SCALAR CURVATURE AND THE EXISTENCE OF GEOMETRIC STRUCTURES ON 3-MANIFOLDS, II

MICHAEL T. ANDERSON

Contents

Abstract. This paper studies the degeneration, i.e curvature blow-up, of sequences of metrics approaching the Sigma constant $\sigma(M)$ on 3-manifolds M with $\sigma(M) \leq 0$. The degeneration is related to the sphere decomposition of M in case M is σ-tame.

0. Introduction.

This paper is a continuation of [1], and is mainly concerned with the Sphere conjecture of [1,§0]. We recall the basic issues and set-up.

Let M be a closed oriented 3-manifold and $\sigma(M)$ the Sigma constant of M, i.e. the supremum of the scalar curvatures of unit volume Yamabe metrics on M. Throughout the paper, it is assumed that

$$\sigma(M) \leq 0; \tag{0.1}$$

this is the case if for instance M has a $K(\pi,1)$ factor in its sphere decomposition, c.f. [11]. Consider the L^2 norm of the negative part of the scalar curvature as a functional on the space of metrics on M, i.e.

$$\mathcal{S}^2_-(g) = \big(v^{1/3}\int_M (s^-)^2 dV_g\big)^{1/2}, \tag{0.2}$$

where $s^- = \min(s,0)$, $s = s_g$ is the scalar curvature of the metric g, v is the volume of (M,g). The volume power in (0.2) is chosen so that $\mathcal{S}^2_-$ is scale invariant. By [2, Prop.3.1],

$$\inf \mathcal{S}^2_- = |\sigma(M)|, \tag{0.3}$$

under the assumption (0.1), so that a minimizing sequence for $\mathcal{S}^2_-$ has similar characteristics to a maximizing sequence of Yamabe metrics on M.

Partially supported by NSF Grants DMS 9802722 and 0072591.

2000 *Mathematics Subject Classification.* Primary 58J60, 57M50; Secondary 58E11, 53C20.

Conjectures I and II of [1] concern the limiting behavior of certain minimizing sequences $\{g_i\}$ for $\mathcal{S}^2_-$ on irreducible 3-manifolds. These conjectures imply the geometrization conjecture for 3-manifolds satisfying (0.1). In [1, Thm.0.2], the validity of these two conjectures was proved for *tame* 3-manifolds, i.e. 3-manifolds M which admit a minimizing sequence of unit volume metrics $\{g_i\}$ for $\mathcal{S}^2_-$ such that, for some $K < \infty$,

$$\mathcal{Z}^2(g_i) = \int_M |z_{g_i}|^2 dV_{g_i} \leq K, \quad \text{as} \quad i \to \infty, \tag{0.4}$$

where $z = r - \frac{s}{3}g$ is the trace-free Ricci curvature. c.f. also [5] for an outline of the proof.

The assumption (0.4) is a rather strong topological assumption and it is not clear which 3-manifolds actually satisfy (0.4). For instance if N_i, $i = 1, 2$, are hyperbolic 3-manifolds then $M = N_1 \# N_2$ does not satisfy (0.4). The fundamental issue concerning Conjectures I and II above is in fact whether the converse holds, i.e. if M is irreducible, is then M necessarily tame.

In this paper, building on the work in [1], substantial progress is made on this problem. To explain this, let

$$\widetilde{\sigma}(g) = \mathcal{S}^2_-(g) - |\sigma(M)|, \tag{0.5}$$

so that $\widetilde{\sigma}(g) \geq 0$, and a sequence $\{g_i\}$ is minimizing for $\mathcal{S}^2_-$ if and only if $\widetilde{\sigma}(g_i) \to 0$.

Definition 0.1. *A 3-manifold M satisfying (0.1) is σ-*tame*, if there exist constants $k < \infty$ and $K < \infty$, and some unit volume minimizing sequence $\{g_i\}$ for $\mathcal{S}^2_-$, such that*

$$\widetilde{\sigma}(g_i)^k \cdot \mathcal{Z}^2(g_i) \leq K, \quad \text{as} \quad i \to \infty. \tag{0.6}$$

Thus, M is σ-tame if there exists a sequence of unit volume metrics $\{g_i\}$ with $\widetilde{\sigma}(g_i) \to 0$ such that the (trace-free) curvature of g_i blows up in L^2 at most as fast as some polynomial in $\widetilde{\sigma}^{-1}$. This is of course a topological condition on M, which is clearly much weaker than the tameness condition (0.4), (which corresponds to the case $k = 0$).

In fact we believe that *any* closed 3-manifold M with $\sigma(M) \leq 0$ is σ-tame, regardless of whether M is irreducible or reducible; this will be pursued in subsequent work since the methods dealing with this question are not related to the main issues here. In [2,3], numerous examples or, more precisely, models for potentially minimizing sequences $\{g_i\}$ for $\mathcal{S}^2_-$ were constructed. All of these models satisfy (0.6) with $k = 1$. Using these models it is not difficult to show that if Conjectures I and II are valid, then any 3-manifold M, $\sigma(M) \leq 0$, is σ-tame with $k = 1$. Conversely, a much simpler proof of the Main Theorem below can be given for σ-tame 3-manifolds with $k = 1$, (based on the work in [2, Thm.C]).

The purpose of this paper is to prove the following result:

Main Theorem. *If M is σ-tame, then Conjectures I and II of [1] are valid.*

As noted above, this implies the geometrization conjecture for σ-tame irreducible 3-manifolds M with $\sigma(M) \leq 0$, as well as the evaluation of their Sigma constant $\sigma(M)$. We refer to [1,5] for background discussion on these Conjectures. Exact statements of the Main Theorem in the cases $\sigma(M) < 0$ and $\sigma(M) = 0$ are given in Theorems 7.6 and 7.7 respectively.

The overall strategy of proof for the Main Theorem is briefly as follows; further details are given at the beginning of each section, c.f. also Remarks 1.3, 1.6 and 2.2. Basically, a modified version of the Sphere conjecture in [1] is proved for σ-tame 3-manifolds. Thus, suppose M is σ-tame but not tame, so that the curvature of any minimizing sequence $\{g_i\}$ for $\mathcal{S}^2_-$ blows up in L^2, i.e. (0.4) fails. By far the most crucial issue of the paper is to locate "natural" 2-spheres S^2 from the geometry of the sequence $\{g_i\}$, as $i \to \infty$. Under two relatively weak hypotheses, namely a degeneration hypothesis and a non-collapse hypothesis, it is shown that such 2-spheres arise as a certain locus where the curvature of $\{g_i\}$ blows-up, i.e. becomes very large, for $\widetilde{\sigma}(g_i)$ small. This core issue of the paper requires most of the technical work.

Given the existence of such 2-spheres, one then needs to prove that they are essential in M, i.e. do not bound 3-balls. This is done by means of the cut and paste or comparison methods already developed and used in [1, §2,§4] just for this purpose. The upshot is that M is then necessarily a reducible 3-manifold.

Recall that Conjectures I and II apply only to irreducible 3-manifolds. Hence, if M is irreducible, one of the two aforementioned hypotheses must fail. In the last section of the paper, §7, this is shown to imply that there exists a minimizing sequence which does not degenerate at all, so that M is in fact tame. This proves the Main Theorem, as discussed preceding (0.4). In sum, if M is σ-tame, but not tame, then M is necessarily reducible.

The crucial issue is thus to locate geometrically natural 2-spheres from the geometry of the degeneration of a suitable minimizing sequence $\{g_i\}$. The approach taken in the previous papers [1,2] was to find such spheres near regions where the curvature is blowing up at a *maximal* rate, since these are geometrically the most obvious places. Blow-up limits of the metric sequence based at such points are solutions of a certain elliptic system of equations, the $\mathcal{Z}_c^2$ equations, c.f. §1, studied in detail in [2]. The Sphere conjecture states that such limit solutions or metrics have an asymptotically flat end, c.f. also Remark 1.6; such an end thus carries a natural 2-sphere.

While we believe the Sphere conjecture above is true, in this paper a somewhat different approach is taken. Namely, 2-spheres in M are detected as a locus where the curvature of $\{g_i\}$ is blowing up, but blowing up at a, comparatively speaking, almost *minimal* rate; (similar issues regarding the rate of curvature blow-up play a central role in [3]). The metric blow-up limits based at such points are *flat*, and the near limiting metric behavior is governed by solutions of linearized equations closely analogous to the linearized $\mathcal{Z}_c^2$ equations. These linearized equations are derived in §2 and their implications for the near limiting geometry are analysed further in §3-§4. The main point, developed in §5, is the identification of 2-spheres in almost flat regions of rescalings (M, g_i') of (M, g_i), which enclose regions where the curvature is much larger. Such 2-spheres serve as a substitute for the spheres at infinity in the Sphere conjecture. Remark 3.6 explains briefly one important reason why this approach is advantageous to the analysis of general $\mathcal{Z}_c^2$ solutions.

Thus, the most significant part of the paper is the work in §5, c.f. in particular Theorems 5.4, 5.9 and 5.12, given the preparatory work in §2-§4 leading up to §5. More detailed descriptions of the contents and the strategy of the arguments are given at the beginning of each section as well as in the Remarks.

1. Background Setting and Results.

In this section, we set the stage and discuss background material needed for the work to follow; most of this is taken from [2,4]. Remarks 1.3 and 1.6 give some preliminary perspective on the overall strategy of the paper.

For a given $\varepsilon > 0$, consider the scale-invariant functional $I_\varepsilon^- = v^{1/3}\varepsilon\mathcal{Z}^2 + \mathcal{S}_-^2$ on the space of metrics on the closed, oriented 3-manifold M, $\sigma(M) \leq 0$, i.e.

$$I_\varepsilon^- = \varepsilon v^{1/3}\int_M |z|^2 dV + \big(v^{1/3}\int_M (s^-)^2 dV\big)^{1/2}. \tag{1.1}$$

Although seemingly complicated at first sight, it is explained in detail in [2, §1], (c.f. also [1]), why this is, in a natural sense, an optimal functional to consider from the point of view of geometrization of M via direct methods in the calculus of variations.

It is proved in [2,Thm.3.9] that, for any $\varepsilon > 0$, there exists a domain Ω_ε and a complete $C^{2,\alpha}\cap L^{3,p}$ smooth Riemannian metric g_ε on Ω_ε with $vol_{g_\varepsilon}\Omega_\varepsilon = 1$, which realizes the infimum of I_ε^-. The domain Ω_ε has a finite number $q = q(\varepsilon, M)$ of components and is weakly embedded in M in the sense that any domain with smooth compact closure in Ω_ε embeds in M. Further, there exists an exhaustion $\{K_j\}$ of Ω_ε by compact subsets such that $M \setminus K_j$ is a graph manifold. The pair $(\Omega_\varepsilon, g_\varepsilon)$

is called a *minimizing pair* for I_ε^-. In brief, $(\Omega_\varepsilon, g_\varepsilon)$ gives a geometric decomposition of M w.r.t. the functional I_ε^-.

As $\varepsilon \to 0$, the metrics g_ε form a minimizing sequence or family for $\mathcal{S}_-^2$, in that

$$\mathcal{S}_-^2(g_\varepsilon) \to |\sigma(M)| \quad \text{and} \quad \varepsilon \mathcal{Z}^2(g_\varepsilon) \to 0, \quad \text{as} \quad \varepsilon \to 0. \tag{1.2}$$

The essential issue is to understand the behavior of sequences of minimizing pairs $(\Omega_\varepsilon, g_\varepsilon)$, where $\varepsilon = \varepsilon_i$ is some sequence with $\varepsilon_i \to 0$.

The domain Ω_ε is empty if and only if M is a graph manifold which does not admit a flat metric; in fact, if Ω_ε is non-empty, then no component of Ω_ε is a graph manifold. Similarly, it is apriori possible that a sequence $(\Omega_{\varepsilon_i}, g_{\varepsilon_i})$ is constant, i.e. independent of ε_i, as $\varepsilon_i \to 0$. In this case of course (0.4) holds, so the Main Theorem is already proved; thus [1, Thm.0.2] implies that $(\Omega_\varepsilon, g_\varepsilon) = (H, g_o)$, where g_o is a constant curvature metric with scalar curvature $\sigma(M) < 0$, with ∂H incompressible in M, or $\Omega_\varepsilon = M$ is flat if $\sigma(M) = 0$. In these situations, the geometric decomposition of M w.r.t. I_ε^- above already gives the geometrization of M, for any $\varepsilon > 0$, c.f. [1, §2] for further details.

Thus, in effect, only the situation where $(\Omega_\varepsilon, g_\varepsilon)$ is not empty and not constant in ε needs to be considered. All of the work in §2-§7 is to prove that this situation can occur only if M is reducible. Hence, if M is irreducible, then the point is to prove that either Ω_ε is empty, (so M is a graph manifold), or $(\Omega_\varepsilon, g_\varepsilon)$ is constant in ε as above, c.f. Theorems 7.2, 7.6, and 7.7.

Abusing terminology slightly, a metric of constant negative curvature will be called hyperbolic, even if the sectional curvature is not -1. The topology of Ω_ε may, apriori, change with ε.

The metrics g_ε satisfy the system of Euler-Lagrange equations for I_ε^-, c.f. [2, Thm.3.3]:

$$L^* w = \varepsilon \nabla \mathcal{Z}^2 + \phi \cdot g, \tag{1.3}$$

$$2\Delta(w - \tfrac{\varepsilon s}{12}) + \tfrac{1}{4} s w = \tfrac{1}{2}\varepsilon |z|^2 - 3c. \tag{1.4}$$

Here $\nabla \mathcal{Z}^2$ is the gradient of $\mathcal{Z}^2 = \int |z|^2 dV$, c.f. (B.4), $\Delta = tr D^2$ where D^2 is the Hessian, and

$$L^* u = D^2 u - \Delta u \cdot g - u \cdot r, \tag{1.5}$$

where r is the Ricci curvature, (all w.r.t. $(\Omega_\varepsilon, g_\varepsilon)$). The subscript ε has been dropped from the notation in (1.3)-(1.4). The function ϕ is given by $\phi = -\frac{1}{4} s w + c$, and the constant c is given by

$$c = \frac{1}{12\sigma} \int (s^-)^2 + \frac{\varepsilon}{6} \int |z|^2, \tag{1.6}$$

where $\sigma = (v^{1/3} \int (s^-)^2)^{1/2}$. Observe that σ is scale invariant and converges to $|\sigma(M)| = -\sigma(M)$ as $\varepsilon \to 0$, by (1.2). Most importantly, the function w is given by

$$w = -\frac{s^-}{\sigma} \geq 0, \tag{1.7}$$

so that the L^2 norm of w over Ω_ε equals 1. The equation (1.4) is the trace of (1.3). The metric g_ε is $C^{2,\alpha} \cap L^{3,p}$, for any $p < \infty$, while the scalar curvature s is Lipschitz. In any region where $s \neq 0$, the metric is C^∞, in fact real-analytic. The function $-w$ was denoted by τ in [1].

Again the equations (1.3)-(1.4) are a complicated nonlinear system of elliptic PDE (in the metric g), but are explained and analysed in detail in [2, §1,3]. Essentially all of the work in this paper depends crucially on the exact form of the terms in these equations; the analysis developed is thus specific to these equations and may not be relevant to general variational problems on the space of metrics.

Let $T = T_\varepsilon = \sup_{\Omega_\varepsilon} w$. Then one has the estimate, (c.f. [2, (3.43)]),

$$1 \leq T \leq (1 + 2\frac{\varepsilon \mathcal{Z}^2}{\sigma})^{1/2}. \tag{1.8}$$

Thus, if $\sigma(M) < 0$, then

$$T_\varepsilon \to 1, \quad \text{as} \quad \varepsilon \to 0, \tag{1.9}$$

and so in particular

$$w_\varepsilon \to 1, a.e. \tag{1.10}$$

in the sense that for any $\mu > 0$, the measure of the set where $|1 - w_\varepsilon| \geq \mu$ converges to 0, as $\varepsilon \to 0$. Moreover, in general for $\sigma(M) \leq 0$, one has

$$\int_{\Omega_\varepsilon} |dw_\varepsilon|^2 dV_{g_\varepsilon} \to 0, \quad \text{as} \quad \varepsilon \to 0, \tag{1.11}$$

c.f. [3, (3.39),(3.46)]. However, if $\sigma(M) = 0$ and M is σ-tame, then (1.9) does not hold. Although it can be shown that T_ε remains bounded as $\varepsilon \to 0$, this will not be used here. Thus to unify these two situations, regardless of the behavior of T, the potential w will *always* be renormalized as

$$u = \frac{w}{T}, \tag{1.12}$$

so that, by fiat,

$$\sup_{\Omega_\varepsilon} u = 1. \tag{1.13}$$

The points $x_\varepsilon \in \Omega_\varepsilon$ where $u(x_\varepsilon) \to 1$ play the central role throughout the paper.

The equations (1.3)-(1.4) then read

$$L^* u = \bar{\varepsilon} \nabla \mathcal{Z}^2 + \bar{\phi} \cdot g, \tag{1.14}$$

$$2\Delta(u - \tfrac{\bar{\varepsilon}s}{12}) + \tfrac{1}{4} s u = \tfrac{1}{2}\bar{\varepsilon}|z|^2 - 3\bar{c}, \tag{1.15}$$

where $\bar{\varepsilon} = \varepsilon/T, \bar{c} = c/T, \bar{\phi} = \phi/T$. The equations (1.14)-(1.15) will be used throughout the paper.

The non-negative function u in (1.12) is of course just a renormalization of the non-positive part of the scalar curvature of g_ε. It is viewed as a potential function on $(\Omega_\varepsilon, g_\varepsilon)$, since it satisfies the potential type equation (1.15). The main point of view of the paper is to understand the degenerating geometry of $(\Omega_\varepsilon, g_\varepsilon)$ as $\varepsilon \to 0$ from analysis on the scalar curvature, or more precisely u, as a solution of the equations (1.14)-(1.15).

While all the terms in (1.14)-(1.15) are important, the two most important are the potential u and the constant term $\bar{c}$. Most of the paper is concerned with the distribution of the values of u over Ω_ε, and its relation to the behavior of the curvature blow-up of g_ε. It will be seen in §3, c.f. Theorem 3.5, that c in (1.4) or $\bar{c}$ in (1.15) gives some important global control on the value distribution of u. Observe that $\bar{c}$ is the *only* global term in the equations (1.14)-(1.15). Lemma 1.7 below relates the two terms of c in (1.4) to the condition that M is σ-tame.

Let $\rho(x) = \rho_\varepsilon(x)$ be the L^2 curvature radius of $(\Omega_\varepsilon, g_\varepsilon)$ at x; $\rho(x)$ is the largest radius such that, for all geodesic balls $B_y(s) \subset B_x(\rho(x)) \subset (\Omega_\varepsilon, g_\varepsilon)$,

$$\frac{s^4}{volB_y(s)} \int_{B_y(s)} |r|^2 \leq c_o, \tag{1.16}$$

where c_o is a fixed constant, c.f. [2,(2.1)], [4,§3]. The radius $\rho(x)$ measures the degree of curvature concentration near x; $\rho(x)$ is small if and only if the L^2 average of the curvature of g_ε is large near x. The function ρ is a Lipschitz function, with $|\nabla\rho| \leq 1$.

We recall the following result from [2,Rmk.7.3], (compare also with [3,Thm.3.3]), which will be used repeatedly throughout the paper.

Proposition 1.1. *There are constants $v_o > 0$ and $\rho_o > \infty$, independent of ε, such that if $x \in U^{v_o} = \{x \in \Omega_\varepsilon : u(x) \geq 1 - v_o\}$, then*

$$\rho(x) \geq \rho_o \cdot \min(t_{v_o}(x), 1), \tag{1.17}$$

where $t_{v_o}(x) = dist(x, L_{v_o})$, and $L_{v_o} = \{x \in \Omega_\varepsilon : u(x) = 1 - v_o\}$ is the $1 - v_o$ level of u. Further, analogous higher order estimates also hold, i.e, for any $k \geq 0$,

$$|\nabla^k r|(x) \leq c(k) \cdot \max(t_{v_o}(x)^{-2-k}, 1), \quad |\nabla^k du|(x) \leq c(k) \cdot \max(t_{v_o}(x)^{-1-k}, 1). \tag{1.18}$$

The constant 1 on the right in (1.17)-(1.18) arises from the scalar curvature of $(\Omega_\varepsilon, g_\varepsilon)$. The estimate (1.17) can be re-expressed scale-invariantly as

$$\rho(x) \geq \rho_o \cdot \min(t_{v_o}(x), \rho_s(x)), \tag{1.19}$$

where ρ_s is the L^2 scalar curvature radius, defined as in (1.16) with s in place of r. Similarly, (1.18) holds when 1 is replaced by the appropriate power of $\rho_s(x)$.

As throughout the previous papers, we use the L^2 Cheeger-Gromov theory, c.f. [7,8], [4, §3] to understand the limiting behavior of the metrics $(\Omega_\varepsilon, g_\varepsilon)$ as $\varepsilon \to 0$. Thus, suppose $x_\varepsilon \in \Omega_\varepsilon$ are base points such that

$$u(x_\varepsilon) \geq u_o, \tag{1.20}$$

for some $u_o > 0$, and

$$\rho(x_\varepsilon) \sim 1, \tag{1.21}$$

i.e. $\rho(x_\varepsilon)$ is bounded away from 0 and ∞ as $\varepsilon = \varepsilon_i \to 0$. Then there two possible behaviors.

Convergence (Non-Collapse). Suppose $vol_{g_\varepsilon} B(\rho(x_\varepsilon)) \geq \nu_o$, for some $\nu_o > 0$. Let $U_j \equiv U_{j,\varepsilon} \subset \Omega_\varepsilon$ be the component of $\{q_\varepsilon \in \Omega_\varepsilon : 2^{-j} \leq \rho(q_\varepsilon) \leq 2^j\}$ containing x_ε, and choose a sequence $j = j(\varepsilon) \to \infty$ as $\varepsilon \to 0$. Then for any sequence $\varepsilon = \varepsilon_i \to 0$, there is a subsequence such that $(U_j, g_\varepsilon, x_\varepsilon)$ converges, (modulo diffeomorphisms), in the C^α topology, $\alpha < \frac{1}{2}$, and weak $L^{2,2}$ topology, uniformly on compact subsets, to a maximal limit (U, g_o, x). In particular, any compact domain in U embeds in U_j, for j sufficiently large, (in the subsequence).

Collapse. Suppose $vol_{g_\varepsilon} B(\rho(x_\varepsilon)) \to 0$ as $\varepsilon = \varepsilon_i \to 0$. Then, for any given $R < \infty$ and ε sufficiently small, the domains $U_j(R) = U_j \cap B_{x_\varepsilon}(R)$ above are either Seifert-fibered spaces over a surface V, or are torus bundles over an interval. In both cases, the diameter of the fiber F_{y_ε} over $y_\varepsilon \in U_j(R)$ satisfies $diam_{g_\varepsilon} F_{y_\varepsilon} \to 0$ as $\varepsilon \to 0$ and the inclusion map of the fiber in $U_j(R)$ induces an injection on π_1. In particular, there are finite covers $\widetilde{U}_j(R)$ of $U_j(R)$, of degree $d_j \to \infty$ as $\varepsilon \to 0$, such that $(\widetilde{U}_j(R), g_\varepsilon)$ does not collapse, and so has a C^α convergent subsequence as above. Letting $R = R_j \to \infty, j = j(\varepsilon)$ as $\varepsilon \to 0$ gives rise to a maximal limit $(\widetilde{U}, g_o, x)$. The limit has a free isometric S^1 or $S^1 \times S^1 = T^2$ action, arising from the unwrapping of the collapse.

There is of course some freedom in the choice of coverings unwrapping the collapse. It will be assumed throughout the paper that such coverings are chosen so that the diameter of the limiting S^1 orbit, or diameter and area of the limiting T^2 orbit, equal 1 at the limit base point x.

Analogous to (1.16) define the volume radius $\omega(x)$ of $(\Omega_\varepsilon, g_\varepsilon)$ by

$$\omega(x) = \sup\{r : \frac{vol B_y(s)}{s^3} \geq \omega_o, \forall B_y(s) \subset B_x(r)\}, \tag{1.22}$$

where ω_o is a fixed (small) positive constant. Both ρ and ω scale as distances. The two possibilities convergence/collapse above correspond to the situations $\omega(x_\varepsilon) \geq v_o \cdot \rho(x_\varepsilon)$ for some $v_o > 0$, or $\omega(x_\varepsilon) << \rho(x_\varepsilon)$ respectively. A similar collapse result holds if $B_{x_\varepsilon}(\rho(x_\varepsilon))$ is sufficiently collapsed, for any fixed ε, i.e. if $\omega(x_\varepsilon) \sim \mu_o \cdot \rho(x_\varepsilon)$, for μ_o small. In this case, there are finite covers as above, of order about μ_o^{-1}, such that in the covering space, $\omega(x_\varepsilon) \sim \rho(x_\varepsilon)$.

Although the collapse case corresponds to a degeneration of the sequence of metrics, this situation is in fact somewhat easier to handle than the convergence situation, since the collapse may be (locally) unwrapped to obtain convergence to a limit which has special properties, namely a free isometric S^1 action. This extra structure is of importance and leads to simplifications not present in the convergence situation in general.

In both cases, one has in fact C^∞ convergence to the limit within regions satisfying (1.20)-(1.21), due to regularity estimates as in (1.18), c.f. [2, Rmk.4.3]. It is elementary but important to note that the Euler-Lagrange equations (1.14)-(1.15) are invariant under passing to covering spaces, as is, (modulo bounded factors), the curvature radius ρ.

The metric boundary ∂U of U is the limiting locus where the curvature of (U_j, g_ε) blows-up. Thus $p \in \partial U$ if and only if there is a Cauchy sequence $p_k \in U$, with $p_k \to p$ in the metric completion $\bar{U}$ of U and such that if $q_\varepsilon(k) \in U_j$ converges to p_k as $\varepsilon \to 0$ then $\rho_\varepsilon(q_\varepsilon(k)) \leq \mu(k)$, where $\mu(k) \to 0$ as $k \to \infty$. Thus, ∂U is typically, although not necessarily, singular; it is possible that U may extend through (parts of) ∂U to a larger smooth manifold.

Since each g_ε satisfies the Euler-Lagrange equations (1.14)-(1.15), the limit metric g_o (of a subsequence) satisfies the same equation with ε set to 0, i.e.

$$L^*(u) + (\tfrac{1}{4}su + \tfrac{1}{12}\frac{\sigma(M)}{T})g = 0, \tag{1.23}$$

$$2\Delta u + \tfrac{1}{4}(su - \frac{\sigma(M)}{T}) = 0. \tag{1.24}$$

These are the Euler-Lagrange equations for $\mathcal{S}^2_-$ at the value $\sigma(M)$. When $\sigma(M) = 0$, the equations (1.23)-(1.24) are the static vacuum Einstein equations with potential u:

$$L^*u = 0, \quad \Delta u = 0. \tag{1.25}$$

(Here one uses the fact from [2,(3.45)] that $s \to 0$ wherever $u \geq 0$).

The following result characterizes these solutions at points x_ε satisfying $u(x_\varepsilon) \to 1$ as $\varepsilon \to 0$.

Proposition 1.2. *Let x_ε be base points in $(\Omega_\varepsilon, g_\varepsilon)$ such that $u(x_\varepsilon) \to 1$, and suppose (1.21) holds. Then any limit (U, g_o, x) constructed above is hyperbolic, with constant scalar curvature $\sigma(M)$ if $\sigma(M) < 0$, or flat if $\sigma(M) = 0$.*

Proof: If $\sigma(M) = 0$, then the maximum principle applied to trace equation in (1.25) implies that $u \equiv 1$ in U, and hence the full equation $L^*u = 0$ implies g_o is flat, c.f. (1.5).

The argument is essentially the same when $\sigma(M) < 0$. Thus, (1.9) implies that in the limit one has $s = -u \cdot |\sigma(M)|$ and so, setting $\omega = u - 1 \leq 0$, the trace equation (1.24) can be written as

$$2\Delta\omega - \tfrac{1}{4}|\sigma(M)|(u+1)\omega = 0. \tag{1.26}$$

Since $\omega(x) = 0 = \max \omega$, the strong maximum principle, c.f. [9, Thm. 3.5], implies that $\omega \equiv 0$ in U and hence from (1.23), the metric g_o is of constant negative curvature. ■

The situation where

$$\rho(x_\varepsilon) \geq \rho_o > 0, \tag{1.27}$$

for all $x_\varepsilon \in (\Omega_\varepsilon, g_\varepsilon)$ as $\varepsilon \to 0$, i.e. when the metrics g_ε do not degenerate anywhere in Ω_ε, is analysed in detail in [1]; in particular (1.27) holds as $\varepsilon \to 0$ if and only if M is tame, i.e. (0.4) holds. As mentioned in §0, the Main Theorem has been proved in this case.

Observe that Proposition 1.1 shows that the metrics g_ε do not degenerate in the region U^{v_o}, provided one stays a fixed distance away from the boundary L^{v_o}. However, while the estimates (1.9)-(1.11) give some useful information on the structure of U^{v_o}, for instance $vol_{g_\varepsilon} U^{v_o} \to 1$ as $\varepsilon \to 0$ in case $\sigma(M) < 0$, it is possible apriori that the level L^{v_o} is δ-dense in $(\Omega_\varepsilon, g_\varepsilon)$, with $\delta \to 0$ as $\varepsilon \to 0$.

By far the most important issue is to understand the situation where

$$\rho(x_\varepsilon) \to 0 \ \text{ as } \ \varepsilon \to 0, \tag{1.28}$$

for at least some x_ε in the region where $u \geq u_o$, for some $u_o > 0$, so that the curvature blows up at or very near x_ε. We will not attempt to understand the global geometry of the degeneration in all of U^{u_o}, but it will be important to understand the local behavior of the degeneration, say in unit balls $B_{x_\varepsilon}(1) \subset (\Omega_\varepsilon, g_\varepsilon)$.

Remark 1.3. To give some perspective, here is a simple model situation. Suppose there are base points x_ε, $u(x_\varepsilon) \to 1$ and satisfying (1.28), but which are isolated in the sense that there exists $d_o > 0$ such that within $B_{x_\varepsilon}(d_o)$, $\rho(y_\varepsilon) \geq dist_{g_\varepsilon}(y_\varepsilon, x_\varepsilon)$. Thus, if there is no collapse in $B(d_o) = B_{x_\varepsilon}(d_o)$, and $u \to 1$ in $B(d_o/2)$, the metrics converge in $B(d_o/2)$ to a constant curvature metric g_o outside $x = \lim x_\varepsilon$, but the curvature blows-up at x_ε. The limit metric g_o extends smoothly through x, giving a constant curvature metric on a 3-ball B^3. In this situation, one has a natural 2-sphere in $B(d_o/2)$ that surrounds x_ε, and it is easy to see, from the arguments in [1,§3], that this 2-sphere is essential, (at least on the "inside"). Unfortunately, it seems very difficult to prove that base points x_ε with this good behavior necessarily exist.

In the situation where (1.28) holds, we always consider the blow-up metrics

$$g'_\varepsilon = \rho(x_\varepsilon)^{-2} \cdot g_\varepsilon, \tag{1.29}$$

so that $\rho'(x_\varepsilon) = 1$, where ρ' is the L^2 curvature radius w.r.t. g'_ε. (The rescaling (1.28) is also used when $\rho(x_\varepsilon) \to \infty$; this situation will arise only at the end of the paper, in §7.3).

In the rescaling (1.29), the equations (1.14)-(1.15) become

$$L^* u = \bar{\alpha} \nabla \mathcal{Z}^2 + \bar{\phi} \cdot g, \tag{1.30}$$

$$2\Delta\left(u - \tfrac{\bar{\alpha} s}{12}\right) + \tfrac{1}{4} s u = \tfrac{1}{2} \bar{\alpha} |z|^2 - 3\bar{c}'. \tag{1.31}$$

All quantities in (1.30)-(1.31) are w.r.t. the g'_ε metric. The constant $\bar{\alpha} = \bar{\alpha}(\varepsilon, x_\varepsilon)$ is given by $\bar{\alpha} = \bar{\varepsilon}/\rho^2, \rho = \rho(x_\varepsilon)$, while $c' = \rho^2 c$. As emphasized in [2], the function u is scale-invariant, i.e. it is considered as a function, and not as the normalized scalar curvature in the scale g'_ε.

The following result from [2, Cor.5.2] implies there is an apriori maximal rate or scale at which the curvature can blow up on $(\Omega_\varepsilon, g_\varepsilon)$ as $\varepsilon \to 0$.

Proposition 1.4. *There is a constant $\kappa > 0$, independent of M and $(\Omega_\varepsilon, g_\varepsilon)$, such that*

$$\rho^2(x_\varepsilon) \geq \kappa \frac{\varepsilon}{T} = \kappa \bar{\varepsilon}, \tag{1.32}$$

for all $x_\varepsilon \in \Omega_\varepsilon$. Equivalently, $\bar{\alpha} \leq \kappa^{-1}$.

The convergence/collapse possibilities, as well as Proposition 1.2, hold for the sequence g'_ε exactly as in the case (1.21) holds; as before, the convergence is modulo diffeomorphisms but this will usually be ignored in the notation. The possible structure of blow-up limits, i.e. limits (N, g', x) of the Riemannian manifolds $(\Omega_\varepsilon, g'_\varepsilon, x_\varepsilon)$ has been analysed in detail in [2]; specifically from [2, §4,§5], one has:

Proposition 1.5. *If $u(x_\varepsilon) \geq u_o$, for some constant $u_o > 0$, then any blow-up limit $(N,\ g',\ x)$ of the metrics $(\Omega_\varepsilon, g_\varepsilon, x_\varepsilon)$ is of the following form:*

(i) a complete non-flat solution of the $\mathcal{Z}^2_c$ equations, with uniformly bounded curvature,

(ii) an incomplete non-flat solution of the static vacuum Einstein equations,

(iii) an incomplete flat solution of the static vacuum Einstein equations.

The static vacuum equations are given by (1.25) while the $\mathcal{Z}^2_c$ equations with potential u are

$$L^* u = \bar{\alpha} \nabla \mathcal{Z}^2, \tag{1.33}$$

$$2\Delta(u - \tfrac{\bar{\alpha}}{12}s) = \tfrac{1}{2}\bar{\alpha}|z|^2, \tag{1.34}$$

where $\bar{\alpha}$ is a positive constant, (the limit of $\bar{\varepsilon}/\rho^2$). These equations are obtained from (1.30)-(1.31) by replacing $\bar{\varepsilon} = \varepsilon/T$ by $\bar{\alpha}$, replacing $\bar{\phi}$ and $\bar{c}$ by 0 and setting $su = 0$. Thus, on a $\mathcal{Z}_c^2$ solution, the support of s and of u are disjoint in their interiors. Note that when $\bar{\alpha} = 0$, the equations (1.33)-(1.34) become the static vacuum equations (1.25). Further, to leading order, the Euler-Lagrange equations (1.14)-(1.15) are identical to the $\mathcal{Z}_c^2$ equations (1.33)-(1.34).

The three cases above are characterized as follows: Case (i) occurs exactly when $\bar{\alpha} > 0$, Case (ii) occurs exactly when $\bar{\alpha} = 0$ and u is non-constant on the limit, while Case (iii) occurs exactly when $\bar{\alpha} = 0$ and $u = const.$ on the limit.

The limit (N, g', x) is a $\mathcal{Z}_c^2$ solution whenever x_ε is a point of locally *maximal* curvature blow-up, i.e. $\rho(x_\varepsilon) \leq \rho(y_\varepsilon)$, for all $y_\varepsilon \in B_{x_\varepsilon}(\delta) \subset (\Omega_\varepsilon, g_\varepsilon)$, for a fixed $\delta > 0$. By [2, Thm.B], whenever M is not tame, i.e. $\rho \to 0$ somewhere on $(\Omega_\varepsilon, g_\varepsilon)$ as $\varepsilon \to 0$, there exist base points x_ε whose blow-up limit is a $\mathcal{Z}_c^2$ solution. The limit is a possibly flat static vacuum solution when the curvature blows up at a rate much slower than the maximal rate. Observe that if the limit (N, g', x) based at x_ε is flat, then at any base points y_ε converging to points in ∂N, the curvature blows up much faster than at x_ε, i.e. $\rho(y_\varepsilon) << \rho(x_\varepsilon)$.

Remark 1.6. Recall the Sphere conjecture from [1]: if (N, g', y) is a complete $\mathcal{Z}_c^2$ solution arising as a blow-up limit of $(\Omega_\varepsilon, g_\varepsilon, y_\varepsilon)$, then (N, g') has an asymptotically flat end. The Sphere conjecture implies Conjectures I-II.

Any asymptotically flat end of (N, g') of course carries natural 2-spheres, namely the spheres S^2 approximating the spheres $S^2(R)$ of large radius in $\mathbb{R}^3$. Since the convergence to the limit is smooth, such spheres also lie in $(\Omega_\varepsilon, g'_\varepsilon)$, and hence also in M. Theorem C of [2] proves the Sphere conjecture in case the potential function u on (N, g') is bounded away from 0 outside a compact set, and the level sets of u are compact. The essential remaining difficulty in proving the conjecture lies in dealing with the situation where the levels of u are non-compact.

In this paper, in place of attempting to locate natural 2-spheres in ends of a $\mathcal{Z}_c^2$ solution as in the Sphere conjecture above, we instead locate 2-spheres directly in *flat* blow-up limits, as in Case (iii) of Proposition 1.5. Of course these two issues may, (and in fact should be), closely related since if a $\mathcal{Z}_c^2$ solution has an asymptotically flat end, then by "blowing it down", i.e. rescaling the metric by factors converging to 0, one obtains a limit which is flat, with an isolated singular point; the 2-sphere near ∞ thus becomes a standard 2-sphere, of size on the order of 1, surrounding the origin in $\mathbb{R}^3$. This blow-down of the limit $\mathcal{Z}_c^2$ solution can be obtained directly as a flat "blow-up" limit of $(\Omega_\varepsilon, g_\varepsilon)$ at base points x_ε nearby but distinct from y_ε. Note that such a 2-sphere in $(\Omega_\varepsilon, g_\varepsilon)$ is very small; it has size on the order of $\rho(x_\varepsilon) \to 0$ and is being rescaled to approximately unit size in $g'_\varepsilon = \rho(x_\varepsilon)^{-2} \cdot g_\varepsilon$; compare with Remark 1.3.

The main concern of the paper is to locate such 2-spheres by carefully analysing the structure of level sets of the potential u, i.e. studying the value distribution theory of u on $(\Omega_\varepsilon, g_\varepsilon)$. In effect, the main point is to show that there exist base points x_ε satisfying $u(x_\varepsilon) \to 1$ and (1.28) such that in rescalings (1.29) converging to a flat limit, the level sets of u have at least one uniformly compact and separated component $\mathcal{C}$, c.f. Remark 2.2. It is then not difficult to show that such components are surrounded in a natural way by 2-spheres.

We conclude this section with the following elementary consequence of the σ-tame condition.

Lemma 1.7. *Suppose M is σ-tame, so that there exist constants $k < \infty$ and $K < \infty$ such that, on the space of unit volume metrics on M,*

$$\liminf_{\widetilde{\sigma} \to 0} \widetilde{\sigma}^k \mathcal{Z}^2 \leq K, \tag{1.35}$$

where $\widetilde{\sigma}$ is as in (0.5). Then there is a sequence $\varepsilon = \varepsilon_i \to 0$, such that on any minimizing pair $(\Omega_\varepsilon, g_\varepsilon), \varepsilon = \varepsilon_i$,

$$\varepsilon \mathcal{Z}^2 \leq (1+K)\varepsilon^{1/(k+1)}, \quad \text{and} \quad \widetilde{\sigma} \leq (1+K)\varepsilon^{1/(k+1)}. \tag{1.36}$$

Proof: For any given ε the metric g_ε is a minimizer of $I_\varepsilon^{\,-}$, so that

$$\varepsilon \mathcal{Z}^2(g_\varepsilon) + \widetilde{\sigma}(g_\varepsilon) \leq \varepsilon \mathcal{Z}^2(g') + \widetilde{\sigma}(g'), \tag{1.37}$$

for any unit volume metric g' on M. By (1.35), there is a sequence of metrics g_i on M with $\widetilde{\sigma}_i = \widetilde{\sigma}(g_i) \to 0$ and

$$\mathcal{Z}^2(g_i) \leq K\widetilde{\sigma}_i^{-k}.$$

Substituting this in (1.37) gives

$$\varepsilon \mathcal{Z}^2(g_\varepsilon) + \widetilde{\sigma}(g_\varepsilon) \leq \varepsilon K \widetilde{\sigma}_i^{-k} + \widetilde{\sigma}_i.$$

Now this estimate holds for any ε. Choose $\varepsilon = \varepsilon_i$ by setting

$$\varepsilon_i = \widetilde{\sigma}_i^{k+1},$$

so that $\varepsilon_i \to 0$ as $i \to \infty$. This gives

$$\varepsilon_i \mathcal{Z}^2(g_{\varepsilon_i}) + \widetilde{\sigma}(g_{\varepsilon_i}) \leq K\varepsilon_i \cdot \varepsilon_i^{-k/(k+1)} + \varepsilon_i^{1/(k+1)} \leq \varepsilon_i^{1/(k+1)}(1+K),$$

which gives the result. ■

The estimate (1.36) is the only consequence of the σ-tame condition that will be used in the paper. Observe that (1.36) gives an estimate for the constant term $\bar{c}$ in (1.15); this will be used to obtain the global control in Theorem 3.5 and Theorem 5.4.

Finally, recall the following formulas for the change of curvature under conformal deformation of the metric, c.f. [6, 1.159ff]. Thus, if $\widetilde{g} = f^2 \cdot g$, then the Ricci curvature $\widetilde{r}$ of $\widetilde{g}$ is given by

$$\widetilde{r} = r - f^{-1}D^2 f + 2(d\ln f)^2 - f^{-1}\Delta f \cdot g. \tag{1.38}$$

Similarly, if $\widetilde{g} = \psi^4 \cdot g$, then

$$\psi^5 \widetilde{s} = -8\Delta\psi + s\psi. \tag{1.39}$$

It should also be understood that throughout the paper, we always pass to subsequences where convergence issues are concerned, without explicitly mentioning this at each use. A sequence is said to sub-converge if a subsequence converges.

2. The Linearized Equations.

In this section, we study in detail the linearization of the metrics g_ε and their rescalings g'_ε in (1.29) about their constant curvature limits given by Proposition 1.2. This forms the foundation for the work to follow in later sections. Throughout the paper, we study the behavior of $(\Omega_\varepsilon, g_\varepsilon, x_\varepsilon)$ at base points x_ε satisfying

$$u_\varepsilon(x_\varepsilon) \to 1, \tag{2.1}$$

on some arbitrary sequence $\varepsilon = \varepsilon_i \to 0$. From §4 on, the sequence ε_i will no longer be arbitrary, but will be required to satisfy (1.36).

A general introductory discussion on such linearizations is given in §2.1, while in §2.2 the three possible forms of the linearized equations and the associated form of the metrics are analysed in detail. These are then summarized and unified in §2.3, c.f. Theorem 2.11.

2.1. The metrics g_ε satisfy the Euler-Lagrange equations (1.14)-(1.15). After a little algebra on the trace equation, these may be rewritten in the form

$$L^* u = \bar{\varepsilon}\nabla\mathcal{Z}^2 + \bar{\phi} \cdot g, \tag{2.2}$$

$$2\Delta(u - \tfrac{\bar{\varepsilon}s}{12}) = \psi_\varepsilon + \tfrac{1}{2}\bar{\varepsilon}|z|^2, \tag{2.3}$$

where

$$\psi_\varepsilon = \tfrac{1}{4}\sigma T(u+1)(u-1) + \bar{d}_\varepsilon, \tag{2.4}$$

and $\bar{d}_\varepsilon$ is a constant, depending on the *global* geometry of $(\Omega_\varepsilon, g_\varepsilon)$, given by

$$\bar{d}_\varepsilon = \tfrac{1}{4}\sigma T(1 - T^{-2}) - \tfrac{1}{2}\bar{\varepsilon}\mathcal{Z}^2. \tag{2.5}$$

All quantities in (2.2)-(2.5) are w.r.t. $(\Omega_\varepsilon, g_\varepsilon)$ and $\bar{\varepsilon}\mathcal{Z}^2 \equiv \bar{\varepsilon}\mathcal{Z}^2(g_\varepsilon)$. Observe that (1.8) implies that $\bar{d}_\varepsilon \leq 0$, so that both summands of ψ_ε are non-positive. The term $\sigma T \to |\sigma(M)|$ when $\sigma(M) < 0$, while $\sigma T \to 0$ when $\sigma(M) = 0$, again by (1.8). Thus, ψ_ε is also uniformly bounded below as $\varepsilon \to 0$, i.e. there exists $K < \infty$ such that

$$-K \leq \psi_\varepsilon \leq 0. \tag{2.6}$$

Further, in regions where $u \to 1$, one has

$$\psi_\varepsilon \to 0. \tag{2.7}$$

Apriori, the L^2 curvature radius $\rho(x_\varepsilon) = \rho_\varepsilon(x_\varepsilon)$ may converge to 0, remain bounded away from 0 and ∞, or diverge to ∞, (as usual in subsequences), as $\varepsilon \to 0$. Regardless of which of these behaviors occurs, we will always work in the scale g'_ε given by

$$g'_\varepsilon = \rho(x_\varepsilon)^{-2} \cdot g_\varepsilon, \tag{2.8}$$

so that $\rho'(x_\varepsilon) = 1$. The metrics g'_ε then satisfy the rescaled Euler-Lagrange equations (1.30)-(1.31).

As discussed in §1, one may then pass, (in subsequences), to a maximal connected (blow-up) limit (F, g'_o, x_o) of the metrics $(\Omega_\varepsilon, g'_\varepsilon)$ based at $\{x_\varepsilon\}$, passing to sufficiently large finite covers if the sequence collapses at x_ε. As noted following Proposition 1.4, the assumption (2.1) and Proposition 1.2, (in the scale (2.8)), imply that the maximal limit (F, g'_o, x_o) at $x_o = \lim x_\varepsilon$ is a constant curvature manifold, i.e. either flat or hyperbolic, with limit potential $u \equiv 1$, provided $\rho(x_\varepsilon) \leq R_o$, for some $R_o < \infty$. If $\rho(x_\varepsilon) \to \infty$, so that g_ε is being 'blown-down', this remains true but does not follow directly from Proposition 1.2; since the proof is out of place here, it is given in Appendix A. (The very special situation when $\rho(x_\varepsilon) \to \infty$ will only arise in §7.3 and so may be ignored until then). Note also that F so defined is not apriori maximal w.r.t. a smooth flat or hyperbolic structure, i.e. F may possibly extend to a larger smooth flat or hyperbolic manifold, (c.f. the discussion preceding (1.23)).

It is of course apriori possible that the metrics $(\Omega_\varepsilon, g_\varepsilon)$ do not degenerate at all, i.e. $\rho(x_\varepsilon) \geq \rho_o > 0$ whenever $u(x_\varepsilon)$ is sufficiently close to 1. The question of whether the metrics g_ε degenerate in this region plays an important role in §4, but is not used until then. The main case of interest is when the metrics degenerate, so that $\rho(x_\varepsilon) \to 0$.

To study the linearization of the metrics g'_ε at the limit constant curvature metric (F, g'_o), write

$$g'_\varepsilon = g'_o + \delta \cdot h + o(\delta), \tag{2.9}$$

where the parameter $\delta = \delta(\varepsilon) \to 0$ as $\varepsilon = \varepsilon_i \to 0$ and h is a symmetric bilinear form, formally corresponding to $\frac{dg'_\varepsilon}{d\delta}|_{\delta=0}$. The choices of the parameter δ and the form h of course depend on each other. While the sequence $\{g'_\varepsilon\}, \varepsilon = \varepsilon_i$, converges smoothly to the limit g'_o, it may not converge smoothly with respect to ε, i.e. $dg'_\varepsilon/d\varepsilon$ may not exist at $\varepsilon = 0$, so that one cannot necessarily choose $\delta = \varepsilon$. (Of course, this is related to the fact that the choice of base points x_ε is far from unique).

The natural value for the parameter δ is the local size of the curvature near the base point x_ε, which gives a measure as to how far away g'_ε is from the limit metric g'_o. For reasons which will only become clear later, we use for this the trace-free Ricci curvature z in place of the full Ricci curvature r. Thus, set

$$\delta = \delta_z = (\fint_{B'_{x_\varepsilon}(\frac{1}{2})} |z'_\varepsilon|^2 dV_{g'_\varepsilon})^{1/2} = (\rho^4(x_\varepsilon) \fint_{B_{x_\varepsilon}(\frac{1}{2}\rho(x_\varepsilon))} |z_\varepsilon|^2 dV_{g_\varepsilon})^{1/2}, \tag{2.10}$$

where $\fint$ denotes the average value, i.e. the integral divided by the volume of the domain. Observe that the right side of (2.10) is scale invariant, so that δ is scale invariant. The quantity δ depends only on the isometry class of g_ε and so is independent of diffeomorphisms 'reparametrizing' g_ε. Clearly $\delta = \delta(x_\varepsilon)$ depends strongly on the choice of the base point. The factor $\frac{1}{2}$ in (2.10) could be replaced by any other fixed constant < 1. The quantities δ_r and δ_s are defined in the same way as δ, by replacing z by the Ricci curvature r or scalar curvature s respectively.

Note that from the definition of ρ and δ,

$$\delta \leq 1$$

always. Further, since the convergence to the limit is smooth in $(B_x(1), g'_o)$ and the limit g'_o is of constant curvature since (2.1) holds, one has

$$\delta(x_\varepsilon) \to 0, \quad \text{as} \quad \varepsilon \to 0. \tag{2.11}$$

On the other hand, only x_ε such that $\delta(x_\varepsilon) > 0$ for ε sufficiently small need to be considered. For if $\delta(x_\varepsilon) = 0$ on some sequence $\varepsilon = \varepsilon_i \to 0$, then g_ε is of constant curvature, (hyperbolic or flat), and $u = $ const, on $B_{x_\varepsilon}(\rho(x_\varepsilon))$. Since g_ε is real-analytic in the region where $u > 0$, g_ε is of constant curvature on the full component C_ε of Ω_ε containing x_ε. In this case, the metrics $(C_\varepsilon, g_\varepsilon) = (C, g_o)$ are constant, i.e. independent of ε, for all ε, and no further analysis is necessary, c.f. also §1.

There are three basic cases for the behavior of δ at base points x_ε, as $\varepsilon = \varepsilon_i \to 0$:

$$\begin{aligned} &(i) \quad \delta >> \rho^2, \\ &(ii) \quad \delta \sim \rho^2, \\ &(iii) \quad \delta << \rho^2. \end{aligned} \tag{2.12}$$

By definition (2.10), δ/ρ^2 is just the L^2 average of z over $B(\frac{1}{2}\rho)$. Observe also that $s_{g'_\varepsilon} = \rho^2 s_{g_\varepsilon} = O(\rho^2)$, so that in all cases

$$\delta_s \leq c \cdot \rho^2. \tag{2.13}$$

It is obvious from (2.11) that $\rho(x_\varepsilon) \to 0$ in Cases (i) and (ii). In Case (iii), either of the behaviors $\rho(x_\varepsilon) \to 0$ or $\rho(x_\varepsilon) \geq \rho_o > 0$ is possible. Thus, in Cases (i) and (ii), the maximal blow-up limit (F, g'_o, x) based at $\{x_\varepsilon\}$ with $x = \lim x_\varepsilon$, is flat, while in Case (iii), it is either flat or hyperbolic.

For δ as in (2.10), the local L^2 norm of h in (2.9) is on the order of 1 in the Cases (i) and (ii) above. More precisely, define

$$h_\varepsilon = \frac{g'_\varepsilon - g'_o}{\delta}. \tag{2.14}$$

Then

$$\fint_{B_{x_\varepsilon}(\frac{1}{2})} |h_\varepsilon|^2 dV_{g'_\varepsilon} \sim 1 \quad \text{as} \quad \delta \to 0. \tag{2.15}$$

This is because the L^2 norm of the curvature r'_ε of g'_ε is bounded by δ in $(B'_{x_\varepsilon}(\frac{1}{2}), g'_\varepsilon)$ and the components of $g'_\varepsilon - g'_o$ in a harmonic coordinate chart on these balls are bounded in the $L^{2,2}$ topology by the L^2 norm of the curvature. Moreover, since g'_ε satisfies the elliptic system (1.30)-(1.31), the covariant derivatives $\nabla^k z_{g'_\varepsilon}/\delta$, $k \geq 0$, are uniformly bounded in L^2, while $\alpha \nabla^k z_{g'_\varepsilon}/\delta \to 0$ as $\delta \to 0$, both on compact subsets of (F, g'_o). The proof of this (technical) result is given in Appendix B.

The estimate (2.15) may not hold in Case (iii), since one might have $\delta_z << \delta_s$, so that $s'_{g_\varepsilon}/\delta$ may be unbounded as $\delta \to 0$. This case must then be treated somewhat more carefully.

In analogy to (2.9), let

$$u_\varepsilon = 1 + \nu \cdot \delta + o(\delta), \tag{2.16}$$

where ν is (formally) $du_\varepsilon/d\delta$ at $\delta = 0$. More precisely, as in (2.14), set

$$\nu_\delta = \frac{u_\varepsilon - 1}{\delta}, \tag{2.17}$$

and let $\nu = \lim \nu_\delta$ if the limit exists. Special care must be taken here, since apriori ν might be infinite.

We complete this subsection with the following characterization of the boundary ∂F.

Proposition 2.1. *For any maximal limit (F, g'_o, x) with $x_\varepsilon \to x$ satisfying (2.1), the boundary ∂F is the Gromov-Hausdorff limit of the level set L^{v_o} from (1.17).*

Proof: Recall that ∂F is formed by limits of sequences where $\rho' \to 0$. By the scale-invariant estimate (1.19), $\rho' \to 0$ implies $t'_{v_o} \to 0$, where $t'_{v_o} = dist_{g'_\varepsilon}(L^{v_o}, \cdot)$. Hence ∂F is contained in the Gromov-Hausdorff limit of L^{v_o}, c.f. [10,Ch.3], or [14,Ch.10] for discussion of Gromov-Hausdorff limits.

To prove the converse, it suffices to prove that the converse of Proposition 1.1 also holds, i.e. there is a constant $R_o < \infty$, independent of ε, such that

$$\rho(x_\varepsilon) \leq R_o \cdot t_{v_o}(x_\varepsilon). \tag{2.18}$$

To see this, the estimate (2.18) is scale-invariant, and so it suffices to prove it in the scale g'_ε where $\rho'(x_\varepsilon) = 1$. The metrics $(\Omega_\varepsilon, g'_\varepsilon, x_\varepsilon)$ converge smoothly to the maximal limit (F, g'_o, x), and the potential u converges smoothly on compact subsets of F to the limit function u_o. Proposition 1.2, or more precisely its proof, shows that $u_o \equiv 1$. It follows that the level set L^{v_o} cannot converge to any interior region of F, and hence $t'_{v_o}(x_\varepsilon)$ is at least 1 in the limit, which gives (2.18).

■

Note that for any flat manifold (F, g'_o), the curvature radius ρ' satisfies

$$\rho' = t', \tag{2.19}$$

where $t'(x) = dist_F(x, \partial F)$; in (1.16) the balls $B_x(\rho'(x))$ are required to be contained in F. Similarly, if the limit is hyperbolic, i.e. of constant negative curvature, then $\rho' \leq t'$. Note that equality, i.e. (2.19), holds when $(s')^2 \leq 3c_o$, where s' is the (constant) scalar curvature of g'_o and c_o is as in (1.16).

The structure of ∂F plays an important role throughout the paper. The metric closure $\bar{F} = F \cup \partial F$ is a length space, c.f. [10, Ch.1], and is viewed as a singular flat (or hyperbolic) manifold. In situations where F is embedded in $\mathbb{R}^3$, (or a quotient of $\mathbb{R}^3$), $\bar{F}$ is then just the closure of a domain in $\mathbb{R}^3$, (or the quotient).

Some of the more technical issues in §3 and §4 arise because unit balls in $\bar{F}$ are not necessarily compact. Consider for instance F of the form of a product $F = \mathbb{R} \times V$, where V is the universal cover of $\mathbb{R}^2 \setminus \{0\}$; V may naturally be identified with the upper half plane. In this setting, the topological boundary of $\mathbb{R} \times V$ is $\mathbb{R}^2$, while the metric boundary is a line, i.e. $\mathbb{R} \times \{0\}$. Hence, ∂F, or even better $\partial V = \{0\}$, acts as a "wormhole", in that points of V that are far apart, when measured by the length of curves a bounded distance away from ∂V, may be close when the curves are allowed to be arbitrarily close to or within ∂V. On the other hand, it will be seen later following Proposition 4.8 that this is essentially the only way that ∂F can act as a wormhole.

Remark 2.2. Much of the work of the paper through §5 can be characterized as follows: find a flat blow-up limit (F, g', y) of $(\Omega_\varepsilon, g_\varepsilon)$ such that ∂F is compact with F non-compact, or at least there is a ball $B_x(R)$ of finite radius R with center $x \in \bar{F}$, and a constant $d_o > 0$ such that

$$B_x(R) \cap \partial F \neq \emptyset, \quad \text{but} \quad A_x(R, (1+d_o)R) \cap \partial F = \emptyset, \tag{2.20}$$

where $A_x(r, s)$ is the geodesic annulus about x of radii r, s. Thus ∂F has a compact region, separated by a definite amount from any other part of ∂F. The relation (2.20) is established in Theorem 5.4. It is in such gap regions that geometrically natural 2-spheres in $(\Omega_\varepsilon, g_\varepsilon)$ will be located.

2.2. The main results of this subsection, (Propositions 2.3-2.4 and Corollary 2.8), describe the form of the linearizations of g'_ε and u at their limits $(g'_o, 1)$. For clarity, the results are separated into three cases according to the three possibilities in (2.12). These three distinct situations will then be unified in §2.3, c.f. Theorem 2.11. An analysis similar to the analysis here was previously carried out in [1, §4], and also in [2, §7] in the context of $\mathcal{Z}_c^2$ solutions; this may serve as an introduction to the discussion below.

The first, and most important case, is the following.

Proposition 2.3. *Suppose the base points x_ε satisfy (2.1) and suppose that*

$$\delta >> \rho^2(x_\varepsilon), \quad \text{as} \quad \varepsilon = \varepsilon_i \to 0. \tag{2.21}$$

Then there is a sequence of affine functions a_δ such that $\nu_\delta - a_\delta$ subconverges to a smooth limit function ν on F. The linearizations $h = \lim_{\varepsilon\to 0} h_\varepsilon$ and ν are a non-trivial solution of the linearized static vacuum Einstein equations

$$r' = \frac{dr}{d\delta}|_{\delta=0} = D^2\nu, \quad \Delta\nu = 0, \tag{2.22}$$

at the flat metric g'_o. Thus, the limit function ν is a non-affine, locally bounded and smooth harmonic function on (F, g'_o).

To 1^{st} order in δ, the metric g'_ε is conformally flat. In fact, for ε small, and on any compact domain $K \subset F$, the metric g'_ε has the expansion, modulo diffeomorphisms,

$$g'_\varepsilon = (1 - 2\nu\delta)(g'_o + \delta\chi) + o(\delta), \tag{2.23}$$

where $\delta \to 0$ as $\varepsilon \to 0$, $|o(\delta)| << \delta$, and χ is an infinitesimal holonomy deformation, i.e. $g'_o + \delta\chi$ is a curve (in δ) of flat metrics on F.

Proof: The metrics g'_ε satisfy the Euler-Lagrange equations (2.2)-(2.3) in this scale, i.e.

$$L^*u = \bar{\alpha}\nabla\mathcal{Z}^2 + (-\tfrac{1}{4}su + \bar{c}_\varepsilon)\cdot g \tag{2.24}$$

$$2\Delta(1 + \tfrac{\varepsilon\sigma}{12})u = \psi'_\varepsilon + \tfrac{\bar{\alpha}}{2}|z|^2.$$

Here and below, the prime and subscript ε are usually omitted from the notation, and $u > 0$. The term ψ'_ε is given by $\psi'_\varepsilon = \rho^2 \cdot \psi_\varepsilon$, $\rho = \rho(x_\varepsilon)$. To obtain the linearization, divide the equations (2.24) by δ, and consider the limit as $\delta \to 0$, (as always in subsequences).

By Proposition 1.5 and the discussion following it, $\bar{\alpha} = \varepsilon/\rho^2(x_\varepsilon) \to 0$ as $\varepsilon \to 0$. Consider first the linearization of the $\bar{\alpha}|z|^2$ term, and write this in the form:

$$\frac{\bar{\alpha}|z|^2}{\delta} = \bar{\alpha}|z|\frac{|z|}{\delta}.$$

Proposition 1.1 implies that $|z| \to 0$ in L^∞ on any compact subset of F, while $\bar{\alpha}|z|/\delta$ is bounded, by the discussion following (2.15), c.f. also Appendix B. It follows that, as $\delta \to 0$,

$$\frac{\bar{\alpha}|z|^2}{\delta} = o(1), \tag{2.25}$$

on compact subsets of F. The same reasoning, again from the discussion following (2.15), implies

$$(\bar{\alpha}\nabla\mathcal{Z}^2)/\delta = o(1). \tag{2.26}$$

Next consider the remaining terms on the right in (2.24). By definition

$$\frac{\psi'_\varepsilon}{\delta} = \frac{\rho^2}{\delta}\psi_\varepsilon.$$

From (2.7), $\psi_\varepsilon \to 0$ as $\varepsilon \to 0$. Since, by assumption, $\delta >> \rho^2$, this term tends to 0 uniformly on compact subsets of F. The same analysis holds for the terms su and $\bar{c}_\varepsilon$; note that $s = g_{g'_\varepsilon} = \rho^2 s_{g_\varepsilon}$, and similarly for the term $\bar{c}_\varepsilon$ in this scale.

Hence, these arguments imply that

$$(L^*u)/\delta = o(1) \quad \text{and} \quad (\Delta u/\delta) = o(1),$$

in L^∞ on compact subsets of F. In particular, by (1.5),

$$u\frac{r}{\delta} = D^2\nu_\delta + o(1), \quad \Delta\nu_\delta = 0 + o(1). \tag{2.27}$$

The equations (2.27), with the $o(1)$ and δ terms removed, are the static vacuum Einstein equations (1.25). Since $u \to 1$, it follows that any limit (h, ν) of (h_δ, ν_δ) is a solution of the linearization of the static vacuum equations (2.22) at the flat limit g'_o, with $u \equiv 1$. To prove the existence of such limits, the regularity discussion following (2.15), (c.f. Appendix B), implies that r/δ converges, in a subsequence, to a limit r' and further h_ε converges, modulo diffeomorphisms, to a limit h on (F, g'_o).

Thus, (2.27) implies that the functions ν_δ are bounded, modulo addition of functions in the kernel of D^2. Since Ker D^2 consists of affine functions on F, there are affine functions a_δ such that $\nu_\delta - a_\delta$ converges to a limit harmonic function ν on (F, g'_o). This proves the first statement.

To prove the second statement, write

$$\frac{u-1}{\delta} = \nu + a_\delta + o(1), \tag{2.28}$$

and set $v = u - a_\delta \cdot \delta = 1 + \nu\delta + o(\delta)$. Clearly $a_\delta \cdot \delta \to 0$ as $\delta \to 0$. Now consider the conformally equivalent metric

$$\widetilde{g}_\varepsilon = v^2 g'_\varepsilon, \tag{2.29}$$

A simple computation using (2.27) and (1.38) shows that

$$\widetilde{r} = 2(d\ln v)^2 - \frac{2}{v}\Delta v \cdot g - \frac{\bar{\alpha}}{v}\nabla\mathcal{Z}^2, \tag{2.30}$$

where all metric quantities on the right in (2.30) are w.r.t. $g = g'_\varepsilon$. Hence,

$$\frac{\widetilde{r}}{\delta} = \frac{2(d\ln v)^2}{\delta} - \frac{2\Delta v}{v\delta} \cdot g - \frac{\bar{\alpha}\nabla\mathcal{Z}^2}{v\delta}.$$

By the arguments in (2.25)-(2.27), the last two terms on the right go to 0 as $\delta \to 0$. Similarly, since $d\ln v \to 0$ as $\delta \to 0$, and $(d\ln v)/\delta = \frac{1}{v}d\nu_\delta + o(1) \to d\nu$ as $\delta \to 0$, it follows that

$$\frac{\widetilde{r}}{\delta} \to 0, \quad \text{as} \quad \delta \to 0. \tag{2.31}$$

Thus, to first order in δ, $\widetilde{g}_\varepsilon$ is flat and so g'_ε is conformally flat. More precisely, one may write

$$\widetilde{g}_\varepsilon = \widetilde{g}_o + \delta\chi + o(\delta) = g'_o + \delta\chi + o(\delta), \tag{2.32}$$

where χ is an infinitesimal flat variation of the metric g'_o. The variation χ is a sum of a trivial term, due to variations of $\widetilde{g}_\varepsilon$ by diffeomorphisms, and a possibly non-trivial term due to variations in the holonomy, i.e. periods, of (F, g'_o). Ignoring the contribution due to diffeomorphisms, it follows that

$$g'_\varepsilon = \upsilon^{-2}(g'_o + \delta\chi) + o(\delta) = (1 - 2\nu \cdot \delta)(g'_o + \delta\chi) + o(\delta),$$

which proves the result. ∎

As noted above in the proof, the affine indeterminacy of ν arises from the fact that Ker D^2 consists of affine functions.

Next we turn to the Case (ii) of (2.12).

Proposition 2.4. *Suppose the base points x_ε satisfy (2.1) and suppose that*

$$\delta \sim \rho^2(x_\varepsilon), \quad \text{as} \quad \varepsilon \to 0. \tag{2.33}$$

Then there exists a sequence of affine functions a_δ such that the functions $\nu_\delta - a_\delta$ converge to a limit function ν. The linearizations h_ε from (2.14) converge to a solution of the linearized $\mathcal{S}^2_-$ equations,

$$z' = D^2\nu, \quad \Delta\nu = 0, \tag{2.34}$$

on (F, g_o). This solution is non-trivial in the sense that $D^2\nu \neq 0$. On compact subsets of F, the metric g'_ε has the expansion, modulo diffeomorphisms,

$$g'_\varepsilon = (1 - 2\nu \cdot \delta)(g'_{-\delta} + \delta\chi) + o(\delta), \tag{2.35}$$

where $g'_{-\delta}$ is the space form of constant curvature $(\sigma(M)r_o/6)\delta$, $r_o = \lim_{\varepsilon\to 0}(\rho^2/\delta)(x_\varepsilon)$ and χ is an infinitesimal holonomy deformation of $(F, g'_{-\delta})$.

Proof: Following the proof of Proposition 2.3, the estimates (2.25)-(2.26) remain valid here. Further, since ρ^2/δ is bounded and $\psi_\varepsilon \to 0$ on F by (2.7), it follows from (2.24) as before that

$$\Delta\nu_\delta = o(1). \tag{2.36}$$

Similarly, as before from the full equation in (2.24), one obtains

$$L^*(\frac{u}{\delta}) = \frac{\sigma T}{3}\frac{\rho^2}{\delta} \cdot g'_\varepsilon + o(1), \tag{2.37}$$

using the fact that $u \to 1$ and $-s \sim \rho^2\sigma T$ in this scale. When $\sigma(M) < 0$, one does not obtain here a solution to the linearized static vacuum equations, since

$$\frac{s_{g'_\varepsilon}}{\delta} = -\frac{\rho^2}{\delta}u\sigma T = \frac{\rho^2}{\delta}\sigma(M) + o(1). \tag{2.38}$$

The equations (2.36) and (2.37) imply, in the limit $\delta \to 0$, the equations

$$z' = D^2\nu, \quad \Delta\nu = 0. \tag{2.39}$$

These equations are the linearization of the equations for critical points of $\mathcal{S}^2_-$, i.e. the linearized $\mathcal{S}^2_-$ equations.

Exactly the same arguments as in Proposition 2.3 shows that the limit h exists, and there exist affine functions a_δ such that $\nu_\delta - a_\delta$ converges to a harmonic function ν on (F, g'_o). Similarly, for $\widetilde{g}_\varepsilon$ as in (2.29), the arguments in (2.30)-(2.31) give here that

$$\widetilde{r}_\varepsilon/\delta = \frac{\sigma(M)}{3} \cdot g'_\varepsilon + o(1), \tag{2.40}$$

and so

$$\widetilde{z}_\varepsilon/\delta = 0 + o(1). \tag{2.41}$$

Thus, to first order in δ, $\widetilde{g}_\varepsilon$ is of constant curvature, and hence g_ε to first order is again conformally flat. More precisely, the formula (1.39) for conformal deformation and the estimate (2.36) give

$$\frac{\widetilde{s}}{\delta} = \frac{s_{g'_\varepsilon}}{\delta}\upsilon^{-2} + o(1) = -\frac{\rho^2}{\delta}\sigma T(1 + o(1)). \tag{2.42}$$

Hence, to first order in δ, $\widetilde{g}_\varepsilon$ is locally isometric to the space form of constant curvature $-\delta(\frac{\rho^2}{\delta}\sigma T)/6$. As always, by passing to subsequences, one may assume that $\rho^2/\delta \to r_o > 0$, while by (1.8) $-\sigma T \to \sigma(M)$. Thus, to first order in δ, $\widetilde{g}_\varepsilon$ is locally isometric to the space form $g'_{-\delta}$ of constant curvature $\delta r_o \sigma(M)/6$. The remainder of the proof, in particular (2.35), then follows as in Proposition 2.3. ■

Next, we turn to the situation where $\delta << \rho^2$, which is somewhat more complicated.

Remark 2.5. Actually, the situation below where $\delta << \rho^2$ or the situation (2.33) where $\delta \sim \rho^2$, although logically needed for the work through §5, are needed only in a comparatively minor way, (in terms of the overall picture). Hence on a first reading, it is advisable to skip at this point to the summary Theorem 2.11, and then onto §3, returning to this case as needed.

(ii). Similarly, given Proposition A, the work to follow in §2.2 - §2.3 also holds when $\rho(x_\varepsilon) \to \infty$ as $\varepsilon \to 0$; this will be used however only in §7.3.

Some preliminary estimates are needed in this case before deriving the analogue of Propositions 2.3 and 2.4. First, return to the equations (2.24), at the blow-up scale g'_ε. The estimate (2.26) also holds here, i.e. on compact subsets of F,

$$\frac{\bar{\alpha}}{\delta}\nabla\mathcal{Z}^2 \to 0, \quad \text{as} \quad \varepsilon \to 0, \tag{2.43}$$

c.f. again Appendix B for the details of the proof. Taking the trace-free part of (2.24) then gives

$$\frac{z}{\delta} = D_o^2\nu_\delta + o(1), \tag{2.44}$$

where $\nu_\delta = (u-1)/\delta$ as before and D_o^2 is the trace-free Hessian.

Observe that the quantity z/δ in (2.44) is both bounded away from 0 and bounded away from ∞, in the blow-up scale. Thus it has a limit, in a subsequence, and the limit is not identically 0.

We return to the estimate (2.44), the trace-free analogue of (2.27) or (2.37) in Corollary 2.8 below, but first need to understand in detail the behavior of the trace equation in (2.24) in this setting.

Referring to ψ_ε in (2.4), let $\psi_\delta = \psi_\varepsilon/\delta$, so that $\psi'_\delta = \rho^2 \cdot \psi_\delta$ in the g'_ε scale (2.8). Hence from (2.43), or more directly from the definition of δ, one has

$$2\Delta\nu_\delta = \rho^2(x_\varepsilon)(\psi_\delta + o(1)) \leq o(1). \tag{2.45}$$

Thus, in any limit, the potential is superharmonic. Of course, since the behavior of u is being examined near its maximal value, this is not unexpected.

The following result gives control on the size of the right side of (2.45) from below.

Lemma 2.6. *Let $x_\varepsilon \in (\Omega_\varepsilon, g_\varepsilon)$ be any base points satisfying (2.1) and suppose one of the following conditions holds: there are positive constants χ_o or ρ_o such that*

$$|u-1|(x_\varepsilon) \geq \chi_o \cdot \bar{\varepsilon}\mathcal{Z}^2, \quad \text{or} \tag{2.46}$$

$$\rho(x_\varepsilon) \geq \rho_o. \tag{2.47}$$

Then there exists a constant $K < \infty$, depending only on χ_o or ρ_o, such that

$$\xi(x_\varepsilon) \equiv \rho^2(x_\varepsilon)\psi_\delta(x_\varepsilon) \geq -K. \tag{2.48}$$

Proof: The proof is by contradiction, and so suppose

$$\xi(x_\varepsilon) = \rho^2(x_\varepsilon)\psi_\delta(x_\varepsilon) \to -\infty, \quad \text{as} \quad \varepsilon \to 0.$$

Renormalize the equations (2.45) and (2.48) by dividing them by $|\xi(x_\varepsilon)|$, so that

$$2\Delta\bar{\nu}_\delta = \bar{\xi} + o(1), \tag{2.49}$$

where $\bar{\nu}_\delta = \nu_\delta/|\xi(x_\varepsilon)|, \bar{\xi} = \xi/|\xi(x_\varepsilon)|$, so that

$$\bar{\xi}(x_\varepsilon) = -1. \tag{2.50}$$

It then follows from (2.44) and the uniform bound on $(z_{g'_\varepsilon}/\delta)$ that

$$D_o^2\bar{\nu}_\delta = o(1), \tag{2.51}$$

on compact subsets of F. Consider first functions ϕ such that

$$D_o^2\phi = 0, \tag{2.52}$$

on the limit (F, g'_o, x), $g'_o = \lim g'_\varepsilon$. If $\rho(x_\varepsilon) \to 0$, the limit F is flat, while if $\rho(x_\varepsilon) \geq \rho_o > 0$, then the limit is either flat or hyperbolic. We treat first the case where (F, g'_o) is flat.

On a flat manifold, the only functions ϕ satisfying (2.52) locally are, modulo affine functions,

$$\phi = b_o \cdot t_o^2, \tag{2.53}$$

where b_o is a constant and $t_o(x) = dist(x, p_o)$, for some base point $p_o \in F$. It follows from (2.51)-(2.53) that $\bar{\nu}_\delta$ converges to a limit function $\bar{\nu}_\infty$ modulo addition of affine functions and functions ϕ of the form (2.53). By (2.51), the limit function $\bar{\nu}_\infty$ necessarily satisfies (2.52) and hence

$$\Delta\bar{\nu}_\infty = 6b_o.$$

The normalization (2.50) gives $b_o = -1/12$ and further $\bar{\xi} = \bar{\xi}_\varepsilon$ converges to the constant function -1 as $\varepsilon \to 0$. Hence, provided $y_\varepsilon \to y \in F$, from (2.4) we have

$$\bar{\xi}(y_\varepsilon) = \frac{1}{|\psi(x_\varepsilon)|}[\tfrac{1}{4}\sigma T(u+1)\nu_\delta + \bar{d}_\delta](y_\varepsilon) \to -1,$$

where $\bar{d}_\delta = \bar{d}_\varepsilon/\delta$ and $\bar{d}_\varepsilon$ is as in (2.5). Via (2.49), this implies that on compact subsets of F,

$$2\Delta\bar{\nu}_\delta = 2\Delta\frac{\nu_\delta}{\rho^2(x_\varepsilon)|\psi(x_\varepsilon)|} = \frac{1}{|\psi(x_\varepsilon)|}[\tfrac{1}{4}\sigma T(u+1)\nu_\delta + \bar{d}_\delta] \to -1. \tag{2.54}$$

Now suppose first (2.46) holds, and consider the right side of (2.54). The term $\hat{d}_\delta \equiv \bar{d}_\delta/|\psi(x_\varepsilon)|$ is constant, and $u \to 1$ uniformly. Hence the oscillation of $\hat{\nu}_\delta \equiv \nu_\delta/|\psi(x_\varepsilon)|$ converges to 0 as $\delta \to 0$, i.e.

$$\sup_B \hat{\nu}_\delta - \inf_B \hat{\nu}_\delta \to 0, \tag{2.55}$$

where B is any compact subset of F. The hypothesis (2.46) and the definition of $\bar{d}_\delta$ imply that $|\nu_\delta|(x_\varepsilon) = -\nu_\delta(x_\varepsilon) \geq -\chi_o \cdot \bar{d}_\delta = \chi_o \cdot |\bar{d}_\delta|$ and so $|\hat{\nu}_\delta|(x_\varepsilon) \geq \chi_o \cdot |\hat{d}_\delta|$. Via (2.55), this now implies that

$$\sup_B \hat{\nu}_\delta / \inf_B \hat{\nu}_\delta \to 1, \tag{2.56}$$

and so of course $\sup_B \bar{\nu}_\delta / \inf_B \bar{\nu}_\delta \to 1$. In turn, this means that $\Delta\bar{\nu}_\delta \to 0$ weakly, which contradicts (2.54). An essentially identical argument holds if (2.47) holds, since in this case, $|\bar{\nu}| \leq \rho_o^{-1}|\hat{\nu}|$.

Next suppose the limit (F, g'_o) is hyperbolic, i.e. of constant negative curvature, so that in particular (2.47) automatically holds. The argument in this case is essentially the same as the flat case, with some minor modifications. Thus, since $u(x_\varepsilon) \to 1$, the scalar curvature of g'_ε is approximately $-\sigma T\rho^2, \rho = \rho(x_\varepsilon)$, so that g'_ε approximates a metric of constant sectional curvature $-\kappa^2, \kappa = \rho(\sigma T/6)^{1/2}$.

The arguments (2.49)-(2.52) hold as before. An elementary computation shows that the only solutions of (2.52) on the limit, (mod constants), are

$$\phi = b_o \cdot \cosh \kappa t_o, \tag{2.57}$$

with t_o again the distance from some point on F. A simple computation then gives

$$\Delta\phi = 3\kappa^2\phi. \tag{2.58}$$

On the other hand, arguing exactly as in (2.54), one has

$$2\Delta\frac{\hat{\nu}_\delta}{\rho^2} \sim \frac{\sigma T}{2}\hat{\nu}_\delta + \hat{d}_\delta,$$

or equivalently

$$\Delta\bar{\nu}_\delta \sim \frac{\sigma T}{4}\rho^2\bar{\nu}_\delta + \frac{1}{2}\hat{d}_\delta. \tag{2.59}$$

Since $\bar{\nu}_\delta$ converges to a solution of (2.52), (mod constants), $\bar{\nu}_\delta$ approaches a solution of (2.58). However, since $\kappa^2 = \rho^2(\sigma T/6)$, this contradicts (2.59), (since $\kappa > 0$ in the limit). ∎

Lemma 2.6 leads to the following definition:

Definition 2.7. *Base points $x_\varepsilon \in (\Omega_\varepsilon, g_\varepsilon)$ are called* allowable *if $u(x_\varepsilon) \to 1$ and $|\Delta\nu_\delta|(x_\varepsilon)$ remains uniformly bounded in the scale $g'_\varepsilon = \rho(x_\varepsilon)^{-2} \cdot g_\varepsilon$, as $\varepsilon = \varepsilon_i \to 0$.*

Thus, the results above imply that x_ε is allowable if either $\delta \geq c \cdot \rho^2$ at x_ε, as $\varepsilon \to 0$, for some $c > 0$, or if $\delta << \rho^2$, then either (2.46) or (2.47) holds. Observe that if x_ε is allowable, then both terms in $\xi(x_\varepsilon)$, i.e. $\rho^2\bar{d}_\delta$ and $\rho^2\nu_\delta(x_\varepsilon)$, are bounded as $\delta \to 0$.

Note that x_ε allowable does *not* imply that the linearized scalar curvature, i.e. $s_{g'_\varepsilon}/\delta$ remains uniformly bounded at x_ε as $\varepsilon \to 0$. On the other hand, by the statement following (2.44) and Lemma 2.6, the full Hessian $|D^2\nu_\delta|(x_\varepsilon)$ does remain bounded on allowable sequences.

The discussion above proves the following partial analogue of Propositions 2.3 and 2.4.

Corollary 2.8. *Suppose the base points x_ε are allowable and suppose that*

$$\delta << \rho^2(x_\varepsilon), \quad \text{as} \quad \varepsilon \to 0. \tag{2.60}$$

Then there exists a sequence of affine functions a_δ such that the functions $\nu_\delta - a_\delta$ converge to a non-affine limit function ν on the maximal limit (F, g'_o, x). The linearization of the metrics g'_ε at $\varepsilon = 0$ gives rise to a solution of the equations,

$$z' = D^2_o\nu, \quad 2\Delta\nu = \xi \leq 0, \tag{2.61}$$

where $\xi = \lim \rho^2(x_\varepsilon)(\frac{1}{2}\sigma T\nu_\delta + \bar{d}_\delta)$ on (F, g'_o).

Proof: The first equation in (2.61) follows from the limit of (2.44). For the second equation in (2.61), as noted above the constant term $\rho^2\bar{d}_\delta$ in ξ is bounded as $\varepsilon \to 0$. Further, the coefficient $\rho^2\sigma T$ of ν_δ in ξ is also bounded, since it approximates the scalar curvature of g'_ε. Hence, elliptic regularity for the trace equation in (2.45) implies that $\xi = \xi_\varepsilon$ is locally bounded on F and so (2.61) follows as the limit of (2.45). The proof that the functions ν_δ converge to ν modulo affine functions follows for the same reasons as before in Proposition 2.3. ∎

The situation in Corollary 2.8 of course differs somewhat from that in Propositions 2.3 and 2.4, where the limit potential ν is harmonic. It is not to be expected that $\xi = 0$ in (2.61) in general in situations where $\delta << \rho^2$. The form of the metric g'_ε, i.e. the analogue of (2.23) or (2.35) in this situation, is discussed below in Theorem 2.11.

Remark 2.9. If the limit (F, g_o) in Corollary 2.8 is hyperbolic, i.e. of constant negative curvature, then the functions ν_δ converge modulo constants to the limit ν; this follows since hyperbolic manifolds do not carry, even locally, any non-constant affine functions.

Remark 2.10. Suppose the base points x_ε are non-allowable, in that both (2.46) and (2.47) fail. The implication that (2.55) implies (2.56) no longer holds, since one may have $|\hat{\nu}_\delta| << |\hat{d}_\delta|$. Hence $\xi = \psi'_\delta$ may become unbounded as $\delta \to 0$. Observe that in this case the dominant term in ξ is $\rho^2 \bar{d}_\delta$. Whether this dominant term $\rho^2 \bar{d}_\delta$ stays bounded or not depends on the relation of ρ^2/δ, which depends only on the local geometry at x_ε, with the global term $\bar{\varepsilon}\mathcal{Z}^2$. There are other, much more fundamental reasons to work only with allowable base points which arise in §4 and §5 below.

2.3. In this section, we complete and unify the three forms of the linearizations, according to the three cases of (2.12). To do this, given allowable base points $x_\varepsilon \in \Omega_\varepsilon$, let $g_{-\kappa}$ be the space form of constant curvature

$$-\kappa^2 = -\frac{T\sigma_{g_\varepsilon}}{6}\rho^2(x_\varepsilon). \tag{2.62}$$

If $\delta(x_\varepsilon) >> \rho^2(x_\varepsilon)$ as $\varepsilon \to 0$, so $\kappa^2 << \delta$, set

$$g'_{-\kappa} \equiv g'_o, \tag{2.63}$$

where g'_o is the flat metric on F. If $\delta(x_\varepsilon) \sim \rho^2(x_\varepsilon)$ or $\delta(x_\varepsilon) << \rho^2(x_\varepsilon)$, set

$$g'_{-\kappa} \equiv g_{-\kappa}. \tag{2.64}$$

Theorem 2.11. *Let x_ε be any sequence of allowable base points. Then there is a sequence of affine functions a_δ, either 0 or diverging to $-\infty$, such that $\nu_\delta - a_\delta$ sub-converges to a non-affine function ν, defined on the maximal constant curvature limit (F, g'_o, x), $x = \lim x_\varepsilon$ of $(\Omega_\varepsilon, g'_\varepsilon, x_\varepsilon)$. The linearization $h = \lim_{\delta\to 0} h_\varepsilon$ from (2.14) exists and gives rise to a non-trivial solution of the equations*

$$z' = D_o^2\nu, \quad \Delta\nu = \lambda\nu + a_\infty \leq 0, \tag{2.65}$$

where λ is a constant, either 0 or positive, and a_∞ is a non-positive affine function, possibly constant. For ε sufficiently small, the metric g'_ε has the form, modulo diffeomorphisms,

$$g'_\varepsilon = (1 - 2\nu\delta)(g'_{-\kappa} + \delta\chi) + o(\delta), \tag{2.66}$$

where χ is an infinitesimal holonomy deformation of $(F, g'_{-\kappa})$.

Proof: The first equation in (2.65) has already been verified in (2.22), (2.34) and (2.61). As in (2.29), consider the conformally equivalent metric $\widetilde{g}_\varepsilon = v^2 g'_\varepsilon$, for v as following (2.28). The trace-free part of (1.38) implies

$$\frac{\widetilde{z}_\varepsilon}{\delta} = \frac{z_\varepsilon}{\delta} - \frac{D_o^2 v_\delta}{v} + 2\frac{(dv_\delta dv)_o}{v^2}, \tag{2.67}$$

where the subscript o denotes trace-free part and $v_\delta = (v-1)/\delta$. Since v_δ converges to the limit ν, dv_δ is bounded, while $dv \to 0$ uniformly on compact subsets of F. Hence, by the first equation in (2.65), one obtains

$$\frac{\widetilde{z}_\varepsilon}{\delta} \to 0, \quad \text{as} \quad \varepsilon \to 0, \tag{2.68}$$

on F, so that $\widetilde{g}_\varepsilon$ is of constant curvature, and hence g'_ε is conformally flat, both to first order in δ.

For the trace equation in (2.65), by (2.45) one has

$$\Delta\nu_\delta = \frac{\rho^2}{8}\sigma T(u+1)\nu_\delta + \frac{\rho^2 \bar{d}_\delta}{2} + o(1), \tag{2.69}$$

where $\bar{d}_\delta = \bar{d}_\varepsilon/\delta$. Recall that $u \to 1$ uniformly on compact subsets of the limit F. Lemma 2.6 implies that both terms on the right side of (2.69) are uniformly bounded as $\varepsilon \to 0$, while (2.5ff) implies both are non-positive. As in (2.28), $\nu_\delta = \nu + a_\delta + o(1)$, where a_δ is either 0 or a divergent sequence of affine functions, so that (2.69) has the form

$$\Delta\nu_\delta = \frac{\rho^2}{4}\sigma T(\nu + a_\delta) + \frac{\rho^2 \bar{d}_\delta}{2} + o(1). \tag{2.70}$$

Suppose first that there exists $c_o > 0$ such that, as $\varepsilon \to 0$,

$$\rho^2\sigma T \geq c_o, \tag{2.71}$$

Then the term a_δ must be uniformly bounded as $\varepsilon \to 0$, and so it may be ignored, i.e. absorbed into ν. In this case, (2.70) limits to

$$\Delta\nu = \lim(\frac{\rho^2}{4}\sigma T)\nu + \lim(\frac{\rho^2\bar{d}_\delta}{2}), \tag{2.72}$$

giving (2.65). On the other hand, if $\rho^2\sigma T \to 0$, then $\rho^2\sigma T\nu \to 0$ also, so the limit is

$$\Delta\nu = \lim(\frac{\rho^2}{4}\sigma T a_\delta) + \lim(\frac{\rho^2\bar{d}_\delta}{2}), \tag{2.73}$$

giving (2.65) with $\lambda = 0$.

Next we relate the scalar curvatures $\widetilde{s}_\varepsilon$ and s'_ε. A short calculation using (1.39) shows that

$$\frac{\widetilde{s}_\varepsilon}{\delta} = -4v^{-3}\Delta\nu_\delta + 2u^{-4}|d\nu_\delta||dv| + v^{-2}\frac{s'_\varepsilon}{\delta}, \tag{2.74}$$

where again the right side of (2.74) is w.r.t the blow-up metric g'_ε. As noted above $d\nu_\delta \to d\nu$, and so is bounded, while $dv \to 0$ as $\varepsilon \to 0$. Hence the second term in (2.74) tends to 0 as $\delta \to 0$. Further,

$$v^{-2}\frac{s'_\varepsilon}{\delta} = -v^{-2}\frac{\rho^2}{\delta}T\sigma u, \tag{2.75}$$

so that (2.74) becomes

$$\frac{\widetilde{s}_\varepsilon}{\delta} = -4u^{-3}\Delta\nu - v^{-2}\frac{\rho^2}{\delta}T\sigma u + o(1). \tag{2.76}$$

Referring to (2.72) and (2.73), $\nu\cdot\delta$, $a_\delta\cdot\delta$ and $d_\delta\cdot\delta$ all converge to 0 as $\delta \to 0$. Hence the second term in (2.76) dominates the first, so that

$$\frac{\widetilde{s}_\varepsilon}{\delta} = -\frac{\rho^2}{\delta}T\sigma(1 + o(1)). \tag{2.77}$$

In particular, one sees that $\widetilde{s}_\varepsilon$ is constant, to 1^{st} order in δ, which of course it must be by (2.68) and the Bianchi identity. The estimates (2.68) and (2.77) thus imply that $\widetilde{g}_\varepsilon$ is isometric, to first order in δ, to the space form $g'_{-\kappa}$, as in (2.67)-(2.68). The remainder of the proof follows as in Proposition 2.3. ■

We point out that if (2.71) holds, i.e. $\lambda > 0$ in (2.65), then the proof shows that ν_δ converges to ν, i.e. there is no constant or affine indeterminacy; in particular, $\nu \leq 0$ in this situation.

Observe that Theorem 2.11 implies that *any* allowable sequence $\{x_\varepsilon\}$ gives rise to a *non-trivial* and *uniformly controlled* solution of the corresponding linearized equations, modulo addition of affine functions. This fact plays a crucial role throughout the rest of the paper.

In particular, at any allowable base point sequence $\{x_\varepsilon\}$, there is a constant $\kappa_o > 0$, independent of ε and x_ε, such that, for most $y_\varepsilon \in B_{x_\varepsilon}(\rho(x_\varepsilon))$,

$$|d\nu_{\delta(x_\varepsilon)}|(y_\varepsilon) \geq \kappa_o/\rho(x_\varepsilon). \tag{2.78}$$

Note that the estimate (2.78) is scale-invariant. There is a minor technical issue here in that (2.78) need not hold for *all* $y_\varepsilon \in B_{x_\varepsilon}(\rho(x_\varepsilon))$; for instance, if y_ε is a critical point of u, then $|d\nu_{\delta(x_\varepsilon)}|(y_\varepsilon) = 0$. However, (2.78) does hold for generic $y_\varepsilon \in B_{x_\varepsilon}(\rho(x_\varepsilon))$, since if not, the limit potential ν on the maximal limit (F, g_o', x) would be constant, (modulo affine functions), contradicting Theorem 2.11. Thus, (2.78) holds for a large percentage of $y_\varepsilon \in B_{x_\varepsilon}(\rho(x_\varepsilon))$, where κ_o depends only on the volume percentage. In particular, one can always slightly adjust the choice of x_ε, within $\rho(x_\varepsilon)$, so that (2.78) holds at $y_\varepsilon = x_\varepsilon$. Such a choice of x_ε, (of course with $u(x_\varepsilon) \to 1$), will be called *generic*, and it will always be assumed, (usually implicitly), in the work to follow that base points are generic.

The uniformity in Theorem 2.11 also implies that for generic base points there is a constant $a_o > 0$, depending only on the collapse ratio $\mu_o = \omega(x_\varepsilon)/\rho(x_\varepsilon)$ such that

$$a_o \cdot \rho^2(x_\varepsilon) \leq area(B_{x_\varepsilon}(\rho(x_\varepsilon) \cap L)) \leq a_o^{-1} \cdot \rho^2(x_\varepsilon), \tag{2.79}$$

where L is the level set of ν_δ through x_ε.

Finally, it is worth pointing out that Theorem 2.11 is the reason for choosing $\delta = \delta_z$ and not $\delta = \delta_r$; linearizations w.r.t. δ_r in the situation $\delta << \rho^2$ always lead to trivial solutions of the linearized equations, at least when $\sigma(M) < 0$.

3. Level Set Measures and Masses.

In this section, we describe natural measures and corresponding mass functions associated to the potential functions ν_δ and their limits ν from §2. These measures are supported on the level sets of u, and it is the distribution of these measures in relation to the geometry of the metrics g_ε, or g_ε', which plays the central role in locating the 2-spheres in M.

First, recall that the choice of δ in (2.10) led to the constant and affine indeterminacy in the construction of the limit potential ν, at any given base point sequence x_ε, (as always satisfying (2.1)). At a later point in the paper in §5, it will be important to remove the affine indeterminacy of the potential ν_δ; however, until then this indeterminacy does not play a major role.

In analogy to the definition (2.10) of δ, define then

$$\delta_a = \delta_a(x_\varepsilon) = \oint_{B_{x_\varepsilon}(\frac{1}{2}) \cap L_\varepsilon} |\nabla u| dA_{g_\varepsilon'} = \rho(x_\varepsilon) \cdot \oint_{B_{x_\varepsilon}(\frac{1}{2}\rho(x_\varepsilon)) \cap L_\varepsilon} |\nabla u| dA_{g_\varepsilon}, \tag{3.1}$$

where the subscript stands for affine, L_ε is the level set of u through the base point x_ε and $\oint$ is the average value, as before. The first integral is w.r.t g_ε' while the second is w.r.t. g_ε, so that δ_a is scale-invariant, as δ is. Theorem 2.11 implies that the potentials $\nu_{\delta(x_\varepsilon)}$, at any allowable base point sequence x_ε, (sub)-converge, modulo affine functions, to non-trivial limit functions. As mentioned in connection with (2.78), it is also assumed henceforth that x_ε is generic. This implies immediately that

$$\delta_a \geq c_1 \cdot \delta, \tag{3.2}$$

for a fixed numerical constant $c_1 > 0$, independent of ε and the generic base point x_ε. The affine indeterminacy of ν_δ corresponds to the possibility that

$$\delta_a >> \delta. \tag{3.3}$$

The reason for preferring δ_a to δ, (at this stage), is that the potential function

$$\nu_{\delta_a(x_\varepsilon)} = (u-1)/\delta_a(x_\varepsilon) = \nu_{\delta(x_\varepsilon)} \cdot (\frac{\delta(x_\varepsilon)}{\delta_a(x_\varepsilon)}) \tag{3.4}$$

converges, modulo addition of constants, to a limit function $\nu = \nu_a$, so that the affine indeterminacy is reduced to a constant indeterminacy. When (3.3) occurs at x_ε, the potential ν_{δ_a} converges to a non-constant affine function ν_a, with $|\nabla \nu_a|(x) = 1$, on the limit (F, g_o', x_ε), while if

$$\delta_a \leq k_o \cdot \delta, \tag{3.5}$$

for some fixed constant $k_o < \infty$, (on base points x_ε), then the functions ν_{δ_a} and ν_δ have uniformly bounded ratios, and thus the limit functions ν and ν_a differ merely by a multiplicative factor. In particular, Theorem 2.11 holds also w.r.t. δ_a in place of δ, with the understanding that the limit function ν_a is affine in case (3.3) holds.

Note the following elementary result.

Lemma 3.1. *Suppose the base points $x_\varepsilon \in (\Omega_\varepsilon, g_\varepsilon)$ are allowable. Then there is a constant $C_o < \infty$, (independent of ε), such that*

$$|du|(x_\varepsilon) \leq C_o \cdot \frac{\delta_a(x_\varepsilon)}{\rho(x_\varepsilon)}. \tag{3.6}$$

Proof: To see this, observe that (3.6) is equivalent to the scale invariant estimate

$$\rho(x_\varepsilon)|d\nu_{\delta_a}|(x_\varepsilon) \leq C_o.$$

Theorem 2.11 implies that the functions ν_{δ_a} (sub)-converge smoothly, modulo constants, to a limit function ν_a on the maximal limit (F, g_o', x), $x = \lim x_\varepsilon$, (unwrapping in the case of collapse as described in §1). By construction and elliptic regularity, the limit satisfies $|d\nu_a|(x) \leq C_o$, with C_o independent of the base points x_ε. This gives (3.6), since the convergence to the limit is smooth. ∎

For the same reasons, if y_ε are points with $y_\varepsilon \to y \in (F, g_o')$, then there is a constant $C_1 < \infty$, depending only on $dist_{g_o'}(y, \partial F)$, and $dist_{g_o'}(x, y)$ such that

$$|du|(y_\varepsilon) \leq C_1 \cdot \frac{\delta_a(x_\varepsilon)}{\rho(x_\varepsilon)}. \tag{3.7}$$

We now introduce Riesz-type measures and mass functions for the potentials ν_{δ_a} and limit potentials ν_a constructed above. Let $L = L_\varepsilon$ be the level set of u through the base point x_ε and define

$$d\mu = d\mu_{\delta_a}(L) = |\nabla \nu_{\delta_a}| dA_{g_\varepsilon}, \tag{3.8}$$

where $\delta_a = \delta_a(x_\varepsilon)$ and dA_{g_ε} is Lebesgue measure w.r.t. g_ε on the level L. This measure depends on the base point, both with regard to δ_a and the level L. Observe that $d\mu$ is unchanged when adding constants to ν_{δ_a}. For any compact set $E \subset F$, one has the associated mass

$$m(E) = \int_{E \cap L} |\nabla \nu_{\delta_a}| dA_{g_\varepsilon}. \tag{3.9}$$

In particular, for any $q \in L$, define the mass function m_q of $d\mu$ by

$$m_q(s) = \int_{L \cap B_q(s)} |\nabla \nu_{\delta_a}| dA_{g_\varepsilon}. \tag{3.10}$$

Similar measures and mass functions are defined on the other level sets $\bar{L}$ of u as in (3.8)-(3.10), with $d\bar{A}$ the Lebesgue measure on $\bar{L}$, and with the same potential ν_{δ_a}. These measures on distinct level sets can be related to each other by means of the divergence theorem applied to the trace equation (2.3).

The mass functions in (3.9)-(3.10) scale as distance. Thus, if $g_\varepsilon' = \rho(x_\varepsilon)^{-2} \cdot g_\varepsilon$, then $m'(E) = \rho(x_\varepsilon)^{-1} m(E)$ and $m_q'(s) = \rho(x_\varepsilon)^{-1} m_q(\rho(x_\varepsilon) \cdot s)$.

When the blow-ups g_ε' (sub)-converge, to the limit (F, g_o', x), the rescaled measure $d\mu'$ and mass m' converge to that of the limit potential ν_a satisfying (2.65). The same holds on each level set $\bar{L}_\varepsilon \to \bar{L} \subset F$. Note that in case $\Delta \nu_a = 0$, the limit measure $d\mu'$ is the Riesz measure of the subharmonic function $\nu_{\bar{L}} = \sup(\nu_a, \nu_a(\bar{L}))$ on F.

However, when the blow-ups g_ε' collapse at x_ε, the mass m' as in (3.10) converges to 0. Thus, if the volume radius of x_ε is small compared with the curvature radius, i.e. $\omega(x_\varepsilon) \leq \mu_o \cdot \rho(x_\varepsilon)$, for some $\mu_o > 0$ small but fixed, unwrap the collapse by passing to large finite covers as described in

§1, so that $\omega(x_\varepsilon) \sim \rho(x_\varepsilon)$, and define the measure $d\widetilde{\mu}$ as in (3.8) in the cover. This is essentially equivalent to renormalizing the mass (3.9) as follows:

$$\widetilde{m}_{\bar{L}}(E) = \frac{\rho^2(x_\varepsilon)}{area_{g_\varepsilon}(L \cap B_{x_\varepsilon}(\frac{1}{2}\rho(x_\varepsilon)))} \int_{E\cap\bar{L}} |\nabla \nu_{\delta_a}| dA_{g_\varepsilon}, \tag{3.11}$$

where L is the level through x_ε. When the sequence based at x_ε is not μ_o-collapsed, then $\widetilde{m}(E) \sim m(E)$, for sets E whose diameter is on the order of $\rho(x_\varepsilon)$, c.f. (2.79). Observe again that the mass (3.11) scales as a distance. In particular, in the blow-up scale $g'_\varepsilon = \rho(x_\varepsilon)^{-2} g_\varepsilon$ attached to x_ε, one has, on $\bar{L}$,

$$\widetilde{m}'(E) = \frac{1}{area_{g'_\varepsilon}(L \cap B'_{x_\varepsilon}(\frac{1}{2}))} \int_{E\cap\bar{L}} |\nabla \nu_{\delta_a}| dA_{g'_\varepsilon}. \tag{3.12}$$

Note from the definitions (3.1), (3.10) and (3.12) that $\widetilde{m}'_{x_\varepsilon}(1/2) = 1$, and correspondingly,

$$\widetilde{m}_{x_\varepsilon}(\tfrac{1}{2}\rho(x_\varepsilon)) = \rho(x_\varepsilon). \tag{3.13}$$

Each of these notions may be defined w.r.t. the 'curvature' δ of (2.10) in place of the 'affine' δ of (3.1). In situations where (3.5) holds, the corresponding measures and mass functions are uniformly bounded, in terms of k_o, with respect to each other. However, when (3.3) holds, the measures and mass functions w.r.t. δ become unbounded; this is why δ_a is used in place of δ, for the time being.

Next, we derive some estimates for the behavior of the limit potential ν_a on F. We begin with the following upper bound on ν_a.

Lemma 3.2. *Let x_ε be an allowable base point sequence satisfying (2.46). Then the potential function ν_a on the maximal limit (F, g'_o, x), normalized so that $\nu_a(x) = 0$, is bounded above on compact subsets of $\bar{F}$, i.e. for any compact set $K \subset \bar{F}$,*

$$\nu_a(y) \leq C, \tag{3.14}$$

for all $y \in K$, where $C = C(K)$. The same estimate holds for the potentials $\nu_{\delta_a(x_\varepsilon)}$, renormalized by additive constants, converging to ν_a.

Proof: If $\delta_a(x_\varepsilon) >> \delta(x_\varepsilon)$, i.e. (3.3) holds, then the limit function ν_a is affine and so (3.14) is immediate. Thus, one may assume that (3.5) holds and so replace δ_a by δ. The functions ν_δ thus sub-converge, modulo constants, to the limit function ν.

Recall from Theorem 2.11 that the potential ν satisfies the trace equation (2.65). If $\lambda > 0$, then as noted following Theorem 2.11, $\nu \leq 0$ and so again (3.14) follows. Hence, one may assume that (2.65) is given by

$$\Delta\nu = \bar{d}_o \leq 0, \tag{3.15}$$

where $\bar{d}_o = \lim(\rho^2(x_\varepsilon)/\delta(x_\varepsilon))\bar{d}_\varepsilon > -\infty$. Note that $\bar{d}_\varepsilon \to 0$, while the ratio $\rho^2(x_\varepsilon)/\delta(x_\varepsilon)$ may remain bounded, or go to ∞; the product remains bounded as $\varepsilon \to 0$, since x_ε is allowable.

Since ν is smooth on F, ν can diverge to $+\infty$ on a compact set K in $\bar{F}$ only on approach to ∂F. Let p be any fixed point in $F \cap K$ and let $p_\varepsilon \in U^{v_o}$ be a sequence converging to p. Let $\hat{\nu}_{\delta(p_\varepsilon)} = \nu_{\delta(p_\varepsilon)} - \nu_{\delta(p_\varepsilon)}(p_\varepsilon)$ be the normalized potentials converging to the limit (normalized) potential ν_p. Then it suffices to prove that $\hat{\nu}_{\delta(p_\varepsilon)}$ is uniformly bounded above on the component $D' = D'_{p_\varepsilon}$ of $B'_{p_\varepsilon}(\frac{10}{9}) \cap U^{v_o}$ containing p_ε; here $\frac{10}{9}$ may be replaced by $1 + \mu$, for any fixed $\mu > 0$, and the ball B' is the ball in the metric $g'_\varepsilon = \rho(p_\varepsilon)^{-2} \cdot g_\varepsilon$ based at p_ε.

Suppose this is not the case, so that there exist $q_\varepsilon \in D'_{p_\varepsilon}$, for some sequence $p_\varepsilon \to p$, such that

$$\hat{\nu}_{\delta(p_\varepsilon)}(q_\varepsilon) \to \infty. \tag{3.16}$$

Then we claim that

$$\frac{\rho^2(q_\varepsilon)}{\delta(q_\varepsilon)} << \frac{\rho^2(p_\varepsilon)}{\delta(p_\varepsilon)}. \tag{3.17}$$

To see this, recall that $\delta(p_\varepsilon)$ is the L^2 average of z on the rescaled ball $B'_{p_\varepsilon}(\frac{1}{2})$ based at p_ε. The relation (3.16) implies that $|D_o^2\nu_{\delta(p_\varepsilon)}| >> 1$, near q_ε and hence by the relation (2.44), $z/\delta(p_\varepsilon) >> 1$ near q_ε, in the scale based at p_ε. Transferring this estimate to the scale based at q_ε gives (3.17).

It follows from (3.17) and the definition of $\bar{d}_o$ in (3.15) that on the limit (F, g'_o, q), based at $q = \lim q_\varepsilon$, $g'_\varepsilon = \rho(q_\varepsilon)^{-2} \cdot g_\varepsilon$, that the limit potential $\nu_q = \lim \hat{\nu}_{\delta(q_\varepsilon)}$ is harmonic, (since $\bar{d}_\varepsilon$ is independent of base point). In particular, ν_q has no local maxima. Hence the approximating potentials $\hat{\nu}_{\delta(q_\varepsilon)}$, restricted to the rescaled balls $B'_{q_\varepsilon}(1)$, have a maximal value a definite amount larger than the value at q_ε. Note that this gives a contradiction if for instance the component of ∂F containing q is compact.

Now observe this argument holds with respect to *any* allowable base point sequence y_ε. Thus, given x_ε, let L_ε be the u-level through x_ε and let R_ε be the component of the region bounded by L^{v_o} and L_ε containing x_ε. For each $y_\varepsilon \in L_\varepsilon \cap \partial R_\varepsilon$, consider the function $\hat{\nu}_{\delta(y_\varepsilon)}$ on D'_{y_ε} as above. The potential $\hat{\nu}_{\delta(y_\varepsilon)}$ has a maximal value on the closure $\bar{D}'_{y_\varepsilon}$ of D'_{y_ε}, and so one thus obtains a function ϕ_ε on L_ε. The function ϕ_ε itself has a maximal value, achieved, (or arbitrarily close to being achieved, since Ω_ε is complete), at say y_ε. Let $q_\varepsilon \in \bar{D}'_{y_\varepsilon}$ be a point realizing the maximum of $\hat{\nu}_{\delta(y_\varepsilon)}$. It follows that (3.16) and (3.17) hold, with y_ε in place of p_ε. As before, this gives a contradiction if q_ε is in the interior D'_{y_ε}. Suppose instead that $q_\varepsilon \in \partial D'_{y_\varepsilon}$. Observe that (3.17) gives

$$\delta(q_\varepsilon) >> \rho'(q_\varepsilon)^2\delta(y_\varepsilon). \tag{3.18}$$

Hence, if y_ε is moved slightly to y'_ε so that $q'_\varepsilon \in D'_{y_\varepsilon}$, then (3.18) is still preserved. It follows that the limit harmonic potential $\nu_q = \lim \hat{\nu}_{\delta(q_\varepsilon)}$ as above has a local maximum at q_ε, giving again a contradiction. This completes the proof of Lemma 3.2. ■

By construction, on the maximal limit (F, g'_o, x), one has $B'_x(1) \subset F$, c.f. also (2.19). Assume for the moment that F is flat, so that $\partial F \cap \partial B'_x(1) \neq \emptyset$. If $B'_x(1)$ is simply connected, then the developing map $\mathcal{D}$ of the flat structure on F, c.f. [17], gives an isometric embedding of $B'_x(1)$ onto the unit ball $B_0(1) \subset \mathbb{R}^3$, where $\mathcal{D}$ is normalized so that $\mathcal{D}(x) = 0$. For any fixed $R_o > 1$, let

$$D(R_o) \subset B'_x(R_o) \subset \bar{F} \tag{3.19}$$

be a maximal domain in $B'_x(R_o)$, containing $B'_x(1)$, such that $\mathcal{D}$ is an embedding of $D(R_o)$ onto its image in $\mathbb{R}^3$. Such a maximal domain is not unique, (as for instance cuts in a Riemann surface are not unique), but the results below are independent of the particular choice of maximal domain.

A similar definition holds if $B'_x(1)$ is not simply connected. Thus, if $\Gamma = \pi_1(B'_x(1))$, then Γ injects in $\pi_1(F)$ since the unit ball $\widetilde{B}'_x(1)$ is convex in the universal cover $\widetilde{F}$ of F. The group Γ is either $\mathbb{Z}$ or $\mathbb{Z} \oplus \mathbb{Z}$, acting by translation or twist on $\widetilde{F}$ or $\mathbb{R}^3$. The developing map descends to a map $\mathcal{D} : \widetilde{F}/\Gamma \to \mathbb{R}^3/\Gamma$, and $\mathcal{D}$ induces an embedding of $B'_x(1)$ onto the unit ball in $\mathbb{R}^3/\Gamma$. Thus, as in (3.19), define $D(R_o)$ to be a maximal domain on which $\mathcal{D}$ is an embedding onto its image in $\mathbb{R}^3/\Gamma$.

The reason for constructing the domains $D(R_o)$ is that they are natural *compact* analogues of the geodesic balls $B'_x(R_o)$; as noted preceding Remark 2.2, the balls $B'_x(R_o)$ may not be compact, (or have compact closure), in general. The same definition holds if the limit F is hyperbolic, using the developing map $\mathcal{D}$ of the hyperbolic structure. Domains $D_y(R_o)$, centered at a point $y \in F$, may be defined in the same way.

The next two results give (local) upper bounds on the mass of ν_a; these will be important in §4, (in Lemma 4.6).

Proposition 3.3. *For (F, g'_o, x, ν_a) as in Lemma 3.2, let $\bar{L}$ be any level set of the potential ν_a. Then there is a constant M_o, depending only on R_o, such that*

$$m(D(R_o) \cap \bar{L}) \leq M_o, \tag{3.20}$$

where m is the mass of the limit potential ν_a on the level set $\bar{L}$.

Proof: Assume the limit F is flat; the proof in the hyperbolic case is the same. Thus, the domain $D(R_o)$ may be viewed as a domain in $\mathbb{R}^3$, (or $\mathbb{R}^3/\Gamma$).

As noted following Theorem 2.11, the function ν_a is uniformly controlled in $B'_x(\frac{1}{2})$. The function ν_a satisfies the elliptic equation (2.65) and is bounded above near ∂F by Lemma 3.2. Standard properties of positive (or negative) solutions of elliptic equations in domains in $\mathbb{R}^3$ imply that ν_a is locally uniformly controlled, in that it satisfies a uniform local Harnack inequality and gradient estimates off ∂F, c.f. [9, Ch.8]. This implies that (3.20) holds for regions of $\bar{L} \cap D(R_o)$ not too close to ∂F.

In general, let $\bar{H}$ be a compact region within $D(R_o) \cap \bar{L}$ and assume w.l.o.g. that $\nu_a(\bar{L}) < \nu_a(L)$, where L is the level through x. The flow lines of $\nabla \nu_a$ starting in $\bar{H}$ then tend to a local maximum of ν_a. (The small set of flow lines ending in critical points of ν_a may be ignored in the following). For those flow lines γ for which $dist(\gamma(s), \partial F) \geq t_o$, for some small $t_o > 0$, the arguments above give a uniform upper bound on $|\nabla \nu_a|(\gamma(s))$. Suppose instead a flow line γ stays near ∂F. Then since ν_a is bounded above, $|\nabla \nu_a|(\gamma(s))$ is small, sufficiently far out along the flow. Hence, the region R consisting of the union of these flow lines which terminate where $|\nabla \nu_a| \leq C$, for some fixed constant C, is a compact set in F. Then (3.20) follows by applying the divergence theorem to the equation (2.65) over the domain K. ■

Proposition 3.3 also holds on the sequence $(\Omega_\varepsilon, g'_\varepsilon, x_\varepsilon)$ converging to (F, g'_o, x). Thus, let $D_\varepsilon(R_o)$ be a sequence of domains in $B'_{x_\varepsilon}(R_o) \cap U^{v_o}$, for U^{v_o} as in (1.18), converging in the Gromov-Hausdorff topology to the domain $D(R_o)$ as in (3.19). In addition, $D_\varepsilon(R_o)$ may be chosen to that if $\partial_s D_\varepsilon(R_o)$ is any subset of $\partial D_\varepsilon(R_o)$ converging to a proper subset of $\partial F \cap \bar{D}(R_o)$, then $\partial_s D_\varepsilon(R_o) \subset L^{v_o}$. These conditions do not uniquely define $D_\varepsilon(R_o)$, (even topologically), but this plays no role. If the sequence $(\Omega_\varepsilon, g'_\varepsilon, x_\varepsilon)$ is collapsing, or is μ_o-collapsed for ε sufficiently small, define D_ε so that in addition $\pi_1(B'_{x_\varepsilon}(1)) \subset \pi_1(D_\varepsilon(R_o))$, so that coverings unwrapping the collapse of $B'_{x_\varepsilon}(1)$ also unwrap $D_\varepsilon(R_o)$.

Let $\bar{L}_\varepsilon$ be any level set of u with $u(\bar{L}_\varepsilon) \to 1$ and let m'_{δ_a} be the mass in the scale g'_ε of $\nu_{\delta_a(x_\varepsilon)}$, lifted to covers if g'_ε is sufficiently collapsed at x_ε.

Corollary 3.4. *There is a constant $M_o < \infty$, such that on the levels $\bar{L}_\varepsilon$ as above,*

$$m'_{\delta_a}(D_\varepsilon(R_o) \cap \bar{L}_\varepsilon) \leq M_o. \tag{3.21}$$

Proof: The proof is identical to the proof of Proposition 3.3; alternately, (3.21) follows from (3.20) and the continuity of the mass to the limit. ■

The results above give local upper bounds for the mass of the potential, within small g_ε-neighborhoods of base points x_ε. The following elementary result gives a crucial upper bound on the *total* mass, on the base scale $(\Omega_\varepsilon, g_\varepsilon)$.

Theorem 3.5. *Let L be any level set of u, $\delta > 0$ and set $\nu_\delta = (u-1)/\delta$. Then on $(\Omega_\varepsilon, g_\varepsilon)$,*

$$m_\delta(L) \equiv \int_L |\nabla \nu_\delta| dA_{g_\varepsilon} \leq \frac{\bar{\varepsilon}\mathcal{Z}^2}{2\delta}. \tag{3.22}$$

Proof: Let $U^+ = \{x \in \Omega_\varepsilon : u(x) \geq u(L)\}$ and apply the divergence theorem to the trace equation (2.3), renormalized by δ, over U^+. A little algebra then gives

$$\int_L |\nabla \nu_\delta| = \int_{U^+} \Big(\frac{\varepsilon \mathcal{Z}^2}{4\delta T} - \frac{1}{8}\frac{\sigma}{T}\Big(\frac{T^2u^2 - 1}{\delta}\Big) - \frac{\varepsilon |z|^2}{4\delta T}\Big) \leq \int_{U^+} \Big(\frac{\varepsilon \mathcal{Z}^2}{4\delta T} - \frac{1}{8}\frac{\sigma}{T}\Big(\frac{T^2u^2 - 1}{\delta}\Big)\Big). \tag{3.23}$$

The left side of (3.23) is the total mass of the potential ν_δ at this base scale g_ε. Recall from §1 that $Tu = w$ and $w = -s/\sigma$, where σ is the L^2 norm of s; thus, the L^2 norm of w equals 1. Since $vol_{g_\varepsilon}\Omega_\varepsilon = 1$, this gives

$$\int_{\Omega_\varepsilon} \Big(\frac{T^2u^2 - 1}{\delta}\Big) = 0.$$

Also, $T^2u^2 - 1 \leq 0$ precisely on the set $W_1 = \{w \leq 1\}$. Hence

$$-\int_{U+}(\frac{T^2u^2-1}{\delta}) \leq -\int_{W_1}(\frac{T^2u^2-1}{\delta}) = \int_{W^1}(\frac{T^2u^2-1}{\delta}),$$

where $W^1 = \{w \geq 1\}$. Since $volW^1 \leq 1$ and $u \leq 1$, it follows that

$$-\frac{1}{8}\frac{\sigma}{T}\int_{U+}(\frac{T^2u^2-1}{\delta}) \leq \frac{1}{8}\frac{\sigma}{T}(\frac{T^2-1}{\delta}) \leq \frac{1}{4}\frac{\varepsilon\mathcal{Z}^2}{\delta T}, \tag{3.24}$$

where the last estimate follows from (1.8). This gives the bound (3.22). ■

Remark 3.6. Here a remark, not actually relevant to this paper, but relevant to the Sphere conjecture and the asymptotic geometry of $\mathcal{Z}_c^2$ solutions discussed in [1,2]. All of the analysis in §2-§3 carries over to the linearization at infinity of $\mathcal{Z}_c^2$ solutions (N, g'), provided $\sup_N u < \infty$. In fact, this situation is considerably easier, since $su \equiv 0$ and there is no constant term in the Euler-Lagrange equations (1.14)-(1.15). In particular, one can define mass functions similar to (3.9), c.f. [2, (7.54)ff]. However, in this case, there is no upper bound estimate available for the mass - Theorem 3.5 is lacking. It is precisely the constant term $c = c_\varepsilon$ of (1.6) in the Euler-Lagrange equations which leads to this bound. This information, which will be seen to be crucial, is thus lost when passing to $\mathcal{Z}_c^2$ limits.

4. Uniformity Estimates.

All of the results in §2 and §3, with the exception of Theorem 3.5, are local; they concern the geometry of blow-ups g'_ε and potentials ν_δ in small neighborhoods of base points x_ε. All of these results depend strongly on the size of δ, and hence on the choice of base point x_ε.

In order to relate these local behaviors at different base points, as well as relate them to the global estimate (3.22), it is important to obtain uniform control on δ on a given level set of u.

The purpose of this section is to prove the existence of base points y_ε where δ, (or more precisely δ_a), is not too small, both at y_ε and at points on the same level set as y_ε and within, say, unit g_ε-distance to y_ε. The main result in this section is Theorem 4.4. Some applications to the structure of ∂F are given in Proposition 4.8-Lemma 4.9.

To begin, the analysis is divided into two cases, according to the following:

Definition 4.1. *The sequence $(\Omega_\varepsilon, g_\varepsilon), \varepsilon = \varepsilon_i$, satisfies the* degeneration hypothesis *if there is a sequence of base points $x_\varepsilon \in \Omega_\varepsilon$ such that, as $\varepsilon \to 0$,*

$$|u-1|(x_\varepsilon) \leq \frac{1}{(\ln\bar{\varepsilon})^2}, \quad \text{and} \quad \rho(x_\varepsilon) \to 0. \tag{4.1}$$

The sequence $(\Omega_\varepsilon, g_\varepsilon)$ satisfies the non-degeneration hypothesis *if there exists a constant $\rho_o > 0$ such that whenever $|u-1|(x_\varepsilon) \leq 1/(\ln\bar{\varepsilon})^2$, then*

$$\rho(x_\varepsilon) \geq \rho_o. \tag{4.2}$$

Thus, the given sequence $(\Omega_\varepsilon, g_\varepsilon), \varepsilon = \varepsilon_i$, satisfies either the degeneration or the non-degeneration hypothesis; if it satisfies the degeneration (or non-degeneration) hypothesis, then so do all subsequences. Of course if $(\Omega_\varepsilon, g_\varepsilon)$ satisfies the degeneration hypothesis, then there still may well exist other base points $z_\varepsilon \in \Omega_\varepsilon$, distinct from x_ε, such that $|u-1|(z_\varepsilon) \leq 1/(\ln\bar{\varepsilon})^2$ but $\rho(z_\varepsilon) \geq \rho_o$, for some $\rho_o > 0$. The specific decay requirement (4.1) on $|u-1|$ is used only in the proof of Proposition 4.8 below and for most purposes can be replaced by the weaker condition that $u(x_\varepsilon) \to 1$ at any rate as $\varepsilon \to 0$.

The analysis in these two cases is quite different, and in particular is much simpler when $(\Omega_\varepsilon, g_\varepsilon)$ satisfies the non-degeneration hypothesis. The non-degeneration case (4.2) will be analysed in §7, and throughout §4-§5, it is assumed that $(\Omega_\varepsilon, g_\varepsilon)$ satisfies the degeneration hypothesis. (It will be used however only in Theorem 4.4).

Further, in this section we use the assumption that M is σ-tame. Thus, it will always be assumed that the sequence $(\Omega_\varepsilon, g_\varepsilon), \varepsilon = \varepsilon_i$, is chosen to satisfy (1.36), i.e.

$$\bar{\varepsilon}\mathcal{Z}^2 \equiv \bar{\varepsilon}\mathcal{Z}^2(g_\varepsilon) \leq \bar{\varepsilon}^{4\mu}, \tag{4.3}$$

for some $\mu < \frac{1}{4}$ and ε sufficiently small; (4.3) follows from (1.36) since $T \geq 1$ and hence $T^\mu \leq$ T. Of course, (4.3) then remains valid for any subsequences.

Now for any given $\varepsilon > 0$ fixed, consider certain special base points as follows:

Definition 4.2. *A sequence of base points x_ε is called* admissible *if it satisfies $\rho(x_\varepsilon) \to 0$ together with the following conditions. Let $L = L_\varepsilon$ be the level set of u containing x_ε. Then*

$$\bar{\varepsilon}^{2\mu} \leq |u(L) - 1| \leq \frac{2}{(\ln \bar{\varepsilon})^2}, \tag{4.4}$$

and, for the affine δ_a defined as in (3.1) and any point $p_\varepsilon \in B_{x_\varepsilon}(1) \cap L$,

$$\delta_a(p_\varepsilon) \geq \delta_o \equiv \bar{\varepsilon}^{2\mu}. \tag{4.5}$$

Observe that by Definition 2.7, admissible base points are allowable. Of course, $\delta_o \to 0$ as $\varepsilon \to 0$ and in fact, from (4.3),

$$\frac{\bar{\varepsilon}\mathcal{Z}^2}{\delta_o} \leq \bar{\varepsilon}^{2\mu} \to 0, \quad \text{as} \quad \varepsilon \to 0. \tag{4.6}$$

Hence $\bar{\varepsilon}\mathcal{Z}^2/\delta(p_\varepsilon) \to 0$, for all $p_\varepsilon \in L \cap B_{x_\varepsilon}(1)$; compare with (3.22). The estimates (4.5) and (4.6) play the central role in the arguments of §5. Although (4.5) is not required to hold globally on L, it is important to recognize that (4.5) holds on a region in L of determined size about the base point x_ε.

The definition of δ_o in (4.5), natural in view of (4.6), is just a convenient choice, and other choices, such as $\delta_o = \bar{\varepsilon}^\mu$, are equally possible. Nevertheless, it is necessary to make one choice, and so we choose (4.5).

An important related class of base points are the following:

Definition 4.3. *A sequence of base points x_ε is called* well-separated *(w.r.t. U^{v_o}), if it satisfies (4.4) together with the following conditions: there exist scales $l_\varepsilon > 0$ and constants $r_o, d_o > 0$, (independent of ε), such that $l_\varepsilon \geq r_o\rho(x_\varepsilon)$, (4.5) holds for all $p_\varepsilon \in B_{x_\varepsilon}^{v_o}(l_\varepsilon) \cap L$, and*

$$A_{x_\varepsilon}^{v_o}(l_\varepsilon, (1 + d_o)l_\varepsilon) \cap L = \emptyset. \tag{4.7}$$

Here $A_x^{v_o}(r, s)$ is the geodesic annulus about x of inner and outer radii r, s, in the metric space (U^{v_o}, g_ε), where $U^{v_o} = \{x_\varepsilon \in \Omega : u(x_\varepsilon) \geq 1 - v_o\}$, as in Proposition 1.1; similarly for the ball $B_x^{v_o}$. While it is possible to work with the (more natural) full annulus A_{x_ε} in $(\Omega_\varepsilon, g_\varepsilon)$, it is somewhat simpler at this stage to work with the restricted annulus $A_{x_\varepsilon}^{v_o}$; it will be seen later (in Lemma 5.7 below) that the difference between these two choices becomes irrelevant.

The region $\mathcal{C} = \mathcal{C}(\varepsilon) = B_{x_\varepsilon}^{v_o}(l_\varepsilon) \cap L$ gives rise to a *compact* and *isolated* region of the level set L of ν_a in the limit (F, g_o', x), where $g_\varepsilon' = l_\varepsilon^{-2} g_\varepsilon$; compare with Remark 2.2. A well-separated base point satisfies all the properties of admissible base points, except that (4.5) may not hold at all points in the 1-ball about x_ε, (and $\rho(x_\varepsilon)$ is not required to go to 0).

Finally, a sequence of base points x_ε is called preferred if it is well-separated with $l_\varepsilon \sim \rho(x_\varepsilon)$, i.e. $l_\varepsilon/\rho(x_\varepsilon)$ is bounded away from 0 and infinity as $\varepsilon \to 0$.

The purpose of this section is to prove the existence of at least one of these two kinds of base points, either admissible or preferred, near any base points x_ε satisfying (4.1). In §5, it will then be proved that if y_ε is an admissible base point, then there exist well-separated base points z_ε

near y_ε. As in Remark 2.2, it is the well-separated base points which are naturally associated with 2-spheres.

Theorem 4.4. *Suppose $(\Omega_\varepsilon, g_\varepsilon)$ satisfies the degeneration hypothesis (4.1), with $\varepsilon = \varepsilon_i$ satisfying (4.3). Then for any base points x_ε satisfying (4.1), there are base points $y_\varepsilon \in B_{x_\varepsilon}(1)$ which are either admissible or preferred.*

The proof will proceed in severals steps, via two Lemmas. Let $L = L(\varepsilon)$ be the the level set of u through x_ε. By Proposition 1.1, $t_{v_o}(x_\varepsilon) \to 0$, where $t_{v_o} = dist(\cdot, L^{v_o})$. Let $L_1 = L_1(\varepsilon)$ be the level set of u defined by

$$u(L_1) = u(L) - \bar{\varepsilon}^{2\mu}, \tag{4.8}$$

so that $u(L_1) \to 1$, as $\varepsilon \to 0$ also. By the converse of Proposition 1.1, c.f. (2.18), it follows that there are base points $y_\varepsilon \in B_{x_\varepsilon}(\frac{1}{10}) \cap L_1$ such that $\rho(y_\varepsilon) \to 0$ as well. Thus, y_ε also satisfies (4.4).

It will be proved that such base points y_ε are either admissible, or there exist preferred base points $q_\varepsilon \in B_{y_\varepsilon}(1)$, i.e. either all $z_\varepsilon \in B_{y_\varepsilon}(1) \cap L_1$ satisfy (4.5) or there exist preferred base points $q_\varepsilon \in B_{y_\varepsilon}(1)$. Note that the Lipschitz property of ρ implies that $\rho(z_\varepsilon) \leq 1$, for all $z_\varepsilon \in B_{y_\varepsilon}(1)$.

Let L_2 be the level set of u defined by

$$u(L_2) = u(L_1) - \bar{\varepsilon}^{\mu}. \tag{4.9}$$

Consider arbitrary (piecewise) C^1 paths γ in U^{v_o}, which start at some point in L^{v_o}, end at a point z_ε in $B_{y_\varepsilon}(1) \cap L_1$, and which satisfy the following cone condition:

$$s \leq C \cdot dist(\gamma(s), L^{v_o}) = C \cdot t_{v_o}(\gamma(s)), \tag{4.10}$$

where s is the arclength parameter of γ and C is any fixed constant. Observe that (4.10) is scale-invariant. It is clear that such paths exist; for instance one may choose γ to be a shortest path in U^{v_o} realizing the distance of z_ε to L^{v_o}. The condition (4.10) roughly means that γ does not zigzag too much across the levels of u. For any such γ, let s_2 be the largest value of s such that $\gamma(s_2) \in L_2$ and let $a_\varepsilon = \gamma(s_2)$. Finally set $\gamma^2 = \gamma|_{[s_2,b]}$, where $\gamma(b) = z_\varepsilon$.

A simple but important idea in the proof is the following:

Lemma 4.5. *For any curve γ as above, there exist points $p_\varepsilon \in \gamma^2$ such that*

$$\delta_a(p_\varepsilon) \geq \delta_o. \tag{4.11}$$

Proof: If not, then there is a curve γ satisfying (4.10) such that, $\forall s \geq s_2$, $\delta_a(\gamma^2(s)) \leq \delta_o$, and hence by Lemma 3.1,

$$|du|(\gamma^2(s)) \leq C_o \frac{\delta_o}{\rho(\gamma^2(s))}. \tag{4.12}$$

Observe again that the estimate (4.12) is scale-invariant. For the argument to follow, we work in the (base) scale g_ε. For t_{v_o} as above, Proposition 1.1 implies that

$$\frac{1}{\rho(\gamma^2(s))} \leq \frac{c_1}{t_{v_o}} \leq \frac{c_2}{s},$$

where the 2nd inequality follows from (4.10). From (4.12), it then follows that, along $\gamma^2(s)$,

$$|\frac{du}{ds}| \leq c_3 \frac{\delta_o}{s}.$$

Integrating this along γ^2 gives

$$|u(L^2) - u(L^1)| \leq c_4 \delta_o \ln(\frac{1}{s_2}). \tag{4.13}$$

Now $s_2 \geq dist_{g_\varepsilon}(a_\varepsilon, L^{v_o}) \equiv \lambda_o$ and again by (2.18), $\lambda_o \geq R_o^{-1} \cdot \rho(a_\varepsilon)$. By Proposition 1.4, $\rho^2 \geq \kappa \cdot \bar{\varepsilon}$ at any base point and so in particular $\lambda_o \geq \bar{\varepsilon}$. Thus, using the definition of δ_o, one has

$$|u(L^2) - u(L^1)| \leq c_5 \bar{\varepsilon}^{2\mu} \ln(\bar{\varepsilon}^{-1}) << \bar{\varepsilon}^{\mu}, \tag{4.14}$$

for ε sufficiently small. This contradicts (4.9), and so proves the result.

■

Observe that the proof shows there are points q_ε on γ^2 such that $\delta_a(q_\varepsilon) >> \delta_o$. More importantly, the proof also holds if γ is a curve satisfying (4.10), with, for instance,

$$C_o = C_o(\varepsilon) \leq \varepsilon^{-\mu/2}, \tag{4.15}$$

in place of the ε-independent C_o of (4.12).

If all the points p_ε from Lemma 4.5 may be chosen to be the end points $z_\varepsilon \in L_1 \cap B_{y_\varepsilon}(1)$ of curves γ, then the result is proved. Thus, suppose not, so that there exists $z_\varepsilon \in L_1 \cap B_{y_\varepsilon}(1)$ such that

$$\delta_a(z_\varepsilon) < \delta_o. \tag{4.16}$$

We then construct a path γ, starting at z_ε, (i.e. in the reverse direction from previously), satisfying (4.10) and an analogue of (4.12). The point is to construct γ with a controlled increase of δ_a, i.e. the local mass, along γ, analogous to (4.15). Given the initial estimate (4.16), it is then shown that this path either terminates on L_2, giving a contradiction as in Lemma 4.5, or terminates on a preferred base point q_ε. This will then complete the proof.

The construction of γ and the growth estimate for δ_a is by an inductive procedure. Thus, let $z_\varepsilon = z_\varepsilon^1$, and let $g_\varepsilon^1 = \rho(z_\varepsilon^1)^{-2} \cdot g_\varepsilon$ be the rescaled metric based at z_ε^1. Let $\delta_1 = \delta_a(z_\varepsilon^1)$ and let ν_{δ_1} be the corresponding potential function. By adding a suitable constant, one may assume that $\nu_{\delta_1}(L_1) = -1$; recall that adding constants to ν_{δ_a} does not affect the measure or mass of the potential. Thus

$$\nu_{\delta_1}(L_2) \sim -\bar{\varepsilon}^{\mu}/\delta_1 << -1, \tag{4.17}$$

for ε small. In particular, it follows that $\rho^1 \to 0$ on L_2, where ρ^1 is the g_ε^1 curvature radius, and so L_2 converges, in the Hausdorff topology, to ∂F, where (F, g_o', z^1) is the maximal limit based at z_ε^1.

Define intermediate level sets L^j by

$$\nu_{\delta_1}(L^{j+1}) = e^{k_o j} \nu_{\delta_1}(L^j), \tag{4.18}$$

with $L^1 = L_1$. Here k_o is a fixed number which will be specified below, (see (4.23) and (4.26)). It follows that for $\varepsilon > 0$ fixed, the level L_2 is reached after at most N iterations, i.e. $L_2 \sim L^N$, where

$$N = \frac{1}{k_o} \ln \frac{\bar{\varepsilon}^{\mu}}{\delta_1}. \tag{4.19}$$

The induction will be on j, i.e. down the level sets L^j, beginning with the levels L^1 and L^2.

To start, recall the compact domains $D_\varepsilon(R_o)$ defined following Proposition 3.3, with base point z_ε^1. If $B^1_{z_\varepsilon^1}(1)$ is μ_o-collapsed, i.e. $\omega(z_\varepsilon^1) \leq \mu_o \rho(z_\varepsilon^1)$, then the collapse is unwrapped in covering spaces, as described following Proposition 3.3. Here μ_o is a fixed parameter, e.g. $\mu_o = 1/100$. For convenience, set $R_o = 2$. Then Corollary 3.4 gives

$$m(D_\varepsilon(2) \cap \bar{L}) \leq M_o, \tag{4.20}$$

for any level set $\bar{L}$ of u, e.g. $\bar{L} = L^1$ or L^2, where $m = m_{\delta_1}$ is the mass w.r.t. δ_1 in the scale g_ε^1.

The basic idea in the inductive step, i.e. in the construction of the next base point z_ε^2 on the level L^2 is to examine the distribution of the measure $d\mu_{\delta_1}$ within $L^2 \cap D_\varepsilon(2)$. By way of illustration, suppose $d\mu$ is a measure on $I = [0,1]$ of mass at most M_o. If the support of $d\mu$ satisfies supp $d\mu = I$, then it is clear that for all $0 < t < 1$, there exists $x \in \mathrm{supp} d\mu$ such that

$$m_{d\mu}([x, x+t]) \leq M_o \cdot t, \tag{4.21}$$

(i.e. the analogue of (4.20) holds on the scale t). If on the other hand the mass of $d\mu$ is concentrated, (for example the Dirac measure at a point in the most extreme case), then (4.21) does not hold. The only situation in which (4.21) can fail is when there are gaps in supp $d\mu$, i.e. intervals $J \subset I$ on which $J \cap \mathrm{supp} d\mu = \emptyset$. This corresponds to the existence of well-separated base points.

Define a level set $\bar{L}$, for instance $\bar{L} = L^2$, to be *d_o-separated* (w.r.t. U^{v_o}), if there is a point $p_\varepsilon \in \bar{L}$ and radius s_ε such that

$$A^{v_o}_{p_\varepsilon}(s_\varepsilon, (1+d_o)s_\varepsilon) \cap \bar{L} = \emptyset, \tag{4.22}$$

compare with (4.7). The condition (4.22) is scale-invariant, although we continue to work in the scale g^1_ε. The relation (4.22) implies that $s_\varepsilon \geq diam_{g^1_\varepsilon} C_o$, where C_o is the component of $\bar{L}$ containing p_ε, and $diam$ is the extrinsic diameter, i.e. the diameter within $(U^{v_o}, g^1_\varepsilon)$. Let $t(x) = t^1(x) = dist_{g^1_\varepsilon}(x, L^{v_o})$. Henceforth, assume that k_o is chosen sufficiently large so that

$$t^1(x) \leq \tfrac{1}{4}\mu_o, \forall x \in L^2 \cap D_\varepsilon(2), \tag{4.23}$$

where μ_o is the collapse parameter, defined preceding (4.20).

Lemma 4.6. *Suppose the level L^2 has no d_o-separated components intersecting $D_\varepsilon(2)$, for $d_o \leq 1$. Then there exist base points $z^2_\varepsilon \in L^2 \cap D_\varepsilon(2)$ such that*

$$m(L^2 \cap D_{z^2_\varepsilon}(2t(z^2_\varepsilon))) \leq \mu_1 M_o (1+d_o) \cdot t(z^2_\varepsilon), \tag{4.24}$$

where $\mu_1 = \mu_1(\mu_o)$.

Proof: Suppose first that $\hat{L}^2 = L^2 \cap D_\varepsilon(2)$ is path connected. Since L^2 has no d_o-separated components intersecting $D_\varepsilon(2)$, the diameter, (within U^{v_o}), satisfies $diam_{g^1_\varepsilon}\hat{L}^2 \geq \mu_o/(1+d_o) \geq \mu_o/2$. A curve τ realizing the intrinsic diameter of $\hat{L}^2$ thus has length $l(\tau) \geq \mu_o/2$.

Now cover the path τ by a collection of smaller domains with disjoint interiors as follows. Let s be the arclength parameter for τ, with $p_0 = \tau(0)$. Let $D_1 = D_{p_1}(2t(p_1))$, where p_1 is the point on τ whose arclength distance to $\tau(0)$ equals $2t(p_1)$; here D_1 is the domain as in (4.20), centered on p_1. Let s_2 be the largest value of s such that $\tau(s) \in D_1$ and set $p_2 = \tau(s_2)$; clearly $s_2 > 2t(p_1)$. Next let $D_2 = D_{p_3}(2t(p_3))$, where p_3 is the point on τ whose arclength distance to p_2 equals $2t(p_3)$. Observe that $D_1 \cap D_2 = p_2$. One then continues inductively in this way until the maximal p_k is defined, for k odd. It follows from (4.23) that the domain $V = \cup D_j$ covers at least $\frac{1}{2}$ the length $l(\tau)$ of τ.

By construction, $m(V) = \sum m(D_j)$. If for each j, $m(D_j) > \mu_1 M_o t(p_j)$, then

$$m(V) > \mu_1 M_o \sum t(p_j).$$

But by constuction, $\sum t(p_j)$ is least $\frac{1}{2}l(\tau) \geq \frac{1}{4}\mu_o$. Thus, if $\mu_1 > 4\mu_o^{-1}$, this contradicts (4.20), and hence there is some j such that $m(D_j) \leq \mu_1 M_o t(p_j)$. This gives (4.24) in this situation, with $z^2_\varepsilon = p_j$.

If $\hat{L}^2$ is not path connected, let τ_k be a path realizing the intrinsic diameter of each path component C_k of $\hat{L}^2$, and let τ be the union of the paths τ_k. Then perform the construction above on each component τ_k. Since there are no d_o-separated subsets, for any k, there must be a C_j and point $q_j \in C_j$ such that $dist_{D_\varepsilon(2)}(q_j, C_k) \leq \frac{1+d_o}{2}(diam_{D_\varepsilon(2)} C_k)$. It follows that $l(\tau) \geq (1+d_o)^{-1}\mu_o$, and the same argument as above then gives (4.24).

■

The domain $D^1_\varepsilon \equiv D_\varepsilon(2)$ converges, (in a subsequence), to a limit domain $D^1 \equiv D_o(2)$ contained in the maximal limit (F, g'_o, z^1), $z^1 = \lim z^1_\varepsilon$; the limit F is either flat or hyperbolic. Since the limit level $L^2 \cap D_o(2) \subset F$ has a definite diameter, bounded away from 0 and ∞, it follows that $L^2 \cap D_\varepsilon(2)$ is d_o-separated only when (4.22) holds with

$$s_\varepsilon \sim 1 \tag{4.25}$$

in the scale g^1_ε, i.e. s_ε is bounded away from 0 and ∞ in this scale.

Lemma 4.6 provides the inductive step. Thus, fix any choice of $d_o \in (0, 1]$ and set

$$k_o = \ln[A_o \mu_1 (1 + d_o) M_o], \tag{4.26}$$

where $A_o = a_o^{-1}$ is the area constant from (2.79). (Note that k_o may be chosen to be large, by choosing M_o large). Suppose L^2 has no d_o-separated components in $D_\varepsilon(2)$. Then Lemma 4.6 gives the existence of the next base point $z^2_\varepsilon \in L^2$ and domain $D^2_\varepsilon \equiv D_{z^2_\varepsilon}(2t(z^2_\varepsilon))$ satisfying (4.24). Observe that, by construction, $dist_{g^1_\varepsilon}(z^2_\varepsilon, z^1_\varepsilon) \leq 2$ and $dist_{g^1_\varepsilon}(z^2_\varepsilon, L^1) \geq \frac{1}{2}$, where the 2^{nd} inequality follows from (4.23). Hence, one may choose a curve $\gamma_1(s) \subset U^{v_o}$ from z^2_ε to z^1_ε, with arclength parameter s, such that

$$s \leq c \cdot dist_{g^1_\varepsilon}(\gamma_1(s), L^2), \tag{4.27}$$

i.e. the analogue of (4.10) holds. Further, by (3.7) one has, along γ_1,

$$|du|(\gamma_1(s)) \leq C \frac{\delta_1}{\rho(\gamma_1(s))}, \tag{4.28}$$

where C is a fixed constant, possibly depending on k_o but not on ε.

Now at the endpoint z^2_ε of γ_1, we claim that for $M_1 = A_o \mu_1 M_o (1 + d_o)$,

$$\delta_2 \equiv \delta_a(z^2_\varepsilon) \leq M_1 \delta_1. \tag{4.29}$$

To see this, recall first that the limit based at z^1_ε is flat or hyperbolic. As noted in (2.19ff), without loss of generality, one may assume that

$$\rho^1(z^2_\varepsilon) = (1 + o(1)) t(z^2_\varepsilon), \tag{4.30}$$

where ρ^1 is the curvature radius in the g^1_ε scale. In the scale $g^2_\varepsilon = \rho(z^2_\varepsilon)^{-2} \cdot g_\varepsilon$ where $\rho^2(z^2_\varepsilon) = 1$, by definition $\delta_a(z^2_\varepsilon) = \oint_{H(\frac{1}{2})} |\nabla u|$, where $H(\frac{1}{2}) = B_{z^2_\varepsilon}(\frac{1}{2}) \cap L^2$. Further, in this scale $D^2_\varepsilon = D_{z^2_\varepsilon}(2)$ and hence (4.24) and (4.30) give

$$m_{\delta_1}(D^2_\varepsilon) = \int_{D^2_\varepsilon \cap L^2} |\nabla u| / \delta_1 \leq \mu_1 M_o (1 + d_o).$$

Since $H(\frac{1}{2}) \subset D^2_\varepsilon$, (again in the g^2_ε scale), it follows that

$$\delta_a(z^2_\varepsilon) \leq \big(\mu_1 M_o (1 + d_o) / area_{g^2_\varepsilon} H(\tfrac{1}{2})\big) / \delta_1.$$

Via (2.79), this gives (4.29).

We now repeat this construction at the level 2 in place of level 1. Thus, as above, at the base point $z^2_\varepsilon \in L^2$, rescale the metric based at z^2_ε, so that $g^2_\varepsilon = \rho(z^2_\varepsilon)^{-2} \cdot g_\varepsilon$ and work with the potential ν_{δ_2} in place of ν_{δ_1}. From (4.24), (4.29) and (4.30), one has the bound, in the g^2_ε scale,

$$m_{\delta_2}(D^2) \leq M_1^2. \tag{4.31}$$

Assuming that there are no d_o-separated subsets within $\hat{L}^3 = L^3 \cap D^2_\varepsilon$, one may then construct again by Lemma 4.6 the next level base point z^3_ε and domain D^3_ε, as well as the next part of γ, i.e. γ_2, from z^2_ε to z^3_ε satisfying the analogue of (4.27). The estimate (4.28) translates to the estimate

$$|du|(\gamma_2(s)) \leq C M_1^2 \frac{\delta_1}{\rho(\gamma_2(s))}.$$

Thus, this process may be continued inductively until the level L_2 is reached, provided at each stage there are no d_o-separated subsets.

Suppose then that this process continues to the level L_2, producing a curve γ starting at L_2 and ending at z_ε^1. Since each part γ_1, γ_2, etc. of γ satisfies the analogue of (4.27), γ satisfies the cone condition (4.10). Further, along γ one has

$$\delta_a(\gamma(s)) \leq C \cdot M_1^N \delta_1, \tag{4.32}$$

where N is given by (4.19). However, together with (4.26) this gives

$$\delta_a(\gamma(s)) \leq C \cdot \bar{\varepsilon}^{\mu/2} \delta_1^{1/2} \leq C \cdot \bar{\varepsilon}^{3\mu/2},$$

everywhere along $\gamma(s)$, where the 2nd inequality uses $\delta_1 \leq \bar{\varepsilon}^{2\mu}$ by (4.16); compare also with (4.15). The proof of Lemma 4.5 carries through as before, and again gives a contradiction.

Hence if (4.16) holds, (for some z_ε), there must be a (first) level L^{j-1}, with base points $z_\varepsilon^{j-1} \subset L^{j-1}$, and domains D_ε^{j-1} such that the next level $\hat{L}^j = L^j \cap D_\varepsilon^{j-1}$ is d_o-separated, i.e. there exist $z_\varepsilon^j \in \hat{L}^j$ such that (4.22) holds, with $s_\varepsilon \sim 1$, in the scale $g_\varepsilon^{j-1} = \rho(z_\varepsilon^{j-1})^{-2} \cdot g_\varepsilon$. Suppose that, for all $p_\varepsilon^j \in \hat{L}^j$,

$$\delta_a(p_\varepsilon^j) \geq \min(\delta_o, \delta_a(z_\varepsilon^{j-1})). \tag{4.33}$$

Then (2.78) implies that the potential function $\nu_{\delta_a^{j-1}} \equiv \nu_{\delta_a(z_\varepsilon^{j-1})}$ satisfies $|\nabla \nu_{\delta_a^{j-1}}|(p_\varepsilon^j) \geq \kappa_o / t(p_\varepsilon^j)$ in the g_ε^{j-1} scale. Since $\nu_{\delta_a^{j-1}}(L^j)/\nu_{\delta_a^{j-1}}(L^{j-1}) = e^{k_o}$ and k_o is fixed, it follows that $\rho^{j-1}(z_\varepsilon^j) \geq \rho_o = \rho_o(\kappa_o) > 0$. This means that $\rho^j(z_\varepsilon^j) \sim 1$, i.e. $s_\varepsilon \sim \rho(z_\varepsilon^j)$ in the g_ε scale, as $\varepsilon \to 0$. It follows that z_ε^j is preferred, which completes the proof in this case.

Suppose instead that for some $p_\varepsilon^j \in \hat{L}^j$,

$$\delta_a(p_\varepsilon^j) < \max(\delta_o, \delta_a(z_\varepsilon^{j-1})).$$

In this case, the analogue of (4.16) holds, with p_ε^j in place of z_ε^1 and one just continues the construction of the curve γ joining z_ε^{j-1} to $z_\varepsilon^j \equiv p_\varepsilon^j$. The point is that one must eventually come to points where (4.33) holds and the only way to do this is to come to a preferred base point, (unless x_ε is admissible).

■

Theorem 4.4 implies that there may be many admissible or preferred base points x_ε. In fact they are 1-dense in the region $S_\varepsilon \subset \Omega_\varepsilon$ defined by (4.4), i.e. $\bar{\varepsilon}^{2\mu} \leq |u-1| \leq 2/(\ln \bar{\varepsilon})^2$, and $\rho \leq \chi_\varepsilon$, where χ_ε is any sequence with $\chi_\varepsilon \to 0$ as $\varepsilon = \varepsilon_i \to 0$. It is obvious from the proof that such points are η-dense in S_ε, for any given $\eta > 0$. This lack of uniqueness plays no role however. Any level set L of u satisfying (4.4) and containing some admissible or preferred base point will be called an admissible level set and denoted by

$$L = L_o. \tag{4.34}$$

Let $\nu_{\delta_o} = (u-1)/\delta_o$, for δ_o as in (4.5), and let $m_o, \widetilde{m}_o$ be the mass of the potential ν_{δ_o}, as in (3.9) and (3.11) respectively. The following uniform lower bound on the mass is essentially immediate.

Proposition 4.7. *Let x_ε be any admissible or preferred base point. Then there is a constant $\kappa_o > 0$, independent of $\varepsilon \leq 1$, such that for any $p_\varepsilon \in L_o \cap B_{x_\varepsilon}(1)$, or $p_\varepsilon = x_\varepsilon$ respectively,*

$$\widetilde{m}_o(\rho(p_\varepsilon)) \geq \kappa_o \cdot \rho(p_\varepsilon). \tag{4.35}$$

In particular, if $(\Omega_\varepsilon, g_\varepsilon)$ is not μ_o collapsed at x_ε, i.e. $\omega(x_\varepsilon) \geq \mu_o \cdot \rho(x_\varepsilon)$, then

$$m_o(\rho(x_\varepsilon)) \geq \kappa_o \cdot \rho(x_\varepsilon), \tag{4.36}$$

where κ_o depends only on μ_o.

Proof: By construction, i.e. (4.5), $\delta_a(p_\varepsilon) \geq \delta_o$, and hence

$$|d\nu_{\delta_a(p_\varepsilon)}| \leq |d\nu_{\delta_o}|,$$

for all p_ε as above. Hence, the mass defined w.r.t. δ_o dominates the mass defined w.r.t. $\delta_a(p_\varepsilon)$. The estimates (4.35) and (4.36) are then an immediate consequence of (3.13).

■

We close this section with some applications of the proof of Theorem 4.4 to the structure of ∂F; these will be useful in §5.3 and §6. Let (F, g_o', x) be a maximal limit, (flat or hyperbolic), of $(\Omega_\varepsilon, g_\varepsilon', x_\varepsilon)$, with x_ε allowable, and let ν be the limit potential, $\nu = \lim \nu_{\delta_a}$, $\delta_a = \delta_a(x_\varepsilon)$. If $(\Omega_\varepsilon, g_\varepsilon)$ is μ_o-collapsed at x_ε, i.e. $\omega(x_\varepsilon) \leq \mu_o \cdot \rho(x_\varepsilon)$, it is assumed that the collapse is unwrapped in covering spaces, so that $\omega(x_\varepsilon) \sim \rho(x_\varepsilon)$. Define

$$I_\infty = \overline{\cap_N \{U^{-N}\}}, \tag{4.37}$$

where U^{-N} is the $-N$ superlevel set of ν, $N \in \mathbb{Z}^+$, and the closure is taken in $\bar{F}$. Let $d\mu_{-N}$ be the measure of ν on the level L^{-N}, as in (3.8), but restricted to $D_x(\infty)$, c.f. (3.19), i.e. the maximal domain on which the developing map is an embedding. Thus $d\mu_{-N}$ is supported on the level set $L_{-N} \cap D_x(\infty)$.

Let $d\mu_\infty$ be the weak limit of the measures $d\mu_{-N}$ in $\bar{F}$ as $N \to \infty$. The limit exists since the measures $d\mu_{-N}$ are related via the divergence theorem applied to the trace equation (2.65). The measure $d\mu_\infty$ is supported on ∂F.

Proposition 4.8. *Suppose the base points x_ε are allowable and satisfy $|u-1|(x_\varepsilon) \leq 2/(\ln \bar{\varepsilon})^2$, and let (F, g_o', x, ν) be the maximal limit as above. Then*

$$\partial F = I_\infty. \tag{4.38}$$

There is a constant m_o such that if $B(r)$ is any r-ball in $\bar{F}$ with center in ∂F, then $\forall r \leq 1$,

$$m_o \cdot r \leq m_\infty(B(r) \cap \partial F), \tag{4.39}$$

where m_∞ is the mass of $d\mu_\infty$. Further for any unit ball $B(1)$ in $\bar{F}$ centered at a point in ∂F,

$$m_\infty(B(1) \cap \partial F) \leq M_o, \tag{4.40}$$

for some constant $M_o < \infty$. The constants m_o and M_o are independent of the base point on ∂F. In particular, $dim_{\mathcal{H}} \partial F \leq 1$ and $\mathcal{H}_1(\partial F \cap B(1)) \leq M_o$, where $\mathcal{H}_1$ is the 1-dimensional Hausdorff measure of the metric space $(\bar{F}, g_o')$.

Proof: It is clear that $I_\infty \subset \partial F$ and so only the reverse inclusion in (4.38) needs to be proved. Via Proposition 2.1, if $y_\varepsilon \in U^{v_o}$ converges to $y \in \bar{F}$ (in the Hausdorff topology) then $\nu_{\delta_a}(y_\varepsilon) \to -\infty$ if and only if $y \in \partial F$. This does not immediately imply (4.38), (since ν_{δ_a} does not apriori converge to ν at ∂F). However, if for any $y \in \partial F$ there exist $y_\varepsilon \to y$ as above such that, for some $m_1 > 0$,

$$\widetilde{m}_{\delta_a}(\tfrac{1}{2}\rho(y_\varepsilon)) \geq m_1 \cdot \rho(y_\varepsilon), \tag{4.41}$$

then (4.38) does follow. Here $\widetilde{m}_{\delta_a}$ is the mass of ν_{δ_a}, $\delta_a = \delta_a(x_\varepsilon)$, on the level through y_ε, as in (3.11). Moreover (4.41) implies (4.39) provided the constant m_1 is independent of $y \in \partial F$. Now if

$$\delta_a(y_\varepsilon) \geq c_o \cdot \delta_a(x_\varepsilon), \tag{4.42}$$

for some $c_o > 0$, then $\widetilde{m}_{\delta_a(y_\varepsilon)} \leq c_o^{-1} \widetilde{m}_{\delta_a(x_\varepsilon)}$. But by (3.13), $\widetilde{m}_{\delta_a(y_\varepsilon)}(\frac{1}{2}\rho(y_\varepsilon)) \geq \rho(y_\varepsilon)$. Hence, (4.41) follows from (4.42).

The proof of (4.42) essentially follows from the proof of Lemma 4.5. Thus, consider paths $\gamma = \gamma_{y,\varepsilon}$ in $(\Omega_\varepsilon, g_\varepsilon)$, parametrized by arclength s, starting at a point $q_\varepsilon \in L^{v_o}$ with $q_\varepsilon \to y$, ending at x_ε, and satisfying the cone condition (4.10). If (4.42) were false along any such curve γ, then there is a constant $C_o < \infty$ such that, for all s sufficiently small and hence for all $s \leq S_o$, $S_o < \infty$,

$$\delta_a(\gamma(s)) \leq C_o \cdot \delta_a(x_\varepsilon).$$

Let L_ε be the u-level set through x_ε. As in (4.12)-(4.14), it then follows that

$$|u(L^{v_o}) - u(L_\varepsilon)| \leq C_1 \delta_a(x_\varepsilon) \cdot |\ln \bar{\varepsilon}|, \tag{4.43}$$

where C_1 depends only on C_o and S_o. However, in general one has the estimate

$$\delta_a(x_\varepsilon) \leq c_2 \cdot |u-1|(x_\varepsilon), \tag{4.44}$$

for generic x_ε in the sense of (2.78), for some uniform $c_2 < \infty$. To see this, (4.44) is equivalent to $|\nu_{\delta_a(x_\varepsilon)}|(x_\varepsilon) \geq c_2^{-1}$. If this estimate failed, i.e. $|\nu_{\delta_a(x_\varepsilon)}|(x_\varepsilon) << 1$, then the limit function ν_a would be 0 at generic base points, contradicting the fact that ν_a is non-trivial, c.f. Theorem 2.11.

Since, by hypothesis, $|u-1|(x_\varepsilon) \leq 2/|\ln \bar{\varepsilon}|^2$, (4.43) and (4.44) give $|u(L^{v_o}) - u(L)| \to 0$ as $\varepsilon \to 0$, a contradiction. This establishes (4.42) and hence (4.38)-(4.39), provided the point $y \in \partial F$ is accessible, in the sense that $y = \lim y_\varepsilon$, with y_ε joined to x_ε by a curve satisfying the cone condition (4.10). Let $\partial_a F \subset \partial F$ denote the subset of accessible points.

By the definition of the measure $d\mu_\infty$, the estimate (4.40) follows immediately from Proposition 3.3 for balls $B_y(1)$ with center $y \in \partial F$ and $dist_{g_o'}(y,x) \leq 2$, (for example). Together with the results above, this implies that $\partial_a F$ has Hausdorff dimension at most 1, with locally finite $\mathcal{H}_1$, within $\bar{B}_x(2)$, c.f. [13, Ch.4] for discussion of Hausdorff measures. In turn, this now implies that all points of ∂F within $\bar{B}_x(2)$ are accessible. Further, the arguments above may be applied at any other base point $x' \in F$, with $x'_\varepsilon \to x'$ and with $\delta_a(x_\varepsilon)$ replaced by $\delta_a(x'_\varepsilon)$. The limit of $\delta_a(x'_\varepsilon)/\delta_a(x_\varepsilon)$ as $\varepsilon \to 0$ varies continuously as x' varies over F, c.f. (the end of) Appendix B, and hence it follows that ∂F everywhere has Hausdorff dimension at most 1, with locally finite $\mathcal{H}_1$.

To see that (4.39) holds globally on F, i.e. m_o is independent of center point of $B(r)$, the local estimates above prove that (4.39)-(4.40) hold in any ball $B_{x'}(2)$, when δ'_a is used in place of δ_a to define the local mass m_∞, and with m_o, M_o then independent of x'. This means that ∂F is not too dense in F, and it follows that there exists $r_o = r_o(m_o, M_o) > 0$ such that any point $x' \in F$ with $dist(x', \partial F) \geq r_o$ may be joined to x via a path γ with $dist(\gamma(s), \partial F) \geq r_o$. Hence, since any $y \in F$ satisfies $y \in B_{x'}(2)$ for some x' as above, there also exist paths γ_ε from y_ε to x_ε satisfying the cone condition (4.10); the constant C in (4.10) of course depends on y, but c_o in (4.42) is independent of y, in the limit $\varepsilon \to 0$.

Finally, to prove that (4.40) holds globally, i.e M_o is independent of the base point, suppose x' is a base point in F such that $m_\infty(\partial F \cap B_{x'}(1)) >> 1$, where of course m_∞ is the mass w.r.t. the potential ν_{δ_a} based at x. This implies that $\delta_a(x'_\varepsilon)/\delta_a(x_\varepsilon) >> 1$ in the limit on F. Interchanging the roles of x and x', it follows that $m'_\infty(\partial F \cap B_x(1)) << 1$, where m'_∞ is the mass w.r.t. the potential $\nu_{\delta'_a}$ based at x'. This contradicts the fact that (4.39) holds globally, (at the base point x' in place of x), which gives (4.40). ∎

We point out that the assumption $|u-1|(x_\varepsilon) \leq 2/(\ln \bar{\varepsilon})^2$ was used only in (4.43)-(4.44). Thus, Proposition 4.8 remains valid under the assumption $\delta_a(x_\varepsilon) \leq 2/(\ln \bar{\varepsilon})^2$.

Recall that a subset of $\mathbb{R}^3$ is *polar* if it is contained in a countable collection of compact sets, each of which has capacity 0. These are precisely the sets where non-positive subharmonic functions in domains in $\mathbb{R}^3$ can become infinite, (i.e. $-\infty$), c.f. [12]. A polar set has Hausdorff dimension at most 1. For the discussion below, suppose that the maximal limit F is flat; the discussion holds equally well for hyperbolic limits, with obvious changes. (The following remarks are not actually needed elsewhere in the paper).

The boundary ∂F may be divided into two parts, a *regular* part $\partial_r F$ and a *singular* part, where $p \in \partial_r F$ if and only if $p \in \partial F$ and there is a neighborhood U of p in $\bar{F}$ such that U embeds in $\mathbb{R}^3$. Note that such an extension of F past $\partial_r F$ is unique. By Proposition 4.8, locally the regular boundary is contained in a closed set of finite $\mathcal{H}_1$ measure in $\mathbb{R}^3$ and is of course a polar set. In particular, if $\partial_s F = \emptyset$ and F is flat, then $\bar{F}$ is a complete, smooth flat manifold and hence

$$\bar{F} = \mathbb{R}^3/\Gamma, \tag{4.45}$$

where $\Gamma \subset Isom(\mathbb{R}^3)$ is a discrete group, maybe trivial, acting freely on $\mathbb{R}^3$.

The local structure of $\partial_s F$ is determined by the holonomy of the developing map $\mathcal{D} : \widetilde{F} \to \mathbb{R}^3$, where $\widetilde{F}$ is the universal cover of F. For simplicity, assume here that F is maximally extended as a smooth flat manifold so that $\partial F = \partial_s F$. For $q \in \partial F$ and r sufficiently small, let T_r denote the component of the r-tubular neighborhood of ∂F containing q, and let $\partial T_r \subset F$ denote its boundary. Also, let $S_q(r')$ denote the geodesic sphere about q, $r' \geq r$. For generic r, r', $S_q(r') \cap \partial T_r$ is a collection of embedded curves. Any such curve γ is either a loop, (linking ∂F), or a complete embedding of $\mathbb{R}$, i.e. γ has no endpoints. The former case occurs when $B_q(r')$ is compact, while the latter occurs when $B_q(r')$ is non-compact. Let $\widetilde{\gamma}$ be a lift of γ to $\widetilde{F}$, so that $\widetilde{\gamma}$ is an embedding of $\mathbb{R}$ in $\widetilde{F}$. Then $\mathcal{D}(\widetilde{\gamma})$ is a closed loop in $\mathbb{R}^3$, and $\mathcal{D}|_{\widetilde{\gamma}}$ acts as a covering projection $\mathbb{R} \to S^1$. The loop $\mathcal{D}(\widetilde{\gamma})$ determines an isometric $\mathbb{Z}$-action on $\mathbb{R}^3$.

Any isometric $\mathbb{Z}$-action on $\mathbb{R}^3$ is either rotational, i.e. rotation about an axis A, or a translation or a twist, i.e. a rotation about and translation along an axis A. In the latter two cases, the corresponding quotient spaces are the product $S^1 \times \mathbb{R}^2$ and twisted product $S^1 \times_\alpha \mathbb{R}^2$ respectively, where α is the rotation angle. These quotients are everywhere smooth and hence do not contribute $\partial_s F$. The only source of singularity to ∂F is that given by rotational holonomies. It follows that $\partial_s F$ is given locally by a collection of lines, corresponding to rotation axes. A neighborhood of each line in the closure $\bar{F}$ is a neighborhood of (0,0) in $\mathbb{R} \times C$, where C is a flat cone of cone angle α; the angle α may assume any value in $(0, \infty]$, the value $\alpha = \infty$ corresponding to the universal cover of $(\mathbb{R}^2 \setminus \{0\})$.

We close this section with the following useful result.

Lemma 4.9. *For F as in Proposition 4.8, let $t(x) = dist_{g_o'}(x, \partial F)$. Then t is unbounded on F.*

Proof: The proof is by contradiction, so suppose that $T_o = \sup_F t < \infty$. By Lemma 3.2, the limit potential ν_a is then bounded above on F, $\nu_a \leq C$, for some $C < \infty$. The potential ν_a satisfies the elliptic equation (2.65). Suppose first that $\Delta \nu_a = 0$, and consider a maximizing sequence $\{y_j\}$ for ν_a, i.e. $\nu_a(y_j) \to \sup_F \nu_a$. The maximum principle implies that ν_a approches a constant function near y_j, for j sufficiently large, which in turn implies that the mass of ν_a in $B_{y_j}(1)$ approaches 0. However, this contradicts the uniform lower bound (4.39) for the mass, again since $t(y_j) \leq T_o < \infty$.

If ν_a is not harmonic, then Theorem 2.11 implies that ν_a is superharmonic. Let h be the harmonic function on F determined by the Riesz measure $d\mu_\infty$ on ∂F. Since ν_a is superharmonic, $h \leq \nu_a \leq C$. Then exactly the same argument as above with h in place of ν_a, gives the same contradiction. ■

5. Mass Gaps and Essential Spheres.

This is the most important part of the paper, and culminates in Theorems 5.9 and 5.12, which prove the existence of essential 2-spheres in M under two mild hypotheses. The first of these is the degeneration hypothesis (4.1) and the second, (which has two parts), is a non-collapse hypothesis.

In §5.1, we describe the non-collapse hypothesis (first part), and derive some elementary consequences for the structure of the level set L_o of admissible base points. Next, the work in §5.2 proves that if x_ε is any admissible base point, then there exists a well-separated base point y_ε on the same level L_o, (naturally constructed from the initial x_ε), c.f. Theorem 5.4. This crucial result is obtained by examining the mass distribution on the level L_o near x_ε, using the global mass estimate from Theorem 3.5 in combination with the local mass bound from Proposition 4.7.

In §5.3, it is then shown that well-separated base points are naturally associated to 2-spheres in M, under the 2$^{\text{nd}}$ non-collapse hypothesis, c.f. Theorem 5.9. This leads to Theorem 5.12, proving that such 2-spheres are essential in M.

It is assumed throughout §5 that M is σ-tame. Thus, the sequence $(\Omega_\varepsilon, g_\varepsilon), \varepsilon = \varepsilon_i$, is chosen to satisfy (4.3).

5.1. Theorem 4.4 gives the existence of special base points on $(\Omega_\varepsilon, g_\varepsilon)$, namely admissible or preferred. In this subsection, we derive some elementary consequences for the geometry near such base points, in the presence of the following assumption.

Non-Collapse Assumption I. *There exists an admissible or preferred base point x_ε and a constant $\mu_o > 0$, independent of ε, such that for all $p_\varepsilon \in L_o \cap B_{x_\varepsilon}(1)$, or $p_\varepsilon = x_\varepsilon$ respectively,*

$$\frac{volB_{p_\varepsilon}(\rho(p_\varepsilon))}{\rho(p_\varepsilon)^3} \geq \mu_o. \tag{5.1}$$

where L_o is the level set through x_ε, as in (4.34).

The bound (5.1) means that $(\Omega_\varepsilon, g_\varepsilon)$ is not volume collapsing within ρ-balls centered at p_ε, i.e. the rescaled metrics $(\Omega_\varepsilon, g'_\varepsilon, p_\varepsilon)$, $g'_\varepsilon = \rho(p_\varepsilon)^{-2} \cdot g_\varepsilon$, have a uniform lower bound on the volume radius at p_ε, and so one has convergence to the limit (F, g'_o, p), (in a subsequence). It is clear that (5.1) implies that there exists $a_o = a_o(\mu_o) > 0$ such that, for $H_{p_\varepsilon}(r) = B_{p_\varepsilon}(r) \cap L_o$,

$$\frac{areaH_{p_\varepsilon}(\rho(p_\varepsilon))}{\rho(p_\varepsilon)^2} \geq a_o. \tag{5.2}$$

As always, (5.2) is understood to hold for generic points p_ε, as in (2.79), (although it follows from (5.5) below that it holds in fact for all p_ε). The ball $B_{x_\varepsilon}(1)$ may be replaced by $B_{x_\varepsilon}(\eta)$, for any fixed $\eta > 0$, throughout §5. In §6, we will deal separately with the possibility of collapse, where (5.1) does not hold, (for *all* admissible or preferred base points). The bound (5.1) is part of the assumptions of *all* of the results to follow in §5, but will not be mentioned explicitly each time.

Referring to Proposition 4.7, the following Corollary is now obvious.

Corollary 5.1. *Let x_ε be an admissible or preferred base point satisfying (5.1). There is a constant $\kappa_1 > 0$, depending only on κ_o and μ_o, and in particular independent of ε, such that for any $q_\varepsilon \in L_o \cap B_{x_\varepsilon}(1)$, respectively $q_\varepsilon = x_\varepsilon$,*

$$m_o(\rho(q_\varepsilon)) \geq \kappa_1 \cdot \rho(q_\varepsilon). \tag{5.3}$$

Proof: This is an immediate consequence of (4.35) and (5.1), c.f. (4.36). ∎

Remark 5.2. Observe that under the assumption (5.1), one has

$$\rho(q_\varepsilon) \leq \tfrac{1}{2}\kappa_1^{-1}\delta_o = \tfrac{1}{2}\kappa_1^{-1}\bar{\varepsilon}^{2\mu} \to 0, \quad \text{as} \quad \varepsilon \to 0, \tag{5.4}$$

for all $q_\varepsilon \in L_o \cap B_{x_\varepsilon}(1)$ or $q_\varepsilon = x_\varepsilon$ respectively. To see this, by (5.3), the mass of ν_{δ_o} in $B_{q_\varepsilon}(\rho(q_\varepsilon)) \cap L_o$ is at least $\kappa_1\rho(q_\varepsilon)$. On the other hand, Theorem 3.5 implies that the total mass m_o of L_o is bounded by $\frac{1}{2}\delta_o$, which gives (5.4). In particular, by (4.5), it follows that

$$\rho^2(q_\varepsilon) << \delta_a(q_\varepsilon), \tag{5.5}$$

corresponding to Case (i) of (2.12). In the opposite direction, Proposition 1.4 gives a uniform lower bound on ρ, i.e. $\rho^2 \geq \bar{\varepsilon}$ everywhere and so in particular on L_o.

Theorem 3.5 and (5.3) also combine to give the following result.

Corollary 5.3. *Let x_ε be an admissible or preferred base point and let C_o be any component of $L_o \cap B_{x_\varepsilon}(1)$, respectively $L_o \cap B_{x_\varepsilon}(\rho(x_\varepsilon))$. Then, $\forall p_\varepsilon \in C_o$, there exists $r_o > 0$ s.t.*

$$diam_{g_\varepsilon} C_o \geq r_o \cdot \rho(p_\varepsilon), \tag{5.6}$$

where diam is the extrinsic diameter in $(\Omega_\varepsilon, g_\varepsilon)$, and r_o is independent of C_o and ε. Hence

$$diam_{g_\varepsilon} C_o \geq r_1\bar{\varepsilon}^{1/2}. \tag{5.7}$$

Further,

$$diam_{g_\varepsilon} C_o \leq 5\kappa_1^{-1}\bar{\varepsilon}^{2\mu} \to 0, \quad \text{as} \quad \varepsilon \to 0. \tag{5.8}$$

Proof: The proof is the same in the admissible and well-separated cases, so assume x_ε is admissible. The estimate (5.6) follows exactly as in (2.79) from the uniform non-triviality of the limit potential ν_a. Note that (5.5) implies that the limit potential is harmonic, and so has no local maxima, (or minima). Thus (2.79) and (5.6) hold for all $p_\varepsilon \in C_o$, and not just generic p_ε in the sense of (2.78). The estimate (5.7) is an immediate consequence of (5.6) and Proposition 1.4.

To prove (5.8), consider the cover $\mathcal{U}$ of C_o by ρ-balls $B_z(\rho(z))$, $z \in C_o$. By a standard covering argument, c.f. [10, B33.3] or [13, §2.1], one may select a finite subcollection of disjoint balls $B_{z_j}(\rho(z_j))$ such that the balls $B_{z_j}(5\rho(z_j))$ cover C_o. Then

$$diam_{g_\varepsilon}(C_o) \leq 5\sum \rho(z_j) \leq 5\kappa_1^{-1}\sum m_o(B_{z_j}(\rho(z_j))) \leq 5\kappa_1^{-1}\bar{\varepsilon}^{2\mu}, \tag{5.9}$$

where the second estimate uses (5.3) and the last uses (3.22) and (4.6). ∎

Of course, combining (5.7) and (5.8) gives an upper bound, *depending on* ε, on the number of components of L_o in the ball $B_{x_\varepsilon}(1) \subset (\Omega_\varepsilon, g_\varepsilon)$.

It is worthwhile to keep in mind that Corollary 5.3 does not assert that $diam_{g_\varepsilon}C_o \sim \rho(q_\varepsilon)$, for some $q_\varepsilon \in C_o$. In fact, it may well be the case that for all $q_\varepsilon \in C_o$, $diam_{g_\varepsilon}C_o >> \rho(q_\varepsilon)$. Nevertheless, Corollary 5.3 does show that all the components of L_o in $B_{x_\varepsilon}(1)$ are being crushed (metrically) to points as $\varepsilon \to 0$ and not to more complicated spaces. However, the *distribution* of these points might be very complicated; for instance they could apriori become arbitrarily dense in $B_{x_\varepsilon}(1)$ as $\varepsilon \to 0$, compare with Remark 1.3.

5.2. We now come to perhaps the central technical result of the paper, proving the existence of gaps of a *definite relative size* (independent of ε) in the level set L_o near an admissible base point. This result implies that if x_ε is admissible, then there is a well-separated base point $y_\varepsilon \in L_o$ near to x_ε, c.f (4.7); in fact x_ε is well-separated w.r.t. all of Ω_ε, and not only w.r.t. U^{v_o}. This allows one to pass from admissible base points to well-separated base points.

Theorem 5.4. *For any admissible base point $x_\varepsilon \in L_o = L_o(\varepsilon)$ there exist points $y_\varepsilon \in B_{x_\varepsilon}(1) \cap L_o$ and scales $\lambda_\varepsilon = \lambda_\varepsilon(y_\varepsilon) \in (0, 1)$, $\lambda_\varepsilon \geq r_o \cdot \rho(y_\varepsilon)$, for a fixed $r_o > 0$, such that*

$$L_o \cap A_{y_\varepsilon}(\lambda_\varepsilon, (1+d_o)\lambda_\varepsilon) = \emptyset, \tag{5.10}$$

where $d_o = d_o(\mu) > 0$ is independent of ε and y_ε.

Proof: Given an admissible base point $x_\varepsilon \in L_o$, consider the unit ball $B = B_{x_\varepsilon}(1)$ w.r.t. g_ε. To each component C_j of L_o contained in B associate a subinterval $I_j = [a_j, b_j]$ of $I = [0,1]$ by setting

$$a_j = min\{r : C_j \cap B_{x_\varepsilon}(r) \neq \emptyset\}, \quad b_j = max\{r : C_j \cap B_{x_\varepsilon}(r) \neq \emptyset\}.$$

By (5.6) and (5.7), the length $l(I_j)$ satisfies

$$l(I_j) = b_j - a_j \geq r_o\rho(q_j) \geq r_1\bar{\varepsilon}^{1/2}, \quad \text{for some } \ q_j \in C_j. \tag{5.11}$$

Further, to each I_j associate a mass m by setting $m(I_j) \equiv m_o(C_j)$. Then the same estimates as in (5.9) give

$$m(I_j) \geq \kappa_2 l(I_j), \tag{5.12}$$

where $\kappa_2 = \kappa_2(\kappa_1)$. The intervals I_j may of course intersect, but this will only add further to the mass. Since the total mass of $L_o(1) \equiv L_o \cap B_{x_\varepsilon}(1)$ is small, i.e.

$$m(L_o(1)) \leq \bar{\varepsilon}^{2\mu}, \tag{5.13}$$

by (3.22), the total Lebesque measure of $I_o \equiv \cup I_j$ is very small in I. Hence there exist gaps between the intervals I_j. However, there may be very many, very small gaps. To estimate the gap size, we essentially consider I_o as obtained from I by the usual Cantor set construction of deleting open intervals from I.

Thus, consider I_o as obtained from I by inductively dividing intervals, starting at I, into two parts obtained by removing an open interval of *relative* length $\leq d_o$; the parameter d_o will be estimated below. Initially then, remove an open interval of length d_1 from I, so there are left two closed intervals, of total length $(1 - d_1)$. Remove from either or both of these again an interval of total relative length d_2, so that total length of removed intervals is then $d_1+d_2(1-d_1) = 1-(1-d_1)(1-d_2)$. Repeat this process inductively N times, until the collection of intervals I_o is reached.

A straightforward computation then shows that the total length of the removed intervals equals $1 - \prod_1^N(1 - d_i)$. It follows that the total length of the remaining closed intervals, and hence the total mass of I_o by (5.12), satisfies

$$m(I_o) \geq \kappa_2 \prod_1^N (1 - d_i). \tag{5.14}$$

Let $\bar{d} = \sup_i d_i$, so that (5.14) gives

$$m(I_o) \geq \kappa_2(1 - \bar{d})^N.$$

Then from (5.13) one obtains the bound

$$(1 - \bar{d})^N \leq \kappa_2^{-1}\bar{\varepsilon}^{2\mu}. \tag{5.15}$$

On the other hand, since each repetition of the process above produces a new closed interval of length $< \frac{1}{2}$ that of its predecessor, some of the final closed intervals I_j in I_o must have length $\leq 2^{-N}$. But by (5.11), the length of each interval I_j is at least $\bar{\varepsilon}$, (in fact $\bar{\varepsilon}^{1/2}$). Hence, one also has

$$\bar{\varepsilon} \leq 2^{-N}. \tag{5.16}$$

An elementary calculation combining (5.15) and (5.16) then gives

$$\bar{d} > 1 - \kappa_2^{-1/N}2^{-2\mu}.$$

This implies that there are gaps in I_o at *some* scale, of relative size at least

$$d_o = 1 - \kappa_2^{-1}2^{-2\mu}. \tag{5.17}$$

By construction, this means that there exist points $y_\varepsilon \in L_o \cap B_{x_\varepsilon}(1)$ and numbers $\lambda_\varepsilon \in (0, 1]$ such that (5.10) holds. The fact that $\lambda_\varepsilon \geq r_o \cdot \rho(y_\varepsilon)$ follows from (5.6).

■

Remark 5.5. It is clear from the proof of Theorem 5.4 that there could be many possible choices of base points y_ε and factors λ_ε satisfying (5.10). For the purposes of this paper, this lack of uniqueness, already occuring in §4, plays no fundamental role.

However, it is not asserted that there exist gaps forming as in (5.10) at some definite or determinable scale, (although see Lemma 5.6). For instance, it is not claimed that one can choose λ_ε to be bounded away from 0 as $\varepsilon \to 0$, i.e. that there are gaps forming in L_o of a definite size w.r.t. g_ε, compare with Remark 1.3. At the other, small-scale extreme, it is not known that gaps form on the scale of $\rho(y_\varepsilon)$, i.e. whether $\lambda_\varepsilon \sim \rho(y_\varepsilon)$, as is the case for preferred base points.

It is the existence of such gaps provided by Theorem 5.4 which is important, and not so much the scale at which they are visible. Without the σ-tame assumption in Theorem 5.4 the size of all relative gaps could approach 0 as $\varepsilon \to 0$, as in the construction of Cantor sets of measure 0 in $[0, 1]$ with Hausdorff dimension 1, c.f. [13, §4.10]. This is in fact the only place in the paper where the σ-tame assumption is used in an essential way; (the arguments in §4 and in §5 preceding Theorem 5.4 can be modified to hold without the σ-tame assumption).

The following (more technical) observation based on the proof of Theorem 5.4 will be used later in the final choice of base points.

Lemma 5.6. *Let x_ε be any admissible base point. Then there exist well-separated base points $y_\varepsilon \in B_{x_\varepsilon}(1) \cap L_o$ satisfying (5.10), together with*

$$\lambda_\varepsilon \leq \bar{\varepsilon}^{3\mu/2}, \tag{5.18}$$

and such that

$$m_o(B_{y_\varepsilon}(\lambda_\varepsilon)) \geq \bar{\varepsilon}^{\mu/2}\lambda_\varepsilon. \tag{5.19}$$

Proof: To prove (5.18), observe that the proof of Theorem 5.4 began with the ball $B_{x_\varepsilon}(1)$. However, it can be applied equally well to balls $B_{x_\varepsilon}(s_\varepsilon)$, with $s_\varepsilon \to 0$ sufficiently slowly as $\varepsilon \to 0$. For example, let $s_\varepsilon = \bar{\varepsilon}^{3\mu/2}$ and rescale the balls $B_{x_\varepsilon}(s_\varepsilon)$ to size 1, i.e. set $\bar{g}_\varepsilon = s_\varepsilon^{-2} \cdot g_\varepsilon$. Since the mass scales as a distance, $\bar{m}_o((B_{x_\varepsilon}(1)) \leq \bar{\varepsilon}^{-3\mu/2}\bar{\varepsilon}^{2\mu} \leq \bar{\varepsilon}^{\mu/2}$, where $\bar{m}_o$ is the mass m_o w.r.t the metric $\bar{g}_\varepsilon$. Now starting with (5.13), with $\bar{\varepsilon}^{\mu/2}$ in place of $\bar{\varepsilon}^{2\mu}$, the same arguments as before in Theorem 5.4 apply to give (5.10), with λ_ε satisfying (5.18). (The estimate for $1 - d_o$ is smaller than that in (5.17) by a factor of 2).

The proof of (5.19) is similar. Thus, let $A_{y_\varepsilon}(\lambda_\varepsilon)$ be an annulus satisfying (5.10) and (5.18) and consider the rescaled metrics $\bar{g}_\varepsilon = \lambda_\varepsilon^{-2} \cdot g_\varepsilon$. If there is any fixed constant $\kappa > 0$, say $\kappa = \mu/2$, s.t.

$$\bar{m}_o(\bar{B}_{y_\varepsilon}(1)) \leq \bar{\varepsilon}^{\kappa}, \tag{5.20}$$

then exactly the same arguments as in the proof of (5.18) above may be repeated at this scale to produce a new smaller annulus $A_{y_\varepsilon^1}(\lambda_\varepsilon^1) \subset B_{y_\varepsilon}(\lambda_\varepsilon), \lambda_\varepsilon^1 << \lambda_\varepsilon$, satisfying the gap condition (5.10), with λ_ε^1 in place of λ_ε; as above, here $d_o = d_o(\kappa) = d_o(\mu)$. In this new scale $g_\varepsilon^1 = (\lambda_\varepsilon^1)^2 g_\varepsilon$, the mass is of course much larger. Hence, by iterating this procedure a finite number of times, (depending on ε), one obtains base points y_ε and scale-factors λ_ε satisfying (5.10) and (5.18)-(5.19). ■

Theorem 5.4 and Lemma 5.6 prove the existence of well-separated base points, (of course under the assumptions (4.1), (5.1) and the σ-tame assumption), satisfying the properties (5.10) and (5.18)-(5.19). Preferred base points automatically satisfy (5.18)-(5.19); for such base points, since $\lambda \sim \rho$, (5.18) follows from (5.4) while (5.19) follows from (5.3). From now on, it is assumed that the well-separated base point y_ε has been chosen to satisfy (5.18)-(5.19). One further observation is needed for the final determination of the base point.

The level L_o disconnects the manifold Ω_ε into two parts,

$$\Omega_\varepsilon = U_o \cup U^o, \tag{5.21}$$

where $U_o = \{x \in \Omega_\varepsilon : u(x) < u(L_o)\}$, $U^o = \{x \in \Omega_\varepsilon : u(x) > u(L_o)\}$; of course neither of these sets is necessarily connected. Hence each component $C_{y_\varepsilon}(\lambda_\varepsilon)$ of any annulus $A_{y_\varepsilon}(\lambda_\varepsilon, (1+d_o)\lambda_\varepsilon)$ as in Theorem 5.4 is contained either in U_o or U^o.

The next result asserts the existence of such components contained in U^o.

Lemma 5.7. *Let $A_{y_\varepsilon}(\lambda_\varepsilon, (1+d_o)\lambda_\varepsilon))$ be any annulus satisfying (5.10), with λ_ε satisfying (5.18). Then there is a component $\mathcal{C}_\varepsilon = \mathcal{C}_{y_\varepsilon}(\lambda_\varepsilon)$ of $A_{y_\varepsilon}(\lambda_\varepsilon, (1+d_o)\lambda_\varepsilon)$ such that*

$$\mathcal{C}_\varepsilon \subset U^o. \tag{5.22}$$

Proof: Let $s_\varepsilon = (1 + \frac{d_o}{2})\lambda_\varepsilon$. By (5.10), the ball $B(s_\varepsilon) = B_{y_\varepsilon}(s_\varepsilon)$ satisfies $\partial B(s_\varepsilon) \cap L_o = \emptyset$, so that any component of $\partial B(s_\varepsilon)$ is then contained either in U_o or U^o. It suffices then to prove that some component of $\partial B(s_\varepsilon)$ is contained in U^o.

To see this, suppose instead the full boundary $\partial B(s_\varepsilon) \subset U_o$. Then some part of the level set $L_o \cap B(s_\varepsilon)$ bounds a connected domain $D = D_\varepsilon \subset U^o$, with $D \subset\subset B(s_\varepsilon), \partial D \subset L_o$. Recall that by (5.4), $\rho \leq \bar{\varepsilon}^{2\mu}$ on $L_o \cap B_{y_\varepsilon}(\frac{1}{2})$, (ignoring constant factors here). Hence, by the Lipschitz property of $\rho, \rho \leq \bar{\varepsilon}^{3\mu/2}$ everywhere in $B(s_\varepsilon)$, and so, at all base points in D,

$$\rho^2/\delta_o \to 0. \tag{5.23}$$

Choose a point $z_\varepsilon \in D$ which realizes the maximal value of u on D, and work in the scale g'_ε where $\rho'(z_\varepsilon) = 1$. Since $\rho \sim dist(L^{v_o}, \cdot)$ and L^{v_o} lies outside D, it follows that $dist_{g'_\varepsilon}(z_\varepsilon, L_o) \leq 2$ for ε sufficiently small. Consider the potential function $\nu_{\delta_o} = (u-1)/\delta_o$. If the mass m_o of ν_{δ_o} w.r.t. g'_ε is uniformly bounded over $B'_{z_\varepsilon}(1/2)$, then ν_{δ_o} subconverges, modulo constants, to a limit ν_o. If the mass m_o is very large, then renormalize it by dividing by $m_o(B'_{z_\varepsilon}(1/2))$. (This has effect of increasing δ). In both cases, the estimate (5.23) shows that Case (i) of (2.12) holds and so Proposition 2.3 implies that the limit function ν_o satisfies $\Delta\nu_o = 0$. Since, ν_o has a maximal value at $z = \lim z_\varepsilon$, it follows that ν_o is constant.

However, the estimate (4.35) on L_o passes to the limit, and shows that the limit potential ν_o has non-zero mass on L_o. This gives a contradiction. ■

Lemma 5.7 applies equally well to preferred base points; such base points are only well-separated w.r.t. U^{v_o}, but Lemma 5.7 implies that there is a full component $\mathcal{C}_\varepsilon$ of the annulus A^{v_o} contained in U^o and hence $\mathcal{C}_\varepsilon \cap L^{v_o} = \emptyset$.

5.3. We are now ready to to fix the choice of the base point and prove, under a 2^{nd} natural non-collapse assumption, that such base points give rise to geometrically natural 2-spheres in M.

Definition 5.8. *A* distinguished *base point is a base point $q_\varepsilon \in \mathcal{C}_\varepsilon$ as in (5.22), with $q_\varepsilon \in A_{y_\varepsilon}((1+\frac{1}{4}d_o)\lambda_\varepsilon, (1+\frac{3}{4}d_o)\lambda_\varepsilon)$, where y_ε, λ_ε satisfy the conclusions of Lemma 5.6.*

For the remainder of §5, we work with a choice of distinguished base point. Again there are apriori many possible choices here, and the results to follow in §5.3 apply to any such choice. To any distinguished base point q_ε is associated the (center) base point y_ε and component $\mathcal{C}_\varepsilon \subset U^o$ of $A_{y_\varepsilon}(\lambda_\varepsilon, (1+d_o)\lambda_\varepsilon)$, with $q_\varepsilon \in \mathcal{C}_\varepsilon$.

Consider the rescaled metric

$$\bar{g}_\varepsilon = \lambda_\varepsilon^{-2} \cdot g_\varepsilon, \tag{5.24}$$

so that $A_\varepsilon(\lambda_\varepsilon) = A_{y_\varepsilon}(\lambda_\varepsilon, (1+d_o)\lambda_\varepsilon)$ now becomes $\bar{A}_\varepsilon(1) = \bar{A}_{y_\varepsilon}(1, 1+d_o)$ and $B_\varepsilon(\lambda_\varepsilon) = B_{y_\varepsilon}(\lambda_\varepsilon)$ becomes $\bar{B}_\varepsilon(1) = \bar{B}_{y_\varepsilon}(1)$. For $\mathcal{C}_\varepsilon \subset \bar{A}_\varepsilon(1)$ as above satisfying (5.22), Proposition 1.1 implies

$$\bar{\rho}(p_\varepsilon) \geq d_o/4, \tag{5.25}$$

for all $p_\varepsilon \in \bar{A}_\varepsilon(1+\frac{1}{4}d_o, 1+\frac{3}{4}d_o) \cap \mathcal{C}_\varepsilon$. On the other hand, since $\lambda \geq r_o \cdot \rho(y_\varepsilon)$,

$$\bar{\rho}(y_\varepsilon) \leq r_o^{-1},$$

and hence by the Lipschitz property of ρ,

$$\bar{\rho}(q_\varepsilon) \leq 2r_o^{-1}.$$

Thus, the rescaling (5.24) is uniformly homothetic to the natural blow-up metric based at q_ε, i.e.

$$g'_\varepsilon = \rho(q_\varepsilon)^{-2} \cdot g_\varepsilon. \tag{5.26}$$

We now make the following 2^{nd} non-collapse assumption.

Non-Collapse Assumption II. *For some distinguished base point q_ε, there exists $\mu_o > 0$ and some sequence $K_\varepsilon \to \infty$, (slowly) as $\varepsilon \to 0$, such that, for all $1 \leq K \leq K_\varepsilon$*

$$\frac{B_{q_\varepsilon}(K\rho(q_\varepsilon))}{(K\rho(q_\varepsilon))^3} \geq \mu_o. \tag{5.27}$$

The condition (5.27) means that the rescalings (5.26) based at q_ε do not collapse on large scales in the metric (5.26), i.e. $\omega'(q_\varepsilon) >> 1$, as $\varepsilon \to 0$. Although similar, the assumption NCA II does not quite follow from NCA I, since $q_\varepsilon \notin L_o$ and the collapse is on scales large compared with $\rho(q_\varepsilon)$. However, the assumption NCA II is much weaker than NCA I, since it applies to the behavior at a single base point q_ε, while NCA I applies to all $p_\varepsilon \in B_{x_\varepsilon}(1) \cap L_o$.

Under the NCA II, the pointed sequence $(\Omega_\varepsilon, g'_\varepsilon, q_\varepsilon)$ (sub)-converges to a maximal flat limit (F, g'_o, q). Similarly, the rescalings $(\Omega_\varepsilon, \bar{g}_\varepsilon, q_\varepsilon)$ (sub)-converge to a maximal flat limit $(F, \bar{g}_o, q)$, and

$(F,\bar g_o)$ is homothetic to (F,g_o'). Recall that the convergence of the blow-ups $(\Omega_\varepsilon, g_\varepsilon', q_\varepsilon)$ to the limit (F,g_o',q) is smooth away from ∂F and that ∂F is formed by the limits of points z_ε where the curvature is concentrating at a higher rate, i.e. $\rho'(z_\varepsilon)\to 0$, c.f. the discussion preceding (1.23). Of course $\partial F \neq \emptyset$ and by Proposition 2.1, ∂F is the q_ε-based Hausdorff limit of the level L^{v_o}; the same holds w.r.t. $(F,\bar g_o)$. The center points y_ε converge to a point $y\in \bar F$ with $q\in \bar B_y(1+d_o)$, while the domains $\mathcal{C}_\varepsilon$ converge to the domain $\mathcal{C}_o \subset \bar A_y(1)$.

The results above now combine easily to give the existence of natural 2-spheres.

Theorem 5.9. *Let q_ε be a distinguished base point of $(\Omega_\varepsilon, g_\varepsilon)$ satisfying the non-collapse assumption NCA II and let (F,g_o',q) be a maximal flat blow-up limit constructed above. Then there is a non-empty compact subset Σ_o of ∂F, such that either $\partial F\setminus\Sigma_o=\emptyset$ or*

$$dist_{g_o'}(\Sigma_o,\partial F\setminus \Sigma_o)\geq d_o>0. \tag{5.28}$$

Further, $\mathcal{C}_o$ is embedded in $\mathbb{R}^3$,

$$\mathcal{C}_o\subset\mathbb{R}^3, \tag{5.29}$$

and $\mathcal{C}_o$ is isometric to a spherical annulus $S^2\times I$ about a point in $\mathbb{R}^3$.

Proof: For convenience we work with the metric $(F,\bar g_o)$ in place of (F,g_o'); as noted above, these metrics are homothetic. Returning to the sequence $(\Omega_\varepsilon,\bar g_\varepsilon)$ as in (5.24), suppose first that the annuli $\bar A_\varepsilon(1,1+d_o)$ are connected, (in a subsequence); more generally, one may suppose that $\bar A_\varepsilon(1,1+d_o)\subset U^o$. Then we claim that $L^{v_o}\cap \bar B_\varepsilon(1)\neq\emptyset$, and so, (by Proposition 1.1 as usual),

$$\bar\rho\to 0, \tag{5.30}$$

somewhere within $\bar B_\varepsilon(1)$. For if (5.30) does not hold, then L^{v_o} lies outside $\bar B_\varepsilon(1)$, and $\bar g_\varepsilon$ converges smoothly on $\bar B_\varepsilon(1)$ to the limit $\bar B_y(1)$. Since $\bar A_\varepsilon(1,1+d_o)\subset U^o, L_o\subset \bar B_\varepsilon(1)$ bounds a domain D_ε in $\bar B_\varepsilon(1)$ and hence ν has a minimal value in D_ε. This gives the same contradiction as in Lemma 5.7 above.

Thus, a singularity forms at a point $z_\varepsilon\in \bar B_\varepsilon(1)$, i.e. there are points z_ε within $\bar B_\varepsilon(1)$ of much higher curvature concentration than that within $\mathcal{C}_\varepsilon$. In particular, there is a non-empty component of ∂F within the limit $\bar B_y(1)$ of $\bar B_\varepsilon(1)$. As noted above, $\bar\rho$ is bounded away from 0 in $\mathcal{C}_\varepsilon$, so that

$$\mathcal{C}_o\cap\partial F=\emptyset. \tag{5.31}$$

Suppose instead $L^{v_o}\cap \bar B_\varepsilon(1)=\emptyset$, so that the balls $\bar B_\varepsilon(1)$ converge smoothly to the limit $\bar B_y(1)$ and surrounding annulus $\bar A_y(1,1+d_o)$ in F. This limit annulus is not connected, since the annuli $\bar A_\varepsilon(1,1+d_o)$ are not connected. Hence, the limit $\bar B_y(1)$ is a geodesic ball in a smooth flat manifold, with disconnected boundary. This can only occur if $\bar B_y(1)$ is isometric to a ball in a flat product $F_{prod}=\mathbb{R}\times S^1\times S^1$. Now by Lemma 4.9, the distance function t to ∂F, (w.r.t. $\bar g_o$), is unbounded on F, while by (2.19), $t=\rho$ on F. Hence, one may choose a sequence of points y_i with $t(y_i)\to\infty$, approximate them by points $y_{\varepsilon_i}\in(\Omega_{\varepsilon_i},g_{\varepsilon_i})$, and consider the rescalings based at y_{ε_i}. This has the effect of blowing down F. However, the flat product structure of $(F,\bar g_o)$ on $\bar B_y(1)$ extends to all of $(F,\bar g_o)$ and thus it follows that the rescalings based at y_{ε_i} are collapsing. This contradicts the NCA II, (5.27). Actually, to apply Lemma 4.9, one needs to know that $|u-1|(q_\varepsilon)\geq\bar\varepsilon^{4\mu}$. By (4.44), it suffices to show that $\delta_a(q_\varepsilon)\geq\bar\varepsilon^{4\mu}$, which follows from (5.41) below; (the proof of (5.41) is independent of Theorem 5.9).

Thus $\Sigma_o=\partial F\cap\bar B_y(1)$ is non-empty and the estimate (5.28), (w.r.t. $\bar g_o$), follows as in (5.25). The domain $\mathcal{C}_o$ from (5.31) is a connected geodesic annulus about the point $y\in\bar F$, separating Σ_o from the rest of ∂F.

Suppose now that $\bar B_y(1+d_o)$ is simply connected. Then the developing map $\mathcal{D}$ is an isometric immersion of $\bar B_y(1+d_o)$ into the ball $B_0(1+d_o)$ about 0 in $\mathbb{R}^3$. To prove (5.29) and the fact that $\mathcal{C}_o$ is a spherical annulus, it suffices to prove that $\mathcal{D}$ is an embedding, or equivalently $\mathcal{D}$ has no holonomy on $\bar B_y(1+d_o)$. As in the discussion following Proposition 4.8, each holonomy transformation is either a rotation, twist or translation. Now $\mathcal{D}$ can have no rotational holonomy, generated by rotation about an axis within $\bar{B}_y(1+d_o)$. For such an axis, necessarily a line, must be contained in ∂F, which is impossible by the gap estimate (5.28). Hence the only holonomy transformations are those given by twists or translations. These both introduce non-trivial loops, and thus $\bar{B}_y(1+d_o)$ could not be simply connected.

On the other hand, if $\bar{B}_y(1+d_o)$ is not simply connected, then the arguments above imply that the developing map $\mathcal{D}$ maps F into a non-trivial quotient $\mathbb{R}^3/\Gamma$, for some Γ acting freely and isometrically on $\mathbb{R}^3$; hence $\Gamma = \mathbb{Z}$ or $\mathbb{Z}\oplus\mathbb{Z}$ acts by twists and/or translations. Via Lemma 4.9 as above, this contradicts the NCA II. This completes the proof. ■

Since the convergence of $(\Omega_\varepsilon, g'_\varepsilon, q_\varepsilon)$ to the limit (F, g'_o, q) is smooth away from ∂F, there are 2-spheres $S^2 = S^2_\varepsilon$ embedded in $(\Omega_\varepsilon, g'_\varepsilon)$ converging smoothly to the limit $S^2 \subset \mathcal{C}_o \subset (F, g'_o)$. The metric g'_ε is just a rescaling of g_ε, and so the metrics g_ε are crushing the 2-spheres S^2_ε to *points.* Of course, since Ω_ε is weakly embedded in M, the spheres S^2_ε are also embedded in M, i.e.

$$S^2_\varepsilon \subset M. \tag{5.32}$$

It follows from Theorem 5.9 that the 2-sphere $S^2_\varepsilon \subset \Omega_\varepsilon$, although it converges to a point as $\varepsilon \to 0$, surrounds the formation of a singularity, in that "within" S^2_ε there are regions of much higher curvature concentration than the curvature concentration on S^2_ε itself.

5.4. In this subsection, we prove the central result of the paper, Theorem 5.12, stating that the 2-spheres S^2_ε constructed above in (5.32) are essential 2-spheres in M.

To begin, one needs to understand the behavior of the potential functions ν_δ on the blow-ups g'_ε and the limit F in Theorem 5.9. To do this, recall that the spheres S^2_ε are contained in the connected region $\mathcal{C}_\varepsilon = A_\varepsilon(\lambda_\varepsilon) = A_{y_\varepsilon}(\lambda_\varepsilon, (1+d_o)\lambda_\varepsilon)$, where q_ε is a distinguished base point; the annulus is connected by the proof of Theorem 5.9. In particular, (5.18)-(5.19) hold. Let m'_o be the mass of the potential $\nu_{\delta_o} = (u-1)/\delta_o, \delta_o = \bar{\varepsilon}^{2\mu}$, on the level set L_o containing y_ε, w.r.t. g'_ε in (5.26). Also let $B'_\varepsilon = B'_{y_\varepsilon}(\lambda'_\varepsilon)$, where $\lambda'_\varepsilon = \lambda_\varepsilon \cdot \rho(q_\varepsilon)^{-1}$. Since the metrics $\bar{g}_\varepsilon$ and g'_ε in (5.24) and (5.26) are uniformly homothetic, $\lambda'_\varepsilon \to \lambda'_o \in (0,\infty)$, as $\varepsilon \to 0$.

Now there are three possibilities for the behavior of $m'_o(B'_\varepsilon)$, as $\varepsilon \to 0$, namely

$$\begin{aligned} &(i) \quad m'_o(B'_\varepsilon) \to \infty, \\ &(ii) \quad n_o \le m'_o(B'_\varepsilon) \le n_o^{-1}, \\ &(iii) \quad m'_o(B'_\varepsilon) \to 0, \end{aligned} \tag{5.33}$$

where n_o is some positive constant. Observe that in case (iii), (5.19) implies $m'_o(B'_\varepsilon) \ge \bar{\varepsilon}^{\mu/2}$. On the other hand, one has the natural potential function w.r.t. the base point q_ε, i.e. as in (2.17),

$$\nu_{\delta(q_\varepsilon)} = (u-1)/\delta(q_\varepsilon), \tag{5.34}$$

and its associated mass function m_{q_ε} on L_o, again in the g_ε scale. It is important to note that here we are using the 'curvature' δ as in (2.10), and not δ_a from (3.1), used throughout §4 and so far in §5. Of course, for any fixed ε, the potentials ν_{δ_o} and $\nu_{\delta(q_\varepsilon)}$ are proportional, as are the mass functions.

We claim first that the two potentials and mass functions above have a uniformly bounded proportionality constant, when (5.33)(ii) holds.

Lemma 5.10. *If (5.33)(ii) holds, then there is a constant $n_1 = n_1(n_o) > 0$, independent of ε, s.t.*

$$n_1\delta_o \le \delta(q_\varepsilon) \le n_1^{-1}\delta_o. \tag{5.35}$$

Proof: By assumption, the mass m'_o of ν_{δ_o} in B'_ε is on the order of 1. On the other hand, the mass m_{δ_a} of $\nu_{\delta_a}, \delta_a = \delta_a(q_\varepsilon)$, is also on the order of 1 in B'_ε; this follows from (3.13), the NCA II in (5.27), together with Proposition 3.3. Hence, from the definitions of the masses, it follows that

$$\delta_a \sim \delta_o, \tag{5.36}$$

i.e. the ratio is uniformly bounded away from 0 and ∞ as $\varepsilon \to 0$. Now recall from (3.2)-(3.3) that $\delta_a \geq c_1 \cdot \delta$ and $\delta_a >> \delta$ if and only if the limit ν_a of ν_{δ_a} is an *affine* function; here of course $\delta = \delta(q_\varepsilon)$. However, an affine function has no gap in the support of the (limit) measure $d\mu$, or equivalently no mass gap. Hence by (5.28) one must have $\delta_a \leq c_1 \cdot \delta$, so that

$$\delta_a \sim \delta. \tag{5.37}$$

Combining the estimates (5.36)-(5.37) gives (5.35). ■

Observe that Lemma 5.10 implies that

$$m_{\delta(q_\varepsilon)}(B'_\varepsilon) \sim 1, \quad \text{as} \quad \varepsilon \to 0, \tag{5.38}$$

so that the part of the level L_o in B'_ε has a definite amount of mass for the potential $\nu_{\delta(q_\varepsilon)}$. Since one has convergence to the limit F, and not collapse, observe that (5.5) holds here also, for the same reasons. Thus,

$$\rho^2(q_\varepsilon) << \delta(q_\varepsilon), \tag{5.39}$$

so that the base points q_ε satisfy Case (i) of (2.12).

In the cases (5.33)(i) or (iii), one needs to renormalize the mass by setting

$$\delta_1 = \delta_o \cdot m'_o(B'_\varepsilon), \tag{5.40}$$

so that the mass of the potential $\nu_{\delta_1} = (u-1)/\delta_1$ in B'_ε is on the order of 1. By the same reasoning as in Lemma 5.10, it then follows that $\delta_1 \sim \delta_a(q_\varepsilon) \sim \delta(q_\varepsilon)$. We claim that in both cases (5.39) still holds. In the situation of (5.33)(i), this is obvious, since then $\delta(q_\varepsilon) \geq \delta_o$. Suppose instead (5.33)(iii) holds. Then applying the estimate (5.19) to (5.40) gives

$$\delta_1 \geq \bar{\varepsilon}^{2\mu+\mu/2}. \tag{5.41}$$

On the other hand, on $L_o \cap B_{y_\varepsilon}(\frac{1}{2})$, in the g_ε scale, one has $\rho \leq \bar{\varepsilon}^{2\mu}$ by (5.4). Since $q_\varepsilon \in B_{y_\varepsilon}((1+d_o)\lambda_\varepsilon)$ and $\lambda_\varepsilon \leq \bar{\varepsilon}^{3\mu/2}$ by (5.18), the Lipschitz property of ρ, c.f. §1, thus gives

$$\rho(q_\varepsilon) \leq \bar{\varepsilon}^{3\mu/2}.$$

Hence, for $\rho = \rho(q_\varepsilon)$,

$$\frac{\rho^2}{\delta_1} \leq \frac{\bar{\varepsilon}^{3\mu}}{\bar{\varepsilon}^{2\mu+\mu/2}} \to 0. \tag{5.42}$$

These arguments thus imply that the distinguished base points q_ε necessarily satisfy Case (i) of (2.12), i.e. satisfy the conclusions of Proposition 2.3. Further, (5.37) holds, so that the potentials ν_δ converge to the limit potential ν modulo constants. In sum, one has:

Corollary 5.11. *For q_ε a distinguished base point as in Theorem 5.9, the linearized equations of the blow-up limit (F, g'_o, ν, q) are the linearized static vacuum Einstein equations*

$$r' = D^2\nu, \quad \Delta\nu = 0. \tag{5.43}$$

For ε sufficiently small, and within $B_y((1+d_o)\lambda'_o) \cap F$, $\lambda'_o = \lim \lambda'_\varepsilon \rho(q_\varepsilon)^{-1}$, the metric g'_ε has the form,

$$g'_\varepsilon = (1 - 2\nu\delta)g'_o + o(\delta), \tag{5.44}$$

modulo diffeomorphisms, where g'_o is the flat metric on F. The potential function ν is a non-constant harmonic function and is the limit of $\nu_{\delta(q_\varepsilon)}$, mod constants.

■

We are now in position to prove the main result of the paper; this result constitutes the most important part in the proof of the Main Theorem of §0.

Theorem 5.12. *Let $(\Omega_\varepsilon, g_\varepsilon)$ be a sequence of minimizing pairs for M, with $\varepsilon = \varepsilon_i \to 0$ satisfying (1.36) and suppose the degeneration hypothesis (4.1) and non-collapse assumptions NCA I, NCA II hold on some subsequence.*

Then, for ε sufficiently small, the 2-sphere S^2_ε in M given by (5.32) is essential in M. In particular, M is a reducible 3-manifold.

Proof: Suppose that $S^2 = S^2_\varepsilon$ is inessential in M, i.e. S^2 bounds a 3-ball in M. By constructing a suitable metric glueing of a 3-ball onto S^2 in M as a comparison metric, we will obtain a contradiction to the fact that g_ε is a minimizer of ${I_\varepsilon}^-$, for ε sufficiently small.

Given the preceding results of §5, most all of the work to prove Theorem 5.12 has already been done, explicitly for this purpose, in [1]. First, as in [1, Thm.3.1], the proof needs to be divided into two cases, according to whether or not S^2 bounds a 3-ball B^3 on the inside or outside. The 2-sphere S^2 is said to bound B^3 on the *inside* if the component B^3 of $M \setminus S^2$ contains the center base point y_ε associated to the distinguished base point q_ε; otherwise S^2 bounds on the *outside.*

Since $\mathcal{C}_\varepsilon \subset U^o$ by (5.22), one has $\nu_\delta > \nu_\delta(L_o)$ on $\mathcal{C}_\varepsilon$, $\delta = \delta(q_\varepsilon)$, while Theorem 5.9 implies that $L_o \cap B'_{y_\varepsilon}(\lambda'_\varepsilon) \neq \emptyset$ and $L^{v_o} \cap B'_{y_\varepsilon}(\lambda'_\varepsilon) \neq \emptyset$. Since $\lambda'_\varepsilon \to \lambda'_o \in (0, \infty)$, the factor λ'_ε will be ignored, (i.e. set to 1), in the following; (this is the same as working with $\bar{g}_\varepsilon$ in place of g'_ε). By Corollary 5.11, the potential functions ν_δ converge, modulo constants, to the limit potential ν on (F, g'_o) and by Proposition 4.8, $\partial F = \{\nu = -\infty\}$. It follows that there is a level set $L' = L'(\varepsilon)$ of ν_δ such that $L' \cap B'_{y_\varepsilon}(1 + d_o) \subset\subset B'_{y_\varepsilon}(1 + d_o)$, i.e.

$$L' \cap \partial B'_{y_\varepsilon}(1 + d_o) = \emptyset, \tag{5.45}$$

with $dist_{g'_\varepsilon}(L_o, L') \geq d_1$, for some fixed $d_1 > 0$.

Theorem 5.9 shows that the ball $B_y(1 + d_o)$ is embedded in $\mathbb{R}^3$, and hence $L' \subset \mathbb{R}^3$ also. Thus L' bounds a compact domain $W \subset \mathbb{R}^3, \partial W = L'$, containing the limit $\partial F \cap B'_y(1)$. In particular, $\nu(x) \leq \nu(L')$, for all $x \in W$. Observe also that W is contained in the 3-ball $B(1 + d_o) \subset \mathbb{R}^3$.

Case I. S^2 bounds on the inside. The proof of this case is given in [1, Thm.4.2], (which was developed just for this purpose). The idea here is to glue the *flat* metric g'_o onto the metric g'_ε along the compact level set L'. Namely the 2-sphere $\bar{S}^2 = S^2(1 + \frac{1}{2}d_o) \subset \mathcal{C}_o \subset F \subset \mathbb{R}^3$ bounds a flat 3-ball $\bar{B}^3 \subset \mathbb{R}^3$ on the inside. Since S^2_ε bounds a 3-ball $B^3 \subset M$ on the inside, one may construct a comparison metric $\bar{g}_\varepsilon$ to g'_ε essentially by replacing $g'_\varepsilon|_{B^3}$ with the flat metric $g'_o|_{\bar{B}^3}$. For technical reasons, it is preferable to carry out this replacement argument on (W, L') in place of (B^3, S^2). This construction is explained in complete detail in [1, (4.26)ff], and we refer there for details. (Note that the level L' above corresponds to L_o in [1] and the level set results related to (5.45) are the same as the assumptions [1,(4.5)-(4.7)]). The upshot is that, for ε sufficiently small, the comparison metric $\bar{g}_\varepsilon$ satisfies ${I_\varepsilon}^-(\bar{g}_\varepsilon) < {I_\varepsilon}^-(g'_\varepsilon)$, contradicting the minimizing property of g'_ε.

Case II. S^2 bounds on the outside. The proof in this situation consists of two (related) subcases, both based on the construction of a comparison metric on the outside. The 2-sphere S^2 in M given by (5.32) bounds a 3-ball in M on the outside, and so in particular S^2 separates M into two components, $M = M' \cup_{S^2} B^3$, the outside one given by B^3. The comparison metrics are essentially the same as those already constructed, in essentially the same circumstances, in [1].

(i). Suppose first that, for some $\varepsilon > 0$ sufficiently small,

$$B^3 \cap \partial F \neq \emptyset, \tag{5.46}$$

so that there are other components of ∂F outside S^2 in F. Of course, $\partial F \cap \mathcal{C}_o = \emptyset$ and $\partial F \cap \partial B^3 = \emptyset$.

We explain the idea first formally on the limit F, and then pass to the metrics $(\Omega_\varepsilon, g_\varepsilon)$. The round 2-sphere $S^2 = S^2(1 + \frac{d_o}{2}) \subset \mathbb{R}^3$ isometrically embeds in a round 3-sphere $S^3(R)$ of large radius R, $R = 10(1 + \frac{d_o}{2})$ for instance. The surface S^2 disconnects $S^3(R)$ into a smaller, inside 3-ball B^- and a large, outside 3-ball B^+, so that, as above, $M = M' \cup_{S^2} B^+$. Let $\bar{g}_o$ be the round metric on $B^+ \subset S^3(R)$. This metric joins continuously to the flat metric g'_o along the seam S^2.

The resulting metric across S^2 is then C^o, but not C^1. However, it is clear that this metric may be smoothed in a small neighborhood of the seam to give a smooth metric $\widetilde{g}_o$ with non-negative scalar curvature, $\widetilde{s}_o \geq 0$.

Observe that $vol_{\widetilde{g}_o} B^+ < \infty$, while $vol_{g'_o}(F \setminus B(1+\frac{d_o}{2})) = \infty$, (since F is unbounded). Thus, the volume of the comparison metric is much smaller than that of (F, g'_o). Similarly, $z \equiv 0$ on $(B^+, \bar{g}_o)$ and the smoothing $\widetilde{g}_o$ has curvature $\widetilde{z}_o$ bounded near the seam, (depending only on the fixed smoothing). On the other hand, by (5.46), the metric g'_o formally has $|z| = \infty$ on ∂F, so that $\mathcal{Z}^2$ is much larger for (F, g_o) than for the comparison $\widetilde{g}_o$. Finally, since the smoothing $\widetilde{g}_o$ satisfies $\widetilde{s}_o \geq 0$, one has $(\widetilde{s}_o)^- \equiv 0$, so that the comparison gives no added contribution to $\mathcal{S}^2_-$. Thus, formally, the comparison metric has a smaller value of $I_\varepsilon{}^-$ than g'_o.

It is now a simple matter to make this formal reasoning rigorous. Thus, choose ε sufficiently small so that g'_ε is very close to the limit (F, g_o) on a large compact domain in F. In particular g'_ε is almost flat away from ∂F. Since (S^2, g'_ε) has Gauss curvature strictly larger than that of $S^3(R)$, the Weyl embedding theorem, c.f. [16], implies that (S^2, g'_ε) isometrically embeds in $S^3(R)$ and hence bounds the domain $B^+ = (B^+)_\varepsilon$ as above. It follows that the metric $\bar{g}_\varepsilon = g'_\varepsilon \cup \bar{g}_o$, with g'_ε restricted to $\Omega_\varepsilon \setminus B^3$, $\partial B^3 = S^2$, is a C^o metric on Ω_ε. As above, this metric may be smoothed to a metric $\widetilde{g}_\varepsilon$, satisfying $(\widetilde{s}_\varepsilon)^- \geq (s'_\varepsilon)^-$ everywhere. By (5.46), the curvature $\mathcal{Z}^2$ of g'_ε is arbitrarily large, for ε sufficiently small. It then follows by the same arguments as above in the limit case that

$$I_\varepsilon{}^-(\widetilde{g}_\varepsilon) < I_\varepsilon{}^-(g'_\varepsilon), \tag{5.47}$$

contradicting the minimizing property of $(\Omega_\varepsilon, g'_\varepsilon)$.

(ii). Thus, one may assume that

$$B^3 \cap \partial F = \emptyset. \tag{5.48}$$

Hence ∂F is compact and, (by Theorem 5.9),

$$F = \mathbb{R}^3 \setminus \partial F, \tag{5.49}$$

with ∂F of Hausdorff dimension at most 1. It follows that the harmonic potential ν extends to a global subharmonic function on $\mathbb{R}^3$ and so has the representation, c.f. [12],

$$\nu(x) = -\int_{\partial F} \frac{1}{|x-y|} d\mu_y + h, \tag{5.50}$$

where h is harmonic on $\mathbb{R}^3$, and $d\mu_y$ is the Riesz measure of ν on ∂F, as preceding Proposition 4.8. Note that all level sets of the Newtonian potential in (5.50) are compact, since ∂F is compact. Similarly, the function ν has a compact level set L' in $B_y(1+d_o)$, by (5.45). It is then elementary to see that h also has a level set with a compact component. Since h is harmonic, this forces h to be constant. Hence, by adding a suitable constant to ν, one has

$$\nu(x) = -\int_{\partial F} \frac{1}{|x-y|} d\mu_y. \tag{5.51}$$

It follows that ν has an expansion of the form

$$\nu = -m/r + O(r^{-2}), \tag{5.52}$$

where $r(x) = |x|$ in $\mathbb{R}^3$ and $m > 0$. In particular, the metric g'_ε has the form (5.44), with ν as in (5.52). The expression (5.51) also gives $|D^k \nu| = O(r^{-k-1})$, for all k. (Note that by rescaling down the flat limit F by arbitrarily large factors, one may obtain a new flat blow-up limit $\hat{F}$ with $\partial\hat{F} = pt$ with $\hat{\nu} = -m/r$; this new limit corresponds *formally* to blow-ups with an asymptotically flat end, c.f. Remark 1.6).

The argument in this case is again by a glueing to a large 3-ball on the outside, similar to (i) above. This situation is a little more complicated however, since there are no components to ∂F outside S^2, so that the comparison of the $\mathcal{Z}^2$ term is more delicate. However, the construction of a comparison metric $\widetilde{g}_\varepsilon$ satisfying (5.47) in this case is exactly that already given in Case II of [1, Thm.3.1], and so we refer there for further details.

In both cases one thus has a contradiction, completing the proof. ■

Theorem 5.12 implies that if M is σ-tame and irreducible then on some sequence of minimizing pairs $(\Omega_{\varepsilon_i}, g_{\varepsilon_i})$ satisfying (4.3), either the degeneration assumption (4.1) fails, or one of the non-collapse assumptions NCA I or NCA II fails. To complete the proof of the Main Theorem, it remains to understand the situation where one of these assumptions fails. This is done in the final two sections.

6. The Collapse Situation.

In this section, we analyse in general the situation where there is collapse at base points $x_\varepsilon \in (\Omega_\varepsilon, g_\varepsilon)$, $u(x_\varepsilon) \to 1$, on the scale of the curvature radius $\rho(x_\varepsilon)$. This situation applies for instance when one of the two non-collapse assumptions NCA I or NCA II does not hold. The main point is that collapse at x_ε implies non-degeneration within bounded g_ε-distance to x_ε, c.f. Theorem 6.5. The remaining analysis of the collapse situation is then completed in §7 in connection with the non-degeneration hypothesis.

Let x_ε be any sequence of base points with $u(x_\varepsilon) \to 1$. As discussed in §2, in this generality one may have $\rho(x_\varepsilon) \to 0$, $\rho(x_\varepsilon) \to \rho_o > 0$, or $\rho(x_\varepsilon) \to \infty$, (in the case $\sigma(M) = 0$). Consider such sequences which collapse on the scale of $\rho(x_\varepsilon)$, i.e. $\omega(x_\varepsilon) << \rho(x_\varepsilon)$ as $\varepsilon \to 0$, where ω is the volume radius. As explained in §1, the collapse of $(\Omega_\varepsilon, g'_\varepsilon)$, $g'_\varepsilon = \rho(x_\varepsilon)^{-2} \cdot g_\varepsilon$, at x_ε may then be unwrapped by passing to sufficiently large, (depending on ε), finite covering spaces and on passing to a convergent subsequence, one obtains a maximal limit (F, g'_o, x). As previously, let $\nu_{\delta_a} = (u-1)/\delta_a(x_\varepsilon)$, for $\delta_a(x_\varepsilon)$ as in (3.1) and recall from (3.2) that $\delta_a \geq c \cdot \delta$. For the remainder of the paper we work with δ_a as in (3.1), in place of δ in (2.10). By Theorem 2.11, the potential functions ν_{δ_a} sub-converge to a limit potential function ν_a, modulo addition of constant functions. The limit (F, g_o) has a free isometric S^1 or T^2 action, leaving the potential ν_a invariant.

The limit (F, g'_o, x) may be either flat or hyperbolic, i.e. of constant negative curvature, c.f. §2.1. The former case occurs whenever $\rho \to 0$, or if ρ is bounded and $\sigma(M) = 0$. Conversely, if $\sigma(M) < 0$, then the limit is hyperbolic precisely when $\rho \geq \rho_o$, for some $\rho_o > 0$. If $\rho \to \infty$, forcing $\sigma(M) = 0$, limits could apriori be either flat or hyperbolic, c.f. §7.3 and Appendix A.

Isometric S^1 actions on flat manifolds F may be classified as follows. First, by lifting to a cover if necessary, the S^1 action may be converted to a proper $\mathbb{R}$-action. Since F is flat, an isometric $\mathbb{R}$-action is completely determined by its behavior in a neighborhood U of any given point $x \in F$. One may assume that U is embedded in $\mathbb{R}^3$ and hence the action on U extends uniquely to an $\mathbb{R}$-action on $\mathbb{R}^3$. As discussed following Proposition 4.8, the $\mathbb{R}$-action on $\mathbb{R}^3$ is generated either by a rotation about an axis A, a twist about an axis A, or a translation. Thus, the action is free everywhere except in the case of rotation, where the axis A consists of fixed points. The same classification holds for S^1 actions on hyperbolic manifolds.

If the collapse at x_ε is a rank 2 collapse, (or a sequence of rank 1 collapses converging to a rank 2 collapse), then the the limit (F, g'_o, x) has a free isometric T^2 action, obtained by unwrapping both collapsing circles. The results to follow hold for any S^1 action with $S^1 \subset T^2$.

We begin with the following result.

Lemma 6.1. *Any collapsing sequence of base points x_ε with $u(x_\varepsilon) \to 1$ is allowable, w.r.t. δ_a, (in place of δ), c.f. Definition 2.7.*

Proof: Base points x_ε are not allowable only if, on a limit (F, g'_o, x), there exists a solution $\bar{\nu}_\infty$ to the equation

$$D^2 \bar{\nu}_\infty = g'_o, \tag{6.1}$$

obtained by renormalizing ν_{δ_a}, c.f. (2.51)-(2.53). However, in the collapse case, $\bar{\nu}_\infty$ must, in addition, be invariant under the isometric S^1-action. The equation (6.1) only has S^1-invariant solutions in the case of a rotational S^1-action, and in this case, the only solution is of the form $\bar{\nu}_\infty = \frac{1}{2}\cdot r^2$, where r is the distance from some point on the rotation axis, (modulo affine functions). However, then $|\nabla\bar{\nu}_\infty|(x) = 1$, which implies, via the renormalization, that $|\nabla\nu_a| \to \infty$ at and near x_ε; this contradicts the definition of ν_{δ_a}, c.f. (3.13). ■

Throughout this section, let

$$L^o = \{x_\varepsilon \in (\Omega_\varepsilon, g_\varepsilon) : |u-1|(x_\varepsilon) = 1/(\ln\bar{\varepsilon})^2\}, \tag{6.2}$$

and U^o the corresponding u-superlevel set where $|u-1| \leq 1/(\ln\bar{\varepsilon})^2$, i.e. $u \geq 1 - 1/(\ln\bar{\varepsilon})^2$, as in (4.1); these should not be confused with the levels L_o from (4.34) and their superlevels as in (5.21).

It is useful to separate the cases of flat and hyperbolic limits. We begin with the flat case.

Proposition 6.2. *Let x_ε be a collapsing sequence of base points in U^o, and let (F, g'_o, x, ν) be a maximal flat limit as above. Then ∂F consists of a finite number of orbits of the free S^1 action, together possibly with the full axis A. In fact, there is a constant $M < \infty$, independent of x, such that*

$$\#(\partial F) \leq M, \tag{6.3}$$

where $\#$ denotes the number of components. In particular, away from A, ∂F is compact.

Proof: Since x_ε is allowable, Proposition 4.8 and Lemma 4.9 hold on the limit (F, g'_o, x). The potential ν is invariant under the S^1 action and by Proposition 4.8, $\partial F = I_\infty$. Hence ∂F is a union of orbits of the action, so that the S^1 action extends to $\bar{F}$.

The measure $d\mu_\infty$ of the potential ν at ∂F is also S^1 invariant, c.f. (4.37ff). Hence, for orbits which are not fixed points of the action, i.e. all cases except the axis in the rotational case, this measure is a multiple of Lebesque measure on a circle. Proposition 4.8 then implies that the number of orbits not in A is locally finite, i.e. there is a uniform bound on the number of orbits in any compact set in F.

To obtain the global bound (6.3), recall from Lemma 4.9 that $t = dist_{g'_o}(\partial F, \cdot)$ is unbounded on F. Hence, one may choose a point $y \in F$ with $t(y) >> 1$, as well as base points $y_\varepsilon \in \Omega_\varepsilon$ such that $y_\varepsilon \to y$. The metrics $g'_\varepsilon = \rho(y_\varepsilon)^{-2}\cdot g_\varepsilon$ then converge to a rescaling (F, g'_1, y) of (F, g'_o, x), (again unwrapping collapse). Applying the arguments above to (F, g'_1, y) gives a uniform bound on the number of orbits in balls of fixed size about y in (F, g'_1, y). Since y is arbitrary, this implies (6.3).

Finally, in the case of a rotational action, note that the full axis A is necessarily contained in ∂F, since the S^1 action is free on F itself. ■

The next result shows that collapse at x_ε as in Proposition 6.2 propagates to collapse at larger scales in a natural way. Recall the definition of the levels L^{v_o} and U^{v_o} from (1.17).

Lemma 6.3. *For x_ε as above, suppose y_ε are base points in the component of U^{v_o} containing x_ε such that $u(y_\varepsilon) \to 1$ and*

$$dist_{g_\varepsilon}(y_\varepsilon, x_\varepsilon) \leq c\cdot dist_{g_\varepsilon}(y_\varepsilon, L^{v_o}), \tag{6.4}$$

for some constant $c > 0$. Then there are constants C_ε, with $C_\varepsilon \to \infty$ as $\varepsilon \to 0$ such that if

$$dist_{g_\varepsilon}(y_\varepsilon, x_\varepsilon) \leq C_\varepsilon, \tag{6.5}$$

then the metrics $(\Omega_\varepsilon, g_\varepsilon)$ also collapse at y_ε, i.e. $\omega(y_\varepsilon) << \rho(y_\varepsilon)$.

Proof: Suppose first that $\rho(x_\varepsilon) \geq \rho_o$, for some $\rho_o > 0$, so that by Proposition 1.1, $dist_{g_\varepsilon}(x_\varepsilon, L^{v_o}) \geq \rho_1$, for some $\rho_1 > 0$. There is a curve γ joining $x_\varepsilon, y_\varepsilon$, satisfying (6.4) and hence $\rho \geq \rho_2 > 0$ everywhere along γ, so that the curvature of $(\Omega_\varepsilon, g_\varepsilon)$ is everywhere bounded in a neighorhood of γ. The result then follows from elementary comparison geometry, c.f. [10, Ch.8C] or [9, Ch.9.1].

If instead $\rho(x_\varepsilon) \to 0$, we work with the metrics $\widetilde{g}_\varepsilon = u^2 g_\varepsilon$, as in the proof of Proposition 2.3. Note that $\widetilde{g}_\varepsilon$ is uniformly quasi-isometric to g_ε within U^{v_o}, and hence the volume radii w.r.t. g_ε and $\widetilde{g}_\varepsilon$ are uniformly equivalent within U^{v_o}. Similarly, Proposition 1.1 implies that the L^2 curvature radii ρ, $\widetilde{\rho}$, are uniformly equivalent and hence it suffices to establish collapse w.r.t. $\widetilde{g}_\varepsilon$. The Ricci curvature of $\widetilde{g} = \widetilde{g}_\varepsilon$ is given by

$$\widetilde{r} = 2(d\ln u)^2 + \tfrac{1}{2}sg + \frac{2}{u}cg + \varepsilon\nabla\mathcal{Z}^2;$$

this follows from (1.38), via the Euler-Lagrange equations (1.14)-(1.15). Hence

$$\widetilde{r} \geq \tfrac{1}{2}sg + \varepsilon\nabla\mathcal{Z}^2. \tag{6.6}$$

From (1.18), one has $|\varepsilon\nabla\mathcal{Z}^2| \leq c\varepsilon t_{v_o}^{-4}$, where $t_{v_o} = dist_{g_\varepsilon}(L^{v_o}, \cdot)$ so that (6.6) gives

$$\widetilde{r} \geq \tfrac{1}{2}su^{-2}\widetilde{g} - c\varepsilon t_{v_o}^{-4} \cdot \widetilde{g}. \tag{6.7}$$

The standard volume comparison theorem, c.f. [10, Ch.5A] or [14, 9.1.3], implies that if the Ricci curvature of $\widetilde{g}_\varepsilon$ is uniformly bounded below, then collapse at x_ε implies collapse at y_ε satisfying (6.4)-(6.5). Note that the scalar curvature s of g_ε in (6.7) is uniformly bounded below, so only the second term in (6.7) needs to be considered. To understand this, if t_{v_o} is sufficiently small, pass to the blow-up scale where $\widetilde{\rho}'(x_\varepsilon) = 1$. Then (6.7) translates to

$$(\widetilde{r})' \geq \tfrac{1}{2}s'u^{-2}(\widetilde{g})' - c\alpha(t'_{v_o})^{-4} \cdot (\widetilde{g})'.$$

In this scale, $t'_{v_o}(x_\varepsilon) \to 1$. Since $\alpha \to 0$, the Ricci curvature $(\widetilde{r})'$ is almost non-negative; its negative part decays as $(t'_{v_o})^{-4}$, for (t'_{v_o}) large. Again, the volume comparison theorem implies that collapse at the base point x_ε implies collapse in cones as in (6.4) for t'_{v_o} arbitrarily large. This procedure may then be iterated until the base scale g_ε is reached. The fact that one may allow $C_\varepsilon \to \infty$ as $\varepsilon \to 0$ then follows again from volume comparison at the base scale g_ε. ■

Lemma 6.3 implies that collapse at some initial $x_\varepsilon \in U^o$ propagates to collapse at any y_ε, $u(y_\varepsilon) \to 1$, in any connected cone region defined by (6.4)-(6.5). The limits (F, g'_o, y) - or (F, g_o, y) when no rescaling is involved - at such base points y_ε may either be flat or hyperbolic. Hence, one may iterate this process to increase t_{v_o}, i.e. replace x_ε by y_ε, etc, provided the limits are flat, and provided the conclusions of Proposition 6.2 hold.

Suppose that all limits based at such y_ε satisfying (6.4)-(6.5) are flat; this is the case for instance if $\sigma(M) = 0$. Hence of course $\rho(y_\varepsilon)/t_{v_o}(y_\varepsilon) \to 1$ as $\varepsilon \to 0$. We then claim that

$$\sup_{U^{v_o}} t_{v_o} \to \infty, \quad \text{as} \quad \varepsilon \to 0; \tag{6.8}$$

in fact (6.8) holds in cone regions about x_ε as in (6.4). This estimate may be viewed as an analogue of Proposition 6.2 at the *base* scale $(\Omega_\varepsilon, g_\varepsilon, x_\varepsilon)$.

To prove the claim, start with an initial collapsing sequence $x_\varepsilon \in U^o$. On the maximal flat limit (F, g'_o, x), Theorem 2.11 shows that either

$$\Delta\nu_a = 0, \quad \text{or} \quad \Delta\nu_a = -d_o < 0, \tag{6.9}$$

where $d_o = \lim \rho^2|\bar{d}_\varepsilon|/\delta_a$; here ρ^2/δ_a is evaluated at x_ε and $\bar{d}_\varepsilon$ is from (2.5). The fact that x_ε is allowable by Lemma 6.1 implies that there is a constant $D_o < \infty$, independent of x_ε, such that $d_o \leq D_o$.

Now if $\Delta\nu_a = 0$ on (F, g'_o, x), so that ν_a is an S^1-invariant harmonic function, then it is a standard consequence of the maximum principle that there are cones $C \subset F$ as in (6.4), (with ∂F in place of L^{v_o}), such that $\nu_a(y) > \nu_a(x)$ for some $y \in C$ with $t'(y) = \rho'(y)$ arbitrarily large. Such points y may be approximated by points $y_\varepsilon \in U^o \subset \Omega_\varepsilon$ and hence the process above increasing t_{v_o} may be iterated at the much larger scale based at y_ε in place of x_ε. This proves the estimate (6.8) in case the first alternative in (6.9) always holds in U^o. In fact, the same argument holds if

$$\Delta\nu_a = -\mu, \tag{6.10}$$

where μ is sufficiently small, since all limits ν_a are non-trivial by Theorem 2.11.

Suppose instead, possibly at some stage of the iteration above, one has collapsing base points $x_\varepsilon \in U^o$ such that $\Delta\nu_a = -d_o < 0$ on the limit (F, g_o', x). Hence for ε small, $(\rho^2/\delta_a)(x_\varepsilon) \geq \frac{1}{2}d_o|\bar{d}_\varepsilon^{-1}|$. On the other hand, since $d_o \leq D_o$, one has $(\rho^2/\delta_a)(y_\varepsilon) \leq 2D_o|\bar{d}_\varepsilon^{-1}|$, for ε small, for any collapsing base points y_ε with $u(y_\varepsilon) \to 1$. Combining these estimates gives

$$\delta_a(y_\varepsilon) \geq \frac{d_o}{4D_o}(\rho')^2(y_\varepsilon)\delta_a(x_\varepsilon), \tag{6.11}$$

where ρ' is the L^2 curvature radius in the scale g_ε' based at x_ε. Hence, $\delta_a(y_\varepsilon)$ is essentially increasing quadratically in the distance ρ'. However, Proposition 6.2 implies that ρ' is unbounded on the limit (F, g_o', x). If y_ε are chosen so that $y_\varepsilon \to y \in F$ with $\rho'(y) >> 1$, then (6.11) implies that the mass of the potential $\nu_y = \lim \nu_{\delta_a(y_\varepsilon)}$ based at y on (F, g_o', y) becomes arbitrarily small. This contradicts Proposition 4.8. Hence (6.8) holds in all cases, as claimed.

As usual, Proposition 1.1 implies that whenever $t_{v_o} \geq t_o$, for any fixed $t_o > 0$, one has

$$\rho \geq \rho_o > 0, \tag{6.12}$$

with ρ_o depending only on t_o. Thus, there is no degeneration in such regions.

Next, we turn to the case of hyperbolic limits of collapsing sequences $\{x_\varepsilon\}$. Under the assumption $\rho(x_\varepsilon) \leq K_o$, for some $K_o < \infty$, this can occur only when $\sigma(M) < 0$ and $\rho(x_\varepsilon) \geq \rho_o$, for some $\rho_o > 0$. Thus, essentially *no* rescalings of g_ε are involved and one may set $g_\varepsilon' = g_\varepsilon$. The following is the analogue of Proposition 6.2.

Proposition 6.4. *Let (F, g_o, x, ν) be a maximal hyperbolic limit arising from a collapse at $x_\varepsilon \in U^o$. Then ∂F consists of a uniformly locally finite number of orbits of the S^1 action, together possibly with the full axis A. Thus, there exists $M < \infty$, independent of x, such that for any geodesic 1-ball $B(1)$ in $\bar{F}$,*

$$\#(\partial F \cap B(1)) \leq M. \tag{6.13}$$

Proof: The proof is identical to the first part of Proposition 6.2.

■

Unlike the flat case, the existence of a global bound in (6.13) is not asserted in the hyperbolic case; the hyperbolic case is not scale-invariant, as is the flat case. However, Lemma 4.9 implies that (6.8) holds also in this case.

The results above may be summarized in the following result on the structure of the base metrics $(\Omega_\varepsilon, g_\varepsilon)$ near collapsing base points.

Theorem 6.5. *Let $x_\varepsilon \in U^o$ be a collapsing sequence of base points. Then for ε small, the metrics $(\Omega_\varepsilon, g_\varepsilon)$ do not degenerate in the component of U^{v_o} containing x_ε where*

$$t_{v_o} \geq t, \quad dist(x_\varepsilon, y_\varepsilon) \leq C_\varepsilon, \tag{6.14}$$

for any $t > 0$ and C_ε as in (6.5). For any $t > 0$, the region defined by (6.14) is non-empty and all limits of $(\Omega_\varepsilon, g_\varepsilon, y_\varepsilon)$ based in this region are constant curvature manifolds (F, g_o, y) with scalar curvature $s_o \equiv \sigma(M)$, and with a free isometric S^1 or T^2 action.

If $\sigma(M) = 0$, all limits at base points in the region (6.14) with $t_{v_o} \in [t_o, T_o]$, for some $T_o < \infty$, are flat. If $\sigma(M) < 0$, then all limits based in the same region are hyperbolic, and such limits have boundary ∂F which is either empty, or consists of a uniformly locally bounded number of orbits of the action.

■

Theorem 6.5 allows one to pass to the situation where one has non-degeneration. The analysis in this situation is carried out next.

7. Non-Degeneration Situations.

In this final section, we complete the proof of the Main Theorem. Throughout this section, it is assumed that M is σ-tame and irreducible. These assumptions are not actually always necessary, but making these assumptions uniformly simplifies the overall logic of the arguments.

Recall from §4 that on $(\Omega_\varepsilon, g_\varepsilon)$, either the non-degeneration hypothesis (4.2) or the degeneration hypothesis (4.1) holds. In addition of course, either $\sigma(M) < 0$ or $\sigma(M) = 0$. The analysis is divided accordingly into three subsections.

In §7.1 we analyse the situation where non-degeneration (4.2) holds with $\sigma(M) < 0$, while §7.2 analyses the situation where degeneration (4.1) holds with $\sigma(M) < 0$. In §7.3 these two cases are handled together when $\sigma(M) = 0$. The Main Theorem is proved in Theorem 7.6, (the case $\sigma(M) < 0$), and in Theorem 7.7, (the case $\sigma(M) = 0$).

7.1.σ(M) < 0.

In this subsection, we examine the structure of $(\Omega_\varepsilon, g_\varepsilon)$ when $\sigma(M) < 0$ and the non-degeneration hypothesis holds. For convenience, recall the non-degeneration hypothesis (4.2): namely there exists a constant $\rho_o > 0$ such that if $u(x_\varepsilon) \geq 1 - 1/(\ln \bar{\varepsilon})^2$, i.e. $x_\varepsilon \in U^o$ as following (6.2), then

$$\rho(x_\varepsilon) \geq \rho_o, \tag{7.1}$$

on $(\Omega_\varepsilon, g_\varepsilon)$; here $\varepsilon = \varepsilon_i$ is some sequence satisfying (4.3), with $\varepsilon_i \to 0$.

If $\rho \geq \rho_o$, for some $\rho_o > 0$, on *all* of $(\Omega_\varepsilon, g_\varepsilon)$, then M is tame and the Main Theorem is proved in [1,Thm.0.2]. In fact in this situation, the sequence $(\Omega_\varepsilon, g_\varepsilon)$ is constant in ε and $g_\varepsilon = g_o$ is hyperbolic, c.f. the discussion in §1. The work below is thus a generalization of that in [1].

Let L^o be the level set as in (6.2), with U^o the superlevel set as above. We claim that $vol_{g_\varepsilon} U^o \to 1$ as $\varepsilon \to 0$. This follows from the following more general result, which will also be needed later. Let L^1 be the level of u given by

$$|u(L^1) - 1| = \bar{\varepsilon}^{2\mu} \tag{7.2}$$

with $U^1 = \{u \geq 1 - \bar{\varepsilon}^{2\mu}\}$ the superlevel set of L_1. Hence $U^1 \subset U^o$, (for ε small).

Lemma 7.1. *For U^1 as above,*

$$vol_{g_\varepsilon} U^1 \to 1, \quad \text{as} \quad \varepsilon \to 0. \tag{7.3}$$

Proof: To see this, (1.8) and (4.3) imply that $1 \leq T \leq 1 + \bar{\varepsilon}^{4\mu}$. Since $w = uT$ and $u \leq 1$, this gives $w \leq 1 + \bar{\varepsilon}^{4\mu}$. Let v_1 be the volume of the set $U_1 = \{u \leq 1 - \bar{\varepsilon}^{2\mu}\}$ and v_2 be the volume of $U^1 = \{u \geq 1 - \bar{\varepsilon}^{2\mu}\}$, so that $v_1 + v_2 = 1$. Then $w \leq 1 - \bar{\varepsilon}^{2\mu} + c\bar{\varepsilon}^{4\mu}$ on U_1 while $w \leq 1 + c\bar{\varepsilon}^{4\mu}$ on U^1. Since, by definition, the L^2 norm of w equals 1, one obtains

$$1 \leq v_1(1 - \bar{\varepsilon}^{2\mu} + c\bar{\varepsilon}^{4\mu})^2 + v_2(1 + c\bar{\varepsilon}^{4\mu})^2, \tag{7.4}$$

which gives $v_1\bar{\varepsilon}^{2\mu} \leq v_2 c\bar{\varepsilon}^{4\mu}$, for ε small. It follows that $v_2 \geq 1 - c\bar{\varepsilon}^{2\mu}$, which gives (7.3).

■

Let V^o be the $\rho_o/2$ tubular neighborhood of U^o. We claim that the filling radius, (c.f. [10,3.35]), of L^o in $U_o = \Omega_\varepsilon \setminus U^o$ is at least $\rho_o/2$, i.e. no collection of components of L^o bound within $V^o \cap U_o$. For if not, then there is a domain $D_\varepsilon \subset V^o \cap U_o$ with $\partial D_\varepsilon \subset L^o$ and hence u achieves a minimal value at some point $q_\varepsilon \in D_\varepsilon$. Note that $\rho \geq \rho_o/4$ everywhere on D_ε. Let (F, g_o, q) be a maximal limit based at q_ε. It follows that the limit potential function ν has a minimal value at q. This contradicts Theorem 2.11, via the trace equation (2.65).

Since $U^o \subset V^o, vol_{g_\varepsilon} V^o \to 1$ also, and so $vol_{g_\varepsilon}(V^o \setminus U^o) \to 0$. It follows in particular that

$$vol_{g_\varepsilon} T^o \to 0, \tag{7.5}$$

where T^o is the $\rho_o/4$ tubular neighborhood of L^o.

Now let x_ε be any base point in U^o, so that $\rho(x_\varepsilon) \geq \rho_o$. Any maximal limit (H_x, g_o, x) of $(\Omega_\varepsilon, g_\varepsilon, x_\varepsilon), \varepsilon = \varepsilon_i$, is of constant curvature, with scalar curvature $s_o \equiv \sigma(M)$ and with $u \equiv 1$, c.f. Proposition 1.2. In case the sequence collapses at x_ε, the limit has a free isometric S^1 or T^2 action, i.e. is a rank 1 or 2 hyperbolic cusp. If $dist_{g_\varepsilon}(x_\varepsilon, L^o) \to \infty$, the limit (H_x, g_o) is complete, while if $dist_{g_\varepsilon}(x_\varepsilon, L^o)$ remains bounded, the limit may either be complete or incomplete; however there is no boundary of H_x within the Hausdorff limit of the region V^o above. Observe also that (7.5) implies that all limits (H_x, g_o, x) for $x_\varepsilon \in T^o$ are necessarily collapsing. Hence, if $(\Omega_\varepsilon, g_\varepsilon)$ is not collapsing at x_ε, then the standard volume comparison theorem, c.f. [10, Ch.5A], or [14, 9.1.3], implies that the limit H_x is complete.

The main result of this subsection is the following:

Theorem 7.2. *Suppose $\sigma(M) < 0$, M is irreducible and σ-tame, and the sequence $(\Omega_\varepsilon, g_\varepsilon)$, $\varepsilon = \varepsilon_i$, satisfies the non-degeneration hypothesis (7.1). Then $(\Omega_\varepsilon, g_\varepsilon)$ is constant in ε, i.e. $(\Omega_\varepsilon, g_\varepsilon) = (H, g_o)$, where $H = \dot{\cup} H_k$ is a finite collection of complete, connected, hyperbolic manifolds H_k, $1 \leq k \leq q$, $q = q(M) < \infty$, and H is embedded in M. The metric g_o is hyperbolic, satisfying*

$$s_{g_o} = \sigma(M), \quad vol_{g_o} H = 1. \tag{7.6}$$

Proof: Suppose first that there is a (complete) component C_ε of Ω_ε such that $C_\varepsilon \subset U^o$, so that the sequence of metrics $g_\varepsilon|_{C_\varepsilon}$ is tame. As mentioned at the beginning of §7, the work in [1,Prop.2.5,Rmk.2.10] then implies that $g_\varepsilon|_{C_\varepsilon}$ is a constant sequence, so that $(C_\varepsilon, g_\varepsilon) = (C_o, g_o)$ is a complete hyperbolic manifold of finite volume.

Thus, g_ε can only vary on the components of $U^o \subset \Omega_\varepsilon$ on which $\partial U^o \neq \emptyset$. In this situation, we claim first that not all of U^o is collapsing, i.e. there are base points $x_\varepsilon \in U^o$ such that $(\Omega_\varepsilon, g_\varepsilon)$ does not collapse at x_ε. For suppose instead all of U^o is collapsing as $\varepsilon \to 0$. Then U^o is a graph manifold, and all based limits, when the collapse is unwrapped, are rank 1 or rank 2 hyperbolic cusps. (Of course, any domain of bounded diameter about any such x_ε has volume converging to 0, due to the collapse; however, there could be a large number of such regions, corresponding to a large number of limit cusps, whose total volume approaches 1).

Observe that the volume of geodesic balls in the expanding end of a hyperbolic cusp is proportional to the volume of the corresponding geodesic sphere. It then follows from (7.3) and (7.5) that most all (in terms of volume) of the components of U^o intersecting ∂U^o are cusps which are expanding into the interior of U^o. However, this implies that, for any $\varepsilon > 0$ sufficiently small, there must exist points realizing the maximal diameter of the orbits (circles or tori) of the F-structure, (i.e. graph manifold structure), within U^o. Choosing such points as base points, it follows that the limiting hyperbolic cusp also has a maximal value for the diameter of the associated S^1 or T^2 orbits. However, this is impossible, and thus proves the claim.

Let x_ε be any non-collapsing sequence of base points in U^o. It follows then from the observation preceding Theorem 7.2 that $(\Omega_\varepsilon, g_\varepsilon, x_\varepsilon)$ sub-converges to a *complete* hyperbolic manifold (H_x, g_o), of finite volume; the level set L^o satisfies $dist_{g_\varepsilon}(L^o, x_\varepsilon) \to \infty$ as $\varepsilon \to 0$ and so has no limit on H_x. The convergence is smooth and uniform on compact subsets. Any such manifold H_x is of finite topological type, with a finite number of cusp ends each of the form $T^2 \times \mathbb{R}^+$. Hence H_x embeds in Ω_ε and also in M. By the Margulis Lemma, c.f. [17], there is a fixed lower bound on the volume of any such H_x.

Now repeat this argument on $\Omega_\varepsilon \setminus H_x$. It follows from the Margulis Lemma that after a finite number of iterations, one obtains a finite number of disjoint limit hyperbolic manifolds H_k, with

$H = \dot{\cup}\, H_k$ satisfying $vol_{g_o} H = 1$, i.e. (7.6) holds. As above, H embeds in Ω_ε and so also in M. Thus, one may view the metrics $g_\varepsilon|_H$ as a sequence of metrics on H.

The next result proves that the metrics g_ε are rigid.

Proposition 7.3. *Under the assumptions of Theorem 7.2, the metrics $g_\varepsilon|_H$ are constant, up to isometry, i.e.*

$$(\Omega_\varepsilon, g_\varepsilon) = (H, g_o). \tag{7.7}$$

Proof: This is a simple application of Theorem 2.11 and the maximum principle. If (7.7) does not hold, then the metrics g_ε vary somewhere on H, say at base points $x_\varepsilon \to x \in H$. Now the limit potential ν on H_x, satisfies (2.65), i.e.

$$\Delta\nu = \lambda\nu - d, \tag{7.8}$$

where $d \geq 0$ and $\lambda = \lim \frac{1}{4}\rho^2\sigma T = \frac{1}{4}|\sigma(M)| \lim \rho^2$, for $\rho = \rho(x_\varepsilon)$, c.f. (2.72). Since $\rho(x_\varepsilon)$ is bounded away from 0 and ∞, observe that $\lambda > 0$. Further, as noted following Theorem 2.11, the potential $\nu_{\delta(y_\varepsilon)} = (u-1)/\delta(y_\varepsilon)$ converges to the limit ν, in that there is no constant or affine indeterminacy. Thus, $\nu \leq 0$. This implies that ν is also uniformly bounded below, i.e. there is a constant $K < \infty$ such that

$$\nu \geq -K > -\infty, \tag{7.9}$$

c.f. the remark following (2.69). Hence, if p_j is a minimizing sequence for ν on H_x, i.e. $\nu(p_j) \to \inf \nu$, then one must have $\Delta\nu(p_j) \geq -\mu_j$, for some sequence $\mu_j \to 0$. It then follows from (7.8) that $d = 0$ and also $\lambda = 0$, giving a contradiction. Hence (7.7) must hold.

This completes the proof of Proposition 7.3 and hence also Theorem 7.2.

■

7.2.σ(M) < 0.

In this section, still assuming $\sigma(M) < 0$, we analyse the case where the degeneration hypothesis (4.2), opposite to (4.1) or (7.1) holds. The main result is that Theorem 7.2 also holds in this context.

Theorem 7.4. *Suppose $\sigma(M) < 0$, M is irreducible and σ-tame, and the sequence $(\Omega_{\varepsilon_i}, g_{\varepsilon_i})$, satisfies the degeneration hypothesis (4.1). Then again the conclusions of Theorem 7.2 hold.*

Proof: In view of Theorem 5.12, one may assume that on the sequence $(\Omega_\varepsilon, g_\varepsilon)$, $\varepsilon = \varepsilon_i$, one of the non-collapse hypotheses NCA I or NCA II fails. The strategy is then to reduce to the proof to that of Theorem 7.2, using the results of §6.

First, recall that the degeneration hypothesis and Theorem 4.4 guarantee the existence of admissible level sets L_o as in (4.34), with corresponding superlevel sets U^o. Such levels are not unique, i.e. different admissible or preferred base points may lie on distinct levels L_o. (Similarly, recall that degeneration at some point $x_\varepsilon \in L_o$ does not imply degeneration everywhere on L_o). However, it is always the case, by (4.4), that $|u(L_o) - 1| \geq \bar{\varepsilon}^{2\mu}$. Thus for L^1 and U^1 as in (7.2), the domain U^1 is contained in all superlevel sets U^o of admissible levels L_o.

Now let $t^1 = dist_{g_\varepsilon}(L^1, \cdot)$ and set

$$U^2 = \{x \in U^1 : t^1(x) \geq \tau_o\}, \tag{7.10}$$

where τ_o is a small parameter. Then we have the following analogue of the non-degeneration hypothesis (7.1), Lemma 7.1 and (7.5) on U^2.

Lemma 7.5. *The domain $U^2 = U^2(\tau_o)$ is non-empty and there is a constant $\rho_o = \rho_o(\tau_o) > 0$ s.t.*

$$\rho(x_\varepsilon) \geq \rho_o, \tag{7.11}$$

$\forall x_\varepsilon \in U^2$. Further, the metrics g_ε collapse everywhere on the boundary $L^2 = \partial U^2$ as $\varepsilon \to 0$, while

$$vol_{g_\varepsilon} U^2 \geq 1 - \tau_1, \tag{7.12}$$

where $\tau_1 \to 0$ as $\tau_o \to 0$, on any sequence $\varepsilon \to 0$.

Proof: The estimate (7.11) is an immediate consequence of the definition (7.10) and Proposition 1.1. The main point is to prove that U^2 is non-empty and that g_ε collapses everywhere on L^2. From this, (7.12) follows easily.

Thus, let x_ε be any sequence of base points on L^1. Suppose first that $\rho(x_\varepsilon) \geq \rho_1$, for some $\rho_1 > 0$ (in a subsequence). If $(\Omega_\varepsilon, g_\varepsilon, x_\varepsilon)$ does not collapse, then it (sub)-converges to a limit hyperbolic manifold (H, g_o, x), containing at least the full geodesic ball $B_x(\rho_1)$. In particular then $vol_{g_\varepsilon} B_{x_\varepsilon}(\rho_1) \geq v_1$, for some $v_1 > 0$. However, a definite part of the volume of $B_{x_\varepsilon}(\rho_1)$ is contained in $U_1 = \Omega_\varepsilon \setminus U^1$, which contradicts (7.3). Hence $(\Omega_\varepsilon, g_\varepsilon, x_\varepsilon)$ collapses as $\varepsilon \to 0$. Theorem 6.5 then implies first that $U^2 = U^2(\tau_o)$ is non-empty, for a fixed τ_o small, and further that $(\Omega_\varepsilon, g_\varepsilon)$ collapses at all points on L^2 within bounded distance to x_ε. An examination of the proof of Lemma 7.1, c.f. (7.4ff), shows that the same conclusion holds if $\rho(x_\varepsilon) \to 0$ sufficiently slowly as $\varepsilon \to 0$, for instance $\rho(x_\varepsilon) \geq \bar{\varepsilon}^{\mu/10}$.

Next, consider base points x_ε on L^1 where $\rho(x_\varepsilon) \to 0$, e.g. $\rho(x_\varepsilon) \leq \bar{\varepsilon}^{\mu/10}$. Apriori all points on L^1 may have this property. As mentioned following the proof of Theorem 4.4, the admissible and preferred base points are 1-dense, in fact η-dense, within the region

$$S_\varepsilon = \{\bar{\varepsilon}^{2\mu} \leq |u-1| \leq 2/(\ln \bar{\varepsilon})^2\} \cap \{\rho \leq \chi_\varepsilon\}, \text{ where } \chi_\varepsilon \to 0 \text{ as } \varepsilon \to 0. \tag{7.13}$$

To be definite, choose $\chi_\varepsilon = \bar{\varepsilon}^{\mu/10}$. Clearly, $x_\varepsilon \in S_\varepsilon$. Choose η to be a fixed number with $\eta < \tau_o/3$.

Now as mentioned above, one of the two non-collapse hypotheses fails. Suppose first that NCA I fails. This means that for *all* admissible or preferred base points y_ε, there is collapse at some $p_\varepsilon \in B_{y_\varepsilon}(\eta) \cap L_o$, on the scale of $\rho(p_\varepsilon)$, where L_o is the u-level through y_ε. The collection of balls $B_{y_\varepsilon}(\eta)$, for y_ε admissible, covers S_ε. Now Lemma 6.3 and the analysis following it implies that the collapse at such p_ε propagates in cones to collapse on L^1. Together with the collapse result above when $\rho(x_\varepsilon) \geq \bar{\varepsilon}^{\mu/10}$, it follows that L^1 collapses on an η-dense set. Again, Lemma 6.3 or Theorem 6.5 then imply that $L^2 = L^2(\tau_o)$ collapses everywhere and U^2 is non-empty.

Suppose instead that NCA I holds, but NCA II fails. When NCA I holds, there are some admissible or preferred base points y_ε which satisfy (5.1) and possibly some that don't. For those base points y_ε that don't satisfy (5.1), the same conclusion as above holds, (i.e. U^2 is non-empty and L^2 collapses within bounded distance to y_ε). Thus, one only needs to consider those base points y_ε where (5.1) holds. For any such base point, *all* distinguished base points q_ε within g_ε-distance η to y_ε collapse, possibly on a scale large compared with $\rho(q_\varepsilon)$. Since $\rho(q_\varepsilon) \to 0$, this means there are points q'_ε, with $dist_{g_\varepsilon}(q_\varepsilon, q'_\varepsilon) \to 0$, $\rho(q'_\varepsilon) \to 0$, such that $\omega(q'_\varepsilon) << \rho(q'_\varepsilon)$. As above, Lemma 6.3 and Theorem 6.5 imply that L^2 is non-empty and collapses within bounded distance to q'_ε. By the same η-density arguments as before, it follows that in all cases, L^2 collapses everywhere w.r.t. g_ε.

Finally, to prove (7.12), all limits based at points in L^2 are hyperbolic manifolds H with free isometric S^1 or T^2 action, with ∂H consisting of a locally uniformly bounded number of S^1 orbits. In particular, ∂H has measure 0. Thus, by the continuity of the volume as $\varepsilon \to 0$, (7.12) follows immediately from Lemma 7.1. ■

Lemma 7.5 proves that one has non-degeneration on U^2, with $vol_{g_\varepsilon} U^2 \geq 1 - \tau$, where τ may be made small by choosing τ_o small, (for all ε sufficiently small). Further, all base points on $L^2 = \partial U^2$ collapse as $\varepsilon \to 0$. The rest of the proof of Theorem 7.4 then proceeds exactly as in the proof of Theorem 7.2, by choosing τ smaller than the Margulis constant.

■

Theorems 7.2 and 7.4 now lead easily to the proof of the Main Theorem in case $\sigma(M) < 0$.

Theorem 7.6. *Suppose $\sigma(M) < 0$, and M is irreducible and σ-tame. Then M has a decomposition, unique up to isotopy, as*

$$M = H \cup_{\mathcal{T}} G, \tag{7.14}$$

where H is a finite collection of complete, connected, hyperbolic manifolds with scalar curvature $\sigma(M)$ and total volume 1, and G is a finite collection of connected graph manifolds. The union is along a finite collection of tori $\mathcal{T} = \partial H = \partial G$, each incompressible in M. The Sigma constant $\sigma(M)$ is given by

$$\sigma(M) = -(6vol_{g_{-1}}H)^{2/3}, \tag{7.15}$$

where $vol_{g_{-1}}H$ is the total volume of H in the metric of constant curvature -1.

Proof: Theorems 7.2 and 7.4 prove the existence of the hyperbolic manifold H embedded in M satisfying (7.7). In particular Ω_ε embeds in M. As mentioned in §1, the complement of Ω_ε in M is then a finite union G of connected graph manifolds, collapsed under a minimizing sequence for ${I_\varepsilon}^-$ for any $\varepsilon > 0$. This gives the decomposition (7.14). The remaining parts of Theorem 7.6, namely that $\mathcal{T} = \partial H = \partial G$ is incompressible in M, that the decomposition (7.14) is unique up to isotopy, and that (7.15) holds, then follow from [1, Thm.0.2]; equivalently, the family $(\Omega_\varepsilon, g_\varepsilon)$ is tame, (since it is constant), and hence Theorem 7.6 follows from [1], as mentioned in §0.

∎

7.3.σ(M) = 0.

In this section, the cases of non-degeneration (4.2) and degeneration on L_o are handled together. The following result proves the Main Theorem in case $\sigma(M) = 0$.

Theorem 7.7. *Suppose $\sigma(M) = 0$, and M is irreducible and σ-tame. Then M is a graph manifold and either $\Omega_\varepsilon = \emptyset$, or $\Omega_\varepsilon = M$ and g_ε is a sequence of flat metrics on M.*

Theorem 7.7 will be proved in a sequence of steps below, but logically is proved by contradiction. Thus, we assume that $(\Omega_\varepsilon, g_\varepsilon)$, $\varepsilon = \varepsilon_i$, is a non-empty sequence of minimizers for ${I_\varepsilon}^-$ as in §1, with g_ε non-flat for all $\varepsilon > 0$. As noted in §1, this already implies that M is not a graph manifold.

Suppose first the non-degeneration hypothesis (7.1) holds on $(\Omega_\varepsilon, g_\varepsilon)$, so that there is non-degeneration everywhere on $U^o = \{x \in \Omega_\varepsilon : u(x) \geq 1/(\ln \bar{\varepsilon})^2\}$. All based limits in this region are flat manifolds (F, g_o, x), $x = \lim x_\varepsilon, x_\varepsilon \in U^o$. Such limits are necessarily non-compact; in fact by Lemma 4.9, $dist_{g_o}(\partial F, \cdot)$ is unbounded. Thus, (F, g_o) has infinite volume and so the sequence $(\Omega_\varepsilon, g_\varepsilon, x_\varepsilon)$ *must* collapse. (This is in strong contrast to the situation when $\sigma(M) < 0$). Hence all such limits have a free isometric S^1 or T^2 action. Further, by Proposition 6.2 or Theorem 6.5, the function $t^o = dist_{g_\varepsilon}(L^o, \cdot)$ is unbounded, i.e. assumes arbitrarily large values as $\varepsilon \to 0$.

Next, suppose instead that the degeneration hypothesis (4.1) holds. Then as in the proof of Theorem 7.4, one of the non-collapse assumptions must fail. In this case, let L^1 and U^1 be as in (7.2). The proof of all of Lemma 7.5 except the volume statement (7.1) carries over without any change and implies that one has non-degeneration on U^2, i.e. (7.11) holds, and there is collapse everywhere on $L^2 = \partial U^2$, for U^2 as in (7.10). Again, all limits (F, g_o, x) of $(\Omega_\varepsilon, g_\varepsilon, x_\varepsilon)$ for $x_\varepsilon \in U^2$ are flat manifolds, collapsed and so with free isometric S^1 or T^2 action. As before, the distance function t^1 from (7.10) is unbounded as $\varepsilon \to 0$.

To unify these two situations, let $\bar{L} = L^o, \bar{U} = U^o$ as above in case (7.1) holds, while let $\bar{L} = L^2$, $\bar{U} = U^2$ as above in case (7.1) fails. Hence, all limits (F, g_o, x) of $(\Omega_\varepsilon, g_\varepsilon, x_\varepsilon), x_\varepsilon \in \bar{U}$ are flat manifolds, with free isometric S^1 or T^2 action, with $\bar{t} = dist_{g_\varepsilon}(\bar{L}, \cdot)$ unbounded as $\varepsilon \to 0$.

As in the previous sections, it is then natural to consider the rescaled metrics

$$g'_\varepsilon = \rho(x_\varepsilon)^{-2} \cdot g_\varepsilon. \tag{7.16}$$

Of course $\rho(x_\varepsilon) \geq \rho_o > 0$ within $\bar{U}$. In the region where $\rho(x_\varepsilon) \to \infty$ as $\varepsilon \to 0$, (i.e. where $\bar{t}(x_\varepsilon) \to \infty$ as $\varepsilon \to 0$), the metrics g_ε are then being "blown-down". As before $\rho'(x_\varepsilon) = 1$, and any sequence $(\Omega_\varepsilon, g'_\varepsilon, x_\varepsilon)$ has a subsequence converging smoothly to a maximal limit (N_x, g', x). Since the unscaled sequence $(\Omega_\varepsilon, g_\varepsilon, x_\varepsilon)$ is collapsing, the rescaled sequence (7.16) is collapsing even more, so that all limits (N_x, g', x) necessarily have a free isometric S^1 or T^2 action.

The metrics g'_ε satisfy the Euler-Lagrange equations (2.24). Here $\bar{\alpha} = \varepsilon/T\rho^2, \rho = \rho(x_\varepsilon)$ so that obviously $\bar{\alpha} \to 0$. Thus, as before in §2, the terms in (2.24) containing $\bar{\alpha}\nabla\mathcal{Z}^2$ and $\bar{\alpha}|z|^2$ converge to 0. On the other hand, the constant term $\bar{c}'_\varepsilon$ in (2.24) is given by $\bar{c}'_\varepsilon = \rho^2 \cdot \bar{c}_\varepsilon$, so that $\bar{c}'_\varepsilon >> \bar{c}_\varepsilon$ as $\varepsilon \to 0$. Recall from (1.6) that

$$\bar{c}_\varepsilon = \frac{1}{12}\bar{\sigma} + \frac{1}{6}\bar{\varepsilon}\mathcal{Z}^2 \to 0, \quad \text{as} \quad \varepsilon \to 0. \tag{7.17}$$

It follows that the limit metric g' is a solution of the equations

$$L^* u = (-\frac{1}{4}su + \bar{c}_\infty)g, \tag{7.18}$$

$$2\Delta u + \frac{1}{4}su = -3\bar{c}_\infty. \tag{7.19}$$

Here s, $(= s_{g'})$, is the scalar curvature of the limit metric g', and the prime has been dropped from the notation. The constant $\bar{c}_\infty$ is given by $\bar{c}_\infty = \lim \bar{c}'_\varepsilon$. Apriori, this could be infinite, but the following simple Lemma shows it to be finite.

Lemma 7.8. *There is a constant $K < \infty$ such that, for any sequence of base points x_ε in $\bar{U}$,*

$$0 \leq \bar{c}_\infty \leq K. \tag{7.20}$$

Proof: To see this, return to (2.24) in the scale (7.16) and note that, by definition, $0 \leq u \leq 1$. Let η be a bounded smooth cutoff function supported in $B'_{x_\varepsilon}(1)$. Then multiply the trace equation (2.24) by η and integrate by parts over $B'_{x_\varepsilon}(1)$. Since $\bar{\alpha} \to 0$, while u and $\Delta\eta$ are bounded, it follows that $\int \eta\bar{c}'_\varepsilon$ is also bounded as $\varepsilon \to 0$, independent of x_ε. Hence $\bar{c}'_\varepsilon$ is also uniformly bounded, which gives (7.20). ∎

As noted in §2.1, Proposition A below shows that all maximal limits (N_x, g', x) of $(\Omega_\varepsilon, g'_\varepsilon, x_\varepsilon), x_\varepsilon \in \bar{U}$, are of constant curvature, either flat or hyperbolic, i.e. of constant negative curvature. As previously, denote such limits as (F, g'_o, x). The limit F is flat when $\bar{c}_\infty = 0$, hyperbolic when $\bar{c}_\infty > 0$. Recall, c.f. Remark 2.5(ii), that Theorem 2.11 holds for such limits, as does Proposition 6.2.

Now, for any given $\varepsilon > 0$, $\bar{c}'_\varepsilon = \rho^2\bar{c}_\varepsilon > 0$, and hence Lemma 7.8 implies that ρ has a uniform upper bound on $\bar{U}$, depending on ε, i.e. for ε sufficienty small,

$$\rho_{max} = \sup_{\bar{U}} \rho(x_\varepsilon) \leq 2(K/\bar{c}_\varepsilon)^{1/2} < \infty. \tag{7.21}$$

On the other hand, $\rho_{max} \to \infty$ as $\varepsilon \to 0$, since $\bar{t}$ is unbounded and $\sigma(M) = 0$. The next result characterizes the geometry of limits at points $x_\varepsilon \in \bar{U}$ (almost) realizing ρ_{max}.

Lemma 7.9. *If x_ε satisfies $\rho(x_\varepsilon)/\rho_{max} \to 1$, then the limit (F, g'_o, x) of $(\Omega_\varepsilon, g'_\varepsilon, x_\varepsilon)$ cannot be flat.*

Proof: If the limit (F, g'_o, x) is flat, then Proposition 6.2 holds and so the distance function t' to ∂F assumes arbitrarily large values on F. Hence, so does ρ', and so there exist $y_\varepsilon \in \Omega_\varepsilon$ such that $\rho(y_\varepsilon) >> \rho(x_\varepsilon)$. If there exist such $y_\varepsilon \in \bar{U}$, then one has a contradiction. On the other hand, this can only not be the case if $\Delta\nu_x \leq -\mu < 0$, as in (6.10), which leads to the same contradiction as following (6.11). ∎

Thus, blow-downs based at points x_ε (almost) realizing ρ_{max} must be hyperbolic. Note that the scalar curvature $s'_\varepsilon = \rho^2 s_\varepsilon$ of g'_ε, $\rho = \rho(x_\varepsilon)$, satisfies $s'_\varepsilon \to const < 0$ on such limits. Of course, since $s_\varepsilon = -\sigma T u$ and $u \to 1$, one has

$$\rho^2\sigma T \to const > 0. \tag{7.22}$$

There are now several ways to complete the proof of Theorem 7.7, but perhaps the simplest is the following. For x_ε as above, the limit (F, g'_o, x) is hyperbolic. Proposition 6.4 (or Theorem 6.5) describes the structure of the limit. By Lemma 4.9, the function $t = dist_{g'_o}(\partial F, \cdot)$ is unbounded on F. Hence, by choosing new base points $y_j \in F$ if necessary, with $t(y_j) \to \infty$, and approximating them by points $y_\varepsilon = y_{\varepsilon_j} \in \Omega_\varepsilon$, one obtains a new hyperbolic limit (F, g'_o, y) with $\partial F = \emptyset$, i.e. (F, y) is complete, without boundary. (Note that ρ' is constant on hyperbolic limits and does not increase with t, as is the case with flat limits).

This brings one exactly to the situation of Proposition 7.3. Namely, the limit potential ν on F satisfies (7.8)-(7.9), and the potentials $\nu_{\delta(y_\varepsilon)} = (u-1)/\delta(y_\varepsilon)$ converge to the limit ν. The same proof as in Proposition 7.3 then gives a contradiction to (7.22). This completes the proof of Theorem 7.7.

■

Appendix A.

In this Appendix, we prove the analogue of Proposition 1.2 when

$$u(x_\varepsilon) \to 1 \ \text{ and } \ \rho(x_\varepsilon) \to \infty. \tag{A.1}$$

Of course, this situation is only possible when $\sigma(M) = 0$ and the metrics $(\Omega_\varepsilon, g_\varepsilon)$ are collapsing at x_ε. For x_ε satisfying (A.1), let

$$g'_\varepsilon = \rho(x_\varepsilon)^{-2} \cdot g_\varepsilon, \tag{A.2}$$

and let (F, g', x) be the maximal limit of a convergent subsequence of $(\Omega_\varepsilon, g'_\varepsilon, x_\varepsilon)$, unwrapped in covering spaces to obtain convergence. From §7.3, g' is a solution of the equations (7.18)-(7.19).

Proposition A. *Let (F, g'_o, x) be a maximal limit constructed above, satisfying the equations (7.18)-(7.19). If $\bar{c}_\infty = 0$, then (F, g_o) is flat, with $u \equiv 1$. If $\bar{c}_\infty > 0$, then (F, g_o) is hyperbolic, i.e. of constant negative sectional curvature, again with $u \equiv 1$.*

Proof: From (7.19) and (2.3), (2.6), at every point of (F, g_o) one has

$$-\tfrac{1}{4}su - 3\bar{c}_\infty \leq 0. \tag{A.3}$$

Further by the definition of u, $-su \geq 0$ on F. Hence, if $\bar{c}_\infty = 0$, then $su \equiv 0$ on F and the equations (7.18)-(7.19) are the static vacuum equations (1.25). The maximum principle then implies that the limit is flat, with $u \equiv 1$.

Thus suppose $\bar{c}_\infty > 0$. The proof in this case is a rather long computation. Suppose first that at the base point x where $u(x) = 1 = \sup u$,

$$-\tfrac{1}{4}s(x) - 3\bar{c}_\infty = 0, \tag{A.4}$$

or equivalently, $\Delta u(x) = 0$, by (7.19) We then claim that essentially the same proof as Proposition 1.2 then shows that g' is hyperbolic and $u \equiv 1$. To see this, recall that by definition on the sequence $(\Omega_\varepsilon, g_\varepsilon), u = -(s^-)/\sigma T = -(s^-)\rho^2/\sigma T\rho^2, \rho = \rho(x_\varepsilon)$. The term $s^-\rho^2$ is the non-positive part of the scalar curvature s'_ε of g'_ε in (A.2). Passing to the limit, it follows that the trace equation (7.19) may be rewritten in the form

$$2\Delta s + \tfrac{1}{4}s^2 - 3c' = 0, \tag{A.5}$$

on the limit (F, g_o) where $c' = \lim \rho^4 \sigma c_\varepsilon$. Similarly, the equation (7.18) becomes

$$L^* s + (\tfrac{1}{4}s^2 + c')g = 0. \tag{A.6}$$

Write (A.5) in the form

$$2\Delta s + \tfrac{1}{4}(s - (12c')^{1/2})(s + (12c')^{1/2}) = 0. \tag{A.7}$$

Let $w = s + (12c')^{1/2}$ and note that (A.3) implies that $w \geq 0$ while (A.4) implies that $w(x) = 0$. Thus (A.7) is of the form

$$2\Delta w + fw = 0, \tag{A.8}$$

with $f = \frac{1}{4}(s - (12c')^{1/2}) \leq 0$. This equation is exactly analogous to (1.26) and the same argument as following (1.26) implies that $w \equiv 0$, so that (A.4) holds everywhere on F. Hence $u \equiv 1$ and (7.18) or (A.6) implies that g' is hyperbolic.

Thus, it remains to prove that (A.4) holds at x. To do this, observe first that (A.6) implies that $D_o^2 s = sz$, where $D_o^2 s = D^2 s - \frac{\Delta s}{3} \cdot g$ is the trace-free Hessian, and hence

$$\delta d D_o^2 s = \delta d(sz). \tag{A.9}$$

Here d is the exterior derivative on vector valued 1-forms defined by the metric, and δ is the adjoint of d, c.f. [6,Ch.1]. All the computations to follow are at the base point x where $u(x) = 1$. Since $\sup u = 1$, one has $ds = 0$ at x and an elementary computation then shows that $\delta dsz = -(\Delta s)z + s\delta dz$. Consequently, $tr(\delta dsz) = 0$ at x, so that

$$tr(\delta d D_o^2 s) = 0. \tag{A.10}$$

at x. The rest of the proof is a computation of $tr(\delta d D_o^2 s)$. First, $dD^2 s(X,Y,Z) = R(X,Y,ds,Z)$, which vanishes at x. Hence (A.6) and the definition of L^* in (1.5) imply that $dr = \delta R = 0$ at x, where the first equality is the Bianchi identity. A straightforward computation then gives

$$\delta d D^2 s = -R \circ (D^2 s),$$

where $R\circ$ is the action of the curvature on symmetric 2-tensors, c.f [6, Ch.1]. Consequently,

$$tr\delta d D^2 s = -tr\big(R \circ (D^2 s)\big) = - < r, D^2 s >,$$

at x. Also, $\delta d(\Delta s \cdot g) = -(\Delta\Delta s) \cdot g$ and so it follows from (A.10) that

$$- < r, D^2 s > + \Delta\Delta s = 0, \tag{A.11}$$

at x. Expanding r as $r = z + \frac{s}{3} g$ and using the equations (A.5)-(A.6), this implies that

$$-s|z|^2 - \tfrac{1}{3} s\Delta s - \tfrac{1}{4} s\Delta s = 0,$$

so that $s|z|^2 = -\frac{7}{12} s\Delta s$. However, $s < 0$, $\Delta s \geq 0$ at x, and so $\Delta s = 0$ and $z = 0$, at x. This proves (A.4), as required. ∎

Appendix B.

Let $u(x_\varepsilon) \to 1$ and let (F, g_o', x) be a maximal limit of $(\Omega_\varepsilon, g_\varepsilon', x_\varepsilon)$, as in §2. In this Appendix, we prove that, for $\delta = \delta(x_\varepsilon)$,

$$\bar{\alpha} \frac{\nabla \mathcal{Z}^2}{\delta} \to 0, \tag{B.1}$$

uniformly in L^2 on compact subsets of F, proving the claims in (2.25) and (2.43). We also indicate that the covariant derivatives $\nabla^k z/\delta, k \geq 0$, remain bounded locally in L^2 on compact subsets of F. Here and below, all metric quantities are w.r.t. g_ε' or g_o'.

We first prove (B.1) at the base point x_ε and within $B_{x_\varepsilon}'(\frac{1}{2})$. Given this, the proof that (B.1) holds near any $y \in F$, with $\delta = \delta(x_\varepsilon)$, follows easily and is given at the end.

To begin, Proposition 1.4 implies that $\bar{\alpha} \to 0$ while Proposition 1.1 implies that for any $k \geq 0$, $\nabla^k z \to 0$ in L^2 on compact subsets of F. Also, by definition, z/δ is uniformly bounded in $L^2(B), B = B_{x_\varepsilon}'(\frac{1}{2})$. Hence

$$\bar{\alpha} \frac{|z|^2}{\delta} \to 0 \ \text{ in } \ L^2(B). \tag{B.2}$$

Suppose first that $\delta \geq c \cdot \rho^2$, for some $c > 0$, i.e. Case (i) or (ii) of (2.12) holds. Then the same argument as in (2.36) implies that

$$\Delta \nu_\delta \to 0, \tag{B.3}$$

strongly in $L^2(B)$. By [6, Ch.4H] or [2,(3.8)],

(B.4) $$\nabla \mathcal{Z}^2 = D^*Dz + \tfrac{1}{3}D^2 s - 2R\circ z + \tfrac{1}{2}(|z|^2 - \tfrac{1}{3}\Delta s)g,$$

so that the Euler-Lagrange equation (2.24) in the scale g'_ε gives

(B.5) $$\frac{\bar\alpha}{\delta}D^*Dz + u\frac{z}{\delta} = D^2\nu_\delta + o(1)$$

in $L^2(B)$. Now integration by parts shows that

(B.6) $$\frac{\bar\alpha}{\delta}D^*Dz \to 0 \ \text{ in } \ L^{-2,2},$$

i.e. it converges to 0 when paired with any sequence of symmetric 2-tensors which are uniformly bounded in $L^{2,2}$, (since $\bar\alpha \to 0$). Thus, from (B.5),

(B.7) $$D^2\nu_\delta = 0(1) \ \text{ in } \ L^{-2,2}.$$

It then follows from (B.3) and elliptic regularity theory that $D^2\nu_\delta$ is bounded in $L^2(B)$ and converges strongly in $L^2(B)$ to a limit $D^2\nu \equiv h$. (Here and in similar arguments below, one must actually go to slightly smaller balls $B' \subset B$; since this is of no consequence in these arguments, this will be left aside). Returning to (B.5), this means that $\bar\alpha D^*Dz/\delta + uz/\delta$ converges strongly in L^2 to h. In particular, since z/δ is bounded in L^2, $\bar\alpha D^*Dz/\delta$ is bounded in L^2, and thus converges weakly to 0 in L^2.

To prove that $\bar\alpha D^*Dz/\delta$ converges strongly to 0 in L^2, i.e. (B.1), since h is fixed, one has

$$< \bar\alpha D^*Dz/\delta, \eta^2 h > \to 0,$$

where $<,>$ denotes the L^2 inner product w.r.t. g'_ε and η is a fixed cutoff function supported in B. Letting $||\cdot||$ denote the L^2 norm, it follows that $< u(z/\delta), \eta^2 h > \to ||\eta h||^2$ and so the Cauchy-Schwarz inequality gives

(B.8) $$||\eta u\frac{z}{\delta}||^2 \to K \geq ||\eta h||^2.$$

On the other hand, $||(\eta\bar\alpha D^*Dz/\delta) + (\eta uz/\delta)||^2 \to ||\eta h||^2$. Expanding out the norm on the left, the cross term has the form

(B.9) $$\frac{\bar\alpha}{\delta^2}\int < D^*Dz, u\eta^2 z > = \frac{\bar\alpha}{\delta^2}\int < Dz, D(u\eta^2 z) > = \frac{\bar\alpha}{\delta^2}\int u\eta^2|Dz|^2$$
$$+\frac{1}{\delta^2}\int < \bar\alpha^{1/2}\eta Dz, \bar\alpha^{1/2}\eta du\otimes z) > + \frac{2}{\delta^2}\int < \bar\alpha^{1/2}\eta Dz, \bar\alpha^{1/2}u d\eta\otimes z) > .$$

The last two terms are estimated as follows. Since z/δ is bounded in L^2 and $u \to 1$, $du \to 0$ smoothly, the Hölder and Young inequalities ($ab \leq \frac{1}{2}(a^2+b^2)$) imply that the 4^{th} term in (B.9) is the sum of two terms, one small compared with the 3^{rd} term and the other of order $\bar\alpha \to 0$. The same holds w.r.t. the last 5^{th} term, since $d\eta$ is bounded.

Thus the cross term is positive, or goes to 0 as $\varepsilon \to 0$. It follows that $||\eta\bar\alpha D^*Dz/\delta||^2 + ||\eta uz/\delta||^2 \leq ||\eta h||^2$ in the limit and hence (B.1) now follows from (B.8).

The argument in case $\delta << \rho^2$ is similar, although further work is required since (B.3) may not hold. Since z/δ is uniformly bounded in $L^2(B)$, $div(z/\delta)$ is uniformly bounded in $L^{-1,2}(B)$. By the Bianchi identity, $div(z/\delta) = \frac{1}{6}d(s/\delta)$, where $s = s'$ is the scalar curvature of g'_ε. Hence there are constants m_ε such that $(s - m_\varepsilon)/\delta$ remains uniformly bounded in $L^2(B)$; the constant m_ε/δ is the mean value of s/δ on B, and may be unbounded as $\varepsilon \to 0$. It follows that

(B.10) $$(\frac{s - m_\varepsilon}{\delta}) \to \zeta,$$

weakly in $L^2(B)$ as $\varepsilon \to 0$, for some $\zeta \in L^2(B)$. Returning to the trace equation (1.31), it follows from (B.2) and (B.10) that there are constants $\hat{c}_\varepsilon$, possibly unbounded as $\varepsilon \to 0$, such that

$$\Delta\nu_\delta = \zeta_\delta + \hat{c}_\varepsilon, \tag{B.11}$$

where ζ_δ is bounded in $L^2(B)$ and converges to ζ weakly in $L^2(B)$. Suppose the limit (F, g'_o, x) is flat. Then as in (2.53), the function $\phi = \frac{1}{6}\hat{c}_\varepsilon r^2$ satisfies $\Delta\phi = \hat{c}_\varepsilon$ on the limit F. Define $\hat{\nu}_\delta = \nu_\delta - \frac{1}{6}\hat{c}_\varepsilon r^2$; it then follows that $\Delta\hat{\nu}_\delta$ is uniformly bounded in $L^2(B)$ and

$$\Delta\hat{\nu}_\delta \to \zeta, \tag{B.12}$$

weakly in $L^2(B)$. The same argument holds for hyperbolic limits, using (2.57) in place of (2.53).

Now (B.6) holds as before and hence (B.7) holds, with $\hat{\nu}_\delta$ in place of ν_δ, or with D^2_o in place of D^2. Together with (B.12), it follows again from elliptic regularity that $D^2\hat{\nu}_\delta$ is uniformly bounded in $L^2(B)$ which in turn implies that $\bar{\alpha}\frac{\nabla\mathcal{Z}^2}{\delta}$ is uniformly bounded in $L^2(B)$.

To prove (B.1), by the same arguments as above, it suffices to prove that the convergence in (B.12) is in the strong L^2 topology. To prove this, return to the trace equation (2.69), i.e.

$$\Delta\nu_\delta = \frac{1}{2}\sigma T\rho^2\nu_\delta + c_\varepsilon + o(1) = \zeta_\delta + \hat{c}_\varepsilon + o(1),$$

where the constants may be unbounded as $\varepsilon \to 0$ and $o(1) \to 0$ in L^2 as $\varepsilon \to 0$. In particular, $\zeta_\delta = \frac{1}{2}\sigma T\rho^2\nu_\delta$, mod additive constants. The coefficient $\sigma T\rho^2$ is uniformly bounded as $\varepsilon \to 0$, since $-\sigma T\rho^2 u$ is the scalar curvature of g'_ε. It follows that $\Delta\zeta_\delta = \frac{1}{2}\sigma T\rho^2\zeta_\delta + \frac{1}{2}\sigma T\rho^2\hat{c}_\varepsilon$. If one pairs this equation with a cutoff function η and uses the self-adjointness of Δ, together with the fact that ζ_δ is weakly bounded in L^2, it follows that $\frac{1}{2}\sigma T\rho^2\hat{c}_\varepsilon$ is bounded as $\varepsilon \to 0$. (This is the same proof as Lemma 7.8). This implies that $\Delta\zeta_\delta$ is bounded in L^2 and hence, by elliptic regularity, ζ_δ converges strongly in L^2 to its limit ζ.

This completes the proof of (B.1) at x_ε within B. To prove that (B.1) holds on all of (F, g'_o, x) choose any $y \in F$ and let $y_\varepsilon \to y$. Choosing y_ε to be new center points, the validity of (B.1) at y_ε implies that $\bar{\alpha}\nabla\mathcal{Z}^2/\delta(y_\varepsilon) \to 0$ in $L^2(B'_{y_\varepsilon}(\frac{1}{2}))$, where $B'_{y_\varepsilon}(\frac{1}{2})$ the the $\frac{1}{2}$-ball in the scale $\rho'(y_\varepsilon) = 1$. Suppose these balls $B'_{x_\varepsilon}(\frac{1}{2}), B'_{y_\varepsilon}(\frac{1}{2})$ have non-empty intersection. If x_ε is allowable, then Theorem 2.11, i.e. (2.64)-(2.65) holds for each on the intersection region. These equations and elliptic regularity imply that $\delta(y_\varepsilon)/\delta(x_\varepsilon) \sim 1$, i.e. the ratio is bounded away from 0 and ∞ as $\varepsilon \to 0$. Hence, a covering argument shows that (B.1) holds on all of F, provided it holds on $B'_{x_\varepsilon}(\frac{1}{2})$. In fact, the arguments above following (B.10) give the same conclusions even if x_ε is not allowable.

Finally, suppose the base points x_ε are allowable, so that, after adding an affine term to ν_δ, ν_δ converges to a limit function ν. It then follows that one obtains C^∞ convergence of z/δ to its limit. This will not be detailed here, (since its not actually needed), but the proof is exactly the same as that given in [2, Thm.4.2, Rmk.4.3].

References

[1] *M. Anderson*, Scalar curvature and the existence of geometric structures on 3-manifolds, I, (preprint, Stony Brook, Sept. 2000), (to appear in Crelle's Journal).

[2] *M. Anderson*, Scalar curvature, metric degenerations and the static vacuum Einstein equations on 3-manifolds, II, Geom. & Funct. Analysis, **11**, (2001), 273-381.

[3] *M. Anderson*, Scalar curvature, metric degenerations and the static vacuum Einstein equations on 3-manifolds, I, Geom. & Funct. Analysis, **9**, (1999), 855-963.

[4] *M. Anderson*, Extrema of curvature functionals on the space of metrics on 3-manifolds, Calc. of Var. & P.D.E., **5**, (1997), 199-265.

[5] *M. Anderson*, Scalar curvature and geometrization conjectures for 3-manifolds, Comparison Geometry, MSRI Publ., **30**, (1997), 49-82.

The papers above are also available at: www.math.sunysb.edu/ $\sim$ anderson

[6] *A. Besse*, Einstein Manifolds, Ergebnisse der Mathematik, 3. Folge, Band 10, Springer Verlag, New York (1987).

[7] *J. Cheeger and M. Gromov*, Collapsing Riemannian manifolds while keeping their curvature bounded I, Jour. Differential Geometry, **23**, (1986), 309-342.

[8] *J. Cheeger and M. Gromov*, Collapsing Riemannian manifolds while keeping their curvature bounded II, Jour. Differential Geometry, **32**, (1990), 269-293.

[9] *D. Gilbarg and N. Trudinger*, Elliptic Partial Differential Equations of Second Order, 2nd Edition, Springer Verlag, New York, (1983).

[10] *M. Gromov*, Metric Structures for Riemannian and Non-Riemannian Spaces, Progress in Math., **152**, Birkhauser Verlag, Boston, (1999).

[11] *M. Gromov and H. Lawson, Jr.*, Positive scalar curvature and the Dirac operator on complete Riemannian manifolds, Publ. Math. I.H.E.S., **58**, (1983), 83-192.

[12] *W. Hayman*, Subharmonic Functions, Academic Press, London, (1976).

[13] *P. Mattila*, Geometry of Sets and Measures in Euclidean Spaces, Cambridge Univ. Press, Cambridge, U.K. (1995).

[14] *P. Petersen*, Riemannian Geometry, Graduate Texts in Math., **171**, Springer Verlag, New York, (1998).

[15] *R. Schoen and S.T.-Yau*, Existence of incompressible minimal surfaces and the topology of three dimensional manifolds with non-negative scalar curvature, Annals of Math., **110**, (1979), 127-142.

[16] *M. Spivak*, A Comprehensive Introduction to Differential Geometry, V, Publish or Perish, Houston, (1979).

[17] *W. Thurston*, Three-Dimensional Geometry and Topology, Princeton Univ. Press, (1997).

December, 2001

Department of Mathematics
S.U.N.Y. at Stony Brook
Stony Brook, N.Y. 11794-3651
anderson@math.sunysb.edu