left(-2\frac{|DH|^2}{H^2} - 2|A|^2 - 2Rc(\nu,\nu) + H^2 \right) \\
&\qquad\qquad - 2D\phi \cdot \frac{DH}{H} + \nabla\phi \cdot \nu H,
\end{aligned} \tag{5.1}$$

for $0 \leq t \leq T$. In integrated form this becomes

$$\begin{aligned}
\int_{N_r^\varepsilon} \phi H^2 = \int_{N_s^\varepsilon} \phi H^2 + \int_r^s \int_{N_t^\varepsilon} \phi \left(2\frac{|DH|^2}{H^2} + 2|A|^2 + 2Rc(\nu,\nu) - H^2 \right) \\
+ 2D\phi \cdot \frac{DH}{H} - \nabla\phi \cdot \nu H,
\end{aligned} \tag{5.2}$$

for $0 \leq r < s \leq T$.

Estimates. Let us estimate each of these terms in turn, with an eye to making the above formula converge as $\varepsilon \to 0$, $L \to \infty$. Fix $T > 0$ and assume $0 \le t \le T$. Our constants may depend on T but not on ε or L.

By (3.7), there is $R(T) > 0$ depending only on the subsolution v such that

$$N_t^\varepsilon \cap (M \times \operatorname{spt}\phi) \subseteq K(T) := (B_{R(T)} \setminus E_0) \times [1,5], \qquad 0 \le t \le T, \tag{5.3}$$

a fixed compact set. Applying the minimizing hull property to the supergraph E_t^ε compared to the perturbation $E_t^\varepsilon \cup K(T)$, we obtain the area estimate

$$|N_t^\varepsilon \cap (M \times \operatorname{spt}\phi)| \le |\partial K(T)| = C(T), \qquad 0 \le t \le T. \tag{5.4}$$

From (3.6), (3.8), the fact that $H_{N_t^\varepsilon} = |\nabla u^\varepsilon|$, and (5.3), we have for $L \ge T+8$,

$$|H| \le C(T) \qquad \text{on } N_t^\varepsilon \cap (M \times \operatorname{spt}\phi), \qquad 0 \le t \le T. \tag{5.5}$$

These two inequalities imply

$$\int_{N_t^\varepsilon} \phi H^2 + |\nabla\phi \cdot \nu H| \le C(T), \qquad 0 \le t \le T. \tag{5.6}$$

Next,

$$\left| 2D\phi \cdot \frac{DH}{H} \right| \le \frac{|D\phi|^2}{\phi} + \phi\frac{|DH|^2}{H^2} \le C + \phi\frac{|DH|^2}{H^2},$$

since ϕ is C^2 of compact support. Putting these two inequalities into (5.1), we obtain

$$\frac{d}{dt}\int_{N_t^\varepsilon} \phi H^2 \le \int_{N_t^\varepsilon} -\phi\left(\frac{|DH|^2}{H^2} + 2|A|^2\right) + C(T), \qquad 0 \le t \le T, \tag{5.7}$$

which implies together with (5.6) and (5.5) that

$$\int_0^T \int_{N_t^\varepsilon \cap (M\times[2,4])} \frac{|DH|^2}{H^2} + |DH|^2 + |A|^2 \le C(T), \tag{5.8}$$

using a ϕ such that $\phi = 1$ on $[2,4]$. For any sequence $\varepsilon_i \to 0$, Fatou's Lemma yields

$$\liminf_{i\to\infty} \int_{N_t^{\varepsilon_i} \cap (M\times[2,4])} \frac{|DH|^2}{H^2} + |DH|^2 + |A|^2 < \infty, \qquad \text{a.e. } t \ge 0. \tag{5.9}$$

Henceforth we shrink ϕ so that $\operatorname{spt}\phi \subseteq [2,4]$.

Convergence. As in the proof of Theorem 3.1, there are subsequences $\varepsilon_i \to 0$, $L_i \to \infty$, $N_t^i = N_t^{\varepsilon_i}$ such that

$$N_t^i \to \tilde{N}_t = N_t \times \mathbf{R} \quad \text{locally in } C^1, \qquad \text{a.e. } t \geq 0, \tag{5.10}$$

where $N_t = \partial E_t$ and $(E_t)_{t>0}$ is the unique solution of (†) for E_0.

We wish to pass (5.2) to limits. Let us first address the $\int_{N_t^i} \phi H^2$ term. By (1.14), this expression is lower semicontinuous as $i \to \infty$ (at least where there is no jump). To make this an equality for a.e. $t \geq 0$, we use growth control and Rellich's theorem. By (5.7), the function $\int_{N_t^i} \phi H^2 - C(T)t$ is monotone, $0 \leq t \leq T$. Therefore by choosing a diagonal subsequence (labelled the same) we may arrange that

$$\lim_{i\to\infty} \int_{N_t^i} \phi H^2 \text{ exists}, \qquad \text{a.e. } t \geq 0. \tag{5.11}$$

Suppose t is such that (5.9) and (5.10) hold. It is possible to write the converging surfaces N_t^i simultaneously as graphs of C^1 functions w_i over some smooth surface W, that is,

$$N_t^i \cap (M \times [2,4]) = \{x + w_i(x)\nu_W(x) \,:\, x \in W\} \cap (M \times [2,4]), \qquad w_i \in C^1_{loc}(W),$$

and $w_i \to w$ locally in C^1, where $\tilde{N}_t$ is the graph of w. (Here addition indicates the normal exponential map.) Then by Rellich's theorem, (1.13), and (5.9), there is a subsequence i_j such that

$$H_{N_t^{i_j}} \to H_{\tilde{N}_t} \qquad \text{in } L^2(W \cap (M \times [2,4]), \tag{5.12}$$

where the curvature function are projected to W for comparison. We conclude by (5.11) that the full sequence converges, namely

$$\int_{N_t^i} \phi H^2 \to \int_{\tilde{N}_t} \phi H^2, \qquad \text{a.e. } t \geq 0. \tag{5.13}$$

Together with (5.6) and the bounded convergence theorem, this implies that

$$\int_r^s \int_{N_t^i} \phi H^2 \to \int_r^s \int_{\tilde{N}_t} \phi H^2. \tag{5.14}$$

for any $0 \leq r < s$.

Next we turn to the $\nabla\phi \cdot \nu H$ term. By (5.5) and (5.4), we have

$$\int_{N_t^i} |\nabla\phi \cdot \nu H| \le C(T) \sup_{N_t^i} |\nabla\phi \cdot \nu| \to 0, \qquad \text{a.e. } t \ge 0.$$

We have used the fact that N_t^i converges locally in C^1 to the vertical cylinder $\tilde{N}_t$, on which ν is perpendicular to $\nabla\phi$. Then by (5.6) and the bounded convergence theorem, we obtain

$$\int_r^s \int_{N_t^i} |\nabla\phi \cdot \nu H| \to 0. \tag{5.15}$$

Next we address the $|DH|^2/H^2$ term.

5.1 Lemma. *For a.e. $t \ge 0$,*

$$H > 0 \qquad \mathcal{H}^{n-1}\text{-a.e. on } N_t.$$

Proof. Clearly $|\nabla u| > 0$ a.e. with respect to the measure $|\nabla u|\,dx$. By the co-area formula, this implies that $|\nabla u|$ exists and $|\nabla u| > 0$ for a.e. $t \ge 0$ and $\mathcal{H}^{n-1}$-a.e. $x \in N_t$. By (1.12), this implies the claimed result. □

In particular the integral $\int_{\tilde{N}_t} |DH|^2/H^2$ makes sense for a.e. t.

5.2 Lower Semicontinuity Lemma. *For each $0 \le r < s$,*

$$\int_r^s \int_{\tilde{N}_t} \phi \frac{|DH|^2}{H^2} \le \liminf_{i\to\infty} \int_r^s \int_{N_t^i} \phi \frac{|DH|^2}{H^2}.$$

Proof. By (5.9), for a.e. $t \ge 0$ there is a subsequence i_j such that

$$\sup_j \int_{N_t^{i_j} \cap (M\times[2,4])} \frac{|DH|^2}{H^2} < \infty. \tag{5.16}$$

Let $\hat{N}$ be a connected component of $\tilde{N}_t \cap (M \times [2,4])$, and let $\hat{N}^j$ be a connected component of $N_t^{i_j} \cap (M \times [2,4])$ converging locally in C^1 to $\hat{N}$.

Let a_j be the median of $\log H$ on $\hat{N}^j$. By (5.16) and the Rellich theorem, there exists $a \in [-\infty, \infty)$, $f \in L^2(\hat{N})$, and a further subsequence such that $a_j \to a$ and

$$\log H_{\hat{N}^j} - a_j \to f \quad \text{in } L^2(W) \text{ and a.e. on } W,$$

where $\hat{N}^j$, $\hat{N}$ are written as C^1 graphs over some nearby smooth surface W, and the functions are projected to W.

If $a = -\infty$, then $\log H_{\hat{N}^j} \to f + a = -\infty$ a.e., so $H_{\hat{N}^j} \to 0$ a.e. on W. By the boundedness of H and the weak convergence (1.13), it follows that $H_{\hat{N}} = 0$ a.e. on $\hat{N}$. But by Lemma 5.1, this case is excluded for a.e. $t \geq 0$.

Therefore we assume $a > -\infty$. Then

$$\log H_{\hat{N}^j} - a_j \to \log H_{\hat{N}} - a \qquad \text{in } L^2(W).$$

It follows that we have weak convergence of DH/H in each chart W. Then by the usual lower semicontinuity, we obtain for a.e. $t \geq 0$,

$$\int_{\tilde{N}_t} \phi \frac{|DH|^2}{H^2} \leq \liminf_{i\to\infty} \int_{N_t^i} \phi \frac{|DH|^2}{H^2},$$

which implies the desired result by Fatou's Lemma. □

The final term of (5.2) converges only weakly to zero.

5.3 Lemma. *For each* $0 \leq r < s$,

$$\int_r^s \int_{N_t^i} D\phi \cdot \frac{DH}{H} \to 0.$$

Proof. As is (5.15), the idea is that $D\phi$ and DH are orthogonal in the cylindrical limit. From the previous lemma, DH/H converges weakly subsequentially for individual times, but it is tricky to control the time direction. To do this, we recognize the inner integral as a time derivative, using the fact that N_t^i is moving rigidly by translation. Defining

$$\begin{aligned} g_i(t) &:= \int_{N_t^i} -\varepsilon_i \phi(z) \frac{\partial}{\partial z} \cdot \frac{DH}{H}\, d\mu_{N_t^i}(x,z) \\ &= \int_{N_0^{\varepsilon_i}} -\varepsilon_i \phi(z - t/\varepsilon_i) \frac{\partial}{\partial z} \cdot \frac{DH}{H}\, d\mu_{N_0^{\varepsilon_i}}(x,z), \end{aligned}$$

we notice that

$$\begin{aligned} \frac{d}{dt} g_i(t) &= \int_{N_0^{\varepsilon_i}} \phi'(z - t/\varepsilon_i) \frac{\partial}{\partial z} \cdot \frac{DH}{H} \\ &= \int_{N_t^i} \phi'(z) \frac{\partial}{\partial z} \cdot \frac{DH}{H} \\ &= \int_{N_t^i} D\phi \cdot \frac{DH}{H} \\ &=: f_i(t). \end{aligned}$$

We have by Cauchy's inequality, (5.4), and (5.8),

$$\sup_i \int_0^T |f_i| < \infty, \qquad \int_0^T |g_i| \le C(T)\varepsilon_i \to 0,$$

for each $T > 0$, hence

$$f_i \rightharpoonup 0 \qquad \text{on } [0, \infty),$$

in the sense of measures, which gives the result. □

Weak Second Fundamental Form. In order to cope with the $|A|^2$ term, we define the second fundamental form of a surface in $W^{2,2} \cap C^1$. See Hutchinson [Hut] for a related definition. Let N be a C^1 hypersurface of an ambient manifold M with induced metric $h = (h_{ij})$, orthogonal projection (h_i^j) from TM to TN, unit normal (ν^i), and weak mean curvature H in $L^1_{loc}(N)$, as defined by the first variation formula (1.11).

Suppose $p = (p_{ij})$ is a smooth, compactly supported symmetric 2-tensor defined on M, and apply the first variation formula to the vector $Y^j = g^{jl} p_{lk} \nu^k$ to derive in the smooth case

$$\int_N H \nu^j p_{jk} \nu^k = \int_N h^{ij} \nabla_i (p_{jk} \nu^k) = \int_N h^{ij} \nabla_i p_{jk} \nu^k + h^{ij} p_{jk} h^{kl} A_{li}, \tag{5.17}$$

recalling that $A_{ij} = h_i^k \nabla_k \nu^l h_{lj}$.

If N is C^1 and H exists as a locally integrable function, we call a locally integrable section $A = (A_{ij})$ of $\mathrm{Sym}^2(T^*N)$ the *second fundamental form* if the far right and left sides of (5.17) are equal for every p in $C^1_c(\mathrm{Sym}^2(T^*M))$.

Assuming N is C^1, it can be verified that $|A| \in L^2_{loc}(N)$ if and only if N can be written locally in coordinates as the graph of a $W^{2,2}$ function w. (The tensor A_{ij} can be written as the Hessian of w, with a correction that depends only on the normal vector and the Christoffel symbols of g. See (7.10).) Inserting $p_{ij} = \phi g_{ij}$, we find that if A exists, then indeed $H = h^{ij} A_{ij}$ a.e., where H is defined by (1.11).

It follows from weak compactness, Riesz Representation, and (1.13) that if $N^i \to N$ locally in C^1 and

$$\sup_i \int_{N^i} |A_{N^i}|^2 < \infty,$$

then A_N exists in $L^2(N)$ with the weak convergence

$$\int_{N^i} p^{jk} A^{N^i}_{jk} \to \int_N p^{jk} A^N_{jk}, \qquad p \in C^0_c(\mathrm{Sym}^2(TM)),$$

and lower semicontinuity

$$\int_N |A_N|^2 \le \liminf_{i\to\infty} \int_{N_i} |A_{N_i}|^2. \tag{5.18}$$

On the other hand, if a C^1 surface N is given satisfying $\int_N |A|^2 < \infty$, and we are allowed to choose N_i by mollification, we may arrange that each N_i is smooth and

$$N_i \to N \qquad \text{strongly in } C^1 \text{ and } W^{2,2}, \tag{5.19}$$

in the sense that N_i, N are represented locally as graphs of functions w_i, w such that $w^i \to w$ in $C^1 \cap W^{2,2}$. (We can even arrange that N_i approaches N from one side.)

In this situation, let λ_1, λ_2 be the eigenvalues of A with respect to h and consider the function in $L^1(N)$ defined by

$$K := K_{12} + \lambda_1\lambda_2,$$

where K_{12} is the sectional curvature function of the ambient manifold evaluated on T_xN. The following lemma is now a consequence of the Gauss and Gauss-Bonnet formulae applied to the approximators N_i.

5.4 Weak Gauss-Bonnet Formula. *Suppose N is a compact C^1 surface in a 3-manifold, satisfying $\int_N |A|^2 < \infty$. Then*

$$\int_N K_{12} + \lambda_1\lambda_2 = 2\pi\chi(N),$$

where χ denotes the Euler characteristic.

From this we have the following.

5.5 $\mathbf{W^{2,2}}$ Lemma. *Let N_t be the limiting surfaces defined in* (5.10). *Then*

$$\int_{N_t} |A|^2 \le C(T), \qquad 0 \le t \le T.$$

Proof. By (5.9) and (5.18),

$$\int_{N_t} |A|^2 < \infty \qquad \text{for a.e. } t \ge 0.$$

Since the surfaces N_t are compact with locally uniform C^1 estimates, $\chi(N_t)$, $|N_t|$, and $\sup_{N_t} |K_{12}|$ are bounded for $0 \le t \le T$. By Lemma 5.4, this yields

$$\int_{N_t} \lambda_1 \lambda_2 \le C(T), \qquad \text{a.e. } t \in [0,T].$$

Since $H = \lambda_1 + \lambda_2$ is bounded, this yields the desired estimate for a.e. $t \in [0,T]$. It follows for all t by (5.18). $\square$

Remark. A similar result can be proven for inverse mean curvature flow in dimensions $n < 8$, by a local version of the Gauss-Bonnet formula suitable for graphs.

We need one more technical lemma.

5.6 Lemma. *Suppose E is precompact, $E' = E$, and ∂E is $C^{1,1}$. Then either ∂E is a smooth minimal surface, or ∂E can be approximated in C^1 from inside by smooth sets of the form ∂E_τ with $H > 0$, $E'_\tau = E_\tau$, and*

$$\sup_\tau \sup_{\partial E_\tau} |A| < \infty, \qquad \int_{\partial E_\tau} H^2 \to \int_{\partial E} H^2 \quad \text{as } \tau \to 0. \tag{5.20}$$

Proof. Note that $H \ge 0$ on ∂E in the weak sense. We employ the ordinary mean curvature flow $\partial x / \partial \tau = -H\nu$ as developed in [Hu] in order to smooth the boundary while preserving nonnegative curvature. By mollifying, there exists a sequence $(Q_i)_{i \ge 1}$ of smooth surfaces with uniformly bounded $|A|$ approximating ∂E from inside in C^1 and strongly in $W^{2,2}$, so that

$$\int_{Q_i} H^2 \to \int_{\partial E} H^2, \qquad \int_{Q_i} H_-^2 \to \int_{\partial E} H_-^2 \qquad \text{as } i \to \infty,$$

where $H_- := \max(0, -H)$. Let $(Q^i_\tau)_{0 \le \tau < \delta_i}$ be the mean curvature flow of Q_i. On account of the inequality

$$\frac{\partial |A|^2}{\partial \tau} \le \Delta |A|^2 + 2|A|^4 + C|Rm||A|^2 + C|\nabla Rm||A| \qquad \text{on } Q^i_\tau,$$

and higher derivative estimates derived in [Hu], Q^i_τ exists with uniformly bounded $|A|$ for a uniform time independent of i. Passing smoothly to limits we obtain a flow $(Q_\tau)_{0<\tau\le\delta}$ with uniformly bounded $|A|$. We have using the first variation formula and (1.2),

$$\frac{d}{d\tau} \int_{Q^i_\tau} H^2 \le \int_{Q^i_\tau} 2H(\Delta H + |A|^2 H + Rc(\nu,\nu)H) \le C \int_{Q^i_\tau} H^2.$$

Integrating and passing to limits, we obtain $\int_{Q_\tau} H^2 \le e^{C\tau} \int_{\partial E} H^2$. Furthermore, because the speed is bounded and using Arzela-Ascoli, we have $Q_\tau \to \partial E$ in C^1. So by lower semicontinuity (1.14), we have completed (5.20).

By a variation of this argument, we also have

$$\int_{Q_\tau} H_-^2 \le e^{Ct} \int_{\partial E} H_-^2 = 0,$$

so $H \ge 0$ on Q_τ, and by the strong maximum principle, either $H \equiv 0$ or $H > 0$ on Q_τ. In the former case, ∂E is a smooth minimal surface. In the latter case, Q_τ smoothly foliates an interior neighborhood of ∂E. Using ν_{Q_τ} as a calibration as in the proof of Lemma 2.3 and using $E' = E$, we see that $E'_\tau = E_\tau$, where E_τ is the precompact set bounded by Q_τ. $\square$

We are now in a position to pass everything to limits to prove the following proposition. M need not be complete.

5.7 $\mathbf{H^2}$ Growth Formula. *Let M be a 3-manifold, E_0 a precompact open set with C^1 boundary satisfying*

$$\int_{\partial E_0} |A|^2 < \infty \tag{5.21}$$

and $(E_t)_{t>0}$ a family of open sets solving (†) with initial condition E_0. Then for each $0 \le r < s$,

$$\begin{aligned} \int_{N_r} H^2 &\ge \int_{N_s} H^2 + \int_r^s \int_{N_t} \left(2\frac{|DH|^2}{H^2} + 2|A|^2 + 2Rc(\nu,\nu) - H^2 \right) \\ &= \int_{N_s} H^2 + \int_r^s -4\pi\chi(N_t) \\ &\quad + \int_r^s \int_{N_t} \left(2\frac{|DH|^2}{H^2} + \frac{1}{2}(\lambda_1 - \lambda_2)^2 + R + \frac{1}{2}H^2 \right). \end{aligned} \tag{5.22}$$

provided E_s is precompact. In particular, the inner integrals make sense for a.e. $t \ge 0$.

The proof consists of passing to limits, using the above lemmas, to get a result for almost every time, then using a restarting argument to get the result for each time.

Proof. 1. In steps 1 and 2, we assume ∂E_0 is smooth and M possesses a smooth subsolution at infinity, so that the above discussion applies.

Combining together (5.13), Lemma 5.2, (5.18), Fatou's Lemma, (5.14), Lemma 5.3, and (5.15), we can pass (5.2) to limits as $i \to \infty$. This proves that (5.2) holds as an inequality, with the limiting cylinders $\tilde{N}_t$ replacing N_t^ε, for a.e. $0 < r < s$. Since $\tilde{N}_t$ is a cylinder for a.e. t, the integral of each geometric quantity $Q(x,z) = Q(x)$ breaks up as

$$\int_{\tilde{N}_t} \phi Q \, d\mu_{\tilde{N}_t} = \int_2^4 \phi \, dz \int_{N_t} Q \, d\mu_{N_t} = \int_{N_t} Q \, d\mu_{N_t},$$

and we obtain the first inequality of (5.22) for a.e. $0 < r < s$. To see the transition to the second inequality of (5.22), write

$$\begin{aligned} 2|A|^2 + 2Rc(\nu,\nu) - H^2 &= (|A|^2 - \frac{1}{2}H^2) + (R - 2K_{12} + |A|^2 - H^2) + \frac{1}{2}H^2 \\ &= \frac{1}{2}(\lambda_2 - \lambda_1)^2 + R - 2(K_{12} + \lambda_1\lambda_2) + \frac{1}{2}H^2 \end{aligned}$$

and apply Lemmas 5.4 and 5.5. Thus (5.22) holds for a.e. $0 < r < s$.

Recall from (1.10) that $N_t \to N_s$ in C^1 as $t \nearrow s$. Then the lower semicontinuity of $\int H^2$ in (1.14) implies (5.22) holds for a.e. $r > 0$ and all $s \geq r$.

2. Next we prove that (5.22) holds at $r = 0$. Recall from Theorem 1.3(iii) and (1.15) that $\partial E_0'$ is $C^{1,1}$ with $H \geq 0$ in a weak sense and

$$\int_{\partial E_0'} H^2 \leq \int_{\partial E_0} H^2.$$

So it suffices to prove (5.22) for E_0'. There are two cases.

Case 1: If $H > 0$ somewhere on $\partial E_0'$, then by Lemma 5.6, there is a family of smooth surfaces of the form ∂E_τ approximating ∂E_0 in C^1, with $H > 0$ and $E_\tau' = E_\tau$. By Theorem 3.1, there exists a proper solution $(E_t^\tau)_{t>0}$ of (†) with $E_0^\tau = E_\tau$. By Lemma 2.4, E_t^τ evolves smoothly for a short time, so (5.22) is valid for $(E_t^\tau)_{t\geq 0}$ at $r = 0$.

By (3.1), (5.20), Remark 1 following Theorem 2.1, and the uniqueness given by Theorem 2.2(iii), u^τ converges uniformly to u on $M \setminus E_0'$ as $\tau \to 0$, with C^1 convergence of the level sets except at jump times. All quantities in (5.22) converge or are lower semicontinuous to their values for u. This convergence employs the easily proven analogues of (5.13), Lemma 5.2, (5.18), and (5.14) for arbitrary sequences of exact solutions in a 3-manifold; we don't need a version of (5.15) or Lemma 5.3. In addition, by (5.20),

$$\int_{\partial E_\tau} H^2 \to \int_{\partial E_0'} H^2,$$

which proves (5.22) at $r = 0$ for E_0', and hence for E_0, in the case that $H > 0$ somewhere on $\partial E_0'$.

Case 2: The other possibility is that $\partial E_0'$ is a smooth minimal surface. Choose a sequence of smooth functions w_i on M such that

$$w_i \to 1 \quad \text{in } C^\infty(M), \qquad w_i = 1 \text{ and } \frac{\partial w_i}{\partial \nu} > 0 \quad \text{along } \partial E_0, \qquad w_i > 1 \quad \text{in } M \setminus E_0,$$

and consider the metric $g_i := w_i g$. Then $H_i > 0$ along ∂E_0 and E_0 is still a strictly minimizing hull with respect to g_i. Therefore by Step 2, (5.22) holds at $r = 0$ for the solution u_i of (††) for ∂E_0 in the metric g_i.

As above, passing $i \to \infty$, using (3.1), the Remarks following Theorem 2.1, uniqueness, and the lower semicontinuity of all the quantitites involved, together with the fact that

$$\int_{\partial E_0} H_i^2 \to \int_{\partial E_0} H^2,$$

we obtain (5.22) at $r = 0$ for E_0', and hence for E_0, in the case of a minimal surface. Together with Case 1, this proves (5.22) at $r = 0$ when the initial surface is smooth and there is a subsolution at infinity.

3. Next, let ∂E_0 be any C^1 surface satisfying (5.21), but retain the subsolution at infinity. By (5.19), we may approximate ∂E_0 in C^1 by smooth surfaces S_i lying in $M \setminus E_0$, such that

$$\int_{S_i} H^2 \to \int_{\partial E_0} H^2, \tag{5.23}$$

as $i \to \infty$. Let $(E_t^i)_{t>0}$ be the proper solution of (†) corresponding to S_i. Since $S_i \to \partial E_0$ in C^1, there is $\delta_i \to 0$ such that $E_0 \subseteq E_0^i \subseteq E_{\delta_i}$, so

$$E_t \subseteq E_t^i \subseteq E_{t+\delta_i} \qquad \text{for all } t \geq 0,$$

by the Comparison Theorem 2.2(ii). In particular, $u_i \to u$ locally uniformly in $M \setminus \bar{E}_0$, and for each fixed $t > 0$, ∂E_t^i is eventually kept uniformly away from S_i, so by estimate (3.1) and (2.1), we have

$$\partial E_t^i \to N_t \quad \text{in } C^1, \qquad \text{a.e. } t > 0.$$

By Step 2, (5.22) is valid for the approximators $(E^i_t)_{t>0}$ at $r = 0$, so passing to limits, using (5.23) at $r = 0$, and lower semicontinuity of all quantities for $t > 0$, we obtain (5.22) at $r = 0$ for any initial surface satisfying (5.21) and any M with a smooth subsolution.

4. Finally, the subsolution at infinity can be arranged by employing a conic modification near the edge of the manifold, as in the proof of Theorem 3.1, to prove (5.22) at $r = 0$ under the stated general hypotheses.

By Lemma 5.5, this result may now be applied at each $r \geq 0$ to establish (5.22) for arbitrary $0 \leq r \leq s$. □

The next Theorem now follows from (5.22), recalling from Lemma 1.6 that

$$|\partial E_t|^{1/2} = e^{t/2}|\partial E_0|^{1/2}$$

provided that E_0 is a minimizing hull. It is precisely at this point that the exterior region hypothesis declares itself, and it does so through the area rather than through higher order quantities.

5.8 Geroch Monotonicity Formula. *Let M be a complete 3-manifold, E_0 a precompact open set with C^1 boundary satisfying (5.21), and $(E_t)_{t>0}$ a solution of (†) with initial condition E_0. If E_0 is a minimizing hull, then*

$$\text{(5.24)} \quad m_H(N_s) \geq m_H(N_r) + \frac{1}{(16\pi)^{3/2}} \int_r^s |N_t|^{1/2} \Bigg[16\pi - 8\pi\chi(N_t) + \\ + \int_{N_t} \big(2|D\log H|^2 + (\lambda_1 - \lambda_2)^2 + R\big)\, d\mu_t \Bigg]\, dt$$

for $0 \leq r < s$, provided E_s is precompact.

Remark. According to Lemma 4.1(ii), an exterior region is simply connected. In such a manifold, recall from Lemma 4.2(ii) that a connected surface ∂E_t remains connected under the flow, so $\chi(N_t) \leq 2$.[5] Therefore, we already have the following monotonicity result in the case of a single black hole.

If $\tilde{M}$ is an exterior region satisfying (0.1) and (0.2), having $R \geq 0$, and with a connected boundary, then the weak inverse mean curvature flow of the boundary exists forever and has monotone nondecreasing Hawking mass.

[5]Note that even if N_0 is a 2-sphere, we cannot exclude the possibility that N_t later develops genus.

§6 Multiple Horizons

The purpose of this section is to derive a monotonicity result in the presence of multiple black holes.

6.1 Geroch Monotonicity (Multiple Boundary Components). *Let $\tilde{M}$ be an exterior region with $R \geq 0$ and satisfying the asymptotic conditions* (0.1) *and* (0.2). *For each connected component N of $\partial\tilde{M}$, there exists a flow of compact $C^{1,\alpha}$ surfaces $(N_t)_{t\geq 0}$, such that $N_0 = N$, $m_H(N_t)$ is monotone nondecreasing for all time, and for sufficiently large t, N_t satisfies the weak inverse mean curvature flow* (†).

The idea is to flow a single boundary component, treating the others as inessential "occlusions" to be slid across, without incurring any loss of Hawking mass. This is achieved by flowing E_t so that it nearly touches the rest of the boundary, then jumping to the strictly minimizing hull F of the union of E_t with some of the other boundary components. This accomplishes

$$|\partial F| \geq |\partial E_{t_1}|, \qquad \int_{\partial F} H^2 \leq \int_{\partial E_{t_1}} H^2, \tag{6.1}$$

which is sufficient to ensure that the Hawking mass does not decrease during the jump.

Unfortunately, our construction only works on one component at a time. It would be very desirable to have a result involving an additive combination of boundary components. One approach would be to keep track of the mass added during a jump. At the end of §8, we give an example that illustrates the difficulty of doing this.

Proof. Suppose the boundary components of $\tilde{M}$ are $N, N_1, \ldots, N_k$. Fill in the boundary by 3-balls

$$E_0, W_1, \ldots, W_k$$

to obtain a smooth, complete, boundaryless manifold M. We will construct a nested family of sets $(E_t)_{t\geq 0}$ and corresponding locally Lipschitz function u on M such that

(i) E_t satisfies the flow except at a finite number of "jump times" $0 < t_1 < \cdots t_p < \infty$,

(ii) Each W_i is contained in $\{u = t_j\}$ for some j, and E_t swallows at least one W_i at each jump time t_j,

(iii) E_t and ∂E_t remain connected,

(iv) The Hawking mass $m_H(\partial E_t)$ is monotone nondecreasing.

Write $W := \cup_{1\le i\le k} W_i$. By Lemma 4.1, M is diffeomorphic to $\mathbf{R}^3$, and $E_0 \cup W$ is a strictly minimizing hull in M.

To start with, let $(E_t)_{t\ge 0}$ be the flow of E_0 by (†). (Later we will modify it past a certain time.) Since M is simply connected and E_0 is connected, Lemma 4.2 implies that E_t and its boundary remain connected.

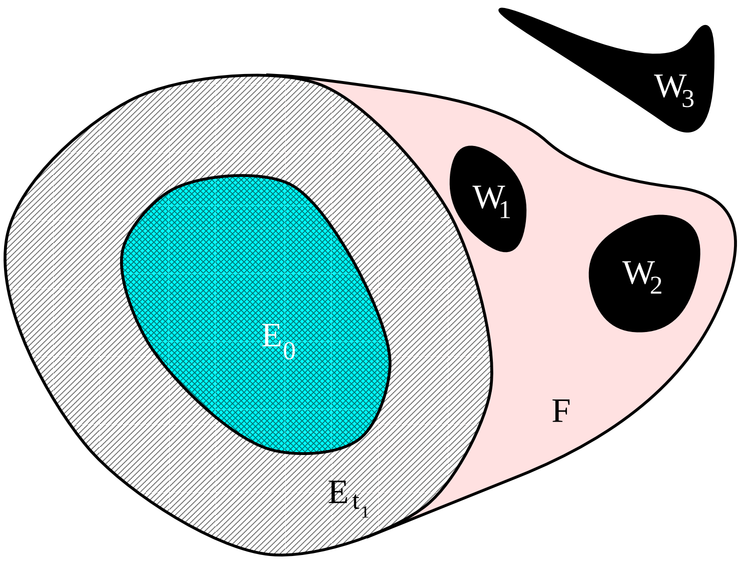
The case when W is empty is covered by the remarks following Theorem 5.8, so we assume that W is nonempty. Left to its own devices, E_t will eventually enter W, and the monotonicity is likely to fail since the scalar curvature there is unknown. So define $s_1 > 0$ to be the supremum of the times when E_t is disjoint from W. By continuity, we have

$$E_{s_1} \cap W = \emptyset, \qquad E^+_{s_1} \cap W \ne \emptyset.$$

Define $t_1 := s_1$ if $\bar{E}_{s_1} \cap \bar{W} = \emptyset$, otherwise let t_1 be slightly less that s_1.

Since ∂E_0 is a minimizing hull, ∂E_t is connected, and E_t stays in the region where $R \ge 0$, we have

$$m_H(\partial E_t) \quad \text{is monotone,} \qquad 0 \le t \le t_1.$$



Time to jump.

Now we construct the jump. Let F be the connected component of $(E_{t_1} \cup W)'$ that contains E_{t_1}. Because E_{t_1} is so close to W, we see that F contains at least one component of W, and we may write (after relabelling)

$$F = (E_{t_1} \cup W_1 \cup \ldots \cup W_j)',$$

for some nonempty union of whole components of W.

Since ∂W is a smooth minimal surface, ∂F is disjoint from ∂W by the strong maximum principle. The first inequality of (6.1) is immediate since E_{t_1} is a minimizing hull by Lemma 1.4(i).

Let us prove the second inequality of (6.1). By (5.19), choose a sequence of sets E_i containing E_{t_i} such that ∂E_i is smooth and converges to ∂E_{t_1} in C^1, and

$$\int_{\partial E_i} H^2 \to \int_{\partial E_{t_1}} H^2$$

We have by (1.15),

$$\int_{\partial E_i'} H^2 \leq \int_{\partial E_i} H^2.$$

Now it can be seen that $E_i' \to F$, and $\partial E_i' \to \partial F$ in C^1 by Theorem 1.3(ii), so passing to limits and recalling the lower semicontinuity given by (1.14), we obtain (6.1). Then (6.1) implies

$$m_H(\partial E_{t_1}) \leq m_H(\partial F).$$

In a similar manner, but using Lemma 6.2 below, we will now show that F is a suitable initial condition for restarting the flow. We may approximate ∂E_{t_1} in C^1 by a sequence of smooth surfaces with uniformly bounded mean curvature, of the form ∂U_i, where U_i contains E_{t_1}. Then by (1.15), $\partial(U_i \cup F)'$ is $C^{1,1}$ and itself has uniformly bounded mean curvature, and one easily checks that $\partial(U_i \cup F)'$ converges to ∂F in C^1. By slightly smoothing, we see that ∂F is approximated in C^1 by smooth surfaces of bounded mean curvature.

Then by Theorem 3.1, (3.1), and Theorem 2.1, there exist sets $(F_t)_{t>0}$ satisfying (††) with initial condition $F_0 = F$. Furthermore, by applying the Gauss-Bonnet Theorem to the approximating sequence, we see

$$\int_{\partial F} |A|^2 < \infty.$$

Also, F is a strictly minimizing hull and F_t remains connected, so as above, Theorem 5.8 applies to show

$$m_H(F_t) \quad \text{is monotone,}$$

at least until F_t in turn encounters one of the remaining components of W.

We now replace E_t by F_t for $t > t_1$. The desired monotonicity holds for a short time past t_1, and E_t has swallowed at least one component of W. Since W has only a finite number of components, we may continue this procedure inductively to obtain a nested family of sets $(E_t)_{t\geq 0}$ satisfying (i)-(iii). This proves the Theorem. □

Remark. We have some freedom in choosing the jump time. For example, we could have defined t_1 to be the first moment when E_t together with some component of W possesses a connected minimizing hull, or even the first moment when there is a connected "stationary hull".

All such schemes implicitly impose grazing (i.e. contact angle zero) boundary conditions for N_t along the boundary ∂W. It would be interesting to resolve the ambiguity by deriving the jumping scheme from a suitable minimization principle.

One possibility is to set $g = \infty$ on W. Like the methods above, this forces $u =$ constant on each F_i, and expels the "potential surfaces" N_t from the interior of W, much like an electric conductor. However, it is unlikely to yield monotonicity of the Hawking mass since there may be negative R concentrating along the boundary of the obstacle.

Now the promised Lemma, which guarantees that N_t always has the necessary smoothness to restart the flow.

6.2 Lemma. *Suppose that ∂E is compact and C^1, E minimizes*

$$J_f(F) := |\partial F| + \int_F f$$

in some open neighborhood U of ∂E, and $\operatorname{ess\,sup}_U |f| \leq M$. Then ∂E may be approximated in C^1 on either side by smooth surfaces with $|H| \leq M + \varepsilon$.

Proof. Choose an open set V with $\partial E_0 \subset\subset V \subset\subset U$. Define g by

$$g := M + \varepsilon \quad \text{in } U \setminus E, \qquad g := -M - \varepsilon \quad \text{in } U \cap E.$$

and select a smooth sequence g_i with $|g_i| \leq M + \varepsilon$, $g_i \to g$ in $L^1(\bar{V})$.

Let E_i be a minimizer of J_{g_i} among all F with $F\setminus\bar{V} = E\setminus\bar{V}$. By lower semicontinuity and the compactness theorem for sets of finite perimeter, there exists a set E_1 and a subsequence with $E^{i_j} \to E_1$, such that E_1 minimizes J_g subject to the same condition. Since

$$J_f(E) \leq J_f(E_1),$$

the definition of g yields the contradiction

$$J_g(E) < J_g(E_1),$$

unless $E = E_1$ a.e. Then by the Regularity Theorem 1.3, $\partial E_i \to \partial E$ in C^1 locally in V. Excluding extraneous components, we may take the convergence to occur in all of M. The smoothness of g_i proves that ∂E_i is smooth with $|H_{\partial E_i}| = |g_i| \leq M + \varepsilon$. To guarantee that E_i contains E (say), it suffices to adjoin the condition that $g_i = -M - \varepsilon$ on $U \cap E$. $\square$

§7 Asymptotic Regime

By a weak blowdown argument, we show that N_t becomes $C^{1,\alpha}$ close to a large coordinate sphere as $t \to \infty$. Then we prove that $\lim_{t\to\infty} m_H(N_t) \leq m_{ADM}(M)$ by expanding the Hawking mass as the ADM mass plus lower order terms, using the montonicity formula itself for the analytic control.

Let Ω be the asymptotically flat end of M, embedded in $\mathbf{R}^n$ as the complement of a compact set K. Let g be the metric of M pulled back to Ω, let δ be the flat metric, and let ∇, $\bar{\nabla}$ be the corresponding connections. Write $B_r(x)$ for the balls with respect to g, and $D_r(x)$ for standard balls. Let u be a solution of (1.5) in Ω, $E^t := \{u < t\} \subseteq \Omega$.

Fix $\lambda > 0$ and define the blown down objects by

$$\Omega^\lambda := \lambda \cdot \Omega, \quad g^\lambda(x) := \lambda^2 g(x/\lambda), \quad u^\lambda(x) := u(x/\lambda), \quad E_t^\lambda := \lambda \cdot E_t,$$

where $\lambda \cdot A := \{\lambda x \,:\, x \in A\}$. By the scaling property $x \mapsto \lambda x$, $t \mapsto t$, u^λ solves (1.5) in Ω^λ.

7.1 Blowdown Lemma. *Suppose the flat metric on Ω satisfies*

$$\text{(7.1)} \qquad |g - \delta| = o(1), \qquad |\bar{\nabla} g| = o\left(\frac{1}{|x|}\right),$$

as $|x| \to \infty$. *Let* u *be a solution of* (1.5) *on* Ω *such that* $\{u = t\}$ *is compact for all sufficiently large* t. *Then for some constants* $c_\lambda \to \infty$,

$$u^\lambda - c_\lambda \to (n-1)\log|x|,$$

locally uniformly in $\mathbf{R}^n \setminus \{0\}$ *as* $\lambda \to 0$, *the standard expanding sphere solution in* $\mathbf{R}^n \setminus \{0\}$.

As a byproduct, we have the following.

7.2 Proposition. *The only solution of* (1.5) *on* $\mathbf{R}^n \setminus \{0\}$ *with compact level-sets is the expanding sphere solution given above.*

Proof of 7.1 and 7.2. 1. Fix t_0 so that $\{u = t\}$ is a compact subset of Ω for all $t \geq t_0$. The asymptotic condition (7.1) and the comments following Definition 3.3 imply that there is $R_0 > 0$ such that

$$\sigma(x) \geq c|x|, \qquad \mathrm{dist}(x, \partial E_{t_0}) \geq c|x|,$$

for $|x| \geq R_0$. By (3.1) with $r = \min(\sigma(x), \mathrm{dist}(x, \partial E_{t_0}))$, we observe

$$|\nabla u(x)| \leq \frac{C}{|x|}, \qquad \text{for all } |x| \geq R_0. \tag{7.2}$$

Next we want to control the eccentricity of N_t for large t. For any surface N, let $[r(N), R(N)]$ be the smallest interval such that N is contained in the annulus $\bar{D}_R \setminus D_r$, and define the eccentricity $\theta(N) := R(N)/r(N)$.

Now (7.1) implies that there exists $A > 0$ and t_1 such that the family $(D_{e^{At}})_{t_1 \leq t < \infty}$ is a subsolution of ($\star$), that is, the surfaces are moving faster than inverse mean curvature would dictate. Using these for comparison, we see that

$$R(N_{t+\tau}) \leq e^{A\tau} R(N_t), \qquad t \geq t_1, \quad \tau \geq 0. \tag{7.3}$$

Next suppose $r = r(N_t) \geq R_0$. Now $u = t$ somewhere on ∂D_r, and so by (7.2), there is C_2 such that $u > t - C_2$ everywhere on ∂D_r. Therefore N_{t-C_2} does not meet ∂D_r. By 4.2(i), N_{t-C_2} cannot have any components outside of D_r, so $R(N_{t-C_2}) \leq r$. Combining with (7.3),

$$R(N_t) \leq e^{AC_2} R(N_{t-C_2}) \leq e^{AC_2} r(N_t), \qquad t \geq t_2, \tag{7.4}$$

for some t_2.

2. Now let λ_i be any sequence converging to zero. The estimates (7.2), (7.3), and (7.4) are scale-invariant, so they are valid for $N_t^{\lambda_i}$, on the complement of a subset shrinking to $\{0\}$ as $i \to \infty$.

By (7.2), Compactness Theorem 2.1, and the Remarks following it, there exists a subsequence (λ_{i_j}), numbers $c_j \to \infty$, and a solution v of (1.5) in $R^n \setminus \{0\}$, such that

$$u^{\lambda_{i_j}} - c_j \to v \qquad \text{locally uniformly in } R^n \setminus \{0\},$$

with local C^1 convergence of the level sets. Define $P_t := \partial\{v < t\}$. By the eccentricity estimate (7.4), each nonempty P_t is a compact subset of $\mathbf{R}^n \setminus \{0\}$ with

$$\theta(P_t) \leq e^{AC_2}.$$

Now (7.3) implies in the limit that v is not constant, so some level-set is nonempty, say P_{t_3}. Then (7.3), together with the fact that N_t is nonempty and compact for all sufficiently large t, implies that P_t is nonempty and compact for $-\infty < t < \infty$.

Thus, v has compact level sets. According to Proposition 7.2, then, P_t is a family of expanding, round spheres and u has the form exhibited. Since this is true for a subsequence of any subsequence, it follows that the full sequence converges, proving Lemma 7.1.

3. It remains to prove Proposition 7.2. Let v be a solution of (1.5) in $\mathbf{R}^n \setminus \{0\}$ with compact level sets P_t. Using expanding spheres as barriers, we see that

$$\theta(P_t) \quad \text{is nonincreasing.}$$

Now, using the estimates above, we can blow up a subsequence $(P_t^{\lambda_i})_{-\infty<t<\infty}$, as $\lambda_i \to \infty$, to obtain a solution $(Q_t)_{-\infty<t<\infty}$ with nonempty, compact level sets and

$$\theta(Q_t) \equiv \theta_0, \qquad -\infty < t < \infty,$$

where $\theta_0 := \sup_t \theta(P_t)$. If $\theta_0 > 1$, then Q_0 lies between ∂D_r and $\partial D_{\theta_0 r}$ for some r, and is tangent to each of them, without being equal to either one. By perturbing ∂D_r outward and $\partial D_{\theta_0 r}$ inward, and applying the strong maximum principle to the smooth flows, and the weak maximum principle, Theorem 2.2(ii), to the nonsmooth flow Q_t, we find that $\theta(Q_t)$ decreases, a contradiction. This proves that $\theta_0 = 1$, which shows that P_t is a round sphere for all t, as claimed in the Proposition. $\square$

The ADM mass is defined in the following lemma. Here U denotes a precompact open set with smooth boundary, ν is the outward unit normal of U with respect to g, and $d\mu$ the surface measure of ∂U with respect to g.

7.3 ADM Lemma. *Suppose $R \geq 0$ on M, and the asymptotic region Ω is embedded as the complement of a compact set in $\mathbf{R}^3$, with flat metric δ.*

(i) (see [ADM]) If δ satisfies

$$c\delta \leq g \leq C\delta \quad in\ \Omega, \qquad \int_\Omega |\bar{\nabla} g|^2 < \infty,$$

then the limit

$$m_{ADM}(g,\delta) := \lim_{U \to M} \frac{1}{16\pi} \int_{\partial U} g^{ij}(\bar{\nabla}_j g_{ik} - \bar{\nabla}_k g_{ij})\nu^k \, d\mu$$

exists, and is finite if and only if $\int_M R < \infty$. Here $\chi_U \to \chi_M$ locally uniformly.

(ii) (Bartnik [Ba2], Chruściel [Ch3]) If δ satisfies

$$|g - \delta| \leq C|x|^{-1/2-\alpha}, \qquad |\bar{\nabla} g| \leq C|x|^{-3/2-\alpha}, \qquad x \in \Omega,$$

for some $\alpha > 0$, then m_{ADM} is a geometric invariant of g, independent of the choice of δ.

Proof. (i) We observe that the vector field

$$Y^k := g^{kl}g^{ij}(\bar{\nabla}_i g_{jl} - \bar{\nabla}_l g_{ij}) = g^{ij}\Gamma^k_{ij} - \Gamma^j_{jl}g^{lk}$$

satisfies

$$\operatorname{div} Y = R \pm C|\Gamma|^2,$$

where the divergence is taken with respect to g. Integrating by parts yields the result.

(ii) See [Ba2,Ch3]. □

In order to effect our mass comparison at infinity, we recall the conditions from the introduction, namely

$$\text{(0.1)-(0.2)} \qquad |g - \delta| \leq \frac{C}{|x|}, \qquad |\bar{\nabla} g| \leq \frac{C}{|x|^2}, \qquad Rc \geq -\frac{Cg}{|x|^2}.$$

Note that this implies directly (without integration by parts as needed with weaker conditions) that the ADM flux integral is finite.

7.4 Asymptotic Comparison Lemma. *Assume that the asymptotic region of M satisfies* (0.1)*, and let $(E_t)_{t\geq t_0}$ be a family of precompact sets weakly solving* ($\star$) *in M. Then*

$$\lim_{t\to\infty} m_H(N_t) \leq m_{ADM}(M).$$

The proof consists of a straightforward linearization of $\int H^2$ in terms of the perturbation from flatness, together with the integration by parts formula (7.15), which is possibly related to the "curious cancellation" in [Ba1,Ch]. We use (7.2) and (5.22) to show that m_H is bounded, which gives us some analytic control of the sequence via Geroch Monotonicity, which allows us to refine the estimates. We ultimately obtain convergence of $|A|^2$ in L^2.

Proof. 1. Define $r = r(t)$ by $|N_t| = 4\pi r^2$. Then $|N_t^{1/r}|_{g^{1/r}} = 4\pi$, so Lemma 7.1 implies that

$$N_t^{1/r(t)} \to \partial D_1 \quad \text{in } C^1 \text{ as } t\to\infty. \tag{7.5}$$

Let h be the restriction of g to the moving surface and let ε be the restriction of the flat metric δ to it. Let ν be the exterior unit normal, ω the unit dual normal, A the second fundamental form, H the mean curvature, all with respect to g. Define $\bar\nu$, $\bar\omega$, $\bar A$, $\bar H$ correspondingly, with respect to δ.

We begin with a series of approximations. Write $p_{ij} = g_{ij} - \delta_{ij}$. Let us restrict our attention to t sufficiently large that $|p| \leq 1/10$ on N_t. Then we have

$$h^{ij} - \varepsilon^{ij} = -h^{ik}p_{kl}h^{lj} \pm C|p|^2, \qquad g^{ij} - \delta^{ij} = -g^{ik}p_{kl}g^{lj} \pm C|p|^2, \tag{7.6}$$

for the inverse matrices. The bars indicate the norm with respect either to g or to δ – it does not matter which. Note that

$$\omega = \frac{\bar\omega}{|\bar\omega|_g}, \qquad \nu^i = g^{ij}\omega_j.$$

Then we have

$$\bar\omega_i = \omega_i \pm C|p|, \qquad \bar\nu^i = \nu^i \pm C|p|, \qquad 1 - |\bar\omega|_g = \frac{1}{2}\nu^i\nu^j p_{ij} \pm C|p|^2, \tag{7.7}$$

$$d\mu - d\bar\mu = \left(\frac{1}{2}h^{ij}p_{ij} \pm C|p|^2\right) d\mu, \tag{7.8}$$

$$\Gamma^k_{ij} = \frac{1}{2}g^{kl}(\nabla_i p_{jl} + \nabla_j p_{il} - \nabla_l p_{ij}) \pm C|p||\nabla p|, \qquad \nabla p = \bar\nabla p \pm C|p||\nabla p|, \tag{7.9}$$

where Γ_{ij}^k are the Christoffel symbols $\nabla - \bar{\nabla}$. We have the formula

$$|\bar{\omega}|_g A_{ij} = \bar{A}_{ij} - \bar{\omega}_k \Gamma_{ij}^k. \tag{7.10}$$

Compute

$$H - \bar{H} = h^{ij} A_{ij} - \varepsilon^{ij} \bar{A}_{ij} = (h^{ij} - \varepsilon^{ij}) A_{ij} + \varepsilon^{ij} A_{ij} (1 - |\bar{\omega}|_g) + \varepsilon^{ij} (|\bar{\omega}|_g A_{ij} - \bar{A}_{ij}).$$

For the second term we have, using (7.7) and (7.6),

$$\varepsilon^{ij} A_{ij} (1 - |\bar{\omega}|_g) = \frac{1}{2} H \nu^i \nu^j p_{ij} \pm C|p|^2 |A|,$$

and for the third, using (7.10), (7.6), (7.7), and (7.9),

$$\begin{aligned}
\varepsilon^{ij} (|\bar{\omega}|_g A_{ij} - \bar{A}_{ij}) &= -\varepsilon^{ij} \bar{\omega}_k \Gamma_{ij}^k \\
&= -\frac{1}{2} h^{ij} \omega_k g^{kl} (\nabla_i p_{jl} + \nabla_j p_{il} - \nabla_l p_{ij}) \pm C|p||\nabla p| \\
&= -h^{ij} \nu^l \nabla_i p_{jl} + \frac{1}{2} h^{ij} \nu^l \nabla_l p_{ij} \pm C|p||\nabla p|.
\end{aligned}$$

Plugging these in and using (7.6), we get

$$\begin{aligned}
H - \bar{H} = -h^{ik} p_{kl} h^{lj} A_{ij} + \frac{1}{2} H \nu^i \nu^j p_{ij} - h^{ij} \nu^l \nabla_i p_{jl} + \frac{1}{2} h^{ij} \nu^l \nabla_l p_{ij} \\
\pm C|p||\nabla p| \pm C|p|^2 |A|,
\end{aligned}$$

and so

$$|H - \bar{H}| \leq C|p||A| + C|\nabla p|, \qquad |H^2 - \bar{H}^2| \leq C|p||A|^2 + C|\nabla p|^2 + C|\nabla p||A|,$$

and using this together with (7.8),

$$\bar{H}^2 (d\mu - d\bar{\mu}) = \left(\frac{1}{2} H^2 h^{ij} p_{ij} \pm C|p|^2 |A|^2 \pm C|\nabla p|^2 \right) d\mu.$$

We are now in a position to estimate $\int H^2$. Our first task is to isolate the leading term 16π from the correction of order $1/r$ which gives the mass. We have from [Wi] that in $\mathbf{R}^3$,

$$\int_N \bar{H}^2 \, d\bar{\mu} \geq 16\pi.$$

From this inequality and the previous three, we find for sufficiently large t,

$$\begin{aligned}\int_{N_t} H^2\,d\mu &= \int_{N_t} \bar{H}^2 d\bar{\mu} + \bar{H}^2(d\mu - d\bar{\mu}) + 2H(H-\bar{H}) - (H-\bar{H})^2\,d\mu \\ &\geq 16\pi + \int_{N_t} \frac{1}{2}H^2 h^{ij}p_{ij} - 2Hh^{ik}p_{kl}h^{lj}A_{ij} + H^2\nu^i\nu^j p_{ij} \\ &\qquad - 2Hh^{ij}\nu^l\nabla_i p_{jl} + Hh^{ij}\nu^l\nabla_l p_{ij} - C|p|^2|A|^2 - C|\nabla p|^2.\end{aligned} \tag{7.11}$$

2. Next we show that $m_H(N_t)$ remains bounded. As remarked in the proof of Lemma 7.1, (0.1) implies that

$$|H| = |\nabla u| \leq \frac{C}{|x|} \leq \frac{C}{r} \qquad \text{on } N_t, \tag{7.12}$$

for a.e. sufficiently large t, where we have recalled (1.12), and used (7.5) to relate $|x|$ to r. Using this with (5.22),

$$\begin{aligned}\int_s^{s+1}\int_{N_t} |A|^2 &= \int_s^{s+1}\int_{N_t} \frac{1}{2}(\lambda_1-\lambda_2)^2 + \frac{1}{2}H^2 \\ &\leq \int_{N_s} H^2 + \int_s^{s+1} 4\pi\chi(N_t) \\ &\leq 4\pi C + 8\pi,\end{aligned}$$

so we may select a subsequence $t_i \to \infty$ such that

$$\sup_i \int_{N_{t_i}} |A|^2 < \infty.$$

Then we estimate from (7.11), (0.1), and (7.12),

$$\int_{N_{t_i}} H^2 \geq 16\pi - C\int_{N_{t_i}} |p||A|^2 + |\nabla p||H| + |\nabla p|^2 \geq 16\pi - \frac{C}{r}$$

so by the definition and monotonicity of $m_H(N_t)$,

$$\sup_t m_H(N_t) < \infty.$$

3. In particular, this implies by the monotonicity formula (5.24),

$$\int_{t_0}^{\infty} e^{t/2}\int_{N_t} (\lambda_1-\lambda_2)^2 + \frac{|DH|^2}{H^2} < \infty,$$

so we can pick a new subsequence $t_\ell \to \infty$ such that, writing $N_\ell := N_{t_\ell}$,

$$\int_{N_\ell} (\lambda_1 - \lambda_2)^2 + \frac{|DH|^2}{H^2} \to 0,$$

and (by scaling) a corresponding result for N_ℓ^{1/r_ℓ}. Since $H \le C/r$, we can use Rellich's theorem as in the proof of 5.12 to show that

$$H_{N_\ell^{1/r_\ell}} \to H_{\partial D_1} = 2 \qquad \text{in } L^2(\partial D_1),$$

where N_ℓ^{1/r_ℓ} is written as a graph over ∂D_1 for large i. Rescaling, this implies

$$H_{N_\ell} = \frac{2}{r} + f_\ell \qquad \text{on } N_\ell, \tag{7.13}$$

where $\int_{N_\ell} f_\ell^2 \to 0$. Also,

$$\int_{N_\ell} \left| A - \frac{H}{2} h \right|^2 = \frac{1}{2} \int_{N_\ell} (\lambda_1 - \lambda_2)^2 \to 0,$$

so

$$A_{N_\ell} = \frac{h}{r} + g_\ell \qquad \text{on } N_\ell, \tag{7.14}$$

where $\int_{N_\ell} g_\ell^2 \to 0$. In particular,

$$\sup_\ell \int_{N_\ell} |A|^2 < \infty.$$

4. Thus we may estimate from (7.11), (7.13) and (7.14),

$$\begin{aligned}
32\pi m_H(N_\ell) &= r_\ell (16\pi - \int_{N_\ell} H^2) \\
&\le \frac{C}{r_\ell} + r_\ell \int_{N_\ell} -\frac{1}{2} H^2 h^{ij} p_{ij} + 2H h^{ik} p_{kl} h^{lj} A_{ij} - H^2 \nu^i \nu^j p_{ij} \\
&\qquad\qquad + 2H h^{ij} \nu^l \nabla_i p_{jl} - H h^{ij} \nu^l \nabla_l p_{ij}, \\
&\le \frac{C}{r_\ell} + \eta_\ell + \int_{N_\ell} -\frac{2}{r_\ell} h^{ij} p_{ij} + \frac{4}{r_\ell} h^{ij} p_{ij} - \frac{4}{r_\ell} \nu^i \nu^j p_{ij} \\
&\qquad\qquad + 4 h^{ij} \nu^l \nabla_i p_{jl} - 2 h^{ij} \nu^l \nabla_l p_{ij},
\end{aligned}$$

where η_ℓ is an error term of the form

$$\eta_\ell = C \int_{N_\ell} |f_\ell||p| + |g_\ell||p| + r_\ell |f_\ell||\nabla p|.$$

By (0.1), we see that $\eta_\ell \to 0$. We observe by the integration by parts formula (5.17), that

$$\int_{N_\ell} h^{ij}\nu^l\nabla_i p_{jl} = \int_{N_\ell} H\nu^j\nu^l p_{jl} - h^{ij}p_{jk}h^{kl}A_{li} = \int_{N_\ell} \frac{2}{r_\ell}\nu^j\nu^l p_{jl} - h^{ij}p_{ij} + \eta'_\ell, \tag{7.15}$$

where again, $\eta'_\ell \to 0$ by (0.1). Applying this to the previous equation we get

$$32\pi m_H(N_\ell) \le \frac{C}{r_\ell} + \eta_\ell + 2\eta'_\ell + \int_{N_\ell} 2h^{ij}\nu^l\nabla_i p_{jl} - 2h^{ij}\nu^l\nabla_l p_{ij}.$$

Using (7.9), this differs from the definition of the ADM mass in Lemma 7.3 by at most

$$\eta''_\ell = C\int_{N_\ell} |p||\nabla p|,$$

which again converges to zero by (0.1). Applying the monotonicity formula (5.24), this yields

$$\sup_{t\ge 0} m_H(N_t) \le m_{ADM}(M).$$

□

§8 Proof of Main Theorem

In this section we prove the rigidity claim, and assemble the pieces to prove the Main Theorem. At the end of the section, we give an example relating to multiple black holes and the nonlocality of mass.

The following lemma is useful for the no-boundary case.

8.1 Lemma. *Let M be a complete, asymptotically flat 3-manifold with no boundary. For any $x \in M$ there is a locally Lipschitz solution u of* (1.5) *on $M \setminus \{x\}$ with $u \to -\infty$ near x and $u \to \infty$ at infinity. Furthermore, $m_H(N_t) \ge 0$ for all t.*

Proof. By Theorem 3.1 and Remark 1 following it, for each $\varepsilon > 0$, there exists u^ε solving (††) with the initial condition $B_\varepsilon(x)$. We may use the eccentricity estimate and other techniques of Lemma 7.1, to show that there exists a subsequence $\varepsilon_i \to 0$, a sequence $c_i \to \infty$, and a function u defined on $M \setminus \{x\}$ such that $u^{\varepsilon_i} - c_i \to u$ locally uniformly,

$$|\nabla u(y)| \le \frac{C}{\mathrm{dist}(y,x)} \qquad \text{in } B_1(x),$$

and N_t is nonempty and compact for all t, with N_t nearly equal to $\partial B_{e^{t/(n-1)}}$ as $t \to -\infty$.

Recalling Theorem 5.8, we have

$$m_H(N^{\varepsilon_i}_{t+c_i}) \geq m_H(\partial B_{\varepsilon_i}(x)), \qquad -c_i \leq t < \infty.$$

Passing $N^{\varepsilon_i}_{t+c_i} \to N_t$ for a.e. t, and recalling the upper semi-continuity of Hawking mass, we obtain $m_H(N_t) \geq 0$, as required. □

Proof of Main Theorem. 1. Let M be an exterior region satisfying (0.1). If M has no boundary, then by Lemma 7.4, the solution given in Lemma 8.1 proves $m_{ADM}(M) \geq 0$.

If M has a boundary, let N be any boundary sphere. By Theorem 6.1, there exists a flow $(N_t)_{t\geq 0}$ such that $N_0 = N$ and $m_H(N_t)$ is monotone. Together with Lemma 7.4 this proves that

$$m_{ADM}(M) \geq \sqrt{\frac{|N|}{16\pi}}$$

which is (0.4).

2. To complete the proof, we must consider the case of equality. We treat the case when ∂M is nonempty; the treatment of the no-boundary case is essentially identical. Then $m_H(N_t)$ equals a constant, m, for all $t \geq 0$. By the construction in Theorem 6.1, N_t solves the flow except for a finite number of times, remains connected, and stays in the regions where $R \geq 0$. So all the growth terms in (5.24) are nonnegative, and therefore zero, for a.e. t. In particular, recalling from Lemma 5.1 that $H > 0$ a.e. on N_t for a.e. t, we have $\int_{N_t} |DH|^2 = 0$ for a.e. t, and therefore by (1.10) and lower semicontinuity,

$$\int_{N_t} |DH|^2 = 0 \qquad \text{for all } t.$$

Therefore

$$H_{N_t}(x) = H(t), \qquad \text{a.e. } x \in N_t, \qquad \text{for all } t \geq 0,$$

that is, N_t has constant mean curvature. Since H is locally bounded on $M \setminus E_0$ and N_t has locally uniform C^1 estimates, it follows by elliptic theory that N_t is smooth for each t, with estimates that are locally uniform for $t \geq 0$. Similar considerations apply to N_t^+, $t \geq 0$.

If there is a jump at time t, either natural or constructed, then $N_t^+ \neq N_t$, and $H = 0$ on a portion of N_t^+, so N_t^+ is a minimal surface, disjoint from ∂M by the strong maximum principle. This contradicts the assumption that M contains no compact minimal surfaces in its interior. Therefore $N_t = N_t^+$ for all $t \geq 0$ (so in particular there is at most one boundary component!).

This shows $H > 0$ for $t > 0$, and a convergence argument shows that $H(t)$ is locally uniformly positive for $t > 0$.

By Lemma 2.4, for each $t > 0$ there is some maximal $T > t$ such that $(N_s)_{t \leq s < T}$ is a smooth evolution. By the regularity derived above, N_s has uniform space and time derivatives as $s \nearrow T$, so the evolution can be continued smoothly past $t = T$. This shows $T = \infty$ and the entire flow is smooth.

By the vanishing of the growth terms in (5.24), we have

$$\lambda_1 = \lambda_2 = \frac{H}{2}, \qquad R = 0,$$

on each N_t. Then (1.3) says

$$\frac{d}{dt} H^2 = -2|A|^2 - 2Rc(\nu, \nu),$$

showing $Rc(\nu, \nu)$ is constant on each N_t, and then

$$K = K_{12} + \lambda_1 \lambda_2 = -Rc(\nu, \nu) + \lambda_1 \lambda_2,$$

showing that K is constant on N_t, and N_t is isometric to a round sphere.

Since the normal speed, $1/H$, is constant on each N_t, the metric on M takes the warped product form

$$g = \frac{1}{H^2} dt^2 + h_{N_t}$$

where $t = u(x)$. We define $r = r(t)$ by imposing

$$h_{N_t} = r^2 \, dS$$

where dS is the standard metric on S^2. Then differentiating the relation

$$e^t |N_0| = |N_t| = 4\pi r^2$$

yields

$$dt = 2 \frac{dr}{r}.$$

Now

$$m = \frac{|N_t|^{1/2}}{(16\pi)^{3/2}} \left(16\pi - \int_{N_t} H^2 \right)$$

from which we obtain

$$H^2 = \frac{4}{r^2} - \frac{8m}{r^3}.$$

Substituting these relation above, we get

$$g = \frac{dr^2}{1 - 2m/r} + r^2 dS$$

which is isometric to $\mathbf{R}^3$ or (taking its exterior region) to one-half of the Schwarzschild manifold (see [HE, p.149]). $\square$

Nonlocal Mass Example. We are now in a position to give an example that sheds light on the nonlocality of mass distribution. In the artificial-jump construction of Theorem 6.1, it is tempting to include a contribution from the other boundary components as they are jumped over. However, the following example shows that not all of the extra mass is acquired at the moment of jumping.

Let M be half of a Schwarzschild manifold of mass $m > 0$ and boundary ∂F_1, and select a point p in M. Consider p as the site of an infinitely small black hole of mass zero. Let $(N_t)_{t>0}$ be the flow starting from p and jumping over ∂F_1 at $t = t_1$ according to the procedure of Theorem 6.1. We have

$$0 = m_H(N_0) < m_H(N_{t_1}) < m_H(N_{t_1}^+) < m_H(N_\infty) = m.$$

The middle inequality represents the jump. The first and third inequalities are strict because of the rigidity statement proven above, for otherwise M would be rotationally symmetric around p, which is impossible.

So, from the point of view of p, part of the mass is located in the "field" fore and aft of the black hole ∂F_1, whereas from the point of view of ∂F_1, all the mass is on the horizon.

This example may easily be modified so that the horizon at p has positive size (by summing two dissimilar Laplace kernels), and similar considerations apply. We would like to thank Professors Friedman, Geroch, and Wald for conversations about this example.

§9 Applications to Quasi-Local Mass

In many circumstances, the double integral on the right-hand side of the Geroch Monotonicity Formula (5.24) can be estimated concretely in terms of local information to yield an explicit lower bound for the ADM mass. In this section, we examine the implications of this idea for the quasi-local mass, or gravitational capacity, put forward by Bartnik. In particular, we show that the Bartnik capacity satisfies the positivity and exhaustion properties conjectured in [Ba3,Ba4,Ba5]. We end with some speculations about how the ADM mass might control the convergence of ADM-mass minimizing sequences of 3-manifolds.

Let us first give an analogy from electrostatics [Ba4, p.123-4]. If E is a bounded set in $\mathbf{R}^n$, $n \geq 3$, consider the class of potentials such that

$$-\Delta u \geq 0 \quad \text{in } \mathbf{R}^n, \qquad u \geq 1 \quad \text{in } E, \qquad u \to 0 \quad \text{as } |x| \to \infty, \tag{9.1}$$

where $\rho \equiv -\Delta u$ is the charge density. The total charge Q is given by

$$Q(u) = \int \rho = \lim_{r\to\infty} \int_{\partial B_r} -u_\nu,$$

analogous to the ADM mass. One definition of the *capacity* of E is the smallest total charge consistent with (9.1). It is attained by the solution of the exterior Dirichlet problem with $u = 1$ on ∂E.

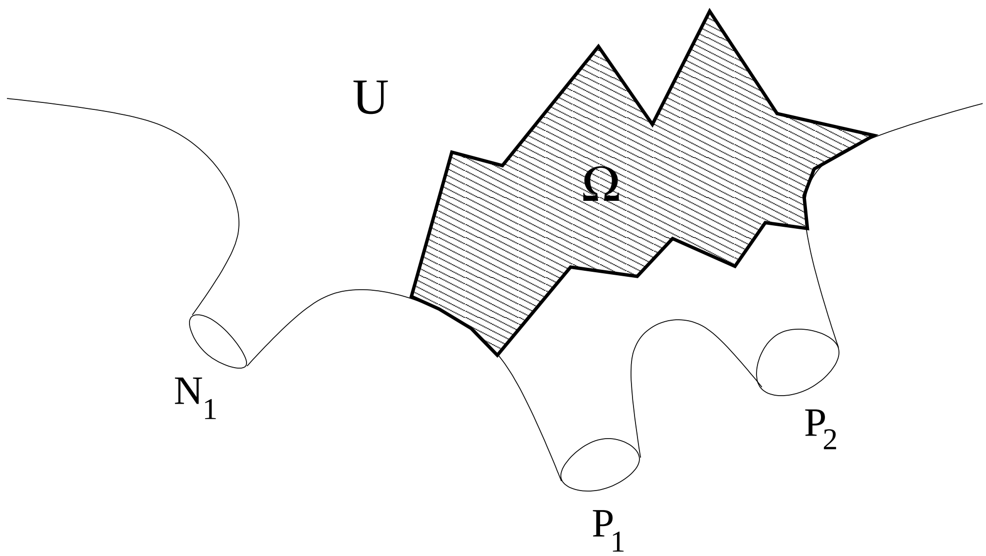
Let (Ω, g) be a Riemannian 3-manifold; we assume that Ω has no boundary and its metric closure is compact. We define an *admissible extension* of Ω to be an isometry between Ω and a bounded open set in an exterior region M as defined in the Main Theorem, to wit: a complete, connected, asymptotically flat 3-manifold with one end, satisfying (0.1)-(0.2), having $R \geq 0$, and possessing a minimal, compact boundary (possibly empty) but no other compact minimal surface. If Ω possesses an admissible extension, we call Ω *admissible.*

In analogy with the electrostatic capacity, we define the *Bartnik gravitational capacity* of Ω by

$$c_B(\Omega) := \inf\{m_{ADM}(M) \,:\, M \text{ is an admissible extension of } \Omega\}.$$

Note that Bartnik's original definition required M to be free of minimal surfaces (and hence diffeomorphic to $\mathbf{R}^3$), whereas our definition allows minimal surfaces on the boundary.

Thus Ω may surround one or more horizons, which contribute to the mass through their effect on the shape of Ω. But we require Ω to lie outside of these horizons; otherwise it would be unclear what part of Ω should be exposed and what part shielded. In addition, our definition permits extra horizons in the infinite component of $M \setminus \Omega$. (See figure.) Though that the ADM mass depends only on the infinite component, the finite components affect the admissibility of M. Our definition is a-priori less than or equal to Bartnik's, and conjecturally equal.



Admissible extension of Ω.

Note that c_B trivially has the monotonicity property [Ba3]

$$\tilde{\Omega} \subseteq \Omega \quad \text{implies} \quad c_B(\tilde{\Omega}) \leq c_B(\Omega).$$

Bartnik conjectured in [Ba3,Ba4,Ba5] that $c_B > 0$ except on subsets of Euclidean space. Now if Ω happens to be a minimizing hull in the extension M, then by the proof of Theorem 6.1 and Lemma 7.4, we have

$$m_{ADM}(M) \geq m_H(\partial\Omega).$$

If we only had to deal with extensions of this kind, then c_B would immediately acquire a positive lower bound, namely m_H. By taking some care about the minimizing hull property, we are able to prove the following.

9.2 Positivity Property. *Let Ω be admissible. Then $c_B(\Omega) > 0$ unless Ω is locally isometric to Euclidean $\mathbf{R}^3$.*

Careful examination of the following proof should give a small, but explicit lower bound for $c_B(\Omega)$ in terms of the scalar curvature in a small region (compare [Ba5, eqn. (50)]). It remains open whether Ω can be isometrically embedded in $\mathbf{R}^3$. (See the conjecture at the end of the section.)

Proof. Let x be a point in Ω where the metric is not flat. Let M_0 be an admissible extension of Ω, and let $(E_t)_{t\in\mathbf{R}}$ be the solution of inverse mean curvature flow in $M_0 \setminus \{x\}$ given by Lemma 8.1. Since Ω is not flat near x, the simple connectedness of $B_r(x)$, (5.24), and the argument used to prove the rigidity in §8 show that

$$m_H(\partial E_t) > 0$$

for any $t \in \mathbf{R}$. It will suffice to show that the initial part of the flow is valid up to a time t_0 that is independent of the extension.

Choose $r > 0$ such that $B_r(x)$ is a topological ball and $B_{3r}(x) \subseteq \Omega$. By the monotonicity formula for minimal surfaces [S, p.85], any properly embedded minimal surface in $B_{3r}(x) \setminus \bar{B}_r(x)$ that meets $\partial B_{2r}(x)$ has a certain minimum area c. Select t_0 so that $E_{t_0} \subseteq B_r(x)$ and $|\partial E_{t_0}| < c$. By Lemma 1.4(i), E_{t_0} is a minimizing hull in Ω. We will prove that that this remains true in any extension.

Let M be any other admissible extension of Ω, and let P be a surface that minimizes area among all surfaces in M that separate E_{t_0} from the asymptotically flat end of M. Then $P \setminus \bar{E}_{t_0}$ is a properly embedded minimal surface in $M \setminus \bar{E}_{t_0}$, and $|P| \le |\partial E_{t_0}| < c$, so P cannot meet $\partial B_{2r}(x)$. Since M contains no closed minimal surfaces besides ∂M, P must lie in $B_{2r}(x) \subseteq \Omega$. Since E_{t_0} is a minimizing hull in Ω, we have

$$|\partial E_{t_0}| \le |P|,$$

which shows that ∂E_{t_0} is a minimizing hull in M.

Therefore, by Lemma 7.4 and the argument of Theorem 6.1, there is a flow starting at E_{t_0} that proves

$$m_H(\partial E_{t_0}) \le m_{ADM}(M).$$

Since M was an arbitrary extension of Ω, this shows

$$0 < m_H(\partial E_{t_0}) \leq c_B(\Omega).$$

$\square$

9.3 Exhaustion Property. *Let M be asymptotically flat satisfying* (0.1). *If Ω_i is a sequence of bounded sets such that $\chi_{\Omega_i} \to \chi_M$ locally uniformly, then $c_B(\Omega_i) \to m_{ADM}(M)$.*

Proof. Select $R_i \to \infty$ such that $D_{R_i} \subseteq \Omega_i$, where D_R denotes the coordinate ball. By (0.1), $D_{R_i} \setminus D_{R_i/25}$ is C^1 close to flat for large i. It follows by an argument similar to the above that $D_{R_i/25}$ is a minimizing hull (in the sense employed above) in any extension $\tilde{M}$ of D_{R_i}. By the argument of Theorem 6.1, there exists a flow in $\tilde{M}$ that proves

$$m_H(D_{R_i/25}) \leq m_{ADM}(\tilde{M}).$$

Taking the infimum over $\tilde{M}$, this proves

$$m_H(D_{R_i/25}) \leq c_B(\Omega_i),$$

which does not exceed $m_{ADM}(M)$. A direct calculation similar to Lemma 7.4 (but easier) proves $m_H(D_{R_i/25}) \to m_{ADM}(M)$, yielding the Proposition. $\square$

Mass Minimization. Let Ω be a fragment of a Riemannian 3-manifold as above. Does there exist an admissible extension M of Ω that achieves the infimum $c_B(M)$, and what Euler-Lagrange equation does it solve? Our discussion is inspired by conversations with R. Bartnik, H. Bray, M. Gromov, and R. Ye.

Bartnik has made the following conjecture, which we have adapted slightly to our setting of manifolds with minimal boundary. Assume that Ω is connected with smooth boundary, possibly consisting of several components. Let U denote the component of $M \setminus \bar{\Omega}$ that contains infinity. According to our definition, ∂U is a disjoint union of $\partial U \cap \partial\Omega$ together with extra boundary horizons $\partial U \cap \partial M$.

Static Minimization Conjecture (Bartnik [Ba4, p.126]. *The infimum $c_B(\Omega)$ is realized by an admissible extension M of Ω such that U is matter-free, static, and contains* *no boundary horizons.*[6] *The metric is smooth on U and $C^{0,1}$ across ∂U, with nonnegative distributional scalar curvature on ∂U.*

Let us explain this heuristically; details can be found in [Ba3,Ba4,Ba5]. Consider (U, g) as time-symmetric initial data for the Einstein equation, and let L be a space-time development. *Matter-free* asserts that the 4-dimensional Ricci curvature of L vanishes; *static* asserts that L possesses a timelike Killing field orthogonal to a foliation by hypersurfaces. The idea is that once the extra mass-energy has been squeezed out, there is nothing left to support matter fields or gravitational dynamics. The fact that U has vanishing second fundamental form allows one to deduce that L is static and that U is one of the hypersurfaces in the foliation (see [Ba5]). Fortified by the Penrose Inequality, we expect that horizons in ∂U would contribute extra ADM mass and therefore do not arise; thus ∂U equals $\partial U \cap \partial\Omega$.

In Riemannian terms, the Euler-Lagrange equation then becomes [Ba4, p.115], [Ba5, (67)-(68)],

$$\begin{cases} R = 0, \qquad Rc = \dfrac{D^2\phi}{\phi} & \text{in } U, \\ H^+ = H^-, \qquad g^+ = g^- & \text{along } \partial U. \end{cases} \tag{9.2}$$

for some function ϕ on U. Here H^+, H^-, g^+, g^- denote the mean curvature and the metric of ∂U as induced from Ω and from U, respectively.

In the electrostatic analogy, a minimizing potential u of the Dirichlet integral on $\mathbf{R}^3 \backslash K$ is harmonic: a static, sourceless solution of the wave equation. Note that u, like g, is not C^1 across the boundary.

Let us now consider how minimizing sequences for $c_B(\Omega)$ might converge. To treat the simplest case, suppose first that M_i is a sequence of asymptotically flat 3-manifolds with nonnegative scalar curvature and

$$m_{ADM}(M_i) \to 0.$$

Based on the equality case of the Positive Mass Theorem, we expect that M_i converges to flat $\mathbf{R}^3$ in some sense.

[6]We have added this last proviso.

But how exactly? In a near-Euclidean Sobolev-space setting, Bartnik shows that the mass controls the metric in a $W^{1,2}$ sense [Ba5, Thm.4]. On the other hand, more globally, we must certainly exclude the region behind the horizons, whose total area is thankfully at most $16\pi m^2$ according to the Penrose inequality in the definitive form given by Bray [B2]. One might then expect that the resulting exterior regions converge in the Gromov-Hausdorff topology, but this is not quite so.

For example, even with no horizons, M_i can contain arbitrarily many long, thin cylindrical spikes that look like $S^2(\varepsilon)\times[0,L]$ capped off by a hemisphere $S^3_+(\varepsilon)$, without violating positive scalar curvature or contributing significantly to the ADM mass. Neither the number nor the volume of these spikes can be bounded. But inspired by the Penrose inequality, we speculate that the spikes are contained in a set whose total perimeter can be controlled by some multiple of the ADM mass. This leads to the following conjecture.

Conjecture. *Suppose M_i is a sequence of asymptotically flat 3-manifolds with ADM mass tending to zero. Then there is a set $Z_i \subseteq M_i$ such that $|\partial Z_i| \to 0$ and $M_i \setminus Z_i$ converges to $\mathbf{R}^3$ in the Gromov-Hausdorff topology.*

The process of resolving this conjecture is likely to produce analytic tools that can be applied to the more general question of the convergence of minimizing sequences in the case $c_B(\Omega) > 0$, and lead to a solution of Bartnik's Static Minimization Conjecture.

References

[A] F. Almgren, *Existence and Regularity Almost Everywhere of Solutions to Elliptic Variational Problems with Constraints,* Mem. Amer. Math. Soc. #165, Providence, 1976.

[ADM] R. Arnowitt, S. Deser and C. Misner, "Coordinate Invariance and Energy Expressions in General Relativity," *Phys. Rev.* **122** (1961) 997–1006.

[AM] A. Ashtekar and A. Magnon, "From i_0 to the $3+1$ Description of Spatial Infinity," *J. Math. Phys.* **25** (1984) 2682–2690.

[Ba1] R. Bartnik, "Existence of Maximal Surfaces in Asymptotically Flat Spacetimes," *Comm. Math. Phys.* **94** (1984) 155–175.

[Ba2] R. Bartnik, "The Mass of an Asymptotically Flat Manifold," *Comm. Pure Appl. Math.* **39** (1986) 661–693.

[Ba3] R. Bartnik, "New Definition of Quasi-Local Mass," *Phys. Rev. Lett.* **62** (1989) 2346.

[Ba4] R. Bartnik, "Results and Conjectures in Mathematical Relativity," *Miniconf. on Geom. and Phys. (Canberra, 1989)*, Proc. Centre Math. Anal., Austr. Nat. Univ., Canberra, 1989, 110–140.

[Ba5] R. Bartnik, "Energy in General Relativity," *Tsing Hua lectures on Geometry and Analysis (Hsinchu, 1990–1991)*, Internat. Press, Cambridge, MA, 1997, 5–27.

[Ba6] R. Bartnik, "Quasi-Spherical Metrics and Prescribed Scalar Curvature," *J. Diff. Geom.* **37** (1995) 31–71.

[BT] R. C. Bassanezi and I. Tamanini, "Subsolutions to the Least Area Problem and the 'Minimal Hull' of a Bounded Set in $\mathbf{R}^n$," *Ann. Univ. Ferrara Sez. VII (N.S.)* **30** (1984) 27–40.

[B1] H. Bray, "The Penrose Inequality in General Relativity and Volume Comparison Theorems Involving Scalar Curvature," thesis, Stanford University, 1997.

[B2] H. Bray, "Proof of the Riemannian Penrose Inequality using the Positive Mass Theorem," *J. Diff. Geom.*, to appear.

[BK] H. Brézis and D. Kinderlehrer, "The Smoothness of Solutions to Nonlinear Variational Inequalities," *Indiana Univ. Math. J.* **23** (1974) 831–844.

[CGG] Y.-G. Chen, Y. Giga, and S. Goto, "Uniqueness and Existence of Viscosity Solutions of Generalized Mean Curvature Flow Equations," *J. Diff. Geom.* **33** (1991) 749–786.

[C] Y. Choquet-Bruhat, "Positive-Energy Theorems," *Relativité, Groupes et Topologie II (Les Houches, 1983)*, ed. B. S. DeWitt and R. Stora, North Holland, Amsterdam, 1984, 739–785.

[CK] D. Christodoulou and S. Klainerman, "The Global Nonlinear Stability of Minkowski Space," *Princeton Math. Series #41*, Princ. Univ. Press, Princeton, 1993.

[CY] D. Christodoulou and S.-T. Yau, "Some Remarks on the Quasi-Local Mass," *Mathematics and General Relativity (Santa Cruz CA 1986),* ed. J. Isenberg, Contemporary Mathematics #71, AMS, 1988, 9–14.

[Ch1] P. T. Chruściel, thesis, Institute of Mathematics, Polish Academy of Sciences, 1985.

[Ch2] P. T. Chruściel, "Sur les Feuilletages 'Conformement Minimaux' des Varietés Riemanniennes de Dimension Trois," *C. R. Acad. Sci. Paris Ser. I Math.* **301** (1985) 609–612.

[Ch3] P. T. Chruściel, "Boundary Conditions at Spatial Infinity from a Hamiltonian Point of View," *Topological Properties and Global Structure of Space-Time (Erice, 1985)*, NATO Adv. Sci. Inst. Ser. B: Phys. #138, Plenum, New York, 1986, 49–59.

[Ch4] P. T. Chruściel, "On Kijowski's Unconstrained Degrees of Freedom of the Gravitational Field," *Proc. of the IV Marcel Grossman Meeting on General Relativity (Rome, 1985)*, North Holland, New York, 1986, 1643–1649.

[Ch5] P. T. Chruściel, "A Remark on the Positive Energy Theorem," *Class. Quan. Grav.* **3** (1986) L115–L121.

[Ch6] P. T. Chruściel, "Sur les Coordonnées p-Harmoniques en Relativité Générale," *C. R. Acad. Sci. Paris Ser. I Math.* **305** (1987) 797–800.

[CIL] M. G. Crandall, H. Ishii, and P.-L. Lions, "User's Guide to Viscosity Solutions of Second Order Partial Differential Equations," *Bull. of the Amer. Math. Soc.* **27** (1992) 1–67.

[E] M. Emmer, "Superfici di Curvatura Media Assegnata con Ostacolo," *Ann. Mat. Pura Appl.* **109** (1976) 371–389.

[ES] L. C. Evans and J. Spruck, "Motion of Level Sets by Mean Curvature I," *J. Diff. Geom.* **33** (1991) 635–681.

[F] H. Friedrich, "Hyperbolic Reductions for Einstein's Equations," *Class. Quant. Grav.* **13** (1996) 1451–69.

[Ga] G. J. Galloway, "On the Topology of Black Holes" *Comm. Math. Phys.* **151** (1993) 53–66.

[Ge] C. Gerhardt, "Flow of Nonconvex Hypersurfaces into Spheres," *J. Diff. Geom.* **32** (1990) 299–314.

[G] R. Geroch, "Energy Extraction," *Ann. New York Acad. Sci.* **224** (1973) 108–17.

[G1] G. W. Gibbons, thesis, Cambridge University, 1973.

[G2] G. W. Gibbons, "The Time Symmetric Initial Value Problem for Black Holes," *Comm. Math. Phys.* **27** (1972) 87–102.

[G3] G. W. Gibbons, "The Isoperimetric and Bogomolny Inequalities for Black Holes," *Global Riemannian Geometry*, T. Willmore and N. Hitchin eds., Ellis Horwood, 1984.

[G4] G. W. Gibbons, "Collapsing Shells and the Isoperimetric Inequality for Black Holes," *Class. Quant. Grav.* **14** (1997) 2905–2915.

[GT] D. Gilbarg and N. Trudinger, *Elliptic Partial Differential Equations of Second Order,* 2nd ed., Springer, New York, 1983.

[GL1] M. Gromov and H. B. Lawson, "Spin and Scalar Curvature in the Presence of a Fundamental Group I," *Ann. of Math.* **111** (1980) 209–230.

[GL2] M. Gromov and H. B. Lawson, "The Classification of Simply Connected Manifolds of Positive Scalar Curvature," *Annals of Math.* **111** (1980) 423–434.

[GL3] M. Gromov and H. B. Lawson, "Positive Scalar Curvature and the Dirac Operator on Complete Riemannian Manifolds," *Inst. Hautes Études Sci. Publ. Math. #58,* 1983, 83–196.

[HZ] R. Hardt and X. Zhou, "An Evolution Problem for Linear Growth Functionals," *Comm. Part. Diff. Eq.* **19** (1994) 1879–1907.

[H1] S. W. Hawking, "Gravitational Radiation from Colliding Black Holes," *Phys. Rev.*

Lett. **26** (1971) 1344.

[H2] S. W. Hawking, "Gravitational Radiation in an Expanding Universe," *J. Math. Phys.* **9** (1968) 598–604.

[H3] S. W. Hawking, "Black Holes in General Relativity," *Comm. Math. Phys.* **25** (1972) 152–166.

[HE] S. W. Hawking and G.F.R. Ellis, *The Large-Scale Structure of Space-Time,* Cambridge Monographs on Mathematical Physics, Cambridge, 1973.

[HP] S. W. Hawking and R. Penrose, "The Singularities of Gravitational Collapse and Cosmology," *Proc. Roy. Soc. A* **314** (1970) 529–548.

[Ha] S. A. Hayward, "General Laws of Black-Hole Dynamics," *Phys. Rev. D* **49** (1994) 6467-6474.

[Hd] M. Heidusch, "Higher Regularity of Inverse Mean Curvature Flow," preprint, Tübingen, 2001.

[Hm] J. Hempel, *3-Manifolds,* Princ. Univ. Press, Princeton, 1976.

[He] M. Herzlich, "A Penrose-like Inequality for the Mass of Riemannian Asymptotically Flat Manifolds," *Comm. Math. Phys.* **188** (1997) 121–133.

[Ho] M. Holder, "Contracting Spacelike Hypersurfaces by their Inverse Mean Curvature," *J. Austral. Math. Soc. Ser. A* **68** (2000) 285–300.

[Hu] G. Huisken, "Contracting Convex Hypersurfaces in Riemannian Manifolds by their Mean Curvature," *Invent. Math.* **84** (1986) 463–480.

[HI1] G. Huisken and T. Ilmanen, "The Riemannian Penrose Inequality," *Int. Math. Res. Not.* **20** (1997) 1045–1058.

[HI2] G. Huisken and T. Ilmanen, "A Note on Inverse Mean Curvature Flow," *Proc. of the Workshop on Nonlinear Part. Diff. Equ. (Saitama Univ., Sept. 1997)*, Saitama Univ., 1997.

[HY] G. Huisken and S.-T. Yau, "Definition of Center of Mass for Isolated Physical Systems and Unique Foliations by Stable Spheres with Constant Mean Curvature," *Invent. Math.* **124** (1996) 281–311.

[Hut] J. E. Hutchinson, "Second Fundamental Form for Varifolds and Existence of Surfaces Minimizing Curvature," *Indiana J. Math.* **35** (1986) 45–71.

[I1] T. Ilmanen, *Elliptic Regularization and Partial Regularity for Motion by Mean Curvature,* Mem. Amer. Math. Soc. #520, Providence, 1994.

[I2] T. Ilmanen, "Lectures on Mean Curvature Flow and Related Equations," lecture notes, ICTP, Trieste, 1995.

[J1] P. S. Jang, "On the Positive Energy Conjecture," *J. Math. Phys.* **17** (1976) 141–145.

[J2] P. S. Jang, "On the Positivity of Energy in General Relativity," *J. Math. Phys.* **19** (1978) 1152–1155.

[J3] P. S. Jang, "Note on Cosmic Censorship," *Phys. Rev. D* **20** (1979) 834–838.

[J4] P. S. Jang, "On the Positivity of the Mass for Black Hole Space-Times," *Comm. Math. Phys.* **69** (1979) 257–266.

[JW] P. S. Jang and R. M. Wald, "The Positive Energy Conjecture and the Cosmic Censor Hypothesis," *J. Math. Phys.* **18** (1977) 41–44.

[Je1] J. Jezierski, "Positivity of Mass for Certain Spacetimes with Horizons," *Class. Quant.*

Grav. **6** (1989) 1535–1539.

[Je2] J. Jezierski, "Stability of Reissner–Nordström Solution with Respect to Small Perturbations of Initial Data," *Class. Quant. Grav.* **11** (1994) 1055–1068.

[Je3] J. Jezierski, "Perturbation of Initial Data for Spherically Symmetric Charged Black Hole and Penrose Conjecture," *Acta Phys. Pol. B* **25** (1994) 1413–17.

[JK1] J. Jezierski and J. Kijowski, "A Simple Argument for Positivity of Gravitational Energy," *Proceedings of the XV International Conference on Differential Geometric Methods in Theoretical Physics (Clausthal, 1986)*, World Sci. Pub., Teaneck NJ, 1987, 187–194.

[JK2] J. Jezierski and J. Kijowski, "Positivity of Total Energy in General Relativity," *Phys. Rev. D* **36** (1987) 1041–1044.

[K1] J. Kijowski, "Unconstrained Degrees of Freedom of Gravitational Field and the Positivity of Gravitational Energy," *Gravitation, Geometry and Relativistic Physics (Aussois, 1984)*, Lect. Notes in Phys. #212, Springer, Berlin, 1984, 40–50.

[K2] J. Kijowski, "On Positivity of Gravitational Energy," *Proc. of the IV Marcel Grossman Meeting on General Relativity (Rome, 1985)*, ed. R. Ruffini, North-Holland, Amsterdam, 1986, 1681–1686.

[LP] J. M. Lee and T. H. Parker, "The Yamabe Problem," *Bull. Amer. Math. Soc.* **17** (1987) 37–91.

[Lc] A. Lichnerowicz, "Spineures Harmoniques," *C.R. Acad. Sci. Paris Ser. A-B* **257** (1963) 7–9.

[LT] A. Lichnewsky and R. Temam, "Pseudosolutions of the Time-Dependent Minimal Surface Problem," *J. Diff. Eq.* **30** (1978) 340–364.

[Ln] F.-H. Lin, "Regularity for a Class of Parametric Obstacle Problems," thesis, Univ. Minnesota, 1985.

[Lo] J. Lohkamp, "Scalar Curvature and Hammocks," *Math. Ann.* **313** (1999) 385–407.

[Lu] S. Luckhaus, "Solutions for the Two-Phase Stefan Problem with the Gibbs-Thompson Law for the Melting Temperature," *Eur. J. Appl. Math.* **1** (1990) 101–111.

[Ma] U. Massari, "Esistenza e Regolarità delle Ipersuperfici di Curvatura Media Assegnata in $\mathbf{R}^n$," *Arch. Rat. Mech. & Anal.* **55** (1974) 357–382.

[MSY] W. H. Meeks III, L. M. Simon, and S.-T. Yau, "Embedded Minimal Surfaces, Exotic Spheres, and Manifolds with Positive Ricci Curvature," *Ann. of Math.* **116** (1982) 621–659.

[MY] W. H. Meeks III and S.-T. Yau, "Topology of Three Dimensional Manifolds and the Embedding Problems in Minimal Surface Theory," *Ann. of Math.* **112** (1980) 441–484.

[Mi] M. Miranda, "Frontiere Minimali con Ostacoli," *Ann. Univ. Ferrara Sez. VII (N.S.)* **16** (1971) 29–37.

[PT] T. H. Parker and C. H. Taubes, "On Witten's Proof of the Positive Energy Theorem," *Comm. Math. Phys.* **84** (1982) 223–238.

[Pa] E. Pasch, "Numerische Verfahren zur Berechnung von Krümmungsflüssen," Ph.D. thesis, Tübingen Univ., 1998.

[P1] R. Penrose, "Gravitational Collapse: The Role of General Relativity," *Rivista del Nuovo Cimento Numero Speciale* **1** (1969) 252–276.

[P2] R. Penrose, "Naked Singularities," *Ann. New York Acad. Sci.* **224** (1973) 125–134.

[P3] R. Penrose, "Some Unsolved Problems in Classical General Relativity," *Seminar on Differential Geometry*, ed. S.-T. Yau, Annals of Math. Stud. #102, Princ. Univ. Press, Princeton, 1982, 631–668.

[PSW] R. Penrose, R.D. Sorkin, and E. Woolgar, "A Positive Mass Theorem based on Focussing and Retardation of Null Geodesics," submitted to *Comm. Math. Phys.*

[R] O. Reula, "Existence Theorems for Solutions of Witten's Equations and Nonnegativity of Total Mass," *J. Math. Phys.* **23** (1982) 810–814.

[RT] O. Reula and K. P. Tod, "Positivity of the Bondi Energy," *J. Math. Phys.* **25** (1984) 1004–1008.

[SSY] R. Schoen, L. M. Simon, and S.-T. Yau, "Curvature Estimates for Minimal Hypersurfaces," *Acta Math.* **134** (1975) 275–288.

[SY1] R. Schoen and S.-T. Yau, "Incompressible Minimal Surfaces, Three-Dimensional Manifolds with Nonnegative Scalar Curvature, and the Positive Mass Conjecture in General Relativity," *Proc. Nat. Acad. Sci.* **75** (1978) 45–76.

[SY2] R. Schoen and S.-T. Yau, "Complete Manifolds with Nonnegative Scalar Curvature and the Positive Action Conjecture in General Relativity," *Proc. Nat. Acad. Sci.* **75** (1978) 45–76.

[SY3] R. Schoen and S.-T. Yau, "On the Proof of the Positive Mass Conjecture in General Relativity," *Comm. Math. Phys.* **65** (1979) 45–76.

[SY4] R. Schoen and S.-T. Yau, "Positivity of the Total Mass of a General Space-Time," *Phys. Rev. Lett.* **43** (1979) 1457–1459.

[SY5] R. Schoen and S.-T. Yau, "On the Structure of Manifolds with Positive Scalar Curvature," *Manuscripta Math.* **28** (1979) 159–183.

[SY6] R. Schoen and S.-T. Yau, "Existence of Incompressible Minimal Surfaces and the Topology of Three Dimensional Manifolds with Non-Negative Scalar Curvature," *Ann. Math.* **110** (1979) 127–142.

[SY7] R. Schoen and S.-T. Yau, "The Energy and the Linear Momentum of Space-Times in General Relativity," *Comm. Math. Phys.* **79** (1981) 47–51.

[SY8] R. Schoen and S.-T. Yau, "Proof of the Positive Mass Theorem II," *Comm. Math. Phys.* **79** (1981) 231–260.

[SY9] R. Schoen and S.-T. Yau, "Proof that the Bondi Mass is Positive," *Phys. Rev. Lett.* **48** (1982) 369–371.

[SY10] R. Schoen and S.-T. Yau, "The Existence of a Black Hole due to Condensation of Matter," *Comm. Math. Phys.* **90** (1983) 575–579.

[SY11] R. Schoen and S.-T. Yau, "The Structure of Manifolds with Positive Scalar Curvature," *Directions in Partial Differential Equations (Madison, 1985)*, Publ. Math. Res. Center Univ. Wisconsin #54, Academic Press, Boston, 1987, 235–242.

[S] L. M. Simon, *Lectures on Geometric Measure Theory,* Proc. of the Centre for Math. Analysis #3, Austr. Nat. Univ., 1983.

[SWZ] P. Sternberg, G. Williams, and W. P. Ziemer, "$C^{1,1}$-Regularity of Constrained Area Minimizing Hypersurfaces," *J. Diff. Eq.* **94** (1991) 83–94.

[T1] I. Tamanini, "Boundaries of Caccioppoli Sets with Holder-Continuous Normal Vector," *J. Reine Angew. Math.* **334** (1982) 27–39.

[T2] I. Tamanini, "Regularity Results for Almost Minimal Oriented Hypersurfaces in $\mathbf{R}^n$," *Quaderni del Dipartimento di Matematica dell'Università di Lecce* **1** (1984) 1–92.

[To] K. P. Tod, "The Hoop Conjecture and the Gibbons-Penrose Construction of Trapped Surfaces," *Class. Quant. Grav.* **9** (1992) 1581–1591.

[Tr] N. S. Trudinger, "Isoperimetric Inequalities for Quermassintegrals," *Ann. Inst. H. Poincaré. Analyse Non-Linéare* **11** (1994) 411–425.

[U] J. Urbas, "On the Expansion of Starshaped Hypersurfaces by Symmetric Functions of Their Principal Curvatures," *Math. Z.* **205** (1990) 355–372.

[V] A. Visintin, "Nucleation and Mean Curvature Flow," *Comm. Part. Diff. Eq.* **23** (1998) 17–53.

[Wa1] R. M. Wald, *General Relativity*, Univ. Chicago Press, Chicago, 1984.

[Wa2] R. M. Wald, "Gravitational Collapse and Cosmic Censorship," *Black Holes, Gravitational Radiation and the Universe,* Fund. Theories Phys. #100, Kluwer Acad. Publ., Dordrecht, 1999, 69–85.

[Wh] B. White, "Partial Regularity of Mean Convex Hypersurfaces Flowing by Mean Curvature," *Int. Math. Res. Not.* **4** (1994) 185–192.

[Wi] T. Willmore, *Total Curvature in Riemannian Geometry,* Wiley, New York, 1982.

[W] E. Witten, "A New Proof of the Positive Energy Theorem," *Comm. Math. Phys.* **80** (1981) 381–402.

[Y] J. W. York, Jr., "Kinematics and Dynamics of General Relativity," *Sources of Gravitational Radiation,* ed. L. Smarr, Camb. Univ. Press, 1979, 83–126.

[Z] X. Zhou, "An Evolution Problem for Plastic Antiplanar Shear," *Appl. Math. Opt.* **25** 263–285.