

## Width and finite extinction time of Ricci flow

TOBIAS H COLDING  
WILLIAM P MINICOZZI II

This is an expository article with complete proofs intended for a general nonspecialist audience. The results are two-fold. First, we discuss a geometric invariant, that we call the width, of a manifold and show how it can be realized as the sum of areas of minimal 2–spheres. For instance, when  $M$  is a homotopy 3–sphere, the width is loosely speaking the area of the smallest 2–sphere needed to ‘pull over’  $M$ . Second, we use this to conclude that Hamilton’s Ricci flow becomes extinct in finite time on any homotopy 3–sphere.

53C44, 53C42; 58E12, 58E20

### Introduction

This is an expository article of our paper [8] with complete proofs intended for a general nonspecialist audience. The results are two-fold. First, we discuss a geometric invariant of a manifold, that we call the width, and show how it can be realized as the sum of areas of minimal 2–spheres. For instance, when  $M$  is a homotopy 3–sphere, the width is loosely speaking the area of the smallest 2–sphere needed to “pull over”  $M$ . Second, we use this to conclude that Hamilton’s Ricci flow becomes extinct in finite time on any homotopy 3–sphere. We have chosen to write this since the results and ideas given here are quite useful and seem to be of interest to a wide audience.

Given a Riemannian metric on a closed manifold  $M$ , sweep  $M$  out by a continuous one-parameter family of maps from  $S^2$  to  $M$  starting and ending at point maps. Pull the sweepout tight by, in a continuous way, pulling each map as tight as possible yet preserving the sweepout. We show the following useful property (see Theorem 1.8 below); cf 12.5 of Almgren [2], Proposition 3.1 of Pitts [26], Proposition 3.1 of Colding and De Lellis [5] and our papers [9; 8]:

Each map in the tightened sweepout whose area is close to the width (ie, the maximal energy of the maps in the sweepout) must itself be close to a collection of harmonic maps. In particular, there are maps in the sweepout that are close to a collection of immersed minimal 2–spheres.

This useful property that *all* almost maximal slices are close to critical points is virtually always implicit in any sweepout construction of critical points for variational problems yet it is not always recorded since most authors are only interested in existence of a critical point.

Similar results hold for sweepouts by curves<sup>1</sup> instead of 2–spheres; cf [9] where we used sweepouts by curves to estimate the rate of change of a 1–dimensional width for convex hypersurfaces in Euclidean space flowing by positive powers of their mean curvatures. The ideas are essentially the same whether one sweeps out by curves or 2–spheres, though the techniques in the curve case are purely ad hoc whereas for sweepouts by 2–spheres additional techniques, developed in the 1980s, have to be used to deal with energy concentration (ie, “bubbling”); cf Sacks and Uhlenbeck [30] and Jost [17]. The basic idea in each of the two cases is a local replacement process that can be thought of as a discrete gradient flow. For curves, this is now known as Birkhoff’s curve shortening process [3; 4].

Local replacement had already been used by H A Schwarz in 1870 to solve the Dirichlet problem in general domains, writing the domain as a union of overlapping balls and using that a solution can be found explicitly on balls by, eg, the Poisson formula [36; 37]. His method, which is now known as Schwarz’s alternating method, continues to play an important role in applied mathematics, in part because the replacements converge rapidly to the solution. The underlying reason why both Birkhoff’s method of finding closed geodesics and Schwarz’s method of solving the Dirichlet problem converge is convexity. We will deviate slightly from the usual local replacement argument and prove a new convexity result for harmonic maps. This allows us to make replacements on balls with small energy, as opposed to balls with small  $C^0$  oscillation. It is, in our view, much more natural to make the replacement based on energy and gives, as a by-product, a new uniqueness theorem for harmonic maps since already in dimension two the Sobolev embedding fails to control the  $C^0$  norm in terms of the energy; see Figure 1.

The second thing we do is explain how to use this property of the width to show that on a homotopy 3–sphere, or more generally closed 3–manifolds without aspherical summands, the Ricci flow becomes extinct in finite time. This was shown by Perelman

---

<sup>1</sup> Finding closed geodesics on the 2–sphere by using sweepouts goes back to Birkhoff in 1917; see Birkhoff [3; 4], Section 2 in Croke [10] and our paper [9]. In the 1980s Sacks and Uhlenbeck [30] found minimal 2–spheres on general manifolds using Morse theoretic arguments that are essentially equivalent to sweepouts; a few years later, Jost explicitly used sweepouts to obtain minimal 2–spheres in [17]. The argument given here works equally well on any closed manifold, but only produces nontrivial minimal objects when the width is positive.

in [25] and by Colding and Minicozzi in [8]; see Perelman [25] also for applications to the elliptic part of geometrization.

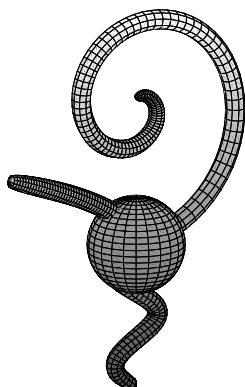


Figure 1: A conformal map to a long thin surface with small area has little energy. In fact, for a conformal map, the part of the map that goes to small area tentacles contributes little energy and will be truncated by harmonic replacement.

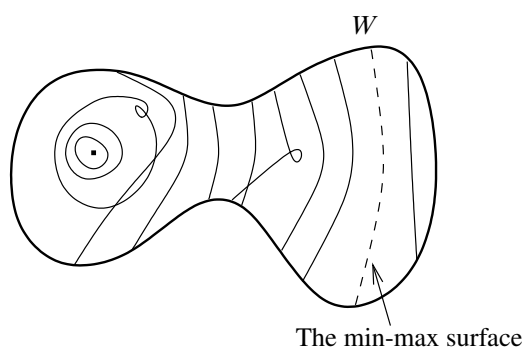


Figure 2: The sweepout, the min-max surface and the width  $W$

## 1 Width and finite extinction

On a homotopy 3–sphere there is a natural way of constructing minimal surfaces and that comes from the min-max argument where the minimizer of all maximal slices of sweepouts is a minimal surface. In [8] we looked at how the area of this min-max surface changes under the flow. Geometrically the area measures a kind of width of the 3–manifold (see Figure 2) and for 3–manifolds without aspherical summands (like a homotopy 3–sphere) when the metric evolve by the Ricci flow, the area becomes zero in finite time corresponding to that the solution becomes extinct in finite time.<sup>2</sup>

<sup>2</sup>It may be of interest to compare our notion of width, and the use of it, to a well-known approach to the Poincaré conjecture. This approach asks to show that for any metric on a homotopy 3–sphere a min-max

## 1.1 Width

Let  $\Omega$  be the set of continuous maps  $\sigma: \mathbf{S}^2 \times [0, 1] \rightarrow M$  so that for each  $t \in [0, 1]$  the map  $\sigma(\cdot, t)$  is in  $C^0 \cap W^{1,2}$ , the map  $t \rightarrow \sigma(\cdot, t)$  is continuous from  $[0, 1]$  to  $C^0 \cap W^{1,2}$ , and finally  $\sigma$  maps  $\mathbf{S}^2 \times \{0\}$  and  $\mathbf{S}^1 \times \{1\}$  to points. Given a map  $\beta \in \Omega$ , the homotopy class  $\Omega_\beta$  is defined to be the set of maps  $\sigma \in \Omega$  that are homotopic to  $\beta$  through maps in  $\Omega$ . We will call any such  $\beta$  a *sweepout*; some authors use a more restrictive notion where  $\beta$  must also induce a degree one map from  $\mathbf{S}^3$  to  $M$ . We will, in fact, be most interested in the case where  $\beta$  induces a map from  $\mathbf{S}^3$  to  $M$  in a nontrivial class<sup>3</sup> in  $\pi_3(M)$ . The reason for this is that the width is positive in this case and, as we will see, equal to the area of a nonempty collection of minimal 2–spheres.

The (energy) width  $W_E = W_E(\beta, M)$  associated to the homotopy class  $\Omega_\beta$  is defined by taking the infimum of the maximum of the energy of each slice. That is, set

$$W_E = \inf_{\sigma \in \Omega_\beta} \max_{t \in [0,1]} \text{Energy}(\sigma(\cdot, t)),$$

where the energy is given by

$$\text{Energy}(\sigma(\cdot, t)) = \frac{1}{2} \int_{\mathbf{S}^2} |\nabla_x \sigma(x, t)|^2 dx.$$

Even though this type of construction is always called min-max, it is really inf-max. That is, for each (smooth) sweepout one looks at the maximal energy of the slices and then takes the infimum over all sweepouts in a given homotopy class. The width is always nonnegative by definition, and positive when the homotopy class of  $\beta$  is nontrivial. Positivity can, for instance, be seen directly using Jost [17]. Namely, page 125 in [17] shows that if  $\max_t \text{Energy}(\sigma(\cdot, t))$  is sufficiently small (depending on  $M$ ), then  $\sigma$  is homotopically trivial.<sup>4</sup>

One could alternatively define the width using area rather than energy by setting

$$W_A = \inf_{\sigma \in \Omega_\beta} \max_{t \in [0,1]} \text{Area}(\sigma(\cdot, t)).$$

The area of a  $W^{1,2}$  map  $u: \mathbf{S}^2 \rightarrow \mathbf{R}^N$  is by definition the integral of the Jacobian

$$J_u = \sqrt{\det(du^T du)},$$

---

type argument produces an *embedded* minimal 2–sphere. Note that in the definition of the width it plays no role whether the minimal 2–sphere is embedded or just immersed, and thus the analysis involved in this was settled a long time ago. This well-known approach has been considered by many people, including Freedman, Meeks, Pitts, Rubinstein, Schoen, Simon, Smith and Yau; see Colding and De Lellis [5].

<sup>3</sup>For example, when  $M$  is a homotopy 3–sphere and the induced map has degree one.

<sup>4</sup>See the remarks after Corollary 3.4 for a different proof.

where  $du$  is the differential of  $u$  and  $du^T$  is its transpose. That is, if  $e_1, e_2$  is an orthonormal frame on  $D \subset \mathbf{S}^2$ , then  $J_u = (|u_{e_1}|^2 |u_{e_2}|^2 - \langle u_{e_1}, u_{e_2} \rangle^2)^{1/2} \leq \frac{1}{2} |du|^2$  and

$$(1.1) \quad \text{Area}(u|_D) = \int_D J_u \leq \text{Energy}(u|_D).$$

Consequently, area is less than or equal to energy with equality if and only if  $\langle u_{e_1}, u_{e_2} \rangle$  and  $|u_{e_1}|^2 - |u_{e_2}|^2$  are zero (as  $L^1$  functions). In the case of equality, we say that  $u$  is *almost conformal*. As in the classical Plateau problem (cf Section 4 of [7]), energy is somewhat easier to work with in proving the existence of minimal surfaces. The next proposition, proven in Appendix D, shows that  $W_E = W_A$  as for the Plateau problem (clearly,  $W_A \leq W_E$  by the discussion above). Therefore, we will drop the subscript and just write  $W$ .

**Proposition 1.2**  $W_E = W_A$ .

## 1.2 Finite extinction

Let  $M^3$  be a smooth closed orientable 3-manifold and  $g(t)$  a one-parameter family of metrics on  $M$  evolving by Hamilton's Ricci flow [13], so

$$(1.3) \quad \partial_t g = -2 \text{Ric}_{M_t}.$$

When  $M$  is prime and nonaspherical, then it follows by standard topology that  $\pi_3(M)$  is nontrivial (see, eg, Colding and Minicozzi [8]). For such an  $M$ , fix a nontrivial homotopy class  $\beta \in \Omega$ . It follows that the width  $W(g(t)) = W(\beta, g(t))$  is positive for each metric  $g(t)$ . This positivity is the only place where the assumption on the topology of  $M$  is used in the theorem below giving an upper bound for the derivative of the width under the Ricci flow. As a consequence, we get that the solution of the flow becomes extinct in finite time (see Paragraph 4.4 of [25] for the precise definition of extinction time when surgery occurs).

**Theorem 1.4** [8] *Let  $M^3$  be a closed orientable prime nonaspherical 3-manifold equipped with a metric  $g = g(0)$ . Under the Ricci flow, the width  $W(g(t))$  satisfies*

$$(1.5) \quad \frac{d}{dt} W(g(t)) \leq -4\pi + \frac{3}{4(t+C)} W(g(t)),$$

*in the sense of the limsup of forward difference quotients. Hence,  $g(t)$  becomes extinct in finite time.*

The  $4\pi$  in (1.5) comes from the Gauss–Bonnet theorem and the  $3/4$  comes from the bound on the minimum of the scalar curvature that the evolution equation implies. Both of these constants matter whereas the constant  $C > 0$  depends on the initial metric and the actual value is not important.

To see that (1.5) implies finite extinction time rewrite (1.5) as

$$\frac{d}{dt}(W(g(t))(t+C)^{-3/4}) \leq -4\pi(t+C)^{-3/4}$$

and integrate to get

$$(1.6) \quad (T+C)^{-3/4} W(g(T)) \leq C^{-3/4} W(g(0)) - 16\pi((T+C)^{1/4} - C^{1/4}).$$

Since  $W \geq 0$  by definition and the right hand side of (1.6) would become negative for  $T$  sufficiently large, we get the claim.

Theorem 1.4 shows, in particular, that the Ricci flow becomes extinct for any homotopy 3–sphere. In fact, we get as a corollary finite extinction time for the Ricci flow on all 3–manifolds without aspherical summands (see 1.5 of [25] or Section 4 of [8] for why this easily follows):

**Corollary 1.7** [8; 25] *Let  $M^3$  be a closed orientable 3–manifold whose prime decomposition has only nonaspherical factors and is equipped with a metric  $g = g(0)$ . Under the Ricci flow with surgery,  $g(t)$  becomes extinct in finite time.*

Part of Perelman’s interest in the question about finite time extinction comes from the following: If one is interested in geometrization of a homotopy 3–sphere (or, more generally, a 3–manifold without aspherical summands) and knew that the Ricci flow became extinct in finite time, then one would not need to analyze what happens to the flow as time goes to infinity. Thus, in particular, one would not need collapsing arguments.

There are two main points when there are surgeries. The first is that the scalar curvature is positive in the surgery regions and, thus, the lower bound  $R \geq -3/(2(t+C))$  for the scalar curvature is still valid in the presence of surgeries. The second is that inequality (1.6) is still valid if there are surgeries. Namely, suppose that  $M$  is as in Corollary 1.7 and in addition irreducible. From a surgery one can construct a degree-one map from the presurgery manifold to a component of the postsurgery manifold (throwing away the irrelevant “geometric” components). With Perelman’s surgery method [24] the map is  $(1+\xi)$ –Lipschitz for a function  $\xi$  which is described in Remark 73.6 of Kleiner

and Lott [18]. The function  $\xi(t)$  depends on the time  $t$  of the presurgery manifold, and vanishes as  $t$  approaches the surgery time.

One of the key ingredients in the proof of Theorem 1.4 is the existence of a sequence of good sweepouts of  $M$ , where each map in the sweepout whose area is close to the width (ie, the maximal energy of any map in the sweepout) must itself be close to a collection of harmonic maps. This will be given by Theorem 1.8 below, but we will first need a notion of closeness and a notion of convergence of maps from  $\mathbf{S}^2$  into a manifold.

### 1.3 Varifold convergence

Fix a closed manifold  $M$  and let  $\Pi: G_k M \rightarrow M$  be the Grassmannian bundle of (unoriented)  $k$ -planes, that is, each fiber  $\Pi^{-1}(p)$  is the set of all  $k$ -dimensional linear subspaces of the tangent space of  $M$  at  $p$ . Since  $G_k M$  is compact, we can choose a countable dense subset  $\{h_n\}$  of all continuous functions on  $G_k M$  with supremum norm at most one (dense with respect to the supremum norm).<sup>5</sup> If  $(X_0, F_0)$  and  $(X_1, F_1)$  are two compact (not necessarily connected) surfaces  $X_0, X_1$  with measurable maps  $F_i: X_i \rightarrow G_k M$  so that each  $f_i = \Pi \circ F_i$  is in  $W^{1,2}(X_i, M)$  and  $J_{f_i}$  is the Jacobian of  $f_i$ , then the varifold distance between them is by definition

$$d_V(F_0, F_1) = \sum_n 2^{-n} \left| \int_{X_0} h_n \circ F_0 J_{f_0} - \int_{X_1} h_n \circ F_1 J_{f_1} \right|.$$

It follows easily that a sequence  $X_i = (X_i, F_i)$  with uniformly bounded areas converges to  $(X, F)$ , iff it converges weakly, that is, if for all  $h \in C^0(G_2 M)$  we have  $\int_{X_i} h \circ F_i J_{f_i} \rightarrow \int_X h \circ F J_f$ . For instance, when  $M$  is a 3-manifold, then  $G_2 M$ ,  $G_1 M$  and  $T^1 M / \{\pm v\}$  are isomorphic. (Here  $T^1 M$  is the unit tangent bundle.) If  $\Sigma_i$  is a sequence of closed immersed surfaces in  $M$  converging to a closed surface  $\Sigma$  in the usual  $C^k$  topology, then we can think of each surface as being embedded in  $T^1 M / \{\pm v\} \cong G_2 M$  by mapping each point to plus-minus the unit normal vector,  $\pm \mathbf{n}$ , to the surface. It follows easily that the surfaces with these inclusion maps converges in the varifold distance. More generally, if  $X$  is a compact surface and  $f: X \rightarrow M$  is a  $W^{1,2}$  map, where  $M$  is no longer assumed to be 3-dimensional, then we let  $F: X \rightarrow G_2 M$  be given by that  $F(x)$  is the linear subspace  $df(T_x X)$ . (When  $M$  is 3-dimensional, then we may think of the image of this map as lying in  $T^1 M / \{\pm v\}$ .) Strictly speaking, this is only defined on the measurable space where  $J_f$  is nonzero; we extend it arbitrarily to all of  $X$  since the corresponding Radon measure on  $G_2 M$  given by  $h \rightarrow \int_X h \circ F J_f$  is independent of the extension.

<sup>5</sup>This is a corollary of the Stone–Weierstrass theorem; see corollary 35 on page 213 of [29].

### 1.4 Existence of good sweepouts

A  $W^{1,2}$  map  $u$  on a smooth compact surface  $D$  with boundary  $\partial D$  is *energy minimizing* to  $M \subset \mathbf{R}^N$  if  $u(x)$  is in  $M$  for almost every  $x$  and

$$\text{Energy}(u) = \inf \{ \text{Energy}(w) \mid w \in W^{1,2}(D, M) \text{ and } (w - u) \in W_0^{1,2}(D) \}.$$

The map  $u$  is said to be *weakly harmonic* if  $u$  is a  $W^{1,2}$  weak solution of the harmonic map equation  $\Delta u \perp TM$ ; see, eg, Lemma 1.4.10 in Hélein [16].

The next result gives the existence of a sequence of good sweepouts.

**Theorem 1.8** *Given a metric  $g$  on  $M$  and a map  $\beta \in \Omega$  representing a nontrivial class in  $\pi_3(M)$ , there is a sequence of sweepouts  $\gamma^j \in \Omega_\beta$  with  $\max_{s \in [0,1]} \text{Energy}(\gamma_s^j) \rightarrow W(g)$ , and so that given  $\epsilon > 0$ , there exist  $\bar{j}$  and  $\delta > 0$  so that if  $j > \bar{j}$  and*

$$\text{Area}(\gamma^j(\cdot, s)) > W(g) - \delta,$$

*then there are finitely many harmonic maps  $u_i: \mathbf{S}^2 \rightarrow M$  with*

$$(1.9) \quad d_V \left( \gamma^j(\cdot, s), \bigcup_i \{u_i\} \right) < \epsilon.$$

In (1.9), we have identified each map  $u_i$  with the varifold associated to the pair  $(u_i, \mathbf{S}^2)$  and then taken the disjoint union of these  $\mathbf{S}^2$ 's to get  $\bigcup_i \{u_i\}$ .

Theorem 1.8 will be proven in Section 2.2.

One immediate consequence of Theorem 1.8 is that if  $s_j$  is any sequence with  $\text{Area}(\gamma^j(\cdot, s_j))$  converging to the width  $W(g)$  as  $j \rightarrow \infty$ , then a subsequence of  $\gamma^j(\cdot, s_j)$  converges to a collection of harmonic maps from  $\mathbf{S}^2$  to  $M$ . In particular, the sum of the areas of these maps is exactly  $W(g)$  and, since the maps are automatically conformal, the sum of the energies is also  $W(g)$ . The existence of at least one nontrivial harmonic map from  $\mathbf{S}^2$  to  $M$  was first proven by Sacks and Uhlenbeck [30], but they allowed for loss of energy in the limit; cf also Struwe [38]. This energy loss was ruled out by Siu and Yau, using also arguments of Meeks and Yau (see Chapter VIII in Schoen and Yau [35]). This was also proven later by Jost in Theorem 4.2.1 of [17] which gives at least one min-max sequence converging to a collection of harmonic maps. The convergence in [17] is in a different topology that, as we will see in Appendix A, implies varifold convergence.



### 1.5 Upper bounds for the rate of change of width

Throughout this subsection, let  $M^3$  be a smooth closed prime and nonaspherical orientable 3–manifold and let  $g(t)$  be a one-parameter family of metrics on  $M$  evolving by the Ricci flow. We will prove Theorem 1.4 giving the upper bound for the derivative of the width  $W(g(t))$  under the Ricci flow. To do this, we need three things.

One is that the evolution equation for the scalar curvature  $R = R(t)$  (see page 16 of [14])

$$(1.10) \quad \partial_t R = \Delta R + 2|\text{Ric}|^2 \geq \Delta R + \frac{2}{3} R^2,$$

implies by a straightforward maximum principle argument that at time  $t > 0$

$$(1.11) \quad R(t) \geq \frac{1}{1/(\min R(0)) - 2t/3} = \frac{3}{2(t+C)}.$$

The curvature is normalized so that on the unit  $\mathbf{S}^3$  the Ricci curvature is 2 and the scalar curvature is 6. In the derivation of (1.11) we implicitly assumed that  $\min R(0) < 0$ . If this was not the case, then (1.11) trivially holds for any  $C > 0$ , since, by (1.10),  $\min R(t)$  is always nondecreasing. This last remark is also used when surgery occurs. This is because by construction any surgery region has large (positive) scalar curvature.

The second thing that we need in the proof is the observation that if  $\{\Sigma_i\}$  is a collection of branched minimal 2–spheres and  $f \in W^{1,2}(\mathbf{S}^2, M)$  with  $d_V(f, \bigcup_i \Sigma_i) < \epsilon$ , then for any smooth quadratic form  $Q$  on  $M$  we have (the unit normal  $\mathbf{n}_f$  is defined where  $J_f \neq 0$ )

$$(1.12) \quad \left| \int_f \text{big}(\text{Tr}(Q) - Q(\mathbf{n}_f, \mathbf{n}_f)) - \sum_i \int_{\Sigma_i} (\text{Tr}(Q) - Q(\mathbf{n}_{\Sigma_i}, \mathbf{n}_{\Sigma_i})) \right| < C \epsilon \|Q\|_{C^1} \text{Area}(f).$$

The last thing is an upper bound for the rate of change of area of minimal 2–spheres. Suppose that  $X$  is a closed surface and  $f: X \rightarrow M$  is a  $W^{1,2}$  map, then using (1.3) an easy calculation gives (cf pages 38–41 of [14])

$$(1.13) \quad \frac{d}{dt}_{t=0} \text{Area}_{g(t)}(f) = - \int_f (R - \text{Ric}_M(\mathbf{n}_f, \mathbf{n}_f)).$$

If  $\Sigma \subset M$  is a closed immersed minimal surface, then

$$(1.14) \quad \frac{d}{dt}_{t=0} \text{Area}_{g(t)}(\Sigma) = - \int_{\Sigma} K_{\Sigma} - \frac{1}{2} \int_{\Sigma} (|A|^2 + R).$$

Here  $K_\Sigma$  is the (intrinsic) curvature of  $\Sigma$ ,  $A$  is the second fundamental form of  $\Sigma$ , and  $|A|^2$  is the sum of the squares of the principal curvatures. To get (1.14) from (1.13), we used that if  $K_M$  is the sectional curvature of  $M$  on the two-plane tangent to  $\Sigma$ , then the Gauss equations and minimality of  $\Sigma$  give  $K_\Sigma = K_M - \frac{1}{2}|A|^2$ . The next lemma gives the upper bound.

**Lemma 1.15** *If  $\Sigma \subset M^3$  is a branched minimal immersion of the 2-sphere, then*

$$(1.16) \quad \frac{d}{dt} \Big|_{t=0} \text{Area}_{g(t)}(\Sigma) \leq -4\pi - \frac{\text{Area}_{g(0)}(\Sigma)}{2} \min_M R(0).$$

**Proof** Let  $\{p_i\}$  be the set of branch points of  $\Sigma$  and  $b_i > 0$  the order of branching. By (1.14)

$$\frac{d}{dt} \Big|_{t=0} \text{Area}_{g(t)}(\Sigma) \leq - \int_\Sigma K_\Sigma - \frac{1}{2} \int_\Sigma R = -4\pi - 2\pi \sum b_i - \frac{1}{2} \int_\Sigma R,$$

where the equality used the Gauss–Bonnet theorem with branch points (this equality also follows from the Bochner type formula for harmonic maps between surfaces given on page 10 of [35] and the second displayed equation on page 12 of [35] that accounts for the branch points). Note that branch points only help in the inequality (1.16).  $\square$

Using these three things, we can show the upper bound for the rate of change of the width.

**Proof of Theorem 1.4** Fix a time  $\tau$ . Below  $\tilde{C}$  denotes a constant depending only on  $\tau$  but will be allowed to change from line to line. Let  $\gamma^j(\tau)$  be the sequence of sweepouts for the metric  $g(\tau)$  given by Theorem 1.8. We will use the sweepout at time  $\tau$  as a comparison to get an upper bound for the width at times  $t > \tau$ . The key for this is the following claim: Given  $\epsilon > 0$ , there exist  $\bar{j}$  and  $\bar{h} > 0$  so that if  $j > \bar{j}$  and  $0 < h < \bar{h}$ , then

$$(1.17) \quad \text{Area}_{g(\tau+h)}(\gamma_s^j(\tau)) - \max_{s_0} \text{Area}_{g(\tau)}(\gamma_{s_0}^j(\tau)) \\ \leq \left( -4\pi + \tilde{C}\epsilon + \frac{3}{4(\tau+C)} \max_{s_0} \text{Area}_{g(\tau)}(\gamma_{s_0}^j(\tau)) \right) h + \tilde{C}h^2.$$

To see why (1.17) implies (1.5), use the equivalence of the two definitions of widths to get

$$W(g(\tau+h)) \leq \max_{s \in [0,1]} \text{Area}_{g(\tau+h)}(\gamma_s^j(\tau)),$$

and take the limit as  $j \rightarrow \infty$  (so that<sup>6</sup>  $\max_{s_0} \text{Area}_{g(\tau)}(\gamma_{s_0}^j(\tau)) \rightarrow W(g(\tau))$ ) in (1.17) to get

$$(1.18) \quad \frac{W(g(\tau+h)) - W(g(\tau))}{h} \leq -4\pi + \tilde{C} \epsilon + \frac{3}{4(\tau+C)} W(g(\tau)) + \tilde{C} h.$$

Taking  $\epsilon \rightarrow 0$  in (1.18) gives (1.5).

It remains to prove (1.17). First, let  $\delta > 0$  and  $\bar{j}$ , depending on  $\epsilon$  (and on  $\tau$ ), be given by Theorem 1.8. If  $j > \bar{j}$  and  $\text{Area}_{g(\tau)}(\gamma_s^j(\tau)) > W(g) - \delta$ , then let  $\bigcup_i \Sigma_{s,i}^j(\tau)$  be the collection of minimal spheres given by Theorem 1.8. Combining (1.13), (1.12) with  $Q = \text{Ric}_M$ , and Lemma 1.15 gives

$$(1.19) \quad \begin{aligned} \frac{d}{dt} \Big|_{t=\tau} \text{Area}_{g(t)}(\gamma_s^j(\tau)) &\leq \frac{d}{dt} \Big|_{t=\tau} \text{Area}_{g(t)} \left( \bigcup_i \Sigma_{s,i}^j(\tau) \right) \\ &\quad + \tilde{C} \epsilon \| \text{Ric}_M \|_{C^1} \text{Area}_{g(\tau)}(\gamma_s^j(\tau)) \\ &\leq -4\pi - \frac{\text{Area}_{g(\tau)}(\gamma_s^j(\tau))}{2} \min_M R(\tau) + \tilde{C} \epsilon \\ &\leq -4\pi + \frac{3}{4(\tau+C)} \max_{s_0} \text{Area}_{g(\tau)}(\gamma_{s_0}^j(\tau)) + \tilde{C} \epsilon, \end{aligned}$$

where the last inequality used the lower bound (1.11) for  $R(\tau)$ . Since the metrics  $g(t)$  vary smoothly and every sweepout  $\gamma^j$  has uniformly bounded energy, it is easy to see that  $\text{Area}_{g(\tau+h)}(\gamma_s^j(\tau))$  is a smooth function of  $h$  with a uniform  $C^2$  bound independent of both  $j$  and  $s$  near  $h = 0$  (cf (1.13)). In particular, (1.19) and Taylor expansion give  $\bar{h} > 0$  (independent of  $j$ ) so that (1.17) holds for  $s$  with  $\text{Area}_{g(\tau)}(\gamma_s^j(\tau)) > W(g) - \delta$ . In the remaining case, we have  $\text{Area}(\gamma_s^j(\tau)) \leq W(g) - \delta$  so the continuity of  $g(t)$  implies that (1.17) automatically holds after possibly shrinking  $\bar{h} > 0$ .  $\square$

### 1.6 Parameter spaces

Instead of using the unit interval,  $[0, 1]$ , as the parameter space for the maps in the sweepout and assuming that the maps start and end in point maps, we could have used any compact finite dimensional topological space  $\mathcal{P}$  and required that the maps are constant on  $\partial\mathcal{P}$  (or that  $\partial\mathcal{P} = \emptyset$ ). In this case, let  $\Omega^{\mathcal{P}}$  be the set of continuous maps  $\sigma: \mathbf{S}^2 \times \mathcal{P} \rightarrow M$  so that for each  $t \in \mathcal{P}$  the map  $\sigma(\cdot, t)$  is in  $C^0 \cap W^{1,2}(\mathbf{S}^2, M)$ , the map  $t \rightarrow \sigma(\cdot, t)$  is continuous from  $\mathcal{P}$  to  $C^0 \cap W^{1,2}(\mathbf{S}^2, M)$ , and finally  $\sigma$  maps  $\partial\mathcal{P}$

<sup>6</sup>This follows by combining that  $\text{Area}_{g(\tau)}(\gamma_{s_0}^j(\tau)) \leq \text{Energy}_{g(\tau)}(\gamma_{s_0}^j(\tau))$  by (1.1),  $\max_{s_0} \text{Energy}_{g(\tau)}(\gamma_{s_0}^j(\tau)) \rightarrow W(g(\tau))$ , and  $W(g(\tau)) \leq \max_{s_0} \text{Area}_{g(\tau)}(\gamma_{s_0}^j(\tau))$  by the equivalence of the two definitions of width.

to point maps. Given a map  $\hat{\sigma} \in \Omega^{\mathcal{P}}$ , the homotopy class  $\Omega_{\hat{\sigma}}^{\mathcal{P}} \subset \Omega^{\mathcal{P}}$  is defined to be the set of maps  $\sigma \in \Omega^{\mathcal{P}}$  that are homotopic to  $\hat{\sigma}$  through maps in  $\Omega^{\mathcal{P}}$ . Finally, the width  $W = W(\hat{\sigma})$  is  $\inf_{\sigma \in \Omega_{\hat{\sigma}}^{\mathcal{P}}} \max_{t \in \mathcal{P}} \text{Energy}(\sigma(\cdot, t))$ . With only trivial changes, the same proof yields Theorem 1.8 for these general parameter spaces.<sup>7</sup>

## 2 The energy decreasing map and its consequences

To prove Theorem 1.8, we will first define an energy decreasing map from  $\Omega$  to itself that preserves the homotopy class (ie, maps each  $\Omega_{\beta}$  to itself) and record its key properties. This should be thought of as a generalization of Birkhoff's curve shortening process that plays a similar role when tightening a sweepout by curves; see Birkhoff [3; 4], Croke [10] and Colding and Minicozzi [9].

Throughout this paper, by a *ball*  $B \subset \mathbf{S}^2$ , we will mean a subset of  $\mathbf{S}^2$  and a stereographic projection  $\Pi_B$  so that  $\Pi_B(B) \subset \mathbf{R}^2$  is a ball. Given  $\rho > 0$ , we will let  $\rho B \subset \mathbf{S}^2$  denote  $\Pi_B^{-1}$  of the ball with the same center as  $\Pi_B(B)$  and radius  $\rho$  times that of  $\Pi_B(B)$ .

**Theorem 2.1** *There is a constant  $\epsilon_0 > 0$  and a continuous function  $\Psi: [0, \infty) \rightarrow [0, \infty)$  with  $\Psi(0) = 0$ , both depending on  $M$ , so that given any  $\tilde{\gamma} \in \Omega$  without nonconstant harmonic slices and  $W > 0$ , there exists  $\gamma \in \Omega_{\tilde{\gamma}}$  so that  $\text{Energy}(\gamma(\cdot, t)) \leq \text{Energy}(\tilde{\gamma}(\cdot, t))$  for each  $t$  and so for each  $t$  with  $\text{Energy}(\tilde{\gamma}(\cdot, t)) \geq W/2$ :*

( $B_{\Psi}$ ) *If  $\mathcal{B}$  is any finite collection of disjoint closed balls in  $\mathbf{S}^2$  with*

$$\int_{\bigcup_{\mathcal{B}} B} |\nabla \gamma(\cdot, t)|^2 < \epsilon_0$$

*and  $v: \bigcup_{\mathcal{B}} \frac{1}{8}B \rightarrow M$  is an energy minimizing map equal to  $\gamma(\cdot, t)$  on  $\bigcup_{\mathcal{B}} \frac{1}{8}\partial B$ , then*

$$\int_{\bigcup_{\mathcal{B}} \frac{1}{8}B} |\nabla \gamma(\cdot, t) - \nabla v|^2 \leq \Psi(\text{Energy}(\tilde{\gamma}(\cdot, t)) - \text{Energy}(\gamma(\cdot, t))).$$

The proof of Theorem 2.1 is given in Section 3. The second ingredient that we will need to prove Theorem 1.8 is a compactness result that generalizes compactness of harmonic maps to maps that are closer and closer to being harmonic (this is Proposition 2.2 below).

<sup>7</sup>The main change is in Lemma 3.24 below where the bound 2 for the multiplicity in (1) becomes  $\dim(\mathcal{P}) + 1$ . This follows from the definition of (covering) dimension; see pages 302 and 303 in [21].

## 2.1 Compactness of almost harmonic maps

Our notion of almost harmonic relies on two important properties of harmonic maps from  $\mathbf{S}^2$  to  $M$ . The first is that harmonic maps from  $\mathbf{S}^2$  are conformal and, thus, energy and area are equal; see (A) below. The second is that any harmonic map from a surface is energy minimizing when restricted to balls where the energy is sufficiently small; see (B) below.

In the proposition,  $\epsilon_{SU} > 0$  (depending on  $M$ ) is the small energy constant from Lemma 3.4 in [30], so that we get interior estimates for harmonic maps with energy at most  $\epsilon_{SU}$ . In particular, any nonconstant harmonic map from  $\mathbf{S}^2$  to  $M$  has energy greater than  $\epsilon_{SU}$ .

**Proposition 2.2** *Suppose that  $\epsilon_0, E_0 > 0$  are constants with  $\epsilon_{SU} > \epsilon_0$  and  $u^j: \mathbf{S}^2 \rightarrow M$  is a sequence of  $C^0 \cap W^{1,2}$  maps with  $E_0 \geq \text{Energy}(u^j)$  satisfying:*

(A)  $\text{Area}(u^j) > \text{Energy}(u^j) - 1/j$ .

(B) *For any finite collection  $\mathcal{B}$  of disjoint closed balls in  $\mathbf{S}^2$  with*

$$\int_{\bigcup_{\mathcal{B}} B} |\nabla u^j|^2 < \epsilon_0$$

*there is an energy minimizing map  $v: \bigcup_{\mathcal{B}} \frac{1}{8}B \rightarrow M$  that equals  $u^j$  on  $\bigcup_{\mathcal{B}} \frac{1}{8}\partial B$  with*

$$\int_{\bigcup_{\mathcal{B}} \frac{1}{8}B} |\nabla u^j - \nabla v|^2 \leq 1/j.$$

*If (A) and (B) are satisfied, then a subsequence of the  $u^j$ 's varifold converges to a collection of harmonic maps  $v^0, \dots, v^m: \mathbf{S}^2 \rightarrow M$ .*

Proposition 2.2 will be proven in Appendix B.

One immediate consequence of Proposition 2.2 is a compactness theorem for sequences of *harmonic* maps with bounded energy. This was proven by Jost in Lemma 4.3.1 in [17]. In fact, Parker proved compactness of bounded energy harmonic maps in a stronger topology, with  $C^0$  convergence in addition to  $W^{1,2}$  convergence; see Theorem 2.2 in Parker [22]. Therefore, it is perhaps not surprising that a similar compactness holds for sequences that are closer and closer to being harmonic in the sense above. However, it is useful to keep in mind that Parker has constructed sequences of maps where the Laplacian is going to zero in  $L^1$  and yet there is no convergent subsequence (see Proposition 4.2 in [22]).

Finally, we point out that Proposition 2.2 can be thought of as a discrete version of Palais–Smale Condition (C). Namely, if we have a sequence of maps where the maximal energy decrease from harmonic replacement goes to zero, then a subsequence converges to a collection of harmonic maps.

## 2.2 Constructing good sweepouts from the energy decreasing map on $\Omega$

Given Theorem 2.1 and Proposition 2.2, we will prove Theorem 1.8. Let  $\mathcal{G}^{W+1}$  be the set of collections of harmonic maps from  $\mathbf{S}^2$  to  $M$  so that the sum of the energies is at most  $W + 1$ .

**Proof of Theorem 1.8** Choose a sequence of maps  $\tilde{\gamma}^j \in \Omega_\beta$  with

$$(2.3) \quad \max_{t \in [0,1]} \text{Energy}(\tilde{\gamma}^j(\cdot, t)) < W + \frac{1}{j},$$

and so that  $\tilde{\gamma}^j(\cdot, t)$  is not harmonic unless it is a constant map.<sup>8</sup> We can assume that  $W > 0$  since otherwise  $\text{Area}(\tilde{\gamma}^j(\cdot, t)) \leq \text{Energy}(\tilde{\gamma}^j(\cdot, t)) \rightarrow 0$  and the theorem follows trivially.

Applying Theorem 2.1 to the  $\tilde{\gamma}^j$ 's gives a sequence  $\gamma^j \in \Omega_\beta$  where each  $\gamma^j(\cdot, t)$  has energy at most that of  $\tilde{\gamma}^j(\cdot, t)$ . We will argue by contradiction to show that the  $\gamma^j$ 's have the desired property. Suppose, therefore, that there exist  $j_k \rightarrow \infty$  and  $s_k \in [0, 1]$  with  $d_V(\gamma^{j_k}(\cdot, s_k), \mathcal{G}^{W+1}) \geq \epsilon > 0$  and  $\text{Area}(\gamma^{j_k}(\cdot, s_k)) > W - 1/k$ . Thus, by (2.3) and the fact that  $\text{Energy}(\cdot) \geq \text{Area}(\cdot)$ , we get

$$(2.4) \quad \begin{aligned} & \text{Energy}(\tilde{\gamma}^{j_k}(\cdot, s_k)) - \text{Energy}(\gamma^{j_k}(\cdot, s_k)) \\ & \leq \text{Energy}(\tilde{\gamma}^{j_k}(\cdot, s_k)) - \text{Area}(\gamma^{j_k}(\cdot, s_k)) \\ & \leq 1/k + 1/j_k \rightarrow 0, \end{aligned}$$

and, similarly,  $\text{Energy}(\gamma^{j_k}(\cdot, s_k)) - \text{Area}(\gamma^{j_k}(\cdot, s_k)) \rightarrow 0$ . Using (2.4) in Theorem 2.1 gives:

<sup>8</sup>To do this, first use Lemma D.1 (density of  $C^2$ -sweepouts) to choose  $\tilde{\gamma}_1^j \in \Omega_\beta$  so  $t \rightarrow \tilde{\gamma}_1^j(\cdot, t)$  is continuous from  $[0, 1]$  to  $C^2$  and  $\max_{t \in [0,1]} \text{Energy}(\tilde{\gamma}_1^j(\cdot, t)) < W + 1/(2j)$ . Using stereographic projection, we can view  $\tilde{\gamma}_1^j(\cdot, t)$  as a map from  $\mathbf{R}^2$ . Now fix a  $j$ . The continuity in  $C^2$  gives a uniform bound  $\sup_{t \in [0,1]} \sup_{B_1} |\nabla \tilde{\gamma}_1^j(\cdot, t)|^2 \leq C$  for some  $C$ . Choose  $R > 0$  with  $4\pi C R^2 \leq 1/(2j)$ . Define a map  $\Phi: \mathbf{R}^2 \rightarrow \mathbf{R}^2$  in polar coordinates by:  $\Phi(r, \theta) = (2r, \theta)$  for  $r < R/2$ ,  $\Phi(r, \theta) = (R, \theta)$  for  $R/2 \leq r \leq R$ , and  $\Phi(r, \theta) = (r, \theta)$  for  $R < r$ . Note that  $\Phi$  is homotopic to the identity, is conformal away from the annulus  $B_R \setminus B_{R/2}$ , and on  $B_R \setminus B_{R/2}$  has  $|\partial_r \Phi| = 0$  and  $|d\Phi| \leq 2$ . It follows that  $\tilde{\gamma}^j(\cdot, t) = \tilde{\gamma}_1^j(\cdot, t) \circ \Phi$  is in  $\Omega_\beta$ , satisfies (2.3), and has  $\partial_r \tilde{\gamma}^j(\cdot, t) = 0$  on  $B_R \setminus B_{R/2}$ . Since harmonic maps from  $\mathbf{S}^2$  are conformal [30, Corollary 1.7], any harmonic  $\tilde{\gamma}^j(\cdot, t)$  is constant on  $B_R \setminus B_{R/2}$  and, thus, constant on  $\mathbf{S}^2$  by unique continuation [31, Theorem 1.1].

(B) If  $\mathcal{B}$  is any collection of disjoint closed balls in  $\mathbf{S}^2$  with

$$\int_{\bigcup_{\mathcal{B}} B} |\nabla \gamma^{j_k}(\cdot, s_k)|^2 < \epsilon_0$$

and  $v: \bigcup_{\mathcal{B}} \frac{1}{8}B \rightarrow M$  is an energy minimizing map that equals  $\gamma^{j_k}(\cdot, s_k)$  on  $\bigcup_{\mathcal{B}} \frac{1}{8}\partial B$ , then

$$\int_{\bigcup_{\mathcal{B}} \frac{1}{8}B} |\nabla \gamma^{j_k}(\cdot, s_k) - \nabla v|^2 \leq \Psi(1/k + 1/j_k) \rightarrow 0.$$

Therefore, we can apply Proposition 2.2 to get that a subsequence of the  $\gamma^{j_k}(\cdot, s_k)$ 's varifold converges to a collection of harmonic maps. However, this contradicts the lower bound for the varifold distance to  $\mathcal{G}^{W+1}$ , thus completing the proof.  $\square$

### 3 Constructing the energy decreasing map

The purpose of this section is to prove Theorem 2.1 by constructing a sort of “discrete gradient flow” on the space  $\Omega$  of sweepouts. The main tool will be a local replacement process, known as harmonic replacement, that is similar in spirit to Birkhoff’s curve shortening process. The next three subsections will be devoted to developing the properties of this local replacement process that will be used in Section 3.4 when we prove Theorem 2.1.

#### 3.1 Harmonic replacement

The energy decreasing map from  $\Omega$  to itself will be given by a repeated replacement procedure. At each step, we replace a map  $u$  by a map  $H(u)$  that coincides with  $u$  outside a ball and inside the ball is equal to an energy-minimizing map with the same boundary values as  $u$ . This is often referred to as *harmonic replacement*.

One of the key properties that makes harmonic replacement useful is that the energy functional is strictly convex on small energy maps. Namely, Theorem 3.1 below gives a uniform lower bound for the gap in energy between a harmonic map and a  $W^{1,2}$  map with the same boundary values; see Appendix C for the proof.

**Theorem 3.1** *There exists a constant  $\epsilon_1 > 0$  (depending on  $M$ ) so that if  $u$  and  $v$  are  $W^{1,2}$  maps from  $B_1 \subset \mathbf{R}^2$  to  $M$ ,  $u$  and  $v$  agree on  $\partial B_1$ , and  $v$  is weakly harmonic with energy at most  $\epsilon_1$ , then*

$$(3.2) \quad \int_{B_1} |\nabla u|^2 - \int_{B_1} |\nabla v|^2 \geq \frac{1}{2} \int_{B_1} |\nabla v - \nabla u|^2.$$

An immediate corollary of Theorem 3.1 is uniqueness of solutions to the Dirichlet problem for small energy maps (and also that any such harmonic map minimizes energy).

**Corollary 3.3** *Let  $\epsilon_1 > 0$  be as in Theorem 3.1. If  $u_1$  and  $u_2$  are  $W^{1,2}$  weakly harmonic maps from  $B_1 \subset \mathbf{R}^2$  to  $M$ , both with energy at most  $\epsilon_1$ , and they agree on  $\partial B_1$ , then  $u_1 = u_2$ .*

### 3.2 Continuity of harmonic replacement on $C^0(\overline{B_1}) \cap W^{1,2}(B_1)$

The second consequence of Theorem 3.1 is that harmonic replacement is continuous as a map from  $C^0(\overline{B_1}) \cap W^{1,2}(B_1)$  to itself if we restrict to small energy maps. (The norm on  $C^0(\overline{B_1}) \cap W^{1,2}(B_1)$  is the sum of the sup norm and the  $W^{1,2}$  norm.)

**Corollary 3.4** *Let  $\epsilon_1 > 0$  be as in Theorem 3.1 and set*

$$\mathcal{M} = \{u \in C^0(\overline{B_1}, M) \cap W^{1,2}(B_1, M) \mid \text{Energy}(u) \leq \epsilon_1\}.$$

*Given  $u \in \mathcal{M}$ , there is a unique energy minimizing map  $w$  equal to  $u$  on  $\partial B_1$  and  $w$  is in  $\mathcal{M}$ . Furthermore, there exists  $C$  depending on  $M$  so that if  $u_1, u_2 \in \mathcal{M}$  with corresponding energy minimizing maps  $w_1, w_2$ , and we set  $\text{Energy} = \text{Energy}(u_1) + \text{Energy}(u_2)$ , then*

$$(3.5) \quad |\text{Energy}(w_1) - \text{Energy}(w_2)| \\ \leq C \|u_1 - u_2\|_{C^0(\overline{B_1})} \text{Energy} + C \|\nabla u_1 - \nabla u_2\|_{L^2(B_1)} \text{Energy}^{1/2}.$$

*Finally, the map from  $u$  to  $w$  is continuous as a map from  $C^0(\overline{B_1}) \cap W^{1,2}(B_1)$  to itself.*

In the proof, we will use that since  $M$  is smooth, compact and embedded, there exists a  $\delta > 0$  so that for each  $x$  in the  $\delta$ -tubular neighborhood  $M_\delta$  of  $M$  in  $\mathbf{R}^N$ , there is a unique closest point  $\Pi(x) \in M$  and so the map  $x \rightarrow \Pi(x)$  is smooth.  $\Pi$  is called *nearest point projection*. Furthermore, for any  $x \in M$ , we have  $|d\Pi_x(V)| \leq |V|$ . Therefore, there is a constant  $C_\Pi$  depending on  $M$  so that for any  $x \in M_\delta$ , we have  $|d\Pi_x(V)| \leq (1 + C_\Pi|x - \Pi(x)|)|V|$ . In particular, we can choose  $\hat{\delta} \in (0, \delta)$  so that  $|d\Pi_x(V)|^2 \leq 2|V|^2$  for any  $x \in M_{\hat{\delta}}$  and  $V \in \mathbf{R}^N$ .

**Proof of Corollary 3.4** The existence of an energy minimizing map  $w \in W^{1,2}(B_1)$  was proven by Morrey in [20]; by Corollary 3.3,  $w$  is unique. The continuity of  $w$  on  $\overline{B_1}$  is the main theorem of [27].<sup>9</sup> It follows that  $w \in \mathcal{M}$ .

<sup>9</sup>Continuity also essentially follows from the boundary regularity of Schoen and Uhlenbeck [33], except that [33] assumes  $C^{2,\alpha}$  regularity of the boundary data.



**Step 1** (Energy( $w$ ) is uniformly continuous) We can assume that  $\|u_1 - u_2\|_{C^0(\overline{B_1})} \leq \widehat{\delta}$ , since (3.5) holds with  $C = 1/\widehat{\delta}$  if  $\|u_1 - u_2\|_{C^0(\overline{B_1})} \geq \widehat{\delta}$ . Define a map  $v_1$  by

$$v_1 = \Pi \circ (w_2 + (u_1 - u_2)) ,$$

so that  $v_1$  maps to  $M$  and agrees with  $u_1$  on  $\partial B_1$ . Using that  $|d\Pi_x(V)| \leq |V|$  for  $x \in M$  and  $w_2$  maps to  $M$ , we can estimate the energy of  $v_1$  by

$$(3.6) \quad \begin{aligned} \text{Energy}(v_1) &\leq (1 + C_\Pi \|u_1 - u_2\|_{C^0(\overline{B_1})})^2 \\ &\times (\text{Energy}(w_2) + 2(\text{Energy}(w_2) \text{Energy}(u_1 - u_2))^{1/2} \\ &\quad + \text{Energy}(u_1 - u_2)) , \end{aligned}$$

where  $C_\Pi$  is the Lipschitz norm of  $d\Pi$  in  $M_\delta$ . Since  $v_1$  and  $w_1$  agree on  $\partial B_1$ , Corollary 3.3 yields  $\text{Energy}(w_1) \leq \text{Energy}(v_1)$ . By symmetry, we can assume that  $\text{Energy}(w_2) \leq \text{Energy}(w_1)$  so that (3.6) implies (3.5).

**Step 2** (The continuity of  $u \rightarrow \widehat{w}$ ) Suppose that  $u, u_j$  are in  $\mathcal{M}$  with  $u_j \rightarrow u$  in  $C^0(\overline{B_1}) \cap W^{1,2}(B_1)$  and  $w$  and  $w_j$  are the corresponding energy minimizing maps.

We will first show that  $w_j \rightarrow w$  in  $W^{1,2}(B_1)$ . To do this, set

$$v_j = \Pi \circ (w + (u_j - u)) ,$$

so that  $v_j$  maps to  $M$  and agrees with  $u_j$  on  $\partial B_1$ . Arguing as in (3.6) and using that  $\text{Energy}(w_j) \rightarrow \text{Energy}(w)$  by Step 1, we get that  $(\text{Energy}(v_j) - \text{Energy}(w_j)) \rightarrow 0$ . Therefore, applying Theorem 3.1 to  $w_j, v_j$  gives that  $\|w_j - v_j\|_{W^{1,2}(B_1)} \rightarrow 0$ . Since  $\|u_j - u\|_{C^0(\overline{B_1}) \cap W^{1,2}(B_1)} \rightarrow 0$  and  $\Pi \circ w = w$ , it follows that  $\|w - v_j\|_{W^{1,2}(B_1)} \rightarrow 0$ . The triangle inequality gives  $\|w - w_j\|_{W^{1,2}(B_1)} \rightarrow 0$ .

Finally, we will argue by contradiction to see that  $w_j \rightarrow w$  in  $C^0(\overline{B_1})$ . Suppose instead that there is a subsequence (still denoted  $w_j$ ) with

$$(3.7) \quad \|w_j - w\|_{C^0(\overline{B_1})} \geq \epsilon > 0 .$$

Using the uniform energy bound for the  $w_j$ 's together with interior estimates for energy minimizing maps of [32] (and the Arzela–Ascoli theorem), we can pass to a further subsequence so that the  $w_j$ 's converge uniformly in  $C^2$  on any compact subset  $K \subset B_1$ . Finally, as remarked in the proof of the main theorem in [27], Proposition 1 and Remark 1 of [27] imply that the  $w_j$ 's are also equicontinuous near  $\partial B_1$ , so Arzela–Ascoli gives a further subsequence that converges uniformly on  $\overline{B_1}$  to a harmonic map  $w_\infty$  that agrees with  $w$  on the boundary. However, (3.7) implies that  $\|w - w_\infty\|_{C^0(\overline{B_1})} \geq \epsilon > 0$  which contradicts the uniqueness of small energy harmonic maps. This completes the proof.  $\square$

Corollary 3.4 gives another proof that the width is positive when the homotopy class is nontrivial or, equivalently, that if  $\max_t \text{Energy}(\sigma(\cdot, t))$  is sufficiently small (depending on  $M$ ), then  $\sigma$  is homotopically trivial. Namely, since  $t \rightarrow \sigma(\cdot, t)$  is continuous from  $[0, 1]$  to  $C^0$ , we can choose  $r > 0$  so that  $\sigma(\cdot, t)$  maps the ball  $B_r(p) \subset \mathbf{S}^2$  into a convex geodesic ball  $B^t$  in  $M$  for every  $t$ . If each  $\sigma(\cdot, t)$  has energy less than  $\epsilon_1 > 0$  given by Corollary 3.4, then replacing  $\sigma(\cdot, t)$  outside  $B_r(p)$  by the energy minimizing map with the same boundary values gives a homotopic sweepout  $\tilde{\sigma}$ . Moreover, the entire image of  $\tilde{\sigma}(\cdot, t)$  is contained in the convex ball  $B^t$  by the maximum principle.<sup>10</sup> It follows that  $\tilde{\sigma}$  is homotopically trivial by contracting each  $\tilde{\sigma}(\cdot, t)$  to the point  $\sigma(p, t)$  via a geodesic homotopy.

### 3.3 Uniform continuity of energy improvement on $W^{1,2}$

It will be convenient to introduce some notation for the next lemma. Namely, given a  $C^0 \cap W^{1,2}$  map  $u$  from  $\mathbf{S}^2$  to  $M$  and a finite collection  $\mathcal{B}$  of disjoint closed balls in  $\mathbf{S}^2$  so the energy of  $u$  on  $\bigcup_{\mathcal{B}} B$  is at most  $\epsilon_1/3$ , let  $H(u, \mathcal{B}): \mathbf{S}^2 \rightarrow M$  denote the map that coincides with  $u$  on  $\mathbf{S}^2 \setminus \bigcup_{\mathcal{B}} B$  and on  $\bigcup_{\mathcal{B}} B$  is equal to the energy minimizing map from  $\bigcup_{\mathcal{B}} B$  to  $M$  that agrees with  $u$  on  $\bigcup_{\mathcal{B}} \partial B$ . To keep the notation simple, we will set  $H(u, \mathcal{B}_1, \mathcal{B}_2) = H(H(u, \mathcal{B}_1), \mathcal{B}_2)$ . Finally, if  $\alpha \in (0, 1]$ , then  $\alpha\mathcal{B}$  will denote the collection of concentric balls but whose radii are shrunk by the factor  $\alpha$ .

In general,  $H(u, \mathcal{B}_1, \mathcal{B}_2)$  is not the same as  $H(u, \mathcal{B}_2, \mathcal{B}_1)$ . This matters in the proof of Theorem 2.1, where harmonic replacement on either  $\frac{1}{2}\mathcal{B}_1$  or  $\frac{1}{2}\mathcal{B}_2$  decreases the energy of  $u$  by a definite amount. The next lemma (see (3.9)) shows that the energy goes down a definite amount regardless of the order that we do the replacements. The second inequality bounds the possible decrease in energy from applying harmonic replacement on  $H(u, \mathcal{B}_1)$  in terms of the possible decrease from harmonic replacement on  $u$ .

**Lemma 3.8** *There is a constant  $\kappa > 0$  (depending on  $M$ ) so that if  $u: \mathbf{S}^2 \rightarrow M$  is in  $C^0 \cap W^{1,2}$  and  $\mathcal{B}_1, \mathcal{B}_2$  are each finite collections of disjoint closed balls in  $\mathbf{S}^2$  so that the energy of  $u$  on each  $\bigcup_{\mathcal{B}_i} B$  is at most  $\epsilon_1/3$ , then*

$$(3.9) \quad \text{Energy}(u) - \text{Energy}(H(u, \mathcal{B}_1, \mathcal{B}_2)) \\ \geq \kappa \left( \text{Energy}(u) - \text{Energy}\left(H(u, \frac{1}{2}\mathcal{B}_2)\right) \right)^2.$$

<sup>10</sup>This follows from Lemma 4.1.3 in [17] which requires that  $\sigma(\cdot, t)$  is homotopic to a map in  $B^t$  and this follows from the small energy bound and the uniform lower bound for the energy of any homotopically nontrivial map from  $\mathbf{S}^2$  given, eg, in the first line of the proof of Proposition 2 on page 143 of [35].

Furthermore, for any  $\mu \in [1/8, 1/2]$ , we have

$$(3.10) \quad \frac{(\text{Energy}(u) - \text{Energy}(H(u, \mathcal{B}_1)))^{1/2}}{\kappa} + \text{Energy}(u) - \text{Energy}(H(u, 2\mu\mathcal{B}_2)) \\ \geq \text{Energy}(H(u, \mathcal{B}_1)) - \text{Energy}(H(u, \mu\mathcal{B}_2)).$$

We will prove Lemma 3.8 by constructing comparison maps with the same boundary values and using the minimizing property of small energy harmonic maps to get upper bounds for the energy. The following lemma will be used to construct the comparison maps.

**Lemma 3.11** *There exists  $\tau > 0$  (depending on  $M$ ) so that if  $f, g: \partial B_R \rightarrow M$  are  $C^0 \cap W^{1,2}$  maps that agree at one point and satisfy*

$$(3.12) \quad R \int_{\partial B_R} |f' - g'|^2 \leq \tau^2,$$

then there exists  $\rho \in (0, R/2]$  and a  $C^0 \cap W^{1,2}$  map  $w: B_R \setminus B_{R-\rho} \rightarrow M$  so that

$$(3.13) \quad w(R - \rho, \theta) = f(R, \theta) \text{ and } w(R, \theta) = g(R, \theta),$$

and  $\int_{B_R \setminus B_{R-\rho}} |\nabla w|^2 \leq 17\sqrt{2} (R \int_{\partial B_R} (|f'|^2 + |g'|^2))^{1/2} (R \int_{\partial B_R} |f' - g'|^2)^{1/2}$ .

**Proof** Let  $\Pi$  and  $\delta > \hat{\delta} > 0$  (depending on  $M$ ) be as in the proof of Corollary 3.4 and set  $\tau = \hat{\delta}/\sqrt{2\pi}$ . Since  $f - g$  vanishes somewhere on  $\partial B_R$ , integrating (3.12) gives  $\max |f - g| \leq \hat{\delta}$ .

Since the statement is scale-invariant, it suffices to prove the case  $R = 1$ . Set  $\rho^2 = \int_{\mathbb{S}^1} |f' - g'|^2 / (8 \int_{\mathbb{S}^1} (|f'|^2 + |g'|^2)) \leq 1/4$  and define  $\hat{w}: B_1 \setminus B_{1-\rho} \rightarrow \mathbf{R}^N$  by

$$\hat{w}(r, \theta) = f(\theta) + \left( \frac{r + \rho - 1}{\rho} \right) (g(\theta) - f(\theta)).$$

Observe that  $\hat{w}$  satisfies (3.13). Furthermore, since  $f - g$  vanishes somewhere on  $\mathbb{S}^1$ , we can use Wirtinger's inequality  $\int_{\mathbb{S}^1} |f - g|^2 \leq 4 \int_{\mathbb{S}^1} |(f - g)'|^2$  to bound

$\int_{\mathcal{B}_1 \setminus \mathcal{B}_{1-\rho}} |\nabla \widehat{w}|^2$  by

$$\begin{aligned} & \int_{\mathcal{B}_1 \setminus \mathcal{B}_{1-\rho}} |\nabla \widehat{w}|^2 \\ & \leq \int_{1-\rho}^1 \left( \frac{1}{\rho^2} \int_0^{2\pi} |f-g|^2(\theta) d\theta + \frac{1}{r^2} \int_0^{2\pi} (|f'|^2 + |g'|^2)(\theta) d\theta \right) r dr \\ & \leq \frac{4}{\rho} \int_0^{2\pi} |f'-g'|^2(\theta) d\theta + 2\rho \int_0^{2\pi} (|f'|^2 + |g'|^2)(\theta) d\theta \\ & = 17/\sqrt{2} \left( \int_{\mathbb{S}^1} |f'-g'|^2 \int_{\mathbb{S}^1} (|f'|^2 + |g'|^2) \right)^{1/2}. \end{aligned}$$

Since  $|f-g| \leq \widehat{\delta}$ , the image of  $\widehat{w}$  is contained in  $\overline{M_{\widehat{\delta}}}$  where we have  $|d\Pi|^2 \leq 2$ . Thus, if we set  $w = \Pi \circ \widehat{w}$ , then the energy of  $w$  is at most twice the energy of  $\widehat{w}$ .  $\square$

**Proof of Lemma 3.8** We will index the balls in  $\mathcal{B}_1$  by  $\alpha$  and use  $j$  for the balls in  $\mathcal{B}_2$ ; ie, let  $\mathcal{B}_1 = \{B_\alpha^1\}$  and  $\mathcal{B}_2 = \{B_j^2\}$ . The key point is that, by Corollary 3.4, small energy harmonic maps minimize energy. Using this, we get upper bounds for the energy of the harmonic replacement by cutting and pasting to construct comparison functions with the same boundary values.

Observe that the total energy of  $u$  on the union of the balls in  $\mathcal{B}_1 \cup \mathcal{B}_2$  is at most  $2\epsilon_1/3$ . Since harmonic replacement on  $\mathcal{B}_1$  does not change the map outside these balls and is energy nonincreasing, it follows that the total energy of  $H(u, \mathcal{B}_1)$  on  $\mathcal{B}_2$  is at most  $2\epsilon_1/3$ .

**Proof of (3.9)** We will divide  $\mathcal{B}_2$  into two disjoint subsets,  $\mathcal{B}_{2,+}$  and  $\mathcal{B}_{2,-}$ , and argue separately, depending on which of these accounts for more of the decrease in energy after harmonic replacement. Namely, set

$$\mathcal{B}_{2,+} = \left\{ B_j^2 \in \mathcal{B}_2 \mid \frac{1}{2} B_j^2 \subset B_\alpha^1 \text{ for some } B_\alpha^1 \in \mathcal{B}_1 \right\} \quad \text{and} \quad \mathcal{B}_{2,-} = \mathcal{B}_2 \setminus \mathcal{B}_{2,+}.$$

Since the balls in  $\mathcal{B}_2$  are disjoint, it follows that

$$\begin{aligned} & \text{Energy}(u) - \text{Energy}(H(u, \frac{1}{2} \mathcal{B}_2)) \\ & = \left( \text{Energy}(u) - \text{Energy}(H(u, \frac{1}{2} \mathcal{B}_{2,-})) \right) + \left( \text{Energy}(u) - \text{Energy}(H(u, \frac{1}{2} \mathcal{B}_{2,+})) \right). \end{aligned}$$

**Case 1** Suppose that

$$\text{Energy}(u) - \text{Energy}(H(u, \frac{1}{2} \mathcal{B}_{2,+})) \geq \left( \text{Energy}(u) - \text{Energy}(H(u, \frac{1}{2} \mathcal{B}_2)) \right) / 2.$$

Since the balls in  $\frac{1}{2}\mathcal{B}_{2,+}$  are contained in balls in  $\mathcal{B}_1$  and harmonic replacements minimize energy, we get

$$\text{Energy}(H(u, \mathcal{B}_1, \mathcal{B}_2)) \leq \text{Energy}(H(u, \mathcal{B}_1)) \leq \text{Energy}(H(u, \frac{1}{2}\mathcal{B}_{2,+})),$$

so that

$$\begin{aligned} (\text{Energy}(u) - \text{Energy}(H(u, \frac{1}{2}\mathcal{B}_{2,+}))) / 2 &\leq \text{Energy}(u) - \text{Energy}(H(u, \frac{1}{2}\mathcal{B}_{2,+})) \\ &\leq \text{Energy}(u) - \text{Energy}(H(u, \mathcal{B}_1, \mathcal{B}_2)). \end{aligned}$$

**Case 2** Suppose now that

$$\begin{aligned} (3.14) \quad \text{Energy}(u) - \text{Energy}(H(u, \frac{1}{2}\mathcal{B}_{2,-})) \\ \geq \frac{1}{2} (\text{Energy}(u) - \text{Energy}(H(u, \frac{1}{2}\mathcal{B}_{2,+}))). \end{aligned}$$

Let  $\tau > 0$  be given by Lemma 3.11. We can assume that

$$(3.15) \quad 9 \int_{\mathbb{S}^2} |\nabla H(u, \mathcal{B}_1) - \nabla u|^2 \leq \tau^2,$$

since otherwise Theorem 3.1 gives (3.9) with  $\kappa = \tau^2/\epsilon_1^2$ . The key is to show for  $B_j^2 \in \mathcal{B}_{2,-}$  that

$$\begin{aligned} (3.16) \quad \int_{B_j^2} |\nabla H(u, \mathcal{B}_1)|^2 - \int_{B_j^2} |\nabla H(u, \mathcal{B}_1, B_j^2)|^2 \\ \geq \int_{\frac{1}{2}B_j^2} |\nabla u|^2 - \int_{\frac{1}{2}B_j^2} \left| \nabla H(u, \frac{1}{2}B_j^2) \right|^2 \\ - C \left( \int_{B_j^2} |\nabla u|^2 + \left| \nabla H(u, \mathcal{B}_1) \right|^2 \right)^{1/2} \left( \int_{B_j^2} \left| \nabla(u - H(u, \mathcal{B}_1)) \right|^2 \right)^{1/2}, \end{aligned}$$

where  $C$  is a universal constant. Namely, summing (3.16) over  $\mathcal{B}_{2,-}$  and using the inequality  $|\sum a_j b_j| \leq (\sum a_j^2)^{1/2} (\sum b_j^2)^{1/2}$ , the bound for the energy of  $u$  in  $\mathcal{B}_1 \cup \mathcal{B}_2$ , and Theorem 3.1 to relate the energy of  $u - H(u, \mathcal{B}_1)$  to  $\text{Energy}(u) - \text{Energy}(H(u, \mathcal{B}_1))$  gives

$$\begin{aligned} (3.17) \quad \text{Energy}(u) - \text{Energy}(H(u, \frac{1}{2}\mathcal{B}_{2,-})) &\leq \text{Energy}(H(u, \mathcal{B}_1)) - \text{Energy}(H(u, \mathcal{B}_1, \mathcal{B}_{2,-})) \\ &\quad + C \epsilon_1^{1/2} (\text{Energy}(u) - \text{Energy}(H(u, \mathcal{B}_1)))^{1/2} \\ &\leq \delta_{\text{Energy}} + C \epsilon_1^{1/2} \delta_{\text{Energy}}^{1/2} \leq (C + 1) \epsilon_1^{1/2} \delta_{\text{Energy}}^{1/2}, \end{aligned}$$

where we have set  $\delta_{\text{Energy}} = \text{Energy}(u) - \text{Energy}(H(u, \mathcal{B}_1, \mathcal{B}_2))$  in the last line and the last inequality used that  $\delta_{\text{Energy}} \leq 2\epsilon_1/3 < \epsilon_1$ . Combining (3.14) with (3.17) gives (3.9).

To complete Case 2, we must prove (3.16). After translation, we can assume that  $B_j^2$  is the ball  $B_R$  of radius  $R$  about 0 in  $\mathbf{R}^2$ . Set  $u_1 = H(u, \mathcal{B}_1)$  and apply the co-area formula to get  $r \in [3R/4, R]$  (in fact, a set of  $r$ 's of measure at least  $R/36$ ) with

$$\begin{aligned} \int_{\partial B_r} |\nabla u_1 - \nabla u|^2 &\leq \frac{9}{R} \int_{3R/4}^R \left( \int_{\partial B_s} |\nabla u_1 - \nabla u|^2 \right) ds \\ &\leq \frac{9}{r} \int_{B_R} |\nabla u_1 - \nabla u|^2, \\ \int_{\partial B_r} (|\nabla u_1|^2 + |\nabla u|^2) &\leq \frac{9}{R} \int_{3R/4}^R \left( \int_{\partial B_s} |\nabla u_1|^2 + |\nabla u|^2 \right) ds \\ &\leq \frac{9}{r} \int_{B_R} (|\nabla u_1|^2 + |\nabla u|^2). \end{aligned}$$

Since  $B_j^2 \in \mathcal{B}_{2,-}$  and  $r > R/2$ , the circle  $\partial B_r$  is not contained in any of the balls in  $\mathcal{B}_1$ . It follows that  $\partial B_r$  contains at least one point outside  $\bigcup_{\mathcal{B}_1} B$  and, thus, there is a point in  $\partial B_r$  where  $u = u_1$ . This and (3.15) allow us to apply Lemma 3.11 to get  $\rho \in (0, r/2]$  and a map  $w: B_r \setminus B_{r-\rho} \rightarrow M$  with  $w(r, \theta) = u_1(r, \theta)$ ,  $w(r - \rho, \theta) = u(r, \theta)$ , and

$$\begin{aligned} (3.18) \quad \int_{B_r \setminus B_{r-\rho}} |\nabla w|^2 \\ \leq C \left( \int_{B_j^2} |\nabla u|^2 + |\nabla H(u, \mathcal{B}_1)|^2 \right)^{1/2} \left( \int_{B_j^2} |\nabla(u - H(u, \mathcal{B}_1))|^2 \right)^{1/2}. \end{aligned}$$

Observe that the map  $x \rightarrow H(u, B_r)(r x / (r - \rho))$  maps  $B_{r-\rho}$  to  $M$  and agrees with  $w$  on  $\partial B_{r-\rho}$ . Therefore, the map from  $B_R$  to  $M$  which is equal to  $u_1$  on  $B_R \setminus B_r$ , is equal to  $w$  on  $B_r \setminus B_{r-\rho}$ , and is equal to  $H(u, B_r)(r \cdot / (r - \rho))$  on  $B_{r-\rho}$  gives an upper bound for the energy of  $H(u_1, B_R)$

$$\int_{B_R} |\nabla H(u_1, B_R)|^2 \leq \int_{B_R \setminus B_r} |\nabla u_1|^2 + \int_{B_r \setminus B_{r-\rho}} |\nabla w|^2 + \int_{B_r} |\nabla H(u, B_r)|^2.$$

Using (3.18) and that  $||\nabla u_1|^2 - |\nabla u|^2| \leq (|\nabla u| + |\nabla u_1|) |\nabla(u - u_1)|$ , we get

$$\begin{aligned} & \int_{B_R} |\nabla u_1|^2 - \int_{B_R} |\nabla H(u_1, B_R)|^2 \\ & \geq \int_{B_r} |\nabla u_1|^2 - \int_{B_r} |\nabla H(u, B_r)|^2 - \int_{B_r \setminus B_{r-\rho}} |\nabla w|^2 \\ & \geq \int_{B_r} |\nabla u|^2 - \int_{B_r} |\nabla H(u, B_r)|^2 \\ & \quad - C \left( \int_{B_r} |\nabla u|^2 + |\nabla u_1|^2 \right)^{1/2} \left( \int_{B_r} |\nabla(u - u_1)|^2 \right)^{1/2}. \end{aligned}$$

Since  $\int_{B_{R/2}} |\nabla H(u, B_{R/2})|^2 \leq \int_{B_{R/2} \setminus B_r} |\nabla u|^2 + \int_{B_r} |\nabla H(u, B_r)|^2$ , we get (3.16).

**Proof of (3.10)** We will argue similarly with a few small modifications that we will describe. This time, let  $\mathcal{B}_{2,+} \subset \mathcal{B}_2$  be the balls  $B_j^2$  with  $\mu B_j^2$  contained in some  $B_\alpha^1 \in \mathcal{B}_1$ . It follows that harmonic replacement on  $\mu \mathcal{B}_{2,+}$  does not change  $H(u, \mathcal{B}_1)$  and, thus,

$$(3.19) \quad \text{Energy}(H(u, \mathcal{B}_1)) = \text{Energy}(H(u, \mathcal{B}_1, \mu \mathcal{B}_{2,+})).$$

Again, we can assume that (3.15) holds. Suppose now that  $B_j^2 \in \mathcal{B}_{2,-}$ . Arguing as in the proof of (3.16) (switching the roles of  $u$  and  $H(u, \mathcal{B}_1)$ ), we get

$$\begin{aligned} & \int_{B_j^2} |\nabla u|^2 - \int_{B_j^2} |\nabla H(u, 2\mu B_j^2)|^2 \\ & \geq \int_{\mu B_j^2} |\nabla H(u, \mathcal{B}_1)|^2 - \int_{\mu B_j^2} |\nabla H(u, \mathcal{B}_1, \mu B_j^2)|^2 \\ & \quad - C \left( \int_{B_j^2} |\nabla u|^2 + |\nabla H(u, \mathcal{B}_1)|^2 \right)^{1/2} \left( \int_{B_j^2} |\nabla(u - H(u, \mathcal{B}_1))|^2 \right)^{1/2}. \end{aligned}$$

Summing this over  $\mathcal{B}_{2,-}$  and arguing as for (3.17) gives

$$\begin{aligned} & \int |\nabla u|^2 - \int |\nabla H(u, 2\mu \mathcal{B}_2)|^2 \geq \int |\nabla H(u, \mathcal{B}_1)|^2 - \int |\nabla H(u, \mathcal{B}_1, \mu \mathcal{B}_{2,-})|^2 \\ & \quad - C \epsilon_1^{1/2} (\text{Energy}(u) - \text{Energy}(H(u, \mathcal{B}_1)))^{1/2}. \end{aligned}$$

Combining this with (3.19) completes the proof. □

### 3.4 Constructing the map from $\tilde{\gamma}$ to $\gamma$

Now that we have developed the basic properties of harmonic replacement, we are ready to prove Theorem 2.1 in this subsection.

We will construct  $\gamma(\cdot, t)$  from  $\tilde{\gamma}(\cdot, t)$  by harmonic replacement on a family of balls in  $\mathbf{S}^2$  varying continuously in  $t$ . The balls will be chosen in Lemma 3.24 below. Throughout this subsection,  $\epsilon_1 > 0$  will be the small energy constant (depending on  $M$ ) given by Theorem 3.1.

Given  $\sigma \in \Omega$  and  $\epsilon \in (0, \epsilon_1]$ , define the maximal improvement from harmonic replacement on families of balls with energy at most  $\epsilon$  by

$$e_{\sigma, \epsilon}(t) = \sup_{\mathcal{B}} \left\{ \text{Energy}(\sigma(\cdot, t)) - \text{Energy}(H(\sigma(\cdot, t), \frac{1}{2}\mathcal{B})) \right\},$$

where the supremum is over all finite collections  $\mathcal{B}$  of disjoint closed balls where the total energy of  $\sigma(\cdot, t)$  on  $\mathcal{B}$  is at most  $\epsilon$ . Observe that  $e_{\sigma, \epsilon}(t)$  is nonnegative, monotone nondecreasing in  $\epsilon$ , and is positive if  $\sigma(\cdot, t)$  is not harmonic.

**Lemma 3.20** *If  $\sigma(\cdot, t)$  is not harmonic and  $\epsilon \in (0, \epsilon_1]$ , then there is an open interval  $I^t$  containing  $t$  so that  $e_{\sigma, \epsilon/2}(s) \leq 2e_{\sigma, \epsilon}(t)$  for all  $s$  in the double interval  $2I^t$ .*

**Proof** By (3.5) in Corollary 3.4, there exists  $\delta_1 > 0$  (depending on  $t$ ) so that if

$$(3.21) \quad \|\sigma(\cdot, t) - \sigma(\cdot, s)\|_{C^0 \cap W^{1,2}} < \delta_1$$

and  $\mathcal{B}$  is a finite collection of disjoint closed balls where both  $\sigma(\cdot, t)$  and  $\sigma(\cdot, s)$  have energy at most  $\epsilon_1$ , then

$$(3.22) \quad \left| \text{Energy}(H(\sigma(\cdot, s), \frac{1}{2}\mathcal{B})) - \text{Energy}(H(\sigma(\cdot, t), \frac{1}{2}\mathcal{B})) \right| \leq e_{\sigma, \epsilon}(t)/2.$$

Here we have used that  $e_{\sigma, \epsilon}(t) > 0$  since  $\sigma(\cdot, t)$  is not harmonic. Since  $t \rightarrow \sigma(\cdot, t)$  is continuous as a map to  $C^0 \cap W^{1,2}$ , we can choose  $I^t$  so that for all  $s \in 2I^t$  (3.21) holds and

$$(3.23) \quad \frac{1}{2} \int_{\mathbf{S}^2} \left| |\nabla \sigma(\cdot, t)|^2 - |\nabla \sigma(\cdot, s)|^2 \right| \leq \min \left\{ \frac{\epsilon}{2}, \frac{e_{\sigma, \epsilon}(t)}{2} \right\}.$$

Suppose now that  $s \in 2I^t$  and the energy of  $\sigma(\cdot, s)$  is at most  $\epsilon/2$  on a collection  $\mathcal{B}$ . It follows from (3.23) that the energy of  $\sigma(\cdot, t)$  is at most  $\epsilon$  on  $\mathcal{B}$ . Combining (3.22) and (3.23) gives

$$\begin{aligned} & \left| \text{Energy}(\sigma(\cdot, s)) - \text{Energy}(H(\sigma(\cdot, s), \frac{1}{2}\mathcal{B})) \right. \\ & \quad \left. - \text{Energy}(\sigma(\cdot, t)) + \text{Energy}(H(\sigma(\cdot, t), \frac{1}{2}\mathcal{B})) \right| \leq e_{\sigma, \epsilon}(t). \end{aligned}$$

Since this applies to any such  $\mathcal{B}$ , we get that  $e_{\sigma, \epsilon/2}(s) \leq 2e_{\sigma, \epsilon}(t)$ .  $\square$



Given a sweepout with no harmonic slices, the next lemma constructs finitely many collections of balls so that harmonic replacement on at least one of these collections strictly decreases the energy. In addition, each collection consists of finitely many pairwise disjoint closed balls.

**Lemma 3.24** *If  $W > 0$  and  $\tilde{\gamma} \in \Omega$  has no nonconstant harmonic slices, then we get an integer  $m$  (depending on  $\tilde{\gamma}$ ),  $m$  collections of balls  $\mathcal{B}_1, \dots, \mathcal{B}_m$  in  $\mathbf{S}^2$  where the balls in each collection  $\mathcal{B}_j$  are pairwise disjoint, and  $m$  continuous functions  $r_1, \dots, r_m: [0, 1] \rightarrow [0, 1]$  so that for each  $t$ :*

- (1) *At most two  $r_j(t)$ 's are positive and  $\sum_{B \in \mathcal{B}_j} \frac{1}{2} \int_{r_j(t)B} |\nabla \tilde{\gamma}(\cdot, t)|^2 < \epsilon_1/3$  for each  $j$ .*
- (2) *If  $\text{Energy}(\tilde{\gamma}(\cdot, t)) \geq W/2$ , then there exists  $j(t)$  so that harmonic replacement on  $(r_{j(t)}/2) \mathcal{B}_{j(t)}$  decreases energy by at least  $e_{\tilde{\gamma}, \epsilon_1/8}(t)/8$ .*

**Proof** Since the energy of the slices is continuous in  $t$ , the set  $I = \{t \mid \text{Energy}(\tilde{\gamma}(\cdot, t)) \geq W/2\}$  is compact. For each  $t \in I$ , choose a finite collection  $\mathcal{B}^t$  of disjoint closed balls in  $\mathbf{S}^2$  with  $\frac{1}{2} \int_{\cup \mathcal{B}^t} |\nabla \tilde{\gamma}(\cdot, t)|^2 \leq \epsilon_1/4$  so

$$(3.25) \quad \text{Energy}(\gamma(\cdot, t)) - \text{Energy}(H(\gamma(\cdot, t), \frac{1}{2} \mathcal{B}^t)) \geq \frac{e_{\tilde{\gamma}, \epsilon_1/4}(t)}{2} > 0.$$

Lemma 3.20 gives an open interval  $I^t$  containing  $t$  so that for all  $s \in 2I^t$

$$(3.26) \quad e_{\tilde{\gamma}, \epsilon_1/8}(s) \leq 2e_{\tilde{\gamma}, \epsilon_1/4}(t).$$

Using the continuity of  $\tilde{\gamma}(\cdot, s)$  in  $C^0 \cap W^{1,2}$  and Corollary 3.4, we can shrink  $I^t$  so that  $\tilde{\gamma}(\cdot, s)$  has energy at most  $\epsilon_1/3$  in  $\mathcal{B}^t$  for  $s \in 2I^t$  and, in addition,

$$(3.27) \quad \left| \text{Energy}(\gamma(\cdot, s)) - \text{Energy}(H(\gamma(\cdot, s), \frac{1}{2} \mathcal{B}^t)) \right. \\ \left. - \text{Energy}(\gamma(\cdot, t)) + \text{Energy}(H(\gamma(\cdot, t), \frac{1}{2} \mathcal{B}^t)) \right| \leq \frac{e_{\tilde{\gamma}, \epsilon_1/4}(t)}{4}.$$

Since  $I$  is compact, we can cover  $I$  by finitely many  $I^t$ 's, say  $I^{t_1}, \dots, I^{t_m}$ . Moreover, after discarding some of the intervals, we can arrange that each  $t$  is in at least one closed interval  $\overline{I^{t_j}}$ , each  $\overline{I^{t_j}}$  intersects at most two other  $\overline{I^{t_k}}$ 's, and the  $\overline{I^{t_k}}$ 's intersecting

$\overline{I^{t_j}}$  do not intersect each other.<sup>11</sup> For each  $j = 1, \dots, m$ , choose a continuous function  $r_j: [0, 1] \rightarrow [0, 1]$  so that:

- $r_j(t) = 1$  on  $\overline{I^{t_j}}$  and  $r_j(t)$  is zero for  $t \notin 2I^{t_j}$ .
- $r_j(t)$  is zero on the intervals that *do not* intersect  $\overline{I^{t_j}}$ .

Property (1) follows directly and (2) follows from (3.25), (3.26) and (3.27).  $\square$

**Proof of Theorem 2.1** Let  $\mathcal{B}_1, \dots, \mathcal{B}_m$  and  $r_1, \dots, r_m: [0, 1] \rightarrow [0, \pi)$  be given by Lemma 3.24. We will use an  $m$  step replacement process to define  $\gamma$ . Namely, first set  $\gamma^0 = \tilde{\gamma}$  and then, for each  $k = 1, \dots, m$ , define  $\gamma^k$  by applying harmonic replacement to  $\gamma^{k-1}(\cdot, t)$  on the  $k$ -th family of balls  $r_k(t) \mathcal{B}_k$ ; i.e, set  $\gamma^k(\cdot, t) = H(\gamma^{k-1}(\cdot, t), r_k(t) \mathcal{B}_k)$ . Finally, we set  $\gamma = \gamma^m$ .

A key point in the construction is that Property (1) of the family of balls gives that only two  $r_k(t)$ 's are positive for each  $t$ . Therefore, the energy bound on the balls given by Property (1) implies that each energy minimizing map replaces a map with energy at most  $2\epsilon_1/3 < \epsilon_1$ . Hence, Corollary 3.4 implies that these depend continuously on the boundary values, which are themselves continuous in  $t$ , so that the resulting map  $\tilde{\gamma}$  is also continuous in  $t$ . Finally, it is clear that  $\tilde{\gamma}$  is homotopic to  $\gamma$  since continuously shrinking the disjoint closed balls on which we make harmonic replacement gives an explicit homotopy. Thus,  $\gamma \in \Omega_{\tilde{\gamma}}$  as claimed.

For each  $t$  with  $\text{Energy}(\tilde{\gamma}(\cdot, t)) \geq W/2$ , Property (2) of the family of balls gives some  $j(t)$  so that harmonic replacement for  $\tilde{\gamma}(\cdot, t)$  on  $(r_j(t)/2) \mathcal{B}_{j(t)}$  decreases the energy by at least  $e_{\tilde{\gamma}, \epsilon_1/8}(t)/8$ . Thus, even in the worst case where  $r_j(t) \mathcal{B}_{j(t)}$  is the second family of balls that we do replacement on at  $t$ , (3.9) in Lemma 3.8 gives

$$(3.28) \quad \text{Energy}(\tilde{\gamma}(\cdot, t)) - \text{Energy}(\gamma(\cdot, t)) \geq \kappa \left( \frac{e_{\tilde{\gamma}, \epsilon_1/8}(t)}{8} \right)^2.$$

To establish  $(B_\Psi)$ , suppose that  $\mathcal{B}$  is a finite collection of disjoint closed balls in  $\mathbf{S}^2$  so that the energy of  $\gamma(\cdot, t)$  on  $\mathcal{B}$  is at most  $\epsilon_1/12$ . We can assume that  $\gamma^k(\cdot, t)$  has energy at most  $\epsilon_1/8$  on  $\mathcal{B}$  for every  $k$  since otherwise Theorem 3.1 implies a positive

<sup>11</sup>We will give a recipe for doing this. First, if  $\overline{I^{t_1}}$  is contained in the union of two other intervals, then throw it out. Otherwise, consider the intervals whose left endpoint is in  $\overline{I^{t_1}}$ , find one whose right endpoint is largest and discard the others (which are anyway contained in these). Similarly, consider the intervals whose right endpoint is in  $\overline{I^{t_1}}$  and throw out all but one whose left endpoint is smallest. Next, repeat this process on  $I^{t_2}$  (unless it has already been discarded), etc. After at most  $m$  steps, we get the desired cover.

lower bound for  $\text{Energy}(\tilde{\gamma}(\cdot, t)) - \text{Energy}(\gamma(\cdot, t))$ . Consequently, we can apply (3.10) in Lemma 3.8 twice (first with  $\mu = 1/8$  and then with  $\mu = 1/4$ ) to get

$$\begin{aligned}
 (3.29) \quad & \text{Energy}(\gamma(\cdot, t)) - \text{Energy}\left(H(\gamma(\cdot, t), \frac{1}{8}\mathcal{B})\right) \\
 & \leq \text{Energy}(\tilde{\gamma}(\cdot, t)) - \text{Energy}\left(H(\tilde{\gamma}(\cdot, t), \frac{1}{2}\mathcal{B})\right) \\
 & \quad + \frac{2}{\kappa} (\text{Energy}(\tilde{\gamma}(\cdot, t)) - \text{Energy}(\gamma(\cdot, t)))^{1/2} \\
 & \leq e_{\tilde{\gamma}, \epsilon_{1/8}}(t) + \frac{2}{\kappa} (\text{Energy}(\tilde{\gamma}(\cdot, t)) - \text{Energy}(\gamma(\cdot, t)))^{1/2}.
 \end{aligned}$$

Combining (3.28) and (3.29) with Theorem 3.1 gives  $(B_\Psi)$  and, thus, completes the proof.  $\square$

## Appendix A Bubble convergence implies varifold convergence

In this appendix, we will recall the notion of bubble convergence for a sequence of maps from  $\mathbf{S}^2$  and then show that bubble convergence implies varifold convergence. This fact will be used in Appendix B in the proof of Proposition 2.2.

### A.1 Bubble convergence and the topology on $\Omega$

We will need a notion of convergence for a sequence  $v^j$  of  $W^{1,2}$  maps to a collection  $\{u_0, \dots, u_m\}$  of  $W^{1,2}$  maps which is similar in spirit to the convergence in Gromov's compactness theorem for pseudo holomorphic curves, [12]. The notion that we will use is a slight weakening of the bubble tree convergence developed by Parker and Wolfson for  $J$ -holomorphic curves in [23] and used by Parker for harmonic maps in [22]. In our applications, the  $v^j$ 's will be approximately harmonic while the limit maps  $u_i$  will be harmonic. We will need the next definition to make this precise.

$S^+$  and  $S^-$  will denote the northern and southern hemispheres in  $\mathbf{S}^2$  and  $p^+ = (0, 0, 1)$  and  $p^- = (0, 0, -1)$  the north and south poles.

**Definition A.1** Given a ball  $B_r(x) \subset \mathbf{S}^2$ , the *conformal dilation* taking  $B_r(x)$  to  $S^-$  is the composition of translation  $x \rightarrow p^-$  followed by dilation of  $\mathbf{S}^2$  about  $p^-$  taking  $B_r(p^-)$  to  $S^-$ .

The standard example of a conformal dilation comes from applying stereographic projection  $\Pi: \mathbf{S}^2 \setminus \{(0, 0, 1)\} \rightarrow \mathbf{R}^2$ , then dilating  $\mathbf{R}^2$  by a positive  $\lambda \neq 1$ , and applying  $\Pi^{-1}$ .

In the definition below of convergence, the map  $u_0$  will be the standard  $W^{1,2}$ -weak limit of the  $v^j$ 's (see (B1)), while the other  $u_i$ 's will arise as weak limits of the composition of the  $v^j$ 's with a divergent sequence of conformal dilations of  $\mathbf{S}^2$  (see (B2)). The condition (B3) guarantees that these limits all arise in genuinely distinct ways, and the condition (B4) means that together the  $u_i$ 's account for all of the energy.

**Definition A.2** (Bubble convergence) We will say that a sequence  $v^j: \mathbf{S}^2 \rightarrow M$  of  $W^{1,2}$  maps converges to a collection of  $W^{1,2}$  maps  $u_0, \dots, u_m: \mathbf{S}^2 \rightarrow M$  if the following hold:

- (B1) The  $v^j$ 's converge weakly to  $u_0$  in  $W^{1,2}$  and there is a finite set  $\mathcal{S}_0 = \{x_0^1, \dots, x_0^{k_0}\} \subset \mathbf{S}^2$  so that the  $v^j$ 's converge strongly to  $u_0$  in  $W^{1,2}(K)$  for any compact  $K \subset \mathbf{S}^2 \setminus \mathcal{S}_0$ .
- (B2) For each  $i > 0$ , we get a point  $x_{\ell_i} \in \mathcal{S}_0$  and a sequence of balls  $B_{r_{i,j}}(y_{i,j})$  with  $y_{i,j} \rightarrow x_{\ell_i}$  and  $r_{i,j} \rightarrow 0$ . Furthermore, if  $D_{i,j}: \mathbf{S}^2 \rightarrow \mathbf{S}^2$  is the conformal dilation taking the southern hemisphere to  $B_{r_{i,j}}(y_{i,j})$ , then the maps  $v^j \circ D_{i,j}$  converge to  $u_i$  as in (B1). Namely,  $v^j \circ D_{i,j} \rightarrow u_i$  weakly in  $W^{1,2}(\mathbf{S}^2)$  and there is a finite set  $\mathcal{S}_i$  so that the  $v^j \circ D_{i,j}$ 's converge strongly in  $W^{1,2}(K)$  for any compact  $K \subset \mathbf{S}^2 \setminus \mathcal{S}_i$ .
- (B3) If  $i_1 \neq i_2$ , then

$$\frac{r_{i_1,j}}{r_{i_2,j}} + \frac{r_{i_2,j}}{r_{i_1,j}} + \frac{|y_{i_1,j} - y_{i_2,j}|^2}{r_{i_1,j} r_{i_2,j}} \rightarrow \infty.$$

- (B4) We get the energy equality  $\sum_{i=0}^m \text{Energy}(u_i) = \lim_{j \rightarrow \infty} \text{Energy}(v^j)$ .

## A.2 Two simple examples of bubble convergence

The simplest nontrivial example of bubble convergence is when each map  $v^j = u \circ \Psi_j$  is the composition of a fixed harmonic map  $u: \mathbf{S}^2 \rightarrow M$  with a divergent sequence of dilations  $\Psi_j: \mathbf{S}^2 \rightarrow \mathbf{S}^2$ . In this case, the  $v^j$ 's converge to the constant map  $u_0 = u(p_+)$  on each compact set of  $\mathbf{S}^2 \setminus \{p_+\}$  and all of the energy concentrates at the single point  $p_+ = \mathcal{S}_0$ . Composing the  $v^j$ 's with the divergent sequence  $\Psi_j^{-1}$  of conformal dilations gives the limit  $u_1 = u$ .

For the second example, let  $\Pi: \mathbf{S}^2 \setminus \{(0, 0, 1)\} \rightarrow \mathbf{R}^2$  be stereographic projection and let  $z = x + iy$  be complex coordinates on  $\mathbf{R}^2 = \mathbf{C}$ . If we set  $f_j(z) = 1/(jz)$  +

$z = (z^2 + 1/j)/z$ , then the maps  $v^j = \Pi^{-1} \circ f_j \circ \Pi: \mathbf{S}^2 \rightarrow \mathbf{S}^2$  are conformal and, therefore, also harmonic. Since each  $v^j$  is a rational map of degree two, we have  $\text{Energy}(v^j) = \text{Area}(v^j) = 8\pi$ . Moreover, the  $v^j$ 's converge away from 0 to the identity map which has energy  $4\pi$ . The other  $4\pi$  of energy disappears at 0 but can be accounted for by a map  $u_1$  by composing with a divergent sequence of conformal dilations;  $u_1$  must also have degree one. In this case, the conformal dilations take  $f_j$  to  $\tilde{f}_j(z) = f_j(z/j) = 1/z + z/j$  which converges to the conformal inversion about the circle of radius one.

### A.3 Bubble convergence implies varifold convergence

**Proposition A.3** *If a sequence  $v^j$  of  $W^{1,2}(\mathbf{S}^2, M)$  maps bubble converges to a finite collection of smooth maps  $u_0, \dots, u_m: \mathbf{S}^2 \rightarrow M$ , then it also varifold converges.*

Before getting to the proof, recall that a sequence of functions  $f_j$  is said to *converge in measure* to a function  $f$  if for all  $\delta > 0$  the measure of  $\{x \mid |f_j - f|(x) > \delta\}$  goes to zero as  $j \rightarrow \infty$ ; see Royden [29, page 95]. Clearly,  $L^1$  convergence implies convergence in measure. Furthermore, if  $f_j \rightarrow f$  in measure and  $h$  is uniformly continuous, then  $h \circ f_j \rightarrow h \circ f$  in measure. Finally, we will use the following general version of the dominated convergence theorem which combines Theorem 17 on page 92 of [29] and Proposition 20 on page 96 of [29]:

(DCT) If  $f_j \rightarrow f$  in measure,  $g_j \rightarrow g$  in  $L^1$  and  $|f_j| \leq g_j$ , then  $\int f_j \rightarrow \int f$ .

We will also use that the map  $\nabla u \rightarrow J_u$  is continuous as a map from  $L^2$  to  $L^1$  and, thus,  $\text{Area}(u)$  is continuous with respect to  $\text{Energy}(u)$ . To be precise, if  $u, v \in W^{1,2}(\mathbf{S}^2, M)$ , then

$$(A.4) \quad |J_u - J_v| \leq \sqrt{2} |\nabla u - \nabla v|^{1/2} \max\{|\nabla u|^{3/2}, |\nabla v|^{3/2}\}.$$

This follows from the linear algebra fact<sup>12</sup> that if  $S$  and  $T$  are  $N \times 2$  matrices, then

$$(A.5) \quad |\det(S^T S) - \det(T^T T)| \leq 2 |T - S| \max\{|S|^3, |T|^3\},$$

where  $|S|^2$  is the sum of the squares of the entries of  $S$  and  $S^T$  is the transpose.

<sup>12</sup>Note that  $|S^T T| \leq |S| |T|$ ,  $|\text{Tr}(S^T T)| \leq |S| |T|$ , and if  $X_t$  is a path of  $2 \times 2$  matrices, then  $\partial_t \det X_t = \text{Tr}(X_t^c \partial_t X_t)$  where  $X_t^c$  is the cofactor matrix given by swapping diagonal entries and multiplying off-diagonals by  $-1$ . Applying this to  $X_t = (S + t(T - S))^T (S + t(T - S))$  and using the mean value theorem gives (A.5).

**Proof of Proposition A.3** For each  $v^j$ , we will let  $V^j$  denote the corresponding map to  $G_2M$ . Similarly, for each  $u_i$ , let  $U_i$  denote the corresponding map to  $G_2M$ .

It follows from (B1)–(B4) that we can choose  $m + 1$  sequences of domains  $\Omega_0^j, \dots, \Omega_m^j \subset \mathbf{S}^2$  that are pairwise disjoint for each  $j$  and so that for each  $i = 0, \dots, m$  applying  $D_{i,j}^{-1}$  to  $\Omega_i^j$  gives a sequence of domains converging to  $\mathbf{S}^2 \setminus \mathcal{S}_i$  and accounts for all the energy, that is,

$$(A.6) \quad \lim_{j \rightarrow \infty} \int_{\mathbf{S}^2 \setminus (\cup_i \Omega_i^j)} |\nabla v^j|^2 = 0.$$

By (A.6), the proposition follows by showing for each  $i$  and any  $h$  in  $C^0(G_2M)$  that

$$(A.7) \quad \begin{aligned} \int_{\mathbf{S}^2} h \circ U_i J_{u_i} &= \lim_{j \rightarrow \infty} \int_{\Omega_i^j} h \circ V^j J_{v^j} \\ &= \lim_{j \rightarrow \infty} \int_{D_{i,j}^{-1}(\Omega_i^j)} h \circ V^j \circ D_{i,j} J_{(v^j \circ D_{i,j})}, \end{aligned}$$

where the last equality is simply the change of variables formula for integration.

To simplify notation in the proof of (A.7), for each  $i$  and  $j$ , let  $v_i^j$  denote the restriction of  $v^j \circ D_{i,j}$  to  $D_{i,j}^{-1}(\Omega_i^j)$  and let  $V_i^j$  denote the corresponding map to  $G_2M$ .

Observe first that  $J_{v_i^j} \rightarrow J_{u_i}$  in  $L^1(\mathbf{S}^2)$  by (A.4). Given  $\epsilon > 0$  and  $i$ , let  $\Omega_\epsilon^i$  be the set where  $J_{u_i} \geq \epsilon$ . Since  $h$  is bounded and  $J_{v_i^j} \rightarrow J_{u_i}$  in  $L^1(\mathbf{S}^2)$ , (A.7) would follow from

$$(A.8) \quad \lim_{j \rightarrow \infty} \int_{\Omega_\epsilon^i} h \circ V_i^j J_{v_i^j} = \int_{\Omega_\epsilon^i} h \circ U_i J_{u_i}.$$

However, given any  $\delta > 0$ ,  $W^{1,2}$  convergence implies that the measure of

$$\left\{ x \in \Omega_\epsilon^i \mid J_{v_i^j} \geq \frac{\epsilon}{2} \text{ and } |V_i^j - U_i| \geq \delta \right\}$$

goes to zero as  $j \rightarrow \infty$ . Since  $L^1$  convergence of Jacobians implies that the measure of  $\{x \in \Omega_\epsilon^i \mid J_{v_i^j} < \epsilon/2\}$  goes to zero, it follows that the maps  $V_i^j$  converge in measure to  $U_i$  on  $\Omega_\epsilon^i$ . Therefore, the  $h \circ V_i^j$ 's converge in measure to  $h \circ U_i$  on  $\Omega_\epsilon^i$ . Consequently, the general version of the dominated convergence theorem (DCT) gives (A.8) and, thus, also (A.7).  $\square$

## Appendix B The proof of Proposition 2.2

The proof of Proposition 2.2 will follow the general structure developed by Parker and Wolfson in [23] and used by Parker in [22] to prove compactness of harmonic maps

with bounded energy. The main difficulty is to rule out loss of energy in the limit (see (B4) in the definition of bubble convergence). The rough idea to deal with this is that energy loss only occurs when there are very small annuli where the maps are “almost” harmonic and the ratio between the inner and outer radii of the annulus is enormous. We will use Proposition B.19 to show that the map must be “far” from being conformal on such an annulus and, thus, condition (A) allows us to rule out energy loss. Here “far” from conformal will mean that the  $\theta$ -energy of the map is much less than the radial energy. To make this precise, it is convenient to replace an annulus  $B_{er_2} \setminus B_{er_1}$  in  $\mathbf{R}^2$  by the conformally equivalent cylinder  $[r_1, r_2] \times \mathbf{S}^1$ . The (noncompact) cylinder  $\mathbf{R} \times \mathbf{S}^1$  with the flat product metric and coordinates  $t$  and  $\theta$  will be denoted by  $\mathcal{C}$ . For  $r_1 < r_2$ , let  $\mathcal{C}_{r_1, r_2} \subset \mathcal{C}$  be the product  $[r_1, r_2] \times \mathbf{S}^1$ .

### B.1 Harmonic maps on cylinders

The main result of this subsection is that harmonic maps with small energy on long cylinders are almost radial. This implies that a sequence of such maps with energy bounded away from zero is uniformly far from being conformal and, thus, cannot satisfy (A) in Proposition 2.2. It will be used to prove a similar result for “almost harmonic” maps in Proposition B.19 and eventually be used when we show that energy will not be lost.

**Proposition B.1** *Given  $\delta > 0$ , there exist  $\epsilon_2 > 0$  and  $\ell \geq 1$  depending on  $\delta$  (and  $M$ ) so that if  $u$  is a (nonconstant)  $C^3$  harmonic map from the flat cylinder  $\mathcal{C}_{-3\ell, 3\ell} = [-3\ell, 3\ell] \times \mathbf{S}^1$  to  $M$  with  $\text{Energy}(u) \leq \epsilon_2$ , then*

$$\int_{\mathcal{C}_{-\ell, \ell}} |u_\theta|^2 < \delta \int_{\mathcal{C}_{-2\ell, 2\ell}} |\nabla u|^2.$$

To show this proposition, we show a differential inequality which leads to exponential growth for the  $\theta$ -energy of the harmonic map on the level sets of the cylinder. Once we have that, the proposition follows. Namely, if the  $\theta$ -energy in the “middle” of the cylinder was a definite fraction of the total energy over the double cylinder, then the exponential growth would force the  $\theta$ -energy of near the boundary of the cylinder to be too large.

The following standard lemma is the differential inequality for the  $\theta$ -energy that leads to exponential growth through Lemma B.4 below.

**Lemma B.2** *For a  $C^3$  harmonic map  $u$  from  $\mathcal{C}_{r_1, r_2} \subset \mathcal{C}$  to  $M \subset \mathbf{R}^N$*

$$\partial_t^2 \int_t |u_\theta|^2 \geq \frac{3}{2} \int_t |u_\theta|^2 - 2 \sup_M |A|^2 \int_t |\nabla u|^4.$$

**Proof** Differentiating  $\int_t |u_\theta|^2$  and integrating by parts in  $\theta$  gives

$$\begin{aligned} \frac{1}{2} \partial_t^2 \int_t |u_\theta|^2 &= \int_t |u_{t\theta}|^2 + \int_t \langle u_\theta, u_{tt\theta} \rangle \\ &= \int_t |u_{t\theta}|^2 - \int_t \langle u_{\theta\theta}, u_{tt} \rangle = \int_t |u_{t\theta}|^2 - \int_t \langle u_{\theta\theta}, (\Delta u - u_{\theta\theta}) \rangle \\ &\geq \int_t |u_{t\theta}|^2 + \int_t |u_{\theta\theta}|^2 - \sup_M |A| \int_t |u_{\theta\theta}| |\nabla u|^2, \end{aligned}$$

where the last inequality used that  $|\Delta u| \leq |\nabla u|^2 \sup_M |A|$  by the harmonic map equation.<sup>13</sup> The lemma follows from applying the absorbing inequality  $2ab \leq a^2/2 + 2b^2$  and noting that  $\int_t u_\theta = 0$ , so Wirtinger's inequality gives  $\int_t |u_\theta|^2 \leq \int_t |u_{\theta\theta}|^2$ .  $\square$

**Remark B.3** The differential inequality in Lemma B.2 immediately implies that Proposition B.1 holds for harmonic functions, ie, when  $|A| \equiv 0$ , even without the small energy assumption. The general case will follow by using the small energy assumption to show that the perturbation terms are negligible.

We will need a simple ODE comparison lemma:

**Lemma B.4** Suppose that  $f$  is a nonnegative  $C^2$  function on  $[-2\ell, 2\ell] \subset \mathbf{R}$  satisfying

$$(B.5) \quad f'' \geq f - a,$$

for some constant  $a > 0$ . If  $\max_{[-\ell, \ell]} f \geq 2a$ , then

$$(B.6) \quad \int_{-2\ell}^{2\ell} f \geq 2\sqrt{2}a \sinh(\ell/\sqrt{2}).$$

**Proof** Fix some  $x_0 \in [-\ell, \ell]$  where  $f$  achieves its maximum on  $[-\ell, \ell]$ . Since the lemma is invariant under reflection  $x \rightarrow -x$ , we can assume that  $x_0 \geq 0$ . If  $x_0$  is an interior point, then  $f'(x_0) = 0$ ; otherwise, if  $x_0 = \ell$ , then  $f'(x_0) \geq 0$ . In either case, we get  $f'(x_0) \geq 0$ . Since  $f(x_0) \geq 2a$ , (B.5) gives  $f''(x_0) \geq a > 0$  and, hence,  $f'$  is strictly increasing at  $x_0$ .

We claim that  $f'(x) > 0$  for all  $x$  in  $(x_0, 2\ell]$ . If not, then there would be a first point  $y > x_0$  with  $f'(y) = 0$ . It follows that  $f' \geq 0$  on  $[x_0, y]$  so that  $f \geq f(x_0) \geq 2a$  on  $[x_0, y]$  and, thus, that  $f'' \geq a > 0$  on  $[x_0, y]$ , contradicting that  $f'(y) \leq f'(x_0)$ .

<sup>13</sup>If  $u^i$  are the components of the harmonic map  $u$ ,  $g_{jk}$  is the metric on  $B$  and  $A_{u(x)}^i$  is the  $i$ -th component of the second fundamental form of  $M$  at the point  $u(x)$ , then page 157 of [35] gives

$$\Delta_M u^i = g^{jk} A_{u(x)}^i (\partial_j u, \partial_k u).$$



By the claim,  $f$  is monotone increasing on  $[x_0, 2\ell]$  so that (B.5) gives

$$(B.7) \quad f'' \geq \frac{1}{2} f \quad \text{on } [x_0, 2\ell].$$

By a standard Riccati comparison argument using  $f'(x_0) \geq 0$  and (B.7) (see, eg, Corollary A.9 in [6]), we get for  $t \in [0, 2\ell - x_0]$

$$(B.8) \quad f(x_0 + t) \geq f(x_0) \cosh(t/\sqrt{2}) \geq 2a \cosh(t/\sqrt{2}).$$

Finally, integrating (B.8) on  $[0, \ell]$  gives (B.6).  $\square$

**Proof of Proposition B.1** Since we will choose  $\ell \geq 1$  and  $\epsilon_2 < \epsilon_{SU}$ , the small-energy interior estimates for harmonic maps (see Lemma 3.4 in [30]; cf [32]) imply that

$$\sup_{C_{-2\ell, 2\ell}} |\nabla u|^2 \leq C_{SU} \int_{C_{-3\ell, 3\ell}} |\nabla u|^2 \leq C_{SU} \epsilon_2.$$

Set  $f(t) = \int_t |u_\theta|^2$ . It follows from Lemma B.2 that

$$(B.9) \quad \begin{aligned} f''(t) &\geq \frac{3}{2} f(t) - 2 \sup_M |A|^2 C_{SU} \epsilon_2 \int_t (|u_\theta|^2 + |u_t|^2) \\ &\geq f(t) - C \epsilon_2 \int_t (|u_t|^2 - |u_\theta|^2), \end{aligned}$$

where  $C = 2 C_{SU} \sup_M |A|^2$  and we have assumed that  $C \epsilon_2 \leq 1/4$  in the second inequality.

We will use that  $\int_t (|u_t|^2 - |u_\theta|^2)$  is constant in  $t$ . To see this, differentiate to get

$$\frac{1}{2} \partial_t \int_t (|u_t|^2 - |u_\theta|^2) = \int_t (\langle u_t, u_{tt} \rangle - \langle u_\theta, u_{t\theta} \rangle) = \int_t \langle u_t, (u_{tt} + u_{\theta\theta}) \rangle = 0,$$

where the second equality used integration by parts in  $\theta$  and the last equality used that  $u_{tt} + u_{\theta\theta} = \Delta u$  is normal to  $M$  while  $u_t$  is tangent.<sup>14</sup> Bound this constant by

$$(B.10) \quad \int_t (|u_t|^2 - |u_\theta|^2) = \frac{1}{4\ell} \int_{C_{-2\ell, 2\ell}} (|u_t|^2 - |u_\theta|^2) \leq \frac{1}{4\ell} \int_{C_{-2\ell, 2\ell}} |\nabla u|^2.$$

<sup>14</sup>In fact, something much stronger is true: The complex-valued function

$$\phi(t, \theta) = (|u_t|^2 - |u_\theta|^2) - 2i \langle u_t, u_\theta \rangle$$

is holomorphic on the cylinder (see page 6 of [35]). This is usually called the Hopf differential.

By (B.9) and (B.10), Lemma B.4 with  $a = \frac{C\epsilon_2}{4\ell} \int_{C_{-2\ell,2\ell}} |\nabla u|^2$  implies that either

$$(B.11) \quad \max_{[-\ell,\ell]} f < 2 \frac{C\epsilon_2}{4\ell} \int_{C_{-2\ell,2\ell}} |\nabla u|^2,$$

$$\text{or} \quad \int_{C_{-2\ell,2\ell}} |u_\theta|^2 = \int_{-2\ell}^{2\ell} f(t) dt \geq 2\sqrt{2} C \epsilon_2 \frac{\sinh(\ell/\sqrt{2})}{4\ell} \int_{C_{-2\ell,2\ell}} |\nabla u|^2.$$

The second possibility cannot occur as long as  $\ell$  is sufficiently large so that we have

$$(B.12) \quad 2\sqrt{2} C \epsilon_2 \frac{\sinh(\ell/\sqrt{2})}{4\ell} > 1.$$

Using the upper bound (B.11) for  $f$  on  $[-\ell, \ell]$  to bound the integral of  $f$  gives

$$\int_{C_{-\ell,\ell}} |u_\theta|^2 \leq 2\ell \max_{[-\ell,\ell]} f < C \epsilon_2 \int_{C_{-2\ell,2\ell}} |\nabla u|^2.$$

The proposition follows by choosing  $\epsilon_2 > 0$  so that  $C \epsilon_2 < \min\{1/4, \delta\}$  and then choosing  $\ell$  so that (B.12) holds.  $\square$

## B.2 Weak compactness of almost harmonic maps

We will need a compactness theorem for a sequence of maps  $u^j$  in  $W^{1,2}(\mathbf{S}^2, M)$  which have uniformly bounded energy and are locally well-approximated by harmonic maps. Before stating this precisely, it is useful to recall the situation for harmonic maps. Suppose therefore that  $u^j: \mathbf{S}^2 \rightarrow M$  is a sequence of harmonic maps with  $\text{Energy}(u^j) \leq E_0$  for some fixed  $E_0$ . After passing to a subsequence, we can assume that the measures  $|\nabla u^j|^2 dx$  converge and there is a finite set  $\mathcal{S}$  of points where the energy concentrates so that:

$$(B.13) \quad \text{If } x \in \mathcal{S}, \text{ then } \inf_{r>0} \left( \lim_{j \rightarrow \infty} \int_{B_r(x)} |\nabla u^j|^2 \right) \geq \epsilon_{SU}.$$

$$(B.14) \quad \text{If } x \notin \mathcal{S}, \text{ then } \inf_{r>0} \left( \lim_{j \rightarrow \infty} \int_{B_r(x)} |\nabla u^j|^2 \right) < \epsilon_{SU}.$$

The constant  $\epsilon_{SU} > 0$  comes from [30], so that (B.14) implies uniform  $C^{2,\alpha}$  estimates on the  $u^j$ 's in some neighborhood of  $x$ . Hence, Arzela–Ascoli and a diagonal argument give a further subsequence of the  $u^j$ 's  $C^2$ -converging to a harmonic map on every compact subset of  $\mathbf{S}^2 \setminus \mathcal{S}$ . We will need a more general version of this, where  $u^j: \mathbf{S}^2 \rightarrow M$  is a sequence of  $W^{1,2}$  maps with  $\text{Energy}(u^j) \leq E_0$  that are  $\epsilon_0$ -almost harmonic in the following sense:

(B<sub>0</sub>) If  $B \subset \mathbf{S}^2$  is any ball with  $\int_B |\nabla u^j|^2 < \epsilon_0$ , then there is an energy minimizing map  $v: \frac{1}{8}B \rightarrow M$  with the same boundary values as  $u^j$  on  $\partial \frac{1}{8}B$  with

$$\int_{\frac{1}{8}B} |\nabla u^j - \nabla v|^2 \leq 1/j.$$

**Lemma B.15** *Let  $\epsilon_0 > 0$  be less than  $\epsilon_{SU}$ . If  $u^j: \mathbf{S}^2 \rightarrow M$  is a sequence of  $W^{1,2}$  maps satisfying (B<sub>0</sub>) and with  $\text{Energy}(u^j) \leq E_0$ , then there exists a finite collection of points  $\{x_1, \dots, x_k\}$ , a subsequence still denoted by  $u^j$ , and a harmonic map  $u: \mathbf{S}^2 \rightarrow M$  so that  $u^j \rightarrow u$  weakly in  $W^{1,2}$  and if  $K \subset \mathbf{S}^2 \setminus \{x_1, \dots, x_k\}$  is compact, then  $u^j \rightarrow u$  in  $W^{1,2}(K)$ . Furthermore, the measures  $|\nabla u^j|^2 dx$  converge to a measure  $\nu$  with  $\epsilon_0 \leq \nu(x_i)$  and  $\nu(\mathbf{S}^2) \leq E_0$ .*

**Proof** After passing to a subsequence, we can assume that:

- The  $u^j$ 's converge weakly in  $W^{1,2}$  to a  $W^{1,2}$  map  $u: \mathbf{S}^2 \rightarrow M$ .
- The measures  $|\nabla u^j|^2 dx$  converge to a limiting measure  $\nu$  with  $\nu(\mathbf{S}^2) \leq E_0$ .

It follows that there are at most  $E_0/\epsilon_0$  points  $x_1, \dots, x_k$  with  $\lim_{r \rightarrow 0} \nu(B_r(x_j)) \geq \epsilon_0$ .

We will show next that away from the  $x_i$ 's the convergence is strong in  $W^{1,2}$  and  $u$  is harmonic. To see this, consider a point  $x \notin \{x_1, \dots, x_k\}$ . By definition, there exist  $r_x > 0$  and  $J_x$  so that  $\int_{B_{r_x}(x)} |\nabla u^j|^2 < \epsilon_0$  for  $j \geq J_x$ . In particular, (B<sub>0</sub>) applies so we get energy minimizing maps  $v_x^j: \frac{1}{8}B_{r_x}(x) \rightarrow M$  that agree with  $u^j$  on  $\partial \frac{1}{8}B_{r_x}(x)$  and satisfy

$$\int_{\frac{1}{8}B_{r_x}(x)} |\nabla v_x^j - \nabla u^j|^2 \leq 1/j.$$

(Here  $\frac{1}{8}B_{r_x}(x)$  is the ball in  $\mathbf{S}^2$  centered at  $x$  so that the stereographic projection  $\Pi_x$  which takes  $x$  to  $0 \in \mathbf{R}^2$  takes  $\frac{1}{8}B_{r_x}(x)$  and  $B_{r_x}(x)$  to balls centered at 0 whose radii differ by a factor of 8.) Since  $\text{Energy}(v_x^j) \leq \epsilon_0 \leq \epsilon_{SU}$ , it follows from Lemma 3.4 in [30] (cf [32]) that a subsequence of the  $v_x^j$ 's converges strongly in  $W^{1,2}(\frac{1}{9}B_{r_x}(x))$  to a harmonic map  $v_x: \frac{1}{9}B_{r_x}(x) \rightarrow M$ . Combining with the triangle inequality and B.2, we get

$$\begin{aligned} \text{(B.16)} \quad & \int_{\frac{1}{9}B_{r_x}(x)} |\nabla u^j - \nabla v_x|^2 \\ & \leq 2 \int_{\frac{1}{9}B_{r_x}(x)} |\nabla u^j - \nabla v_x^j|^2 + 2 \int_{\frac{1}{9}B_{r_x}(x)} |\nabla v_x^j - \nabla v_x|^2 \rightarrow 0. \end{aligned}$$

Similarly, this convergence, the triangle inequality, B.2 and the Dirichlet Poincaré inequality (theorem 3 on page 265 of [11]; this applies since  $v_x^j$  equals  $u^j$  on  $\partial \frac{1}{8} B_{r_x}(x)$ ) give

$$(B.17) \quad \int_{\frac{1}{9} B_{r_x}(x)} |u^j - v_x|^2 \leq 2 \int_{\frac{1}{8} B_{r_x}(x)} |u^j - v_x^j|^2 + 2 \int_{\frac{1}{9} B_{r_x}(x)} |v_x^j - v_x|^2 \rightarrow 0.$$

Combining (B.16) and (B.17), we see that the  $u^j$ 's converge to  $v_x$  strongly in  $W^{1,2}(\frac{1}{9} B_{r_x}(x))$ . In particular,  $u|_{\frac{1}{9} B_{r_x}(x)} = v_x$ . We conclude that  $u$  is harmonic on  $\mathbf{S}^2 \setminus \{x_1, \dots, x_k\}$ . Furthermore, since any compact  $K \subset \mathbf{S}^2 \setminus \{x_1, \dots, x_k\}$  can be covered by a finite number of such ninth-balls, we get that  $u^j \rightarrow u$  strongly in  $W^{1,2}(K)$ .

Finally, since  $u$  has finite energy, it must have removable singularities at each of the  $x_i$ 's and, hence,  $u$  extends to a harmonic map on all of  $\mathbf{S}^2$  (see Theorem 3.6 in [30]).  $\square$

### B.3 Almost harmonic maps on cylinders

The main result of this subsection, Proposition B.19 below, extends Proposition B.1 from harmonic maps to “almost harmonic” maps. Here “almost harmonic” is made precise in Definition B.18 below and roughly means that harmonic replacement on certain balls does not reduce the energy by much.

**Definition B.18** Given  $\nu > 0$  and a cylinder  $\mathcal{C}_{r_1, r_2}$ , we will say that a  $W^{1,2}(\mathcal{C}_{r_1, r_2}, M)$  map  $u$  is  $\nu$ -almost harmonic if for any finite collection of disjoint closed balls  $\mathcal{B}$  in the conformally equivalent annulus  $B_{e^{r_2}} \setminus B_{e^{r_1}} \subset \mathbf{R}^2$  there is an energy minimizing map  $v: \bigcup_{\mathcal{B}} \frac{1}{8} B \rightarrow M$  that equals  $u$  on  $\bigcup_{\mathcal{B}} \frac{1}{8} \partial B$  and satisfies

$$\int_{\bigcup_{\mathcal{B}} \frac{1}{8} B} |\nabla u - \nabla v|^2 \leq \frac{\nu}{2} \int_{\mathcal{C}_{r_1, r_2}} |\nabla u|^2.$$

We have used a slight abuse of notation, since our sets will always be thought of as being subsets of the cylinder; ie, we identify Euclidean balls in the annulus with their image under the conformal map to the cylinder.

In this subsection and the two that follow it, given  $\delta > 0$ , the constants  $\ell \geq 1$  and  $\epsilon_2 > 0$  will be given by Proposition B.1; these depend only on  $M$  and  $\delta$ .

**Proposition B.19** Given  $\delta > 0$ , there exists  $\nu > 0$  (depending on  $\delta$  and  $M$ ) so that if  $m$  is a positive integer and  $u$  is  $\nu$ -almost harmonic from  $\mathcal{C}_{-(m+3)\ell, 3\ell}$  to  $M$  with  $\text{Energy}(u) \leq \epsilon_2$ , then

$$\int_{\mathcal{C}_{-m\ell, 0}} |u_\theta|^2 \leq 7\delta \int_{\mathcal{C}_{-(m+3)\ell, 3\ell}} |\nabla u|^2.$$

We will prove Proposition B.19 by using a compactness argument to reduce it to the case of harmonic maps and then appeal to Proposition B.1. A key difficulty is that there is no upper bound on the length of the cylinder in Proposition B.19 (ie, no upper bound on  $m$ ), so we cannot directly apply the compactness argument. This will be taken care of by dividing the cylinder into subcylinders of a fixed size and then using a covering argument.

#### B.4 The compactness argument

The next lemma extends Proposition B.1 from harmonic maps on  $\mathcal{C}_{-3\ell,3\ell}$  to almost harmonic maps. The main difference from Proposition B.19 is that the cylinder is of a fixed size in Lemma B.20.

**Lemma B.20** *Given  $\delta > 0$ , there exists  $\mu > 0$  (depending on  $\delta$  and  $M$ ) so that if  $u$  is a  $\mu$ -almost harmonic map from  $\mathcal{C}_{-3\ell,3\ell}$  to  $M$  with  $\text{Energy}(u) \leq \epsilon_2$ , then*

$$\int_{\mathcal{C}_{-\ell,\ell}} |u_\theta|^2 \leq \delta \int_{\mathcal{C}_{-3\ell,3\ell}} |\nabla u|^2.$$

**Proof** We will argue by contradiction, so suppose that there exists a sequence  $u^j$  of  $(1/j)$ -almost harmonic maps from  $\mathcal{C}_{-3\ell,3\ell}$  to  $M$  with  $\text{Energy}(u^j) \leq \epsilon_2$  and

$$(B.21) \quad \int_{\mathcal{C}_{-\ell,\ell}} |u_\theta^j|^2 > \delta \int_{\mathcal{C}_{-3\ell,3\ell}} |\nabla u^j|^2.$$

We will show that a subsequence of the  $u^j$ 's converges to a nonconstant harmonic map that contradicts Proposition B.1. We will consider two separate cases, depending on whether or not  $\text{Energy}(u^j)$  goes to 0.

Suppose first that  $\limsup_{j \rightarrow \infty} \text{Energy}(u^j) > 0$ . The upper bound on the energy combined with being  $1/j$ -almost harmonic (and the compactness of  $M$ ) allows us to argue as in Lemma B.15 to get a subsequence that converges in  $W^{1,2}$  on compact subsets of  $\mathcal{C}_{-3\ell,3\ell}$  to a *nonconstant* harmonic map  $\tilde{u}: \mathcal{C}_{-3\ell,3\ell} \rightarrow M$ . Furthermore, using the  $W^{1,2}$  convergence on  $\mathcal{C}_{-\ell,\ell}$  together with the lower semicontinuity of energy, (B.21) implies that

$$\int_{\mathcal{C}_{-\ell,\ell}} |\tilde{u}_\theta|^2 \geq \delta \int_{\mathcal{C}_{-3\ell,3\ell}} |\nabla \tilde{u}|^2.$$

This contradicts Proposition B.1.

Suppose now that  $\text{Energy}(u^j) \rightarrow 0$ . Replacing  $u^j$  by

$$v^j = (u^j - u^j(0))/(\text{Energy}(u^j))^{1/2}$$

gives a sequence of maps to  $M_j = (M - u^j(0)) / \text{Energy}(u^j)^{1/2}$  with  $\text{Energy}(u^j) = 1$  and, by (B.21),

$$\int_{\mathcal{C}_{-\ell,\ell}} |v_\theta^j|^2 > \delta > 0.$$

Furthermore, the  $v^j$ 's are also  $1/j$ -almost harmonic (this property is invariant under dilation), so we can still argue as in Lemma B.15 to get a subsequence that converges in  $W^{1,2}$  on compact subsets of  $\mathcal{C}_{-3\ell,3\ell}$  to a harmonic map  $v: \mathbf{S}^2 \rightarrow \mathbf{R}^N$  (we are using here that a subsequence of the  $M_j$ 's converges to an affine space). As before, (B.21) implies that

$$\int_{\mathcal{C}_{-\ell,\ell}} |v_\theta|^2 \geq \delta \int_{\mathcal{C}_{-3\ell,3\ell}} |\nabla v|^2.$$

This time our normalization gives

$$\int_{\mathcal{C}_{-\ell,\ell}} |v_\theta|^2 \geq \delta$$

so that  $v$  contradicts Proposition B.1 (see Remark B.3), completing the proof.  $\square$

## B.5 The proof of Proposition B.19

**Proof of Proposition B.19** For each integer  $j = 0, \dots, m$ , let  $\mathcal{C}(j) = \mathcal{C}_{-(j+3)\ell, (3-j)\ell}$  and let  $\mu > 0$  be given by Lemma B.20. We will say that the  $j$ -th cylinder  $\mathcal{C}(j)$  is *good* if the restriction of  $u$  to  $\mathcal{C}(j)$  is  $\mu$ -almost harmonic; otherwise, we will say that  $\mathcal{C}(j)$  is *bad*.

On each good  $\mathcal{C}(j)$ , we apply Lemma B.20 to get

$$\int_{\mathcal{C}_{-(j+1)\ell, (1-j)\ell}} |u_\theta|^2 \leq \delta \int_{\mathcal{C}(j)} |\nabla u|^2,$$

so that summing this over the good  $j$ 's gives

$$(B.22) \quad \sum_{j \text{ good}} \int_{\mathcal{C}_{-(j+1)\ell, (1-j)\ell}} |u_\theta|^2 \leq \delta \sum_{j \text{ good}} \int_{\mathcal{C}(j)} |\nabla u|^2 \leq 6\delta \int_{\mathcal{C}_{-(m+3)\ell, 3\ell}} |\nabla u|^2,$$

where the last inequality used that each  $\mathcal{C}_{i,i+1}$  is contained in at most 6 of the  $\mathcal{C}(j)$ 's.

We will complete the proof by showing that the total energy (not just the  $\theta$ -energy) on the bad  $\mathcal{C}(j)$ 's is small. By definition, for each bad  $\mathcal{C}(j)$ , we can choose a finite collection of disjoint closed balls  $\mathcal{B}_j$  in  $\mathcal{C}(j)$  so that if  $v: \frac{1}{8}\mathcal{B}_j \rightarrow M$  is any energy-minimizing map that equals  $u$  on  $\partial\frac{1}{8}\mathcal{B}_j$ , then

$$(B.23) \quad \int_{\frac{1}{8}\mathcal{B}_j} |\nabla u - \nabla v|^2 \geq a_j > \mu \int_{\mathcal{C}(j)} |\nabla u|^2.$$

Since the interior of each  $\mathcal{C}(j)$  intersects only the  $\mathcal{C}(k)$ 's with  $0 < |j - k| \leq 5$ , we can divide the bad  $\mathcal{C}(j)$ 's into ten subcollections so that the interiors of the  $\mathcal{C}(j)$ 's in each subcollection are pair-wise disjoint. In particular, one of these disjoint subcollections, call it  $\Gamma$ , satisfies

$$\sum_{j \in \Gamma} a_j \geq \frac{1}{10} \sum_{j \text{ bad}} a_j \geq \frac{1}{10} \sum_{j \text{ bad}} \mu \int_{\mathcal{C}(j)} |\nabla u|^2,$$

where the last inequality used (B.23).

However, since  $\bigcup_{j \in \Gamma} \mathcal{B}_j$  is itself a finite collection of disjoint closed balls in the entire cylinder  $\mathcal{C}_{-(m+3)\ell, 3\ell}$  and  $u$  is  $\nu$ -almost harmonic on  $\mathcal{C}_{-(m+3)\ell, 3\ell}$ , we get that

$$(B.24) \quad \frac{\mu}{10} \sum_{j \text{ bad}} \int_{\mathcal{C}(j)} |\nabla u|^2 \leq \nu \int_{\mathcal{C}_{-(m+3)\ell, 3\ell}} |\nabla u|^2.$$

To get the proposition, combine (B.22) with (B.24) to get

$$\int_{\mathcal{C}_{-m\ell, 0}} |u_\theta|^2 \leq \left(6\delta + \frac{10\nu}{\mu}\right) \int_{\mathcal{C}_{-(m+3)\ell, 3\ell}} |\nabla u|^2.$$

Finally, choosing  $\nu$  sufficiently small completes the proof.  $\square$

## B.6 Bubble compactness

We will now prove Proposition 2.2 using a variation of the renormalization procedure developed in [23] for pseudo-holomorphic curves and later used in [22] for harmonic maps. A key point in the proof will be that the uniform energy bound, (A) and (B) are all dilation invariant, so they apply also to the compositions of the  $u^j$ 's with any sequence of conformal dilations of  $\mathbf{S}^2$ .

**Proof of Proposition 2.2** We will use the energy bound and (B) to show that a subsequence of the  $u^j$ 's converges in the sense of (B1), (B2) and (B3) of Definition A.2 to a collection of harmonic maps. We will then come back and use (A) and (B) to show that the energy equality (B4) also holds. Hence, the subsequence bubble converges and, thus by Proposition A.3, also varifold converges.

Set  $\delta = 1/21$  and let  $\ell \geq 1$  and  $\epsilon_2 > 0$  be given by Proposition B.1. Set  $\epsilon_3 = \min\{\epsilon_0/2, \epsilon_2\}$ .

**Step 1** (Initial compactness) Lemma B.15 gives a finite collection of singular points  $\mathcal{S}_0 \subset \mathbf{S}^2$ , a harmonic map  $v_0: \mathbf{S}^2 \rightarrow M$  and a subsequence (still denoted  $u^j$ ) that converges to  $v_0$  weakly in  $W^{1,2}(\mathbf{S}^2)$  and strongly in  $W^{1,2}(K)$  for any compact subset

$K \subset \mathbf{S}^2 \setminus \mathcal{S}_0$ . Furthermore, the measures  $|\nabla u^j|^2 dx$  converge to a measure  $\nu_0$  with  $\nu_0(\mathbf{S}^2) \leq E_0$  and each singular point in  $x \in \mathcal{S}_0$  has  $\nu_0(x) \geq \epsilon_0$ .

**Step 2** (Renormalizing at a singular point) Suppose that  $x \in \mathcal{S}_0$  is a singular point from Step 1. Fix a radius  $\rho > 0$  so that  $x$  is the only singular point in  $B_{2\rho}(x)$  and  $\int_{B_\rho(x)} |\nabla \nu_0|^2 < \epsilon_3/3$ . For each  $j$ , let  $r_j > 0$  be the smallest radius so that

$$\inf_{y \in B_{\rho-r_j}(x)} \int_{B_\rho(x) \setminus B_{r_j}(y)} |\nabla u^j|^2 = \epsilon_3,$$

and choose a ball  $B_{r_j}(y_j) \subset B_\rho(x)$  with  $\int_{B_\rho(x) \setminus B_{r_j}(y_j)} |\nabla u^j|^2 = \epsilon_3$ . Since the  $u^j$ 's converge to  $\nu_0$  on compact subsets of  $B_\rho(x) \setminus \{x\}$ , we get that  $y_j \rightarrow x$  and  $r_j \rightarrow 0$ . For each  $j$ , let  $\Psi_j: \mathbf{R}^2 \rightarrow \mathbf{R}^2$  be the "dilation" that takes  $B_{r_j}(y_j)$  to the unit ball  $B_1(0) \subset \mathbf{R}^2$ . By dilation invariance, the dilated maps  $\tilde{u}_1^j = u^j \circ \Psi_j^{-1}$  still satisfy (B) and have the same energy. Hence, Lemma B.15 gives a subsequence (still denoted by  $\tilde{u}_1^j$ ), a finite singular set  $\mathcal{S}_1$ , and a harmonic map  $\nu_1$  so that the  $\tilde{u}_1^j \circ \Pi$ 's converge to  $\nu_1$  weakly in  $W^{1,2}(\mathbf{S}^2)$  and strongly in  $W^{1,2}(K)$  for any compact subset  $K \subset \mathbf{S}^2 \setminus \mathcal{S}_1$ . Moreover, the measures  $|\nabla \tilde{u}_1^j \circ \Pi|^2 dx$ 's converge to a measure  $\nu_1$ .

The choice of the balls  $B_{r_j}(y_j)$  guarantees that  $\nu_1(\mathbf{S}^2 \setminus \{p^+\}) \leq \nu_0(x)$  and  $\nu_1(S^-) \leq \nu_0(x) - \epsilon_3$ . (Recall that stereographic projection  $\Pi$  takes the open southern hemisphere  $S^-$  to the open unit ball in  $\mathbf{R}^2$ .) The key point for iterating this is the following claim:

( $\star$ ) The maximal energy concentration at any  $y \in \mathcal{S}_1 \setminus \{p^+\}$  is at most  $\nu_0(x) - \epsilon_3/3$ .

Since the energy at a singular point or the energy for a nontrivial harmonic map is at least  $\epsilon_0 > \epsilon_3$ , the only one way that ( $\star$ ) could possibly fail is if  $\nu_1$  is constant,  $\mathcal{S}_1$  is exactly two points  $p^+$  and  $y$  and at most  $\epsilon_3/3$  of  $\nu_0(x)$  escapes at  $p^+$ . However, this would imply that all but at most  $2\epsilon_3/3$  of the  $\int_{B_\rho(x)} |\nabla u^j|^2$  is in  $B_{t_j}(y_j)$  with  $t_j/r_j \rightarrow 0$  which contradicts the minimality of  $r_j$ .

**Step 3** (Repeating this) We repeat this blowing up construction at the remaining singular points in  $\mathcal{S}_0$ , as well as each of the singular points  $\mathcal{S}_1$  in the southern hemisphere, etc., to get new limiting harmonic maps and new singular points to blow up at. It follows from ( $\star$ ) that this must terminate after at most  $3E_0/\epsilon_3$  steps.

**Step 4** (The necks) We have shown that the  $u^j$ 's converge to a collection of harmonic maps in the sense of (B1), (B2) and (B3). It remains to show (B4), ie, that the  $\nu_k$ 's accounted for all of the energy in the sequence  $u^j$  and no energy was lost in the limit.

To understand how energy could be lost, it is useful to re-examine what happens to the energy during the blow up process. At each stage in the blow up process, energy is "taken from" a singular point  $x$  and then goes to one of two places:



- It can show up in the new limiting harmonic map of to a singular point in  $\widehat{\mathbf{S}^2} \setminus \{p^+\}$ .
- It can disappear at the north pole  $p^+$  (ie,  $\nu_1(\mathbf{S}^2 \setminus \{p^+\}) < \nu_0(x)$ ).

In the first case, the energy is accounted for in the limit or survives to a later stage. However, in the second case, the energy is lost for good, so this is what we must rule out.

We will argue by contradiction, so suppose that  $\nu_1(\mathbf{S}^2 \setminus \{p^+\}) < \nu_0(x) - \widehat{\delta}$  for some  $\widehat{\delta} > 0$ . (Note that we must have  $\widehat{\delta} \leq \epsilon_3$ .) Using the notation in Step 1, suppose therefore that  $A_j = B_{s_j}(y_j) \setminus B_{t_j}(y_j)$  are annuli with:

$$(B.25) \quad s_j \rightarrow 0, \frac{t_j}{r_j} \rightarrow \infty \quad \text{and} \quad \int_{A_j} |\nabla u_j|^2 \geq \widehat{\delta} > 0.$$

There is obviously quite a bit of freedom in choosing  $s_j$  and  $t_j$ . In particular, we can choose a sequence  $\lambda_j \rightarrow \infty$  so that the annuli  $\widetilde{A}_j = B_{\rho/2}(y_j) \setminus B_{t_j/\lambda_j}(y_j)$  also satisfies this, ie,  $\lambda_j s_j \rightarrow 0$  and  $t_j/(\lambda_j r_j) \rightarrow \infty$ . It follows from (B.25) and the definition of the  $r_j$ 's that  $\int_{\widetilde{A}_j} |\nabla u^j|^2 \leq \epsilon_3 \leq \epsilon_2$ . However, combining this with Proposition B.19 (with  $\delta = 1/21$ ) shows that the area must be strictly less than the energy for  $j$  large, contradicting (A), and thus completing the proof.  $\square$

## Appendix C The proof of Theorem 3.1

### C.1 An application of the Wente lemma

The proof of Theorem 3.1 will use the following  $L^2$  estimate for  $h\zeta$  where  $\zeta$  is a  $L^2(B_1)$  holomorphic function and  $h$  is a  $W^{1,2}$  function vanishing on  $\partial B_1$ .

**Proposition C.1** *If  $\zeta$  is a holomorphic function on  $B_1 \subset \mathbf{R}^2$  and  $h \in W_0^{1,2}(B_1)$ , then*

$$(C.2) \quad \int_{B_1} h^2 |\zeta|^2 \leq 8 \left( \int_{B_1} |\nabla h|^2 \right) \left( \int_{B_1} |\zeta|^2 \right).$$

The estimate (C.2) does not follow from the Sobolev embedding theorem as the product of functions in  $L^2$  and  $W^{1,2}$  is in  $L^p$  for  $p < 2$ , but not necessarily for  $p = 2$ . To get around this, we will use the following lemma of H Wente [39] (cf Hélein [16, Theorem 3.1.2]).

**Lemma C.3** *If  $B_1 \subset \mathbf{R}^2$  and  $u, v \in W^{1,2}(B_1)$ , then there exists  $\phi \in C^0 \cap W_0^{1,2}(B_1)$  with  $\Delta\phi = \langle (\partial_{x_1} u, \partial_{x_2} u), (-\partial_{x_2} v, \partial_{x_1} v) \rangle$  so that*

$$\|\phi\|_{C^0} + \|\nabla\phi\|_{L^2} \leq \|\nabla u\|_{L^2} \|\nabla v\|_{L^2}.$$

**Proof of Proposition C.1** Let  $f$  and  $g$  be the real and imaginary parts, respectively, of the holomorphic function  $\zeta$ , so that the Cauchy–Riemann equations give

$$(C.4) \quad \partial_{x_1} f = \partial_{x_2} g \text{ and } \partial_{x_2} f = -\partial_{x_1} g.$$

Since  $B_1$  is simply connected, (C.4) gives functions  $u$  and  $v$  on  $B_1$  with  $\nabla u = (g, f)$  and  $\nabla v = (f, -g)$ . We have

$$|\nabla u|^2 = |\nabla v|^2 = \langle (\partial_{x_1} u, \partial_{x_2} u), (-\partial_{x_2} v, \partial_{x_1} v) \rangle = |\zeta|^2.$$

Therefore, Lemma C.3 gives  $\phi$  with  $\Delta\phi = |\zeta|^2$ ,  $\phi|_{\partial B_1} = 0$ , and

$$(C.5) \quad \|\phi\|_{C^0} + \|\nabla\phi\|_{L^2} \leq \int |\zeta|^2.$$

Applying Stokes' theorem to  $dv(h^2\nabla\phi)$  and using Cauchy–Schwarz gives

$$(C.6) \quad \int h^2 |\zeta|^2 = \int h^2 \Delta\phi \leq \int |\nabla h^2| |\nabla\phi| \leq 2 \|\nabla h\|_{L^2} \left( \int h^2 |\nabla\phi|^2 \right)^{1/2}.$$

Applying Stokes' theorem to  $dv(h^2\phi\nabla\phi)$ , noting that  $\Delta\phi \geq 0$ , and using (C.6) gives

$$\int h^2 |\nabla\phi|^2 \leq \int |\phi| (h^2 \Delta\phi + |\nabla h^2| |\nabla\phi|) \leq 4 \|\phi\|_{C^0} \|\nabla h\|_{L^2} \left( \int h^2 |\nabla\phi|^2 \right)^{1/2},$$

so that  $(\int h^2 |\nabla\phi|^2)^{1/2} \leq 4 \|\nabla h\|_{L^2} \|\phi\|_{C^0}$ . Finally, substituting this bound back into (C.6) and using (C.5) to bound  $\|\phi\|_{C^0}$  gives the proposition.  $\square$

## C.2 An application to harmonic maps

**Proposition C.7** *Suppose that  $M \subset \mathbf{R}^N$  is a smooth closed isometrically embedded manifold. There exists a constant  $\epsilon_0 > 0$  (depending on  $M$ ) so that if  $v: B_1 \rightarrow M$  is a  $W^{1,2}$  weakly harmonic map with energy at most  $\epsilon_0$ , then  $v$  is a smooth harmonic map. In addition, for any  $h \in W_0^{1,2}(B_1)$ , we have*

$$(C.8) \quad \int_{B_1} |h|^2 |\nabla v|^2 \leq C \left( \int_{B_1} |\nabla h|^2 \right) \left( \int_{B_1} |\nabla v|^2 \right).$$

**Proof** The first claim follows immediately from F Hélein's 1991 regularity theorem for weakly harmonic maps from surfaces; see Hélein [15; 16, Theorem 4.1.1].

We will show that (C.8) follows by combining estimates from the proof of Theorem 4.1.1 in [16]<sup>15</sup> with Proposition C.1. Following [16], we can assume that the pullback  $v^*(TM)$  of the tangent bundle of  $M$  has orthonormal frames on  $B_1$  and, moreover, that there is a finite energy harmonic section  $e_1, \dots, e_n$  of the bundle of orthonormal frames for  $v^*(TM)$  (the frame  $e_1, \dots, e_n$  is usually called a *Coulomb gauge*). Set  $\alpha^j = \langle \partial_{x_1} v, e_j \rangle - i \langle \partial_{x_2} v, e_j \rangle$  for  $j = 1, \dots, n$ . Since  $e_1, \dots, e_n$  is an orthonormal frame for  $v^*(TM)$ , we have

$$(C.9) \quad |\nabla v|^2 = \sum_{j=1}^n |\alpha^j|^2.$$

On pages 181 and 182 of [16], Hélein uses that the frame  $e_1, \dots, e_n$  is harmonic to construct an  $n \times n$  matrix-valued function  $\beta$  (ie, a map  $\beta: B_1 \rightarrow GL(n, \mathbf{C})$ ) with  $|\beta| \leq C$ ,  $|\beta^{-1}| \leq C$ , and with  $\partial_{\bar{z}}(\beta^{-1}\alpha) = 0$  (where the constant  $C$  depends only on  $M$  and the bound for the energy of  $v$ ; see also Lemma 3 on page 461 in [27] where this is also stated). In particular, we get an  $n$ -tuple of holomorphic functions  $(\zeta^1, \dots, \zeta^n) = \zeta = \beta^{-1}\alpha$ , so that

$$(C.10) \quad C^{-2} |\zeta|^2 \leq |\alpha|^2 = |\beta \zeta|^2 \leq C^2 |\zeta|^2.$$

The claim (C.8) now follows from Proposition C.1. Namely, using (C.9), the second inequality in (C.10), and then applying Proposition C.1 to the  $n$  holomorphic functions  $\zeta^1, \dots, \zeta^n$  gives

$$\int |h|^2 |\nabla v|^2 \leq C^2 \int |h|^2 |\zeta|^2 \leq 8 C^2 \int |\nabla h|^2 \int |\zeta|^2 \leq 8 C^4 \int |\nabla h|^2 \int |\nabla v|^2,$$

where the last inequality used the first inequality in (C.10) and (C.9).  $\square$

### C.3 The proof of Theorem 3.1

**Proof of Theorem 3.1** Use Stokes' theorem and that  $u$  and  $v$  are equal on  $\partial B_1$  to get

$$\int |\nabla u|^2 - \int |\nabla v|^2 - \int |\nabla(u-v)|^2 = -2 \int \langle (u-v), \Delta v \rangle \equiv \Psi.$$

To show (3.2), it suffices to bound  $|\Psi|$  by  $\frac{1}{2} \int |\nabla v - \nabla u|^2$ .

<sup>15</sup>Alternatively, one could use the recent results of T Rivière [28].

The harmonic map in Footnote 13 implies that  $\Delta v$  is perpendicular to  $M$  and

$$|\Delta v| \leq |\nabla v|^2 \sup_M |A|.$$

We will need the elementary geometric fact that there exists a constant  $C$  depending on  $M$  so that whenever  $x, y \in M$ , then

$$(C.11) \quad |(x - y)^N| \leq C |x - y|^2,$$

where  $(x - y)^N$  denotes the normal part of the vector  $(x - y)$  at the point  $x \in M$  (the same bound holds at  $y$  by symmetry). The point is that either  $|x - y| \geq 1/C$  so (C.11) holds trivially or the vector  $(x - y)$  is “almost tangent” to  $M$ .

Using that  $u$  and  $v$  both map to  $M$ , we can apply (C.11) to get  $|(u - v)^N| \leq C |u - v|^2$ , where the normal projection is at the point  $v(x) \in M$ . Putting all of this together gives

$$(C.12) \quad |\Psi| \leq C \int |v - u|^2 |\nabla v|^2,$$

where  $C$  depends on  $\sup_M |A|$ . As long as  $\epsilon_1$  is less than  $\epsilon_0$ , we can apply Proposition C.7 with  $h = |u - v|$  to get

$$(C.13) \quad \int |v - u|^2 |\nabla v|^2 \leq C' \left( \int |\nabla |u - v||^2 \right) \left( \int |\nabla v|^2 \right) \\ \leq C' \epsilon_1 \int |\nabla u - \nabla v|^2.$$

The lemma follows by combining (C.12) and (C.13) and then taking  $\epsilon_1$  sufficiently small.  $\square$

Combining Corollary 3.3 and the regularity theory of Morrey [20] or Schoen and Uhlenbeck [32], for energy minimizing maps recovers Hélein’s theorem that weakly harmonic maps from surfaces are smooth. Note, however, that we used estimates from Hélein [16] in the proof of Theorem 3.1.

## Appendix D The equivalence of energy and area

By (1.1), Proposition 1.2 follows once we show that  $W_E \leq W_A$ . The corresponding result for the Plateau problem is proven by taking a minimizing sequence for area and reparametrizing to make these maps conformal, ie, choosing isothermal coordinates. There are a few technical difficulties in carrying this out since the pullback metric may be degenerate and is only in  $L^1$ , while the existence of isothermal coordinates requires that the induced metric be positive and bounded; see, eg, Proposition 5.4

in [34]. We will follow the same approach here, the difference is that we need the reparametrizations to vary continuously with  $t$ .

### D.1 Density of smooth mappings

The next lemma observes that the regularization using convolution of Schoen–Uhlenbeck in the proposition in Section 4 of [33] is continuous.

**Lemma D.1** *Given  $\gamma \in \Omega$  and  $\epsilon > 0$ , there exists a regularization  $\tilde{\gamma} \in \Omega_\gamma$  so that*

$$(D.2) \quad \max_t \|\tilde{\gamma}(\cdot, t) - \gamma(\cdot, t)\|_{W^{1,2}} \leq \epsilon,$$

*each slice  $\tilde{\gamma}(\cdot, t)$  is  $C^2$ , and the map  $t \rightarrow \tilde{\gamma}(\cdot, t)$  is continuous from  $[0, 1]$  to  $C^2(\mathbf{S}^2, M)$ .*

**Proof** Since  $M$  is smooth, compact and embedded, there exists a  $\delta > 0$  so that for each  $x$  in the  $\delta$ -tubular neighborhood  $M_\delta$  of  $M$  in  $\mathbf{R}^N$ , there is a unique closest point  $\Pi(x) \in M$  and so the map  $x \rightarrow \Pi(x)$  is smooth.  $\Pi$  is called *nearest point projection*.

Given  $y$  in the open ball  $B_1(0) \subset \mathbf{R}^3$ , define  $T_y: \mathbf{S}^2 \rightarrow \mathbf{S}^2$  by  $T_y(x) = (x-y)/|x-y|$ . Since each  $T_y$  is smooth and these maps depend smoothly on  $y$ , it follows that the map  $(y, f) \rightarrow f \circ T_y$  is continuous from  $B_1(0) \times C^0 \cap W^{1,2}(\mathbf{S}^2, \mathbf{R}^N) \rightarrow C^0 \cap W^{1,2}(\mathbf{S}^2, \mathbf{R}^N)$  (this is clear for  $f \in C^1$  and follows for  $C^0 \cap W^{1,2}$  by density). Therefore, since  $T_0$  is the identity, given  $f \in C^0 \cap W^{1,2}(\mathbf{S}^2, \mathbf{R}^N)$  and  $\mu > 0$ , there exists  $r > 0$  so that  $\sup_{|y| \leq r} \|f \circ T_y - f\|_{C^0 \cap W^{1,2}} < \mu$ . Applying this to  $\gamma(\cdot, t)$  for each  $t$  and using that  $t \rightarrow \gamma(\cdot, t)$  is continuous to  $C^0 \cap W^{1,2}$  and  $[0, 1]$  is compact, we get  $\bar{r} > 0$  with

$$(D.3) \quad \sup_{t \in [0,1]} \sup_{|y| \leq \bar{r}} \|T_y \gamma(\cdot, t) - \gamma(\cdot, t)\|_{C^0 \cap W^{1,2}} < \mu.$$

Next fix a smooth radial mollifier  $\phi \geq 0$  with integral one and compact support in the unit ball in  $\mathbf{R}^3$ . For each  $r \in (0, 1)$ , define  $\phi_r(x) = r^{-3} \phi(x/r)$  and set

$$\gamma_r(x, t) = \int_{B_r(0)} \phi_r(y) \gamma(T_y(x), t) dy = \int_{B_r(x)} \phi_r(x-y) \gamma\left(\frac{y}{|y|}, t\right) dy.$$

We have the following standard properties of convolution with a mollifier (see, eg, Section 5.3 and Appendix C.4 in Evans [11]): First, each  $\gamma_r(\cdot, t)$  is smooth and for each  $k$  the map  $t \rightarrow \gamma_r(\cdot, t)$  is continuous from  $[0, 1]$  to  $C^k(\mathbf{S}^2, \mathbf{R}^N)$ . Second,

$$(D.4) \quad \begin{aligned} \|\gamma_r(\cdot, t) - \gamma(\cdot, t)\|_{C^0}^2 &\leq \sup_{|y| \leq r} \|T_y \gamma(\cdot, t) - \gamma(\cdot, t)\|_{C^0}^2, \\ \|\nabla \gamma_r(\cdot, t) - \nabla \gamma(\cdot, t)\|_{L^2}^2 &\leq \sup_{|y| \leq r} \|T_y \gamma(\cdot, t) - \gamma(\cdot, t)\|_{L^2}^2. \end{aligned}$$

It follows from (D.4) and (D.3) that for  $r \leq \bar{r}$  and all  $t$  we have

$$(D.5) \quad \|\gamma_r(\cdot, t) - \gamma(\cdot, t)\|_{C^0 \cap W^{1,2}} < \mu.$$

The map  $\gamma_r(\cdot, t)$  may not land in  $M$ , but it is in  $M_\delta$  when  $\mu$  is small by (D.5). Hence, the map  $\tilde{\gamma}(x, t) = \Pi \circ \gamma_r(x, t)$  satisfies (D.2), each slice  $\tilde{\gamma}(\cdot, t)$  is  $C^2$ , and  $t \rightarrow \tilde{\gamma}(\cdot, t)$  is continuous from  $[0, 1]$  to  $C^2(\mathbf{S}^2, M)$ . Finally,  $s \rightarrow \tilde{\gamma}_{sr}$  is an explicit homotopy connecting  $\tilde{\gamma}$  and  $\gamma$ .  $\square$

## D.2 Equivalence of energy and area

We will also need the existence of isothermal coordinates, taking special care on the dependence on the metric. Let  $\mathbf{S}_{g_0}^2$  denote the round metric on  $\mathbf{S}^2$  with constant curvature one.

**Lemma D.6** *Given a  $C^1$  metric  $\tilde{g}$  on  $\mathbf{S}^2$ , there is a unique orientation preserving  $C^{1,1/2}$  conformal diffeomorphism  $h_{\tilde{g}}: \mathbf{S}_{g_0}^2 \rightarrow \mathbf{S}_{\tilde{g}}^2$  that fixes 3 given points.*

Moreover, if  $\tilde{g}_1$  and  $\tilde{g}_2$  are two  $C^1$  metrics that are both  $\geq \epsilon g_0$  for some  $\epsilon > 0$ , then

$$\|h_{\tilde{g}_1} - h_{\tilde{g}_2}\|_{C^0 \cap W^{1,2}} \leq C \|\tilde{g}_1 - \tilde{g}_2\|_{C^0},$$

where the constant  $C$  depends on  $\epsilon$  and the maximum of the  $C^1$  norms of the  $\tilde{g}_i$ 's.

**Proof** The Riemann mapping theorem for variable metrics (see Theorem 3.1.1 and Corollary 3.1.1 in Jost [17]; cf Ahlfors and Bers [1] or Morrey [19]) gives the conformal diffeomorphism  $h_{\tilde{g}}: \mathbf{S}_{g_0}^2 \rightarrow \mathbf{S}_{\tilde{g}}^2$ .

We will separately bound the  $C^0$  and  $W^{1,2}$  norms. First, Lemma 17 in [1] gives

$$\|h_{\tilde{g}_1} - h_{\tilde{g}_2}\|_{C^0} \leq C_1 \|\tilde{g}_1 - \tilde{g}_2\|_{C^0},$$

where  $C_1$  depends on  $\epsilon$  and the  $C^0$  norms of the metrics. Second, Theorem 8 in [1] gives a uniform  $L^p$  bound for  $\nabla(h_{\tilde{g}_1} - h_{\tilde{g}_2})$  on any unit ball in  $\mathbf{S}^2$  where  $p > 2$  by (8) in [1]

$$\|\nabla(h_{\tilde{g}_1} - h_{\tilde{g}_2})\|_{L^p(B_1)} \leq C_2 \|\tilde{g}_1 - \tilde{g}_2\|_{C^0(\mathbf{S}^2)},$$

where  $C_2$  depends on  $\epsilon$  and the  $C^0$  norms of the metrics. The desired energy bound follows from covering  $\mathbf{S}^2$  by a finite collection of unit balls and applying Hölder's inequality.  $\square$

We can now prove the equivalence of the two widths.

**Proof of Proposition 1.2** By (1.1), we have that  $W_A \leq W_E$ . To prove that  $W_E \leq W_A$ , given  $\epsilon > 0$ , let  $\gamma \in \Omega_\beta$  be a sweepout with  $\max_{t \in [0,1]} \text{Area}(\gamma(\cdot, t)) < W_A + \epsilon/2$ . By Lemma D.1, there is a regularization  $\tilde{\gamma} \in \Omega_\beta$  so that each slice  $\tilde{\gamma}(\cdot, t)$  is  $C^2$ , the map  $t \rightarrow \tilde{\gamma}(\cdot, t)$  is continuous from  $[0, 1]$  to  $C^2(\mathbf{S}^2, M)$ , and (also by (A.4))

$$\max_t \text{Area}(\tilde{\gamma}(\cdot, t)) < W_A + \epsilon.$$

The maps  $\tilde{\gamma}(\cdot, t)$  induce a continuous one-parameter family of pullback (possibly degenerate)  $C^1$  metrics  $g(t)$  on  $\mathbf{S}^2$ . Lemma D.6 requires that the metrics be non-degenerate, so define perturbed metrics  $\tilde{g}(t) = g(t) + \delta g_0$ . For each  $t$ , Lemma D.6 gives  $C^{1,1/2}$  conformal diffeomorphisms  $h_t: \mathbf{S}_{g_0}^2 \rightarrow \mathbf{S}_{\tilde{g}(t)}^2$  that vary continuously in  $C^0 \cap W^{1,2}(\mathbf{S}^2, \mathbf{S}^2)$ . The continuity of  $t \rightarrow \tilde{\gamma}(\cdot, t) \circ h_t$  as a map from  $[0, 1]$  to  $C^0 \cap W^{1,2}(\mathbf{S}^2, M)$  follows from this, the continuity of  $t \rightarrow \tilde{\gamma}(\cdot, t)$  in  $C^2$  and the chain rule.

We will now use the conformality of the map  $h_t$  to control the energy of the composition as in Proposition 5.4 of [34]. Namely, we have that

$$\begin{aligned} \text{Energy}(\tilde{\gamma}(\cdot, t) \circ h_t) &= \text{Energy}(h_t: \mathbf{S}_{g_0}^2 \rightarrow \mathbf{S}_{\tilde{g}(t)}^2) \leq \text{Energy}(h_t: \mathbf{S}_{g_0}^2 \rightarrow \mathbf{S}_{\tilde{g}(t)}^2) \\ &= \text{Area}(\mathbf{S}_{\tilde{g}(t)}^2) \\ \text{(D.7)} \quad &= \int_{\mathbf{S}^2} (\det(g_0^{-1} g(t)) + \delta \text{Tr}(g_0^{-1} g(t)) + \delta^2)^{1/2} d\text{vol}_{g_0} \\ &\leq \text{Area}(\mathbf{S}_{\tilde{g}(t)}^2) + 4\pi (\delta^2 + 2\delta \sup_t |g_0^{-1} g(t)|)^{1/2}. \end{aligned}$$

Choose  $\delta > 0$  so that  $4\pi (\delta^2 + 2\delta \sup_t |g_0^{-1} g(t)|)^{1/2} < \epsilon$ .

We would be done if  $\tilde{\gamma}(\cdot, t) \circ h_t$  was homotopic to  $\tilde{\gamma}$ . However, the space of orientation preserving diffeomorphisms of  $\mathbf{S}^2$  is homotopic to  $\mathbf{RP}^3$  by Smale’s theorem. To get around this, note that  $t \rightarrow \|\tilde{\gamma}(\cdot, t)\|_{C^2}$  is continuous and zero when  $t = 1$ , thus for some  $\tau < 1$

$$\text{(D.8)} \quad \sup_{t \geq \tau} \|\tilde{\gamma}(\cdot, t)\|_{C^2} \leq \frac{\epsilon}{\sup_{t \in [0,1]} \|h_t\|_{W^{1,2}}^2}.$$

Consequently, if we set  $\tilde{h}_t$  equal to  $h_t \equiv h(t)$  on  $[0, \tau]$  and equal to  $h(\tau(1-t)/(1-\tau))$  on  $[\tau, 1]$ , then (D.7) and (D.8) imply that  $\max_{t \in [0,1]} \text{Energy}(\tilde{\gamma}(\cdot, t) \circ \tilde{h}_t) \leq W_A + 2\epsilon$ . Moreover, the map  $\tilde{\gamma}(\cdot, t) \circ \tilde{h}_t$  is also in  $\Omega$ . Finally, replacing  $\tau$  by  $s\tau$  and taking  $s \rightarrow 0$  gives an explicit homotopy in  $\Omega$  from  $\tilde{\gamma}(\cdot, t) \circ \tilde{h}_t$  to  $\tilde{\gamma}(\cdot, t)$ .  $\square$

**Acknowledgements** We would like to thank Frédéric Hélein, Bruce Kleiner and John Lott for their comments. The authors were partially supported by NSF Grants DMS 0606629 and DMS 0405695.

## References

- [1] **L Ahlfors, L Bers**, *Riemann's mapping theorem for variable metrics*, Ann. of Math. (2) 72 (1960) 385–404 MR0115006
- [2] **F J Almgren**, *The theory of varifolds*, Mimeographed notes, Princeton (1965)
- [3] **G D Birkhoff**, *Dynamical systems with two degrees of freedom*, Trans. Amer. Math. Soc. 18 (1917) 199–300 MR1501070
- [4] **G D Birkhoff**, *Dynamical systems*, American Math. Soc. Coll. Publ. IX, American Mathematical Society (1927) MR0209095
- [5] **T H Colding, C De Lellis**, *The min-max construction of minimal surfaces*, from: “Surveys in differential geometry, Vol. VIII (Boston, MA, 2002)”, Int. Press, Somerville, MA (2003) 75–107 MR2039986
- [6] **T H Colding, C De Lellis, W P Minicozzi II**, *Three circles theorems for Schrödinger operators on cylindrical ends and geometric applications*, Comm. Pure Appl. Math. 61 (2008) 1540–1602
- [7] **T H Colding, W P Minicozzi II**, *Minimal surfaces*, Courant Lecture Notes in Math. 4, New York University Courant Inst. of Math. Sciences, New York (1999) MR1683966
- [8] **T H Colding, W P Minicozzi II**, *Estimates for the extinction time for the Ricci flow on certain 3-manifolds and a question of Perelman*, J. Amer. Math. Soc. 18 (2005) 561–569 MR2138137
- [9] **T H Colding, W P Minicozzi II**, *Width and mean curvature flow*, Geom. Topol. 12 (2008) 2517–2535
- [10] **C B Croke**, *Area and the length of the shortest closed geodesic*, J. Differential Geom. 27 (1988) 1–21 MR918453
- [11] **L C Evans**, *Partial differential equations*, Graduate Studies in Math. 19, Amer. Math. Soc. (1998) MR1625845
- [12] **M Gromov**, *Pseudoholomorphic curves in symplectic manifolds*, Invent. Math. 82 (1985) 307–347 MR809718
- [13] **R S Hamilton**, *Three-manifolds with positive Ricci curvature*, J. Differential Geom. 17 (1982) 255–306 MR664497
- [14] **R S Hamilton**, *The formation of singularities in the Ricci flow*, from: “Surveys in differential geometry, Vol. II (Cambridge, MA, 1993)”, Int. Press, Cambridge, MA (1995) 7–136 MR1375255
- [15] **F Hélein**, *Régularité des applications faiblement harmoniques entre une surface et une variété riemannienne*, C. R. Acad. Sci. Paris Sér. I Math. 312 (1991) 591–596 MR1101039



- [16] **F Hélein**, *Harmonic maps, conservation laws and moving frames*, second edition, Cambridge Tracts in Math. 150, Cambridge University Press (2002) MR1913803 Translated from the 1996 French original, With a foreword by J Eells
- [17] **J Jost**, *Two-dimensional geometric variational problems*, Pure and Applied Math. (New York), Wiley-Interscience Publ. John Wiley & Sons Ltd., Chichester (1991) MR1100926
- [18] **B Kleiner, J Lott**, *Notes on Perelman's papers*, Geom. Topol. 12 (2008) 2587–2855
- [19] **C B Morrey, Jr.**, *On the solutions of quasi-linear elliptic partial differential equations*, Trans. Amer. Math. Soc. 43 (1938) 126–166 MR1501936
- [20] **C B Morrey, Jr.**, *Multiple integrals in the calculus of variations*, Die Grund. der math. Wissenschaften 130, Springer, New York (1966) MR0202511
- [21] **J R Munkres**, *Topology: a first course*, Prentice-Hall, Englewood Cliffs, N.J. (1975) MR0464128
- [22] **T H Parker**, *Bubble tree convergence for harmonic maps*, J. Differential Geom. 44 (1996) 595–633 MR1431008
- [23] **T H Parker, J G Wolfson**, *Pseudo-holomorphic maps and bubble trees*, J. Geom. Anal. 3 (1993) 63–98 MR1197017
- [24] **G Perelman**, *Ricci flow with surgery on three-manifolds* arXiv:math.DG/0303109
- [25] **G Perelman**, *Finite extinction time for the solutions to the Ricci flow on certain three-manifolds* arXiv:math.DG/0307245
- [26] **J T Pitts**, *Existence and regularity of minimal surfaces on Riemannian manifolds*, Math. Notes 27, Princeton University Press (1981) MR626027
- [27] **J Qing**, *Boundary regularity of weakly harmonic maps from surfaces*, J. Funct. Anal. 114 (1993) 458–466 MR1223710
- [28] **T Rivière**, *Conservation laws for conformally invariant variational problems*, Invent. Math. 168 (2007) 1–22 MR2285745
- [29] **H L Royden**, *Real analysis*, The Macmillan Co., New York (1963) MR0151555
- [30] **J Sacks, K Uhlenbeck**, *The existence of minimal immersions of 2-spheres*, Ann. of Math. (2) 113 (1981) 1–24 MR604040
- [31] **J H Sampson**, *Some properties and applications of harmonic mappings*, Ann. Sci. École Norm. Sup. (4) 11 (1978) 211–228 MR510549
- [32] **R Schoen, K Uhlenbeck**, *A regularity theory for harmonic maps*, J. Differential Geom. 17 (1982) 307–335 MR664498
- [33] **R Schoen, K Uhlenbeck**, *Boundary regularity and the Dirichlet problem for harmonic maps*, J. Differential Geom. 18 (1983) 253–268 MR710054
- [34] **R Schoen, J Wolfson**, *Minimizing area among Lagrangian surfaces: the mapping problem*, J. Differential Geom. 58 (2001) 1–86 MR1895348

- [35] **R Schoen, S T Yau**, *Lectures on harmonic maps*, Conference Proceedings and Lecture Notes in Geometry and Topology, II, International Press, Cambridge, MA (1997) MR1474501
- [36] **H A Schwarz**, *Über einen Grenzübergang durch altirneirendes Verfahren*, Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich 15 (1870) 272–286
- [37] **H A Schwarz**, *Gesammelete Mathematische Abhandlungen*, Springer, Berlin (1890)
- [38] **M Struwe**, *On the evolution of harmonic mappings of Riemannian surfaces*, Comment. Math. Helv. 60 (1985) 558–581 MR826871
- [39] **H C Wente**, *An existence theorem for surfaces of constant mean curvature*, J. Math. Anal. Appl. 26 (1969) 318–344 MR0243467

*Department of Mathematics, MIT*

*77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA*

*and*

*Courant Institute of Mathematical Sciences*

*251 Mercer Street, New York, NY 10012, USA*

*Department of Mathematics, Johns Hopkins University*

*3400 N Charles St, Baltimore, MD 21218, USA*

*colding@math.mit.edu, minicozz@math.jhu.edu*

Proposed: Ben Chow

Seconded: Colin Rourke, Martin Bridson

Received: 30 June 2007

Accepted: 10 October 2008