Illinois Journal of Mathematics
Volume 45, Number 4, Winter 2001, Pages 1073–1117
S 0019-2082

RECOGNIZING THE 3-SPHERE

S. V. IVANOV

ABSTRACT. A modification of the Rubinstein–Thompson criterion for a 3-manifold to be the 3-sphere is proposed. Special cell decompositions, called Q-triangulations and irreducible Q-triangulations, for closed compact orientable 3-manifolds are introduced. It is shown that if a closed compact orientable 3-manifold M^3 is given by a triangulation (or by a Q-triangulation) then one can effectively decompose M^3 into a connected sum of finitely many 3-manifolds some of which are given by irreducible Q-triangulations and others are 2-sphere bundles over a circle. Furthermore, it is shown that the problem whether a 3-manifold given by an irreducible Q-triangulation is homeomorphic to the 3-sphere is in **NP**, and the problem whether a Q-triangulation of a 3-manifold is irreducible is in **coNP**.

0. Introduction

In 1992, Rubinstein [R92] (see also [R95], [R97]) proposed an elegant algorithm that detects whether a triangulated 3-manifold is the 3-sphere[1]. A proof that Rubinstein's algorithm works was given by Thompson [T94] and later by Matveev [Ma95] (note that in [Ma95] handle decompositions instead of triangulations were used). We note that for 3-manifolds given by Heegaard splittings of genus two the recognition problem for the 3-sphere had been solved by Birman and Hilden [BH73] and by Homma, Ochiai and Takahashi [HOT80].

The first aim of this article is to present a modification of the Rubinstein–Thompson criterion for a 3-manifold to be the 3-sphere. This modification deals with irreducible cell decompositions of 3-manifolds (which are compact and orientable throughout this article) that contain a single 0-cell and whose 2-cells are biangles or triangles (Theorem 1). Such cell decompositions whose

Received August 5, 1999; received in final form December 17, 2001.
2000 *Mathematics Subject Classification.* Primary 57M40. Secondary 57M05, 57M50, 68Q25.
Supported in part by NSF grants DMS 98-01500, 00-99612.

[1]A different and much more complicated algorithm to recognize the 3-sphere had been proposed earlier by Haken [Ha87], but proving that this algorithm works turned out to be an extraordinarily difficult task that was never completed.

3-cells are tetrahedra and degenerate tetrahedra, called here Q-triangulations, seem to be very convenient for algorithmic problems. In particular, we will use Q-triangulations to investigate the complexity of the recognition problem for the 3-sphere and show that this problem for 3-manifolds given by irreducible Q-triangulations is in **NP** (Theorem 3). The problem whether a given Q-triangulation is irreducible is shown to be in **coNP** (Theorem 4). We will also give an effective procedure that decomposes a triangulated closed 3-manifold into a connected sum of finitely many 3-manifolds some of which are given by irreducible Q-triangulations and others are 2-sphere bundles over a circle (Theorem 2). This procedure together with the above results provides another solution to the recognition problem for the 3-sphere in the class of all triangulated 3-manifolds.

Recall that Rubinstein also announced in [R95], [R97] that every closed irreducible orientable triangulated 3-manifold, up to Dehn twistings about embedded incompressible tori, has finitely many isotopy classes of strongly irreducible Heegaard splittings of given genus all of which can be effectively constructed as almost normal surfaces. (Some comments on the status of Rubinstein's program [R97] to prove this statement can be found in [Ma99] and some related results are obtained in [S00].)

We hope that the new concepts of irreducible Q-triangulations and almost normal surfaces introduced here, as well as the techniques developed in this article, will have broader applications, and, in particular, will be useful in implementing Rubinstein's program for atoroidal 3-manifolds.

Before stating the main results, we introduce the basic notation and definitions that are used throughout this article.

Let

$$M^3 = M^3_\Omega$$

be a connected closed (compact and orientable) 3-manifold, equipped with a finite cell decomposition Ω that contains $r_i > 0$ open i-cells whose union we denote by

$$\mathcal{E}^i, \quad i = 0, 1, 2, 3.$$

By

$$\mathcal{F}^3 = \mathcal{F}^3(\Omega)$$

we denote a set of r_3 pairwise disjoint closed polyhedra (that is, closed 3-balls with given cell decompositions of their boundaries $\partial\mathcal{F}^3$) which enable us to define a cellular continuous map

$$\alpha : \ \mathcal{F}^3 \to M^3_\Omega$$

such that α preserves dimension (that is, i-cells of $\mathcal{F}^3$ are sent by α to i-cells of Ω) and $\alpha(\mathcal{F}^3) = M^3_\Omega$. For example, if all polyhedra in $\mathcal{F}^3$ are tetrahedra and for every tetrahedron T^3 in $\mathcal{F}^3$ its α-image $\alpha(T^3)$ is nonsingular (that is, the restriction $\alpha|_{T^3}$ of α on T^3 is an embedding), then Ω is called a *triangulation*.

Recall that the existence of triangulations for compact 3-manifolds was first established by Moise [Mo52] (see also [B59]). The closures of 0-, 1-, and 2-cells of $\mathcal{F}^3$ (or, less formally, closed 0-, 1-, and 2-cells of $\mathcal{F}^3$) will be called, respectively, *vertices*, *edges*, and *faces*.

Consider the following two properties of a cell decomposition $\Omega = \Omega(\mathcal{F}^3)$.

(A1) There is a single 0-cell E^0 in M^3_Ω, that is, $r_0 = 1$ (and r_1, r_2, r_3 are all positive).

(A2) If e is an edge of $\mathcal{F}^3$ then e is nonsingular, that is, e is not a loop.

If a cell decomposition Ω does not have properties (A1) and (A2) then certain reductions (as described in Section 2) apply to Ω and produce finitely many 3-manifolds

$$M^3_{1,\Omega_1}, \dots, M^3_{k,\Omega_k}, \quad k \geq 0,$$

such that M^3_Ω is the connected sum of $M^3_{1,\Omega_1}, \dots, M^3_{k,\Omega_k}$ and cell decompositions $\Omega_1, \dots, \Omega_k$ have properties (A1) and (A2). (The case $k = 0$ here means that M^3_Ω is the 3-sphere.)

By $M^3_\Omega(j)$, $j = 0, 1, 2$, we denote the j-spine of M^3_Ω, i.e.,

$$M^3_\Omega(j) = \mathcal{E}^0 + \cdots + \mathcal{E}^j.$$

Similarly, $\mathcal{F}^3(j)$, $j = 0, 1, 2$, denotes the j-spine of $\mathcal{F}^3$.

We now give the key definitions of normal and $A1$-normal disks and surfaces (cf. [Ha61], [Ha68], [JR89], [Hn92], [R95], [R97], [T94], [S00]).

A disk d^2 properly embedded in $\mathcal{F}^3$ (and its image $\alpha(d^2) \subset M^3_\Omega$) is called *normal* if the curve ∂d^2 is in general position with respect to the 1-spine $\mathcal{F}^3(1)$ (that is, ∂d^2 is disjoint from $\mathcal{F}^3(0)$ and intersects edges of $\mathcal{F}^3$ in finitely many piercing points), the intersection $\partial d^2 \cap \mathcal{F}^3(1)$ is nonempty and ∂d^2 intersects each edge of $\mathcal{F}^3$ in at most one point. A curve in $\partial \mathcal{F}^3$ which bounds a normal disk in $\mathcal{F}^3$ is also termed *normal*. A normal disk $d^2 \subset \mathcal{F}^3$ is called *simple* if ∂d^2 is the link of a vertex of $\mathcal{F}^3$.

A disk d^2 properly embedded in $\mathcal{F}^3$ (as well as its image $\alpha(d^2) \subset M^3_\Omega$ and the curve $\partial d^2 \subset \partial \mathcal{F}^3$) is called *$A1$-normal* if the curve ∂d^2 is in general position with respect to the 1-spine $\mathcal{F}^3(1)$, ∂d^2 intersects each edge of $\mathcal{F}^3$ in at most two points, and the following holds. If

$$\{g_1, \dots, g_{k_g}\}$$

is the set of all edges of $\mathcal{F}^3$ that are crossed by ∂d^2 twice then $k_g > 0$ and there is a partition of this set into two nonempty subsets

$$\{e_1, \dots, e_{k_e}\}, \quad \{f_1, \dots, f_{k_f}\}$$

such that for every

$$i \in \{1, \dots, k_e\} \text{ and } j \in \{1, \dots, k_f\}$$

the points $f_i \cap \partial d^2$ lie in distinct connected components of $\partial d^2 - g_j$; see Fig. 1.

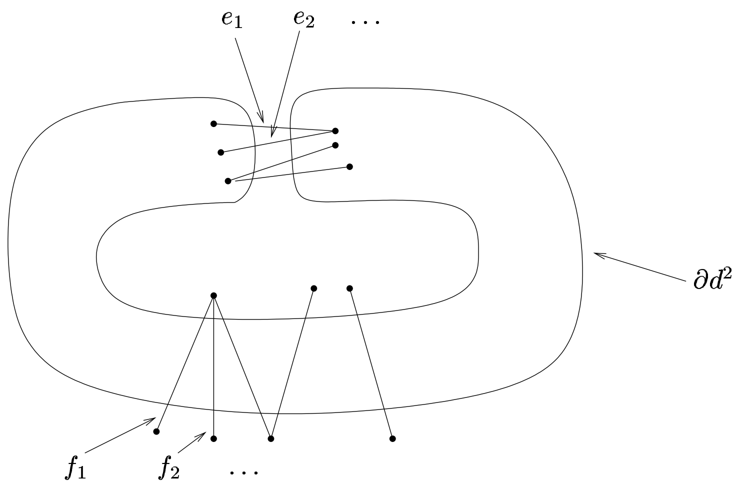


FIGURE 1

A compact closed surface U^2 (not necessarily connected) embedded in M^3_Ω is called *normal* if U^2 is in general position with respect to $M^3_\Omega(2)$ and all connected components of $\alpha^{-1}(U^2)$ are normal disks.

A compact closed surface U^2 (not necessarily connected) embedded in M^3_Ω is called *$A1$-normal* if U^2 is in general position with respect to $M^3_\Omega(2)$, there is exactly one $A1$-normal disk in $\alpha^{-1}(U^2)$, and all other connected components of $\alpha^{-1}(U^2)$ are normal disks.

As an example, we observe that property (A2) holds if and only if the link $\operatorname{Link} E^0$ of every 0-cell E^0 of M^3_Ω is a normal 2-sphere in M^3_Ω.

Two normal (resp. $A1$-normal) surfaces U^2_1, U^2_2 in M^3_Ω are called *normally parallel* if there is an isotopy which turns U^2_1 into U^2_2 so that the normal (resp. $A1$-normal) structure of U^2_1 never changes during the isotopy.

We say that a cell decomposition Ω of M^3_Ω is *irreducible* if Ω has properties (A1)–(A2) and every normal 2-sphere in M^3_Ω is normally parallel to the link $\operatorname{Link} E^0$ of the vertex E^0 of M^3_Ω.

A cell decomposition Ω of M^3_Ω is called a *Q-triangulation* if it has properties (A1)–(A2) and, in addition, one of the following properties holds:

(a) $\mathcal{F}^3$ consists of a single polyhedron with two vertices, two edges and two faces (and so M^3_Ω is the real projective 3-space).

(b) Every polyhedron F^3 in $\mathcal{F}^3$ is either a tetrahedron (that is, there are 4 vertices and 3 nonsingular triangle faces in F^3; see Fig. 2(a)) or a degenerate tetrahedron, by which we mean a polyhedron with 3 vertices, 2 nonsingular triangle faces and $b(F^3) \geq 0$ biangle faces; see Fig. 2(b).

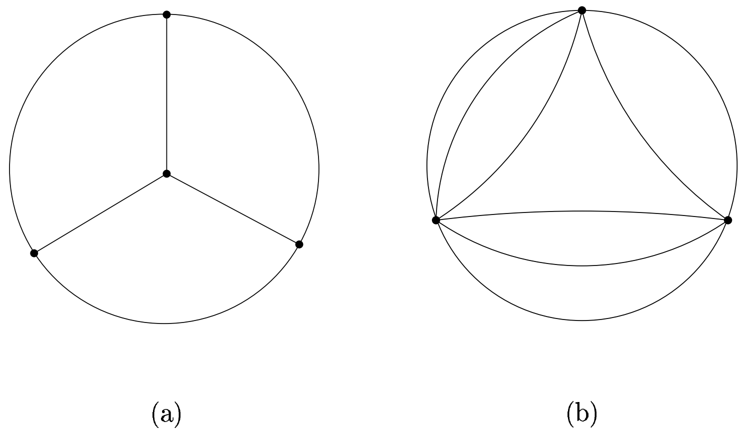


FIGURE 2

We now state the main results of this article.

THEOREM 1. *Suppose that a closed compact orientable 3-manifold M^3_Ω is given by an irreducible cell decomposition Ω and every face in $\mathcal{F}^3 = \mathcal{F}^3(\Omega)$ is a biangle or triangle. Then M^3_Ω is homeomorphic to the 3-sphere if and only if there is an $A1$-normal 2-sphere in M^3_Ω.*

The proof of Theorem 1 occupies Section 1 and is based on the Rubinstein–Thompson idea [R95], [R97], [T94] of using efficient sweepouts or thin positions in order to construct $A1$-normal surfaces in 3-manifolds. (The original idea of thin positions is due to Gabai [G87].) More applications of this fruitful idea are described in [R95], [R97], and details can be found in [S00].

Note that Theorem 1 applies only to 3-manifolds given by irreducible cell decompositions whose 2-cells are biangles and triangles. For this reason, we will describe in Section 2 an effective procedure which, given a triangulated 3-manifold M^3_Ω, decomposes M^3_Ω into a connected sum of finitely many 3-manifolds some of which are given by irreducible Q-triangulations and the others are 2-sphere bundles over a circle. The first step of this procedure is to construct 3-manifolds

$$M^3_{1,\Omega_1}, \dots, M^3_{k,\Omega_k}, \quad k \geq 0,$$

given by Q-triangulations $\Omega_1, \dots, \Omega_k$ (where the case $k = 0$ means that M^3_Ω is the 3-sphere) such that M^3_Ω is the connected sum of $M^3_{1,\Omega_1}, \dots, M^3_{k,\Omega_k}$. (This first step is polynomially fast in terms of the size of the input which is Ω.) Then, similarly to [K29], [R97], [T94], we consider a maximal system of nonparallel normal 2-spheres in each M^3_{i,Ω_i}, $i = 1, \dots, k$, and, using these systems for all $i = 1, \dots, k$, construct a decomposition for M^3_Ω as required.

Define a parameter $N(\Omega)$ of a triangulation (or a Q-triangulation) Ω of M^3_Ω as $7N_T + 3N_D$, where N_T is the number of tetrahedra in $\mathcal{F}^3 = \mathcal{F}^3(\Omega)$ and N_D is the number of degenerate tetrahedra in $\mathcal{F}^3$. Summarizing, we have the following result.

THEOREM 2. *Let M^3_Ω be a closed compact orientable 3-manifold given by a triangulation (or by a Q-triangulation) Ω. Then one can effectively construct a decomposition of M^3_Ω into the connected sum of $k_1 \geq 0$ 2-sphere bundles over a circle and $k_2 \geq 0$ 3-manifolds $M^3_{1,\Omega_1}, \dots, M^3_{k_2,\Omega_{k_2}}$ given by irreducible Q-triangulations $\Omega_1, \dots, \Omega_{k_2}$ so that*

$$N(\Omega_1) + \cdots + N(\Omega_{k_2}) \leq N(\Omega).$$

(As before, the case $k_1 + k_2 = 0$ means that M^3_Ω is homeomorphic to the 3-sphere.)

Theorem 2 effectively reduces the problem whether a triangulated 3-manifold is homeomorphic to the 3-sphere to the corresponding problem for a 3-manifold given by an irreducible Q-triangulation. To investigate the complexity of the latter problem we will use Theorem 1, the Haken theory of normal surfaces and a recent technical result of Hass, Lagarias and Pippenger [HLP99] (see Lemma 2.1 in Section 2; recall that, by [HLP99], the unknotting problem is known to be in **NP**) and establish that this problem is also in **NP**, that is, decidable in nondeterministic polynomial time in terms of the size of Ω.

THEOREM 3. *The problem whether a 3-manifold M^3_Ω given by an irreducible Q-triangulation Ω is homeomorphic to the 3-sphere is in* **NP**.

We will also estimate the complexity of the problem whether a Q-triangulation Ω of a 3-manifold M^3_Ω is irreducible (which is of interest in view of Theorems 2 and 3).

THEOREM 4. *The problem whether a Q-triangulation Ω of a 3-manifold M^3_Ω is irreducible is in* **coNP**.

The author is very grateful to Wolfgang Haken for many useful conversations and helpful comments on an earlier version of this article.

1. Proof of Theorem 1

If $X \subseteq M^3_\Omega$ then $\alpha^{-1}(X)$ denotes the full preimage of X in $\mathcal{F}^3$. The boundary, closure, interior, and the number of connected components of X will be denoted by ∂X, $\operatorname{Cls} X$, $\operatorname{Int} X$, and $|X|$, respectively. The cardinality of a finite set Y will also be denoted by $|Y|$. The union of two sets X, Y will be denoted by $X + Y$ (rather than by $X \cup Y$). The notation $X - Y$ means the closure of the set-theoretical difference $X \setminus Y$. A regular neighborhood of X in Y is denoted by $\mathcal{N}_Y(X)$.

Recall that a cell decomposition Ω of a 3-manifold M^3_Ω is called irreducible if Ω has properties (A1)–(A2) and every normal 2-sphere in M^3_Ω is *trivial*, that is, normally parallel to the link of the 0-cell of M^3_Ω.

LEMMA 1.1. *Suppose that a cell decomposition Ω of a 3-manifold M^3_Ω is irreducible. Then the 3-manifold M^3_Ω is irreducible. In particular, every 2-sphere in M^3_Ω is separating.*

Proof. Arguing on the contrary, assume that M^3_Ω is reducible. That is, M^3_Ω contains a 2-sphere U^2 which does not bound a 3-ball in M^3_Ω. Without loss of generality we can suppose that U^2 is in general position with respect to the 2-spine $M^3_\Omega(2)$ of M^3_Ω. Applying the standard process of normalization to U^2 (for details see [JR89], [Hn92]), at the end of the process we obtain a normal 2-sphere U^2_0 which does not bound a 3-ball either. Clearly, such a 2-sphere U^2_0 is nontrivial, contrary to the assumption that Ω is irreducible. □

LEMMA 1.2. *Suppose that a cell decomposition Ω of a 3-manifold M^3_Ω is irreducible and there is an $A1$-normal 2-sphere A^2 in M^3_Ω. Then M^3_Ω is homeomorphic to the 3-sphere.*

Proof. By Lemma 1.1, we can write

$$M^3_\Omega - A^2 = M^3_{\Omega,0} + M^3_{\Omega,1},$$

where $M^3_{\Omega,0}$ contains the 0-cell E^0. Let A^2_0 (resp. A^1_1) be an $A1$-normal 2-sphere in $M^3_{\Omega,0}$ (resp. $M^3_{\Omega,1}$) that is normally parallel to $\partial M^3_{\Omega,0}$ (resp. $\partial M^3_{\Omega,1}$).

We apply the standard process of normalization to A^2_1 inside $M^3_{\Omega,1}$. It is easy to see that this process will result in a system $\mathcal{B}^2$ of pairwise disjoint 2-spheres such that if $B^2 \in \mathcal{B}^2$ then either B^2 sits in a 3-cell of M^3_Ω (that is, B^2 is disjoint from $M^3_\Omega(2)$) or for every 3-cell $E^3 \in \mathcal{E}^3$ the intersection $E^3 \cap B^2$ consists of finitely many open disks the boundaries of whose closures intersect $M^3_\Omega(1)$ (and no further normalization of B^2 is possible). Let us show that the second case is actually impossible. Arguing on the contrary, let B^2 be a 2-sphere of the second case. Since $M^3_{\Omega,1}$ contains no normal 2-spheres, B^2 is not normal and so there is a disk

$$b^2 \in \alpha^{-1}(B^2)$$

which is not normal. That is, the curve $b^1 = \partial b^2$ intersects an edge e of $\mathcal{F}^3$ at least twice. Let o_1, o_2 be consecutive along e points in $b^1 \cap e$ and e_{12} be the connected component of $e - (o_1 + o_2)$ that connects o_1 and o_2. If $\alpha(e_{12})$ is disjoint from the $A1$-normal 2-sphere A^2 then the normalization process is incomplete and one more compression is possible (which eliminates the points $\alpha(o_1)$, $\alpha(o_2)$). This remark shows that e_{12} is crossed by curves in

$\alpha^{-1}(A^2) \cap \partial\mathcal{F}^3$. Since the curve b^1 is disjoint from $\alpha^{-1}(A^2)$ and bounds a disk $d^2 \subset \partial\mathcal{F}^3$ with $e_{12} \subset \partial d^2$, it follows from the definition of an $A1$-normal surface that e_{12} is crossed only by the boundary $c_0^1 = \partial c_0^2$ of the $A1$-normal disk c_0^2 of $\alpha^{-1}(A^2)$.

Let

$$c_0^1 \subset F^2 = \partial F^3,$$

where F^3 is a polyhedron of $\mathcal{F}^3$, and

$$F^2 - c_0^1 = G^2 + G_0^2,$$

where $\partial e \subset G_0^2$; see Fig. 3.

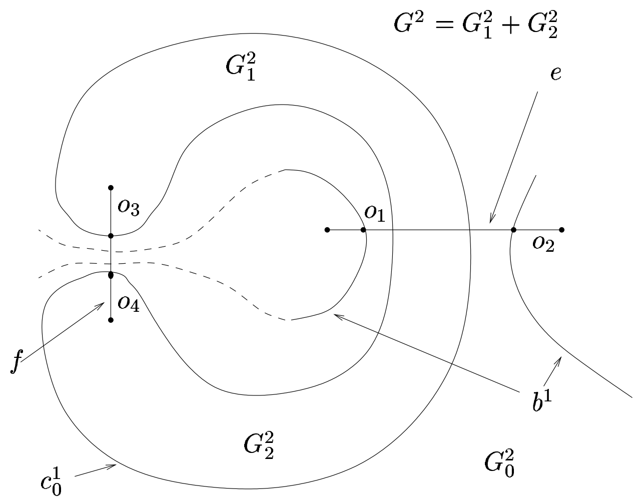


FIGURE 3

We also let

$$G^2 - e = G_1^2 + G_2^2.$$

It follows from the definition of an $A1$-normal disk that there is an edge f of F^3 with one vertex of ∂f in G_1^2 and the other one in G_2^2; see Fig. 3. Let

$$f \cap c_0^1 = o_3 + o_4$$

and let f_{34} be the connected component of $f - (o_3 + o_4)$ that connects the points o_3, o_4. Note that $\operatorname{Int} f_{34}$ is disjoint from $\alpha^{-1}(A^2)$ and $\operatorname{Int} f_{34}$ is crossed by the curve b^1 at least twice. Observe that since A^2 is $A1$-normal, $\operatorname{Int} f_{34}$ is disjoint from $\alpha^{-1}(A^2)$ and b^1 crosses f_{34} at least twice. Now we can conclude as above that one more compression of B^2 is possible (which eliminates two points in $\alpha(b^1 \cap f_{34})$), that is, the normalization process is incomplete. This contradiction proves that all 2-spheres in $\mathcal{B}^2$ are in 3-cells of M_Ω^3, as required.

Consider the following propositions for a finite system $\mathcal{B}_1^2$ of pairwise disjoint 2-spheres embedded in $M^3_{\Omega,1}$:

(B1) For every $B^2 \in \mathcal{B}_1^2$ the connected component of $M^3_{\Omega,1} - B^2$ that contains no $\partial M^3_{\Omega,1}$ (this connected component exists by Lemma 1.1) is a 3-ball.

It is easy to observe that the truth value of proposition (B1) does not change in the process of normalization of the 2-sphere A_1^2 inside $M^3_{\Omega,1}$. Since at the end of the normalization process proposition (B1) is obviously true, it must also be true in the beginning. This means that $M^3_{\Omega,1}$ is a 3-ball.

Analogously, we normalize the 2-sphere A_0^2 inside $M^3_{\Omega,0}$ and prove that this normalization process results in a collection of 2-spheres disjoint from $M^3_\Omega(2)$ and finitely many trivial normal 2-spheres. Letting (B0) denote proposition (B1) with $M^3_{\Omega,1}$ replaced by $M^3_{\Omega,0}$, we observe as before that the truth value of proposition (B0) does not change during the normalization process. Since proposition (B0) is true at the end of the process, it is also true at the beginning. Therefore, $M^3_{\Omega,0}$ is also a 3-ball, whence M^3_Ω is a 3-sphere. This completes the proof of Lemma 1.2. □

We now begin the proof of the converse of Lemma 1.2, which will be completed in Lemma 1.9.

From now on, assume that M^3_Ω is the 3-sphere and Ω is its irreducible cell decomposition all of whose 2-cells are biangles or triangles (that is, all faces in $\mathcal{F}^3$ are biangles or triangles).

Consider a singular foliation of M^3_Ω by $S^2(t)$, $0 \le t \le 1$, where $S^2(t)$ is a 2-sphere if $t \in (0,1)$, $S^2(0) = E^0$ and $S^2(1) = P^0$ are points, E^0 is the 0-cell of M^3_Ω and P^0 is a point in a 3-cell of M^3_Ω.

Without loss of generality, we can assume that the 2-spine $M^3_\Omega(2)$ of M^3_Ω is in general position with respect to the foliation $S^2(t)$, $t \in (0,1)$. In addition, applying isotopic deformations to $M^3_\Omega(2)$ that fix a regular neighborhood of E^0, one can assume that the following properties hold (cf. [G87], [T94]).

(C1) For every $t \in (0,1)$, the intersection

$$S^2(t) \cap M^3_\Omega(1)$$

consists of finitely many points at which 1-cells of M^3_Ω pierce $S^2(t)$ and at most one tangency point at which a 1-cell of M^3_Ω is tangent to $S^2(t)$.

(C2) There are finitely many levels

$$t_1^*, \dots, t_{\ell_t^*}^* \in (0,1), \quad t_1^* < \dots < t_{\ell_t^*}^*,$$

called *critical*, at which $S^2(t_i^*)$ does have a tangency point

$$P^0(t_i^*) \in S^2(t_i^*) \cap M^3_\Omega(1).$$

For convenience, set $t_0^* = 0$ and $t_{\ell_t^*+1}^* = 1$.

(C3) Let $N(t_i^*)$ denote the number of points in

$$M_\Omega^3(1) \cap S^2((t_i^* + t_{i+1}^*)/2),$$

$i = 0, \dots, \ell_t^*$. Then the number ℓ_t^* of critical levels is minimal (over all isotopy deformations of $M_\Omega^3(2)$ as specified above) and, if this number ℓ_t^* is fixed, then the sum

$$N(t_0^*) + \cdots + N(t_{\ell_t^*}^*)$$

is minimal.

As an example, we note that $N(t_0^*) = 2r_1 = 2|\mathcal{E}^1|$ (for $S^2(\varepsilon)$ with sufficiently small $\varepsilon > 0$ is a trivial normal 2-sphere) and $N(t_{\ell_t^*}^*) = 0$ (for $S^2(1-\varepsilon)$ is inside a 3-cell of M_Ω^3).

Applying more isotopy deformations to $M_\Omega^3(2)$, as specified above, that now, in addition, fix $M_\Omega^3(1)$, we obtain the following additional properties:

(C4) For every critical level t_i^*, $i = 1, \dots, \ell_t^*$, the following is true. Let $n = n(t_i^*)$ be the number of connected components in $\mathcal{N}^3_{M_\Omega^3}(P^0(t_i^*)) \cap \mathcal{E}^2$. Consider a union UP of n half planes

$$z = \tan(\tfrac{i\pi}{n+1})x, \quad z \geq 0, \quad i \in \{1, \dots, n\},$$

in the Euclidian 3-space $\mathbf{E}^3$. Also, consider a surface

$$z = -(x^2 + y^2) + \delta(t - t_i^*), \quad \delta \in \{\pm 1\},$$

in $\mathbf{E}^3$. Then there is a sufficiently small neighborhood $t^1 = t^1(t_i^*)$ of t_i^* in $(0,1)$ such that for $t \in t^1$ the intersection

$$\mathcal{N}^3_{M_\Omega^3}(P^0(t_i^*)) \cap M_\Omega^3(2) \cap S^2(t)$$

looks like the intersection of UP (which corresponds to $M_\Omega^3(2)$) with the surface $z = -(x^2+y^2) + \delta(t - t_i^*)$ (which corresponds to $S^2(t)$) in a small neighborhood of the origin (which corresponds to $P^0(t_i^*)$; see Fig. 4, where the case $\delta = 1$ is depicted).

(C5) For every level t, $t \in (0,1)$, that is not critical, the intersection $\mathcal{E}^2 \cap S^2(t)$ consists of finitely many isolated double (open) arcs and curves at which $S^2(t)$ transversally intersects $\mathcal{E}^2$ and at most one exceptional connected component that is either a single point at which $S^2(t)$ is tangent to $\mathcal{E}^2$. (Locally this looks like the intersection of the surface $z = x^2 + y^2$ and xy-plane at the origin) or which contains one simple saddle point (this locally looks like the intersection of the surface $z = x^2 - y^2$ and xy-plane at the origin.)

(C6) For every $i \in \{0, 1, \dots, \ell_t^*\}$ there are finitely many levels

$$t_{i,1}^* < t_{i,2}^* < \cdots < t_{i,k(i,t)}^*$$

in (t_i^*, t_{i+1}^*), called *secondary critical levels*, at which $S^2(t_{i,j}^*)$ does have a tangency or saddle point

$$P^0(t_{i,j}^*) \in \mathcal{E}^2 \cap S^2(t_{i,j}^*).$$

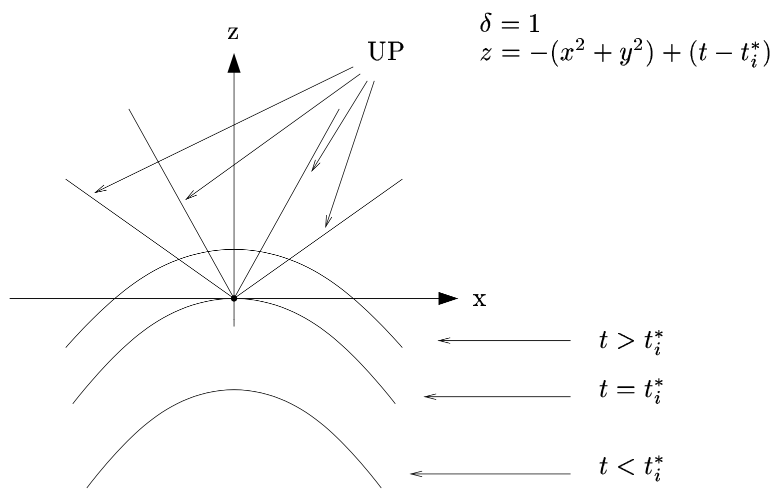


FIGURE 4

A critical level t_i^*, $i \in \{1, \dots, \ell_t^*\}$, is called an *ℓ-level* if the number

$$|M_\Omega^3(1) \cap S^2((t_{i-1}^* + t_i^*)/2)| = N(t_{i-1}^*)$$

is less than $N(t_i^*)$ (by two). If

$$N(t_i^*) = N(t_{i-1}^*) - 2$$

then t_i^* is a *u-level.* We say that $t \in (0,1)$ is a *noncritical level* if t is not a critical level and not a secondary critical level.

We now define upper and lower disks at a level t, $t \in (0,1)$, or for a level surface $S^2(t)$. Note that this definition is quite similar to the corresponding definition of Thompson [T94].

Let D^2 be a disk embedded in M_Ω^3 so that

$$\partial D^2 = A^1 + B^1,$$

where A^1 is a closed arc of a 1-cell of M_Ω^3, B^1 is a closed arc in $S^2(t)$ with $\partial B^1 = \partial A^1 = A^1 \cap B^1$, and $\operatorname{Int} D^2 + \operatorname{Int} B^1$ be disjoint from $M_\Omega^3(1)$. Denote the connected components of $M_\Omega^3 - S^2(t)$ by

$$M_\Omega^3(S^2(t), -), \quad M_\Omega^3(S^2(t), +),$$

where $E^0 \in M^3_\Omega(S^2(t), -)$. If a regular neighborhood $\mathcal{N}^2_{D^2}(B^1)$ of B^1 in D^2 has the property that

$$\mathcal{N}^2_{D^2}(B^1) \subset M^3_\Omega(S^2(t), +)$$

then the disk D^2 with $\partial D^2 = A^1 + B^1$ is called an *upper disk* at level t (or for the level 2-sphere $S^2(t)$) and A^1 is the *upper arc* of D^2, B^1 is the *projection arc* of D^2. Analogously, if

$$\mathcal{N}^2_{D^2}(B^1) \subset M^3_\Omega(S^2(t), -)$$

then the disk D^2 with $\partial D^2 = A^1 + B^1$ is called a *lower disk* at level t, A^1 is the *lower arc* of D^2, and B^1 is the *projection arc* of D^2.

For example, if t_i^* is an ℓ-level then it is easy to show that there is a lower disk $D^2_L(t)$ at any level $t \in (t_i^*, t_{i+1}^*)$ and if t_i^* is a u-level then there is an upper disk $D^2_U(t)$ at any level $t \in (t_{i-1}^*, t_i^*)$; see Fig. 5(a)–(b).

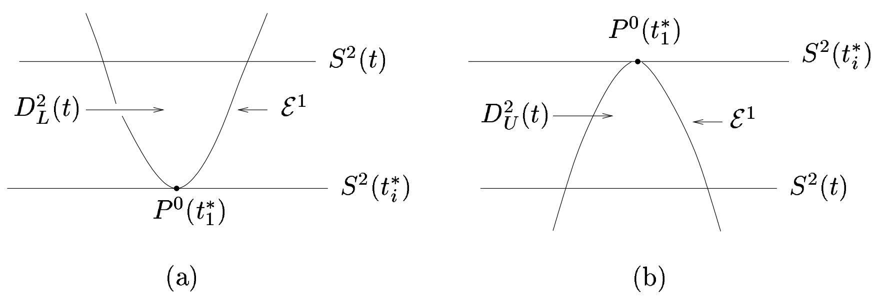


FIGURE 5

LEMMA 1.3. *The first critical level t_1^* is an ℓ-level.*

Proof. Arguing on the contrary, assume that t_1^* is a u-level. Let the tangency point $P_1^0(t_1^*)$ (of the first critical level t_1^*) belong to a 1-cell $E^1 \in \mathcal{E}^1$. Then the closure $\operatorname{Cls} E^1 = E^1 + E^0$ is unknotted in M^3_Ω, that is, there is a disk U^2 embedded in M^3_Ω such that $\partial U^2 = E_1^1 + E^0$ and $\operatorname{Int} U^2$ is disjoint from $M^3_\Omega(1)$. Taking a sufficiently small $\varepsilon > 0$ (note that ε can be chosen arbitrarily small), we may assume that $S^2(\varepsilon)$ is a trivial normal 2-sphere and the intersection

$$D^2 = U^2 \cap M^3(S^2(\varepsilon), +)$$

is an upper disk at level ε with $\partial D^2 = A^1 + B^1$, where $A^1 \subset E^1$ is the upper arc of D^2 and $B^1 \subset S^2(\varepsilon)$ is the projection arc of D^2. Furthermore, isotopically deforming D^2, we can suppose that the intersection

$$D^2 \cap M^3_\Omega(2)$$

consists of A^1 and finitely many double arcs $H_1^1, H_2^1, \ldots$ whose endpoints are in $\operatorname{Int} B^1$ and double curves in $\operatorname{Int} D^2$ at which 2-cells of M_Ω^3 transversally intersect D^2; see Fig. 6. Double curves in $D^2 \cap M_\Omega^3(2)$ can easily be eliminated by using the standard innermost argument. More specifically, we pick an innermost on a 2-cell E^2 of M_Ω^3 curve C^1 in $D^2 \cap M_\Omega^3(2)$ and compress D^2 along the disk $C^2 \subset E^2$ bounded by C^1. By repeating such compressions we can eliminate all double curves in $D^2 \cap M_\Omega^3(2)$ (without introducing new double arcs in $D^2 \cap M_\Omega^3(2)$).

Consider an outermost on D^2 arc, say H_1^1, in $D^2 \cap M_\Omega^3(2)$, that is, an arc H_1^1 such that if K_1^1 is the connected component of $B^1 - \partial H_1^1$ that connects ∂H_1^1 then $H_1^1 + K_1^1$ bounds a disk $K_1^2 \subset D^2$ with $\operatorname{Int} K_1^2$ disjoint from $M_\Omega^3(2)$; see Fig. 6.

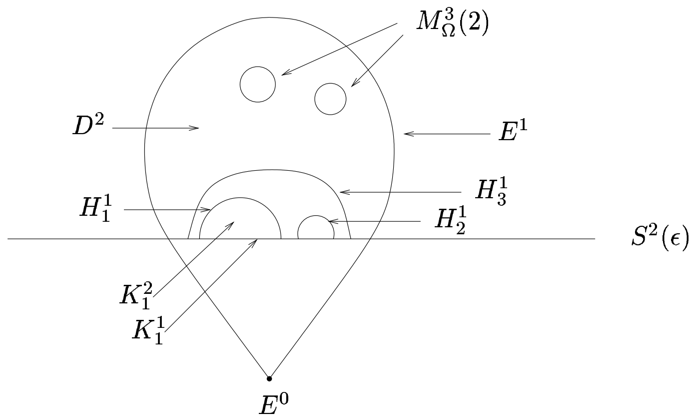


FIGURE 6

Switching to $\mathcal{F}^3$, let G^3 be the polyhedron in $\mathcal{F}^3$ that contains the disk k_1^2 with $\alpha(k_1^2) = K_1^2$. Set $\partial k_1^2 = h_1^1 + k_1^1$, where

$$\alpha(h_1^1) = H_1^1,\ \alpha(k_1^1) = K_1^1,$$

and let G^2 be the face of G^3 that contains the arc h_1^1.

First suppose that the points ∂h_1^1 lie on the same connected component of $G^2 \cap \alpha^{-1}(S^2(\varepsilon))$, say on c^1. Let c_0^1 be the connected component of $c^1 - \partial h_1^1$ that connects points ∂h_1^1. By $G_0^2 \subset G^2$ denote the disk bounded by h_1^1 and c_0^1 (note that $\operatorname{Int} h_1^1$ and c_0^1 are disjoint for $H_1^1 \cap S^2(\varepsilon) = \partial H_1^1$; see Fig. 7).

If $\operatorname{Int} G_0^2$ is disjoint from $\alpha^{-1}(D^2)$ then we can compress the disk D^2 along $\alpha(G_0^2)$ to eliminate the arc H_1^1. That is, we push a regular neighborhood $\mathcal{N}_{D^2}^2(H_1^1)$ along $\alpha(G_0^2)$ to $\alpha(c_0^1)$ and then slightly off $\alpha(c_0^1)$ into $M_\Omega^3(S^2(\varepsilon), -)$ which results in splitting the disk $D^2 \subset M_\Omega^3(S^2(\varepsilon), +)$ into two disks

$$D_1^2, D_2^2 \subset M_\Omega^3(S^2(\varepsilon), +)$$

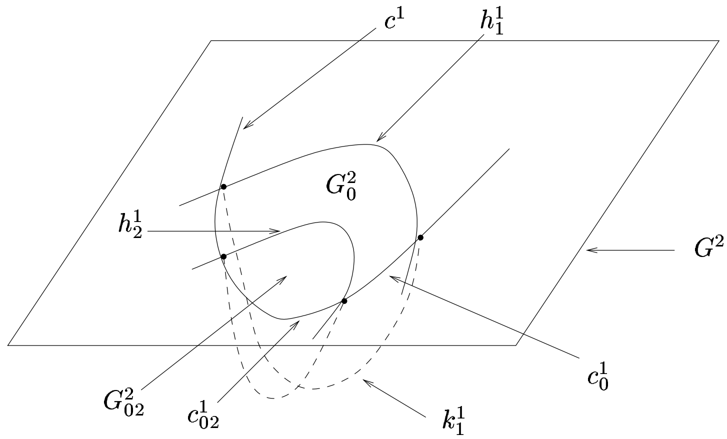


FIGURE 7

one of which, similar to D^2, is an upper disk (with the same upper arc as that of D^2) and is called the result of compression of D^2 along $\alpha(G^2_0)$.

Now suppose that the disk G^2_0 is not disjoint from $\alpha^{-1}(D^2)$, that is, $\operatorname{Int} G^2_0$ contains some double arcs in

$$\alpha^{-1}(D^2 \cap M^3_\Omega(2))$$

(which are properly embedded in G^2_0 with boundary points in $\operatorname{Int} c^1_0$; see Fig. 7). Choosing an outermost on G^2_0 such arc, say h^1_2 with $\alpha(h^1_2) = H^1_2$, denote by $G^2_{02} \subset G^2_0$ a disk bounded by h^1_2 and the connected component c^1_{02} of $c^1_0 - \partial h^1_2$ that connects points ∂h^1_2. Now we can compress the disk D^2 along the disk G^2_{02} to eliminate the arc H^1_2 (and possibly other, outer arcs on D^2 relative to H^1_2). Repeating such compressions will eliminate all arcs of

$$\alpha^{-1}(D^2 \cap M^3_\Omega(2))$$

in $\operatorname{Int} G^2$. If the arc H^1_1 is still in $D^2 \cap M^3_\Omega(2)$ (we keep the same notation for new disks obtained by a series of compressions from D^2) then we can compress D^2 along G^2_0 to eliminate H^1_1.

Thus, we can assume that either for every outermost on D^2 arc

$$H^1_1 \in D^2 \cap M^3_\Omega(2)$$

the two points ∂h^1_1 lie on distinct connected components

$$c^1_1, c^1_2 \in G^2 \cap \alpha^{-1}(S^2(c))$$

or the intersection

$$\operatorname{Int} D^2 \cap M^3_\Omega(2)$$

is empty.

In the first case, we observe that c_1^1, c_2^1 are arcs of the same curve

$$c^1 \in \partial G^3 \cap \alpha^{-1}(S^2(\varepsilon))$$

because the points $\partial h_1^1 \in (c_1^1 + c_2^1)$ are connected in G^3 by an arc in $\alpha^{-1}(B^1)$ (which is k_1^1 in the notation of Fig. 7). Since G^2 is a biangle or triangle, it follows that the arcs c_1^1, c_2^1 have endpoints on the same edge of G^2, whence the curve c^1 crosses an edge of G^2 at least twice. This, however, contradicts the normality of $S^2(\varepsilon)$.

In the second case, there is a disk $d^2 \subset G^3$ with $\alpha(d^2) = D^2$ and so, remembering that $S^2(\varepsilon)$ is a trivial normal 2-sphere, we have that the arc a^1 of ∂d^2 with $\alpha(a^1) = A^1$ belongs to a singular edge of $G^3 \in \mathcal{F}^3$. This contradiction to property (A2) of Ω completes the proof of Lemma 1.3. □

LEMMA 1.4. *Suppose that D^2 is an upper (resp. lower) disk at a noncritical level t, $\partial D^2 = A^1 + B^1$, where A^1 is the upper (resp. lower) arc of D^2, B^1 is the projection arc of D^2. Then the arc A^1 contains precisely one of the tangency points $P^0(t_i^*)$ of critical levels t_i^*, $i = 1, \ldots, \ell_t^*$.*

Proof. It is obvious that A^1 contains some of tangency points $P^0(t_i^*)$ of critical levels t_i^*, $i = 1, \ldots, \ell_t^*$, and we only have to show that A^1 contains at most one such a point. Suppose, on the contrary, that A^1 contains at least two such points. For definiteness, assume that D^2 is upper. Let us push the arc $A^1 \subset E^1$, $E^1 \in \mathcal{E}^1$, (together with other parts of $M_\Omega^3(2)$) through the disk D^2 into B^1 and then slightly isotope a regular neighborhood $\mathcal{N}^1_{M_\Omega^3(1)}(B^1)$ to create a single tangency point in $\mathcal{N}^1_{M_\Omega^3(1)}(B^1)$ at level t, thus turning t into a critical level; see Fig. 8.

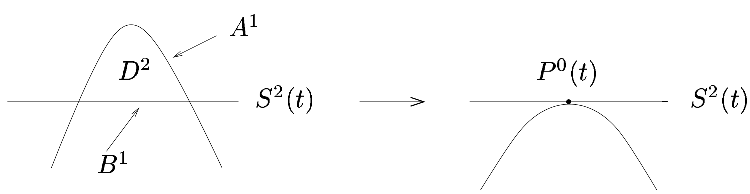


FIGURE 8

As a result, the number ℓ_t^* of critical levels decreases. This contradiction to property (C3) proves Lemma 1.4. □

LEMMA 1.5. *Suppose that D^2 is an upper (resp. lower) disk at a noncritical level t,*

$$\partial D^2 = A^1 + B^1,$$

where A^1 is the upper (resp. lower) arc of D^2 and B^1 is the projection arc of D^2. Then $\operatorname{Int} A^1$ *is disjoint from $S^2(t)$.*

Proof. Observe that if $C^1 \subset E^1$, $E^1 \in \mathcal{E}^1$, is an arc properly embedded in

$$M^3_\Omega(S^2(t),+)$$

(or in $M^3_\Omega(S^2(t),-)$) then C^1 contains at least one tangency point $P^0(t^*_i)$, where $t^*_i > t$ (or $t^*_i < t$, respectively). Therefore, if Lemma 1.5 is false and D^2 with $\partial D^2 = A^1 + B^1$ is a counterexample to it then there are at least two tangency points in A^1. This, however, is impossible by Lemma 1.4. □

LEMMA 1.6. *Suppose that D^2_1, D^2_2 are lower and upper disks at a noncritical level t, A^1_1, A^1_2 are their lower and upper arcs, and B^1_1, B^1_2 are their projection arcs. Then there is at least one point in the intersection* $\operatorname{Int} B^1_1 \cap \operatorname{Int} B^1_2$.

Proof. Suppose, on the contrary, that $\operatorname{Int} B^1_1$, $\operatorname{Int} B^1_2$ are disjoint. First assume that B^1_1, B^1_2 are not disjoint, that is, $B^1_1 \cap B^1_2$ is one point in $\partial B^1_1 \cap \partial B^1_2$. Then we can push arcs A^1_1 and A^1_2 through disks D^2_1, D^2_2 into B^1_1 and B^1_2, respectively, and then slightly isotope a regular neighborhood

$$\mathcal{N}^1_{M^3_\Omega(1)}(B^1_1 + B^1_2)$$

to eliminate all tangency points in $\mathcal{N}^1_{M^3_\Omega(1)}(B^1_1 + B^1_2)$; see Fig. 9(a). An application of Lemma 1.3 shows that this isotopy decreases the number ℓ^*_t of critical levels. A contradiction to property (C3) shows that the arcs B^1_1, B^1_2 must also be disjoint.

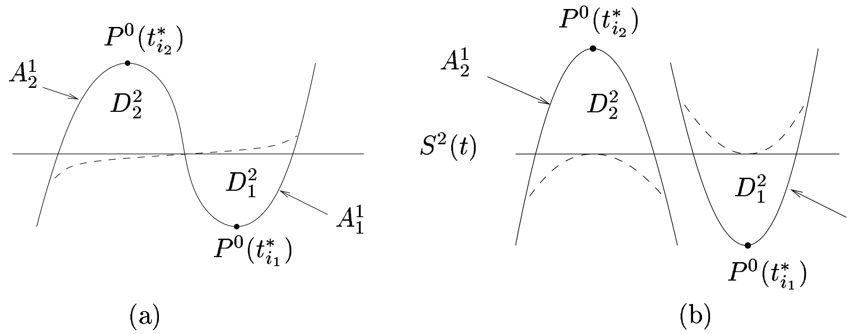


FIGURE 9

By Lemma 1.3, we can write

$$P^0(t^*_{i_1}) \in A^1_1, \quad P^0(t^*_{i_2}) \in A^1_2,$$

where $i_1 < i_2$ (recall that A^1_1 is lower and A^1_2 is upper). We push A^1_1 through D^2_1 into B^1_1 and then slightly deform a regular neighborhood $\mathcal{N}^1_{M^3_\Omega(1)}(B^1_1)$ to create a critical ℓ-level just above t, that is, at level $t+\varepsilon$ with sufficiently small $\varepsilon > 0$; see Fig. 9(b). Similarly, we push A^1_2 through D^2_2 into B^1_2 and then slightly deform a regular neighborhood $\mathcal{N}^1_{M^3_\Omega(1)}(B^1_2)$ to create a critical u-level just below t; see Fig. 9(b). By Lemma 1.3, this isotopy does not change the number ℓ^*_t.

Let us determine what happens to the sum

$$\sum_{i=0}^{\ell^*_t} N(t^*_i)$$

under this isotopy. To simplify our computations, put $\sigma_i = 2$ if t^*_i, $i \in \{1, \dots, \ell^*_t\}$, is an ℓ-level and $\sigma_i = -2$ if t^*_i is a u-level. It is clear that $\sigma_{i_1} = 2$ and $\sigma_{i_2} = -2$. It is also clear that

$$\sum_{i=0}^{\ell^*_t} N(t^*_i) = 2r_1(\ell^*_t + 1) + \sigma_1 \cdot \ell^*_t + \sigma_2(\ell^*_t - 1) + \dots + \sigma_{\ell^*_t}.$$

Let $t \in (t^*_j, t^*_{j+1})$. Observe that the sequence

$$\sigma_1, \dots, \sigma_{i_1-1}, \sigma_{i_1} = 2, \dots, \sigma_j, \sigma_{j+1}, \dots, \sigma_{i_2} = -2, \sigma_{i_2+1}, \dots, \sigma_{\ell^*_t}$$

after the isotopy described above turns into

$$\sigma_1, \dots, \sigma_{i_1-1}, \sigma_{i_1+1}, \dots, \sigma_j, -2, 2, \sigma_{j+1}, \dots, \sigma_{i_2-1}, \sigma_{i_2+1}, \dots, \sigma_{\ell^*_t}.$$

Now we can see that the numbers $N(t^*_i)$ are identical before and after the isotopy, for any $i \in \{0, \dots, i_1 - 1, i_2, \dots, \ell^*_t\}$. Consequently, the difference between the parameters

$$\sum_{i=0}^{\ell^*_t} N(t^*_i)$$

before and after the isotopy equals the difference of the sums

$$\sum_{i=i_1}^{i_2-1} (N(t^*_i) - N(t^*_{i_1-1}))$$

before and after the isotopy, which equals

$$\begin{aligned}
&[2(i_2 - i_1) + (i_2 - i_1 - 1) \cdot \sigma_{i_1+1} + \dots \\
&\qquad + (i_2 - j) \cdot \sigma_j + (i_2 - (j+1))\sigma_{j+1} + \dots + \sigma_{i_2-1}] \\
&\quad - [(i_2 - i_1)\sigma_{i_1+1} + (i_2 - i_1 - 1)\sigma_{i_1+2} + \dots + (i_2 - j + 1) \cdot \sigma_j \\
&\qquad - 2 + (i_2 - j - 2) \cdot \sigma_{j+1} + \dots + \sigma_{i_2-2}] = \\
&\qquad = 2(i_2 - i_1) + 2 - (\sigma_{i_1} + \dots + \sigma_j) + \sigma_{j+1} + \dots + \sigma_{i_2-1}.
\end{aligned}$$

Since $\sigma_i = \pm 2$, this difference is at least

$$2(i_2 - i_1) + 2 - 2(i_2 - i_1) \geq 2.$$

This, however, is impossible by property (C3), so the proof of Lemma 1.6 is complete. □

LEMMA 1.7. *Let t be a noncritical level such that there are both upper and lower disks at t, let k^2 be a connected component of $\alpha^{-1}(S^2(t))$ such that there are two distinct connected components h_1^1, h_2^1 in ∂k^2 that both intersect an edge e of $\mathcal{F}^3$ in points o_1, o_2, and let e_{12} denote the connected component of $e-(o_1+o_2)$ that connects o_1, o_2. Then there is no disk d^2 in $\mathcal{F}^3$ such that*

$$\partial d^2 = e_{12} + b^1,$$

where b^1 is an arc properly embedded in k^2.

Proof. Arguing on the contrary, assume the existence of such a disk d^2 with $\partial d^2 = e_{12} + b^1$. Isotopically deforming d^2 (to eliminate saddle points of the intersection $d^2 \cap k^2$ in $\operatorname{Int} b^1$), we can suppose that

$$b^1 = \mathcal{N}^2_{d^2}(b^1) \cap k^2.$$

Now it is clear that $\alpha(d^2)$ is an upper (resp. lower) disk at level t, $\alpha(e_{12})$ is its upper (resp. lower) arc and $\alpha(b^1)$ is its projection arc. For definiteness, let $\alpha(d^2)$ be upper. It follows from the hypothesis of the lemma that there is a lower disk D^2 at level t with

$$\partial D^2 = A^1 + B^1,$$

where A^1 is its lower arc, B^1 is its projection arc.

Let G^3 be the polyhedron in $\mathcal{F}^3$ that contains the surface k^2, let h_1^2 be the connected component of $\partial G^3 - h_1^1$ that is disjoint from

$$\operatorname{Int} \mathcal{N}^2_{e_{12}}(e_{12} \cap h_1^1),$$

and let u^2 be a disk which lies in a regular neighborhood $\mathcal{N}^3_{G^3}(h_1^2)$ and has the properties

$$u^2 \subset \operatorname{Int} G^3, \quad \mathcal{N}^2_{u^2}(\partial u^2) \cap k^2 = \partial u^2;$$

see Fig. 10.

We can also assume that the intersection $\alpha^{-1}(D^2) \cap u^2$ consists of finitely many double curves and arcs. Applying standard innermost and outermost arguments, we compress D^2 to get u^2 disjoint from $\alpha^{-1}(D^2)$.

By Lemma 1.6, the intersection

$$\operatorname{Int} \alpha(b^1) \cap \operatorname{Int} B^1$$

is nonempty. Let $b^0 \in b^1$ be a point such that

$$\alpha(b^0) \in \operatorname{Int} \alpha(b^1) \cap \operatorname{Int} B^1$$

and if b_1^1 is the connected component of $b^1 - b^0$ which connects points o_1 and b^0 then $\operatorname{Int} b_1^1$ is disjoint from $\alpha^{-1}(D^2)$; see Fig. 10.

Note that, by Lemma 1.5, $\operatorname{Int} e_{12}$ is disjoint from $\alpha^{-1}(S^2(t))$. Without loss of generality, we can assume that the intersection $u^2 \cap b^1$ is a single point u^0; see Fig. 10.

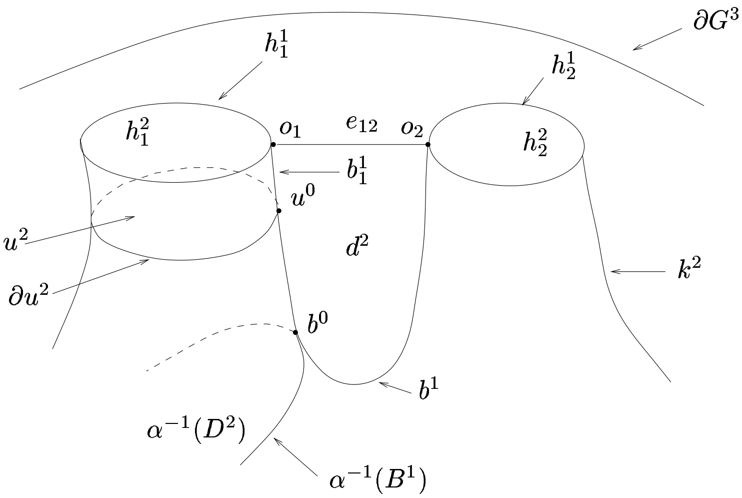


FIGURE 10

Now we push a regular neighborhood $\alpha(\mathcal{N}^2_{u^2}(u^0))$ along $\alpha(b_1^1)$ towards the point $\alpha(b^0)$ and then merge the resulting deformed disk $\alpha(u^2)$ with $\alpha(D^2)$. Doing this results in changing the lower disk D^2 (but not its lower arc A^1) and in elimination of the point $b^0 \in \operatorname{Int}\alpha(b^1) \cap \operatorname{Int} B^1$. By repeating this operation we can eliminate all points in $\operatorname{Int}\alpha(b^1) \cap \operatorname{Int} B^1$. This contradicts Lemma 1.6 and completes the proof of Lemma 1.7. □

Before stating and proving our next lemma we give a few more definitions.

Let D^2 be an upper (resp. lower) disk at level t, $\partial D^2 = A^1 + B^1$, where A^1 is its upper (resp. lower) arc and B^1 is its projection arc. We call D^2 *simple* if

$$D^2 \cap M^3_\Omega(2) = A^1.$$

We say that D^2 is *simplest* if D^2 is a part of $M^3_\Omega(2)$. Note that if D^2 is a simplest upper (resp. lower) disk then there is a simple upper (resp. lower) disk $\widetilde{D}^2$ in a regular neighborhood $\mathcal{N}^3_{M^3_\Omega}(D^2)$ with the same upper (resp. lower) arc (to get $\widetilde{D}^2$ it suffices to push D^2 into a 3-cell of M^3_Ω keeping its upper (resp. lower) arc fixed). Also, note that D^2 is simple if and only if there are a polyhedron G^3 in $\mathcal{F}^3$ and a disk $d^2 \subset G^3$ such that $\alpha(d^2) = D^2$ and $\alpha(d^2 \cap \partial G^3) = A^1$. Clearly, we can write $\partial d^2 = a^1 + b^1$, where $\alpha(a^1) = A^1$, $\alpha(b^1) = B^1$, and we denote d^2, a^1, and b^1 by $\alpha^{-1}(D^2)$, $\alpha^{-1}(A^1)$, and $\alpha^{-1}(B^1)$, respectively (whenever this notation is not ambiguous).

Let c^1 be a curve in $\partial\mathcal{F}^3$ which is in general position with respect to the 1-spine $\mathcal{F}^3(1)$ of $\mathcal{F}^3$ (that is, c^1 intersects $\mathcal{F}^3(1)$ in finitely many piercing points (if any) which are not vertices of $\mathcal{F}^3$). Such a curve c^1 is called *quasinormal* if either c^1 is disjoint from edges of $\mathcal{F}^3$ (in which case c^1 is called *trivial*) or any two consecutive along c^1 points o_1 and o_2 in $c^1 \cap \mathcal{F}^3(1)$ do not belong to the same edge of $\mathcal{F}^3$.

By Lemma 1.3, t_1^* is an ℓ-level and $t_{\ell_t^*}^*$ is obviously a u-level. Therefore, there is a $j \in \{1, \dots, \ell_{t-1}^*\}$ such that t_j^* is an ℓ-level and t_{j+1}^* is a u-level.

LEMMA 1.8. *Let t_j^* be an ℓ-level and t_{j+1}^* be a u-level. Then there is a noncritical level $\hat{t}$, $\hat{t} \in (t_j^*, t_{j+1}^*)$, such that the intersection*

$$\alpha^{-1}(S^2(\hat{t})) \cap \partial\mathcal{F}^3$$

consists of quasinormal curves.

Proof. Consider the secondary critical levels

$$t_j^* < t_{j,1}^* < \cdots < t_{j,k(j,t)}^* < t_{j+1}^*$$

between the primary critical levels t_j^* and t_{j+1}^* (see property (C6)). Let

$$t_1 \in (t_j^*, t_{j,1}^*), \quad t_i \in (t_{j,i-1}^*, t_{j,i}^*),$$

$i = 2, \dots, k(j,t)$, and

$$t_{k(j,t)+1} \in (t_{j,k(j,t)}^*, t_{j+1}^*).$$

It follows from property (C4) that there is a simplest lower disk at level t_1 and that there is a simplest upper disk at level $t_{k(j,t)+1}$. By Lemma 1.6, at any level t there cannot exist both a simplest lower disk and a simplest upper disk. Therefore, there is an $i' \in \{1, \dots, k(j,t)\}$ such that one of the following cases holds.

(D1) There is no simplest lower disk and no simplest upper disk at level $t_{i'}$.

(D2) There is a simplest lower disk at level $t_{i'}$ and there is a simplest upper disk at level $t_{i'+1}$.

If (D1) holds then the intersection

$$\alpha^{-1}(S^2(t_{i'})) \cap \partial\mathcal{F}^3$$

consists of quasinormal curves, and Lemma 1.8 is proved.

In case (D2), there must be a saddle point $P^0(t_{j,i'}^*)$ at the secondary critical level $t_{j,i'}^* \in (t_{i'}, t_{i'+1})$. Let E^2 be the 2-cell of M_Ω^3 that contains $P^0(t_{j,i'}^*)$. Then $\operatorname{Cls} E^2$ contains a simplest lower disk D_L^2 at level $t_{i'}$ which disappears when passing through $t_{j,i'}^*$ in positive direction. Also, $\operatorname{Cls} E^2$ contains a simplest upper disk D_U^2 at level $t_{i'+1}$ which disappears when passing through $t_{j,i'}^*$ in negative direction; see Fig. 11.

Now we can see that there are upper and lower disks at any level

$$t \in (t_{i'}, t_{i'+1})$$

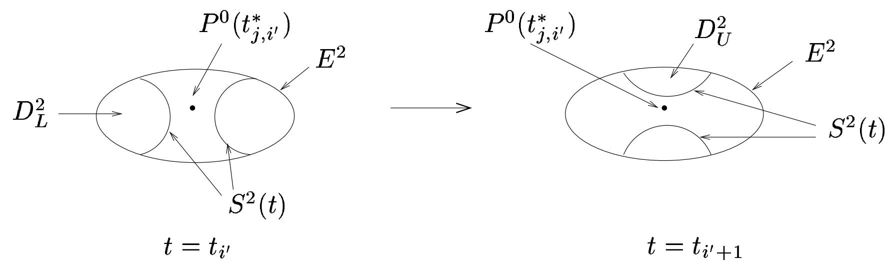


FIGURE 11

whose projection arcs B^1_1, B^1_2 have the property that

$$B^1_1 \cap B^1_2 \subset \partial B^1_1 \cap \partial B^1_2.$$

A contradiction to Lemma 1.6 shows that case (D2) is impossible and the proof of Lemma 1.8 is complete. □

We say that a compact surface U^2 embedded in M^3_Ω is *quasinormal* if U^2 is in general position with respect to $M^3_\Omega(2)$, the intersection $U^2 \cap M^3_\Omega(1)$ is nonempty, and connected components of $\alpha^{-1}(U^2)$ are disks whose boundaries are quasinormal curves in $\partial\mathcal{F}^3$.

Let us investigate the 2-sphere $S^2(\hat{t})$ of Lemma 1.8 in more detail. First we observe that $S^2(\hat{t})$ can be described as follows.

(E1) There are finitely many pairwise disjoint 2-spheres $U^2_1, \ldots, U^2_{k_U}$ embedded in M^3_Ω each of which is either quasinormal or disjoint from $M^3_\Omega(2)$.

(E2) There are pairwise disjoint annuli T^2_i, which could be thought of as 'thin tubes', and pairwise disjoint disks $d^2_1(T^2_i)$, $d^2_2(T^2_i)$ in $U^2_1 + \cdots + U^2_{k_U}$, $i = 1, \ldots, k_T$, such that

$$\partial d^2_1(T^2_i) + \partial d^2_2(T^2_i) = \partial T^2_i.$$

In addition, for every i, $i = 1, \ldots, k_T$, the surface

$$d^2_1(T^2_i) + d^2_2(T^2_i) + T^2_i$$

is a nonsingular 2-sphere which bounds a 3-ball T^3_i disjoint from $M^3_\Omega(1)$, and the intersection $T^3_i \cap M^3_\Omega(2)$ consists of disks properly embedded in T^3_i whose boundaries are essential curves in $\operatorname{Int} T^2_i$.

(E3) In the notation of parts (E1)–(E2),

$$S^2(\hat{t}) = \sum_{i=1}^{k_U} U^2_i - \sum_{j=1}^{k_T} (d^2_1(T^2_j) + d^2_2(T^2_j)) + \sum_{j=1}^{k_T} T^2_j.$$

Note that the tubes $T_1^2, \dots, T_{k_T}^2$ can run through 2-cells of M_Ω^3, thus producing α-images of trivial quasinormal curves in

$$\alpha^{-1}(S^2(\hat{t})) \cap \partial\mathcal{F}^3;$$

they can also run through each other, be knotted, linked, etc.

LEMMA 1.9. *Precisely one of the quasinormal 2-spheres $U_1^2, \dots, U_{k_U}^2$ is A1-normal (and each of the others is either normal or disjoint from $M_\Omega^3(2)$).*

Proof. Since $\hat{t} \in (t_j^*, t_{j+1}^*)$, t_j^* is an ℓ-level and t_{j+1}^* is a u-level, it follows that for any $t \in (t_j^*, t_{j+1}^*)$ there are both upper and lower disks at level t.

Consider an upper disk D_U^2 at level $\hat{t}$ with

$$\partial D_U^2 = A_U^1 + B_U^1,$$

where A_U^1 is the upper arc of D^2 and B_U^1 is the projection arc of D_U^2. Isotopically deforming and compressing D_U^2, we can assume that the intersection $D_U^2 \cap M_\Omega^3(2)$ consists of finitely many double arcs $H_1^1, H_2^1, \dots$ with

$$\partial(H_1^1 + H_2^1 + \dots) \subset \operatorname{Int} B_U^1.$$

Let H_1^1 be such an outermost arc on D_U^2, let K_1^1 be the connected component of $B_U^1 - \partial H_1^1$ that contains ∂H_1^1, and let $K_1^2 \subset D_U^2$ be a disk bounded by H_1^1, K_1^1. Then $\operatorname{Int} K_1^2$ is disjoint from $M_\Omega^3(2)$.

Let E^2 be the 2-cell of M_Ω^3 that contains the arc H_1^1. If the points ∂H_1^1 lie on the same connected component of $E^2 \cap S^2(\hat{t})$ then we can compress the disk D_U^2 (along a disk in E^2) to eliminate the arc H_1^1 as in the proof of Lemma 1.3. Hence, we can assume that the two points ∂H_1^1 lie on distinct connected components of $E^2 \cap S^2(\hat{t})$, say, on C_1^1, C_2^1.

Suppose that one of C_1^1, C_2^1 is a curve in E^2. Then we push a regular neighborhood $\mathcal{N}_{E^2}^2(H_1^1)$ through the disk K_1^2 to K_1^1 and then slightly off K_1^1 (to the other side of $S^2(\hat{t})$). As a result, the connected components C_1^1, C_2^1 of $E^2 \cap S^2(\hat{t})$ merge into a single open arc or curve and the double arc

$$H_1^1 \subset D_U^2 \cap E^2$$

gets eliminated. Observe that this isotopy of $E^2 \subset M_\Omega^3(2)$ does not affect properties (C1)–(C4) of the foliation of M_Ω^3 by 2-spheres $S^2(t)$, $t \in (0,1)$. To restore properties (C5)–(C6), we adjust the deformed part of $M_\Omega^3(2)$ by arbitrarily small isotopic deformations. Properties (E1)–(E3) of $S^2(\hat{t})$ are also retained. In order to see this, we note that after the deformations the intersection

$$\alpha^{-1}(S^2(\hat{t})) \cap \partial\mathcal{F}^3$$

still consists of quasinormal curves (we just lose one trivial curve), and so we can look at $S^2(\hat{t})$ as before to get properties (E1)–(E3).

In view of the above argument, we can assume that the connected components C_1^1, C_2^1 of $E^2 \cap S^2(\hat{t})$ that contain points ∂H_1^1 are distinct open arcs.

Let G^3 be the polyhedron in $\mathcal{F}^3$ that contains the disk d_1^2 with $\alpha(d_1^2) = K_1^2$ and $\partial d_1^2 = h_1^1 + k_1^1$, where $\alpha(h_1^1) = H_1^1$, $\alpha(k_1^1) = K_1^1$. Let G^2 be the face of G^3 that contains the arc h_1^1. Also, let c_1^1, c_2^1 denote the connected components of $\alpha^{-1}(S^2(\hat{t})) \cap G^2$ that contain points ∂h_1^1 so that

$$\alpha(\operatorname{Int} c_1^1) = C_1^1, \quad \alpha(\operatorname{Int} c_2^1) = C_2^1.$$

Since G^2 is a biangle or triangle and curves in $\alpha^{-1}(S^2(\hat{t})) \cap \partial\mathcal{F}^3$ are quasinormal, it follows that the arcs c_1^1, c_2^1 originate from the same edge e of G^2. Now we can see the existence of a simple upper disk at level $\hat{t}$ of the form $\alpha(d^2)$, where $d^2 \subset G^3$ is a disk such that $\partial d^2 = e_{12} + b_1^1$, where $e_{12} \subset e$, $\partial e_{12} \subset \partial(c_1^1 + c_2^1)$, and $b_1^1 \subset \alpha^{-1}(S^2(\hat{t}))$; see Fig. 12.

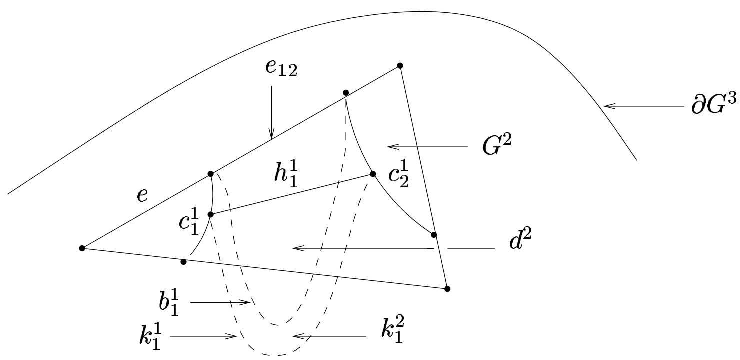


FIGURE 12

In view of Lemma 1.7, the existence of this disk $\alpha(d^2)$ proves that the arcs c_1^1, c_2^1 belong to the same curve c^1 in $\alpha^{-1}(S^2(\hat{t})) \cap \partial G^3$.

Referring to the structure of the 2-sphere $S^2(\hat{t})$ described in properties (E1)–(E2), we pick the 2-sphere among

$$U_1^1, \ldots, U_{k_U}^2,$$

say $U_{i_2}^2$, that contains the curve $\alpha(c^1)$. It is clear that there is also a simple upper disk R_U^2 at level $\hat{t}$ whose upper arc is $\alpha(e_{12})$ and whose projection arc belongs to the quasinormal disk P_U^2 of $U_{i_2}^2$ whose boundary is $\alpha(c^1)$.

In a completely analogous manner, we establish the existence of a simple lower disk R_L^2 at level $\hat{t}$ whose projection arc belongs to a quasinormal disk P_L^2 of a 2-sphere $U_{i_1}^2$. It follows from Lemma 1.6 that $U_{i_1}^2 = U_{i_2}^2$ and $P_U^2 = P_L^2$. In addition, we observe that if c_0^1 is a curve in $\alpha^{-1}(U_i^2) \cap \partial\mathcal{F}^3$ which crosses an edge g of $\mathcal{F}^3$ at least twice then there is a simple upper (or lower) disk whose upper (lower) arc is a part of $\alpha(g)$ and whose projection arc belongs to the quasinormal disk of U_i^2 bounded by $\alpha(c_0^1)$. This observation together with Lemma 1.6 implies that every curve in

$$\alpha^{-1}(U_1^2 + \cdots + U_{k_U}^2) \cap \partial\mathcal{F}^3$$

different from c^1 bounds a normal disk and that c^1 bounds an $A1$-normal disk. Now we can conclude that $U_{i_2}^2$ is an $A1$-normal 2-sphere and every other 2-sphere in

$$U_1^2, \dots, U_{k_U}^2$$

is either normal (and hence trivial) or disjoint from $M_\Omega^3(2)$. This completes the proof of Lemma 1.9. □

Theorem 1 is now immediate from Lemmas 1.2 and 1.9.

2. Proof of Theorem 2

We continue to use the basic notation and definitions introduced in Sections 0 and 1.

Let M_Ω^3 be a connected compact closed orientable 3-manifold given by an arbitrary finite cell decomposition $\Omega = \Omega(\mathcal{F}^3)$ with $r_0, r_1, r_2, r_3 > 0$. We first show how to obtain properties (A1)–(A2).

If property (A1) fails, that is, if $r_0 = |\mathcal{E}^0| > 1$, then, one by one, we can eliminate by contraction those 1-cells in $\mathcal{E}^1$ that connect two distinct 0-cells in $\mathcal{E}^0$, merging these two 0-cells into one.

Now suppose that property (A1) holds and property (A2) fails; that is, there is an edge e in $\mathcal{F}^3$ with ∂e a single vertex. Consider a 1-complex

$$K^1(e) \subseteq \mathcal{F}^3(1)$$

which consists of all edges f of $\mathcal{F}^3$ with $\alpha(f) = \alpha(e)$. Suppose that p is a shortest closed path in $\partial\mathcal{F}^3$ (if there is one) which consists of edges in $K^1(e)$. Clearly, p represents a nonsingular curve in $\partial\mathcal{F}^3$.

If p does not bound a face of $\mathcal{F}^3$ then we draw a disk p^2 properly embedded in $\mathcal{F}^3$ with $\partial p^2 = p$ and cut the polyhedron G^3 which contains p^2 into two halves along p^2. Repeating such cuts we can assume that every shortest closed path in $\partial\mathcal{F}^3$ which consists of edges in $K^1(e)$ bounds a face. Now we contract all edges in $K^1(e)$ into vertices. We also contract all faces on $\partial\mathcal{F}^3$ whose boundaries consist of edges in $K^1(e)$ into points and eliminate those (possible) connected components of $\mathcal{F}^3$ which turn into vertices after this series of contractions.

As a result, we obtain a new collection $\widetilde{\mathcal{F}}^3$ which may be empty (in which case M_Ω^3 is obviously the 3-sphere) or may split into several subcollections

$$\widetilde{\mathcal{F}}_1^3, \dots, \widetilde{\mathcal{F}}_{\ell'}^3, \quad \ell' \geq 1,$$

which define 3-manifolds

$$\widetilde{M}^3_{1,\widetilde{\Omega}_1}, \dots, \widetilde{M}^3_{\ell',\widetilde{\Omega}_{\ell'}}$$

so that M^3_Ω is the connected sum of $\widetilde{M}^3_{1,\widetilde{\Omega}_1}, \dots, \widetilde{M}^3_{\ell',\widetilde{\Omega}_{\ell'}}$. Note that the total number of 1-cells in $\widetilde{M}^3_{1,\widetilde{\Omega}_1}, \dots, \widetilde{M}^3_{\ell',\widetilde{\Omega}_{\ell'}}$ is $r_1 - 1$.

It can be shown that $\widetilde{M}^3_{1,\widetilde{\Omega}_1}, \dots, \widetilde{M}^3_{\ell',\widetilde{\Omega}_{\ell'}}$ have property (A1). Alternatively, we can just restore property (A1) as above by further decreasing the number of 1-cells.

Repeating the above reductions (which decrease each time the total number of 1-cells) yields a finite collection of 3-manifolds

$$M^3_{1,\Omega_1}, \dots, M^3_{\ell,\Omega_\ell}$$

such that M^3_Ω is the connected sum of $M^3_{1,\Omega_1}, \dots, M^3_{\ell,\Omega_\ell}$ and cell decompositions $\Omega_1, \dots, \Omega_\ell$ have properties (A1)–(A2) (where the case $\ell = 0$ means, as above, that M^3_Ω is the 3-sphere). Clearly, the number ℓ of such manifolds and the total number of their 1-cells is at most r_1.

Consider the following additional property of Ω.

(A3) If G^2 is a biangle face in $\mathcal{F}^3$ and $\partial G^2 = e_1 + e_2$, where e_1, e_2 are edges, then $\alpha(e_1) = \alpha(e_2)$ in M^3_Ω (recall that e_1, e_2 are not oriented).

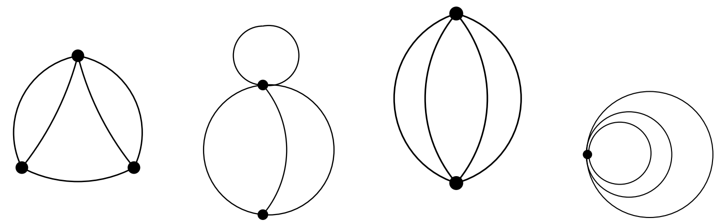
Suppose that a cell decomposition Ω of M^3_Ω has properties (A1)–(A2) and property (A3) fails for a biangle face G^2 with

$$\partial G^2 = e_1 + e_2,$$

so $\alpha(e_1) \neq \alpha(e_2)$ in M^3_Ω. Then we can just eliminate the 2-cell $E^2 \in \mathcal{E}^2$ with $\alpha(\operatorname{Int} G^2) = E^2$ by pushing $\alpha(\operatorname{Int} e_1)$ through E^2 to $\alpha(\operatorname{Int} e_2)$ and attaching $\alpha(\operatorname{Int} e_1)$ to $\alpha(\operatorname{Int} e_2)$. We also make a corresponding change to $\mathcal{F}^3$ (which is to close out two biangle faces in $\mathcal{F}^3$). As a result, we still have properties (A1)–(A2), but a fewer number r_1 of 1-cells. Hence, more such reductions will yield property (A3).

Now suppose that a 3-manifold M^3_Ω is given by a triangulation Ω (that is, all polyhedra in $\mathcal{F}^3$ are tetrahedra and the restriction $\alpha|_{F^3}$ of α on each $F^3 \in \mathcal{F}^3$ is an embedding). Let us consider what happens to $\mathcal{F}^3$ when the reductions used above to obtain properties (A1)–(A3) are applied to Ω.

First, after obtaining property (A1), every polyhedron G^3 in the new $\mathcal{F}^3$ either is a standard tetrahedron T^3 or can be obtained from a standard tetrahedron T^3 by collapsing 1, 2, or 3 edges of T^3 into vertices. Specifically, if one edge of T^3 is collapsed then we have a 1-degenerate tetrahedron with 3 vertices, 2 triangle faces and 2 biangle faces; see Fig. 13(a). If 2 edges of T^3 are collapsed then we obtain a 2-degenerate tetrahedron which has 2 vertices and either 1 triangle face, 2 biangle faces and 1 monoangle face (see Fig. 13(b1)) or 4 biangle faces (see Fig. 13(b2)). Finally, if 3 edges of T^3 are collapsed then we have a 3-degenerate tetrahedron with a single vertex, 2 biangle and 2 monoangle faces, Fig. 13(c). It is easy to see that we need to collapse at most 3 edges of T^3 in order to obtain property (A1).



(a) (b1) (b2) (c)

FIGURE 13

Next, when making reductions to obtain property (A2), we can get new polyhedra that result from cuts along disks properly embedded in $\mathcal{F}^3$, bounded by shortest closed paths that bound no faces on $\mathcal{F}^3$. Such cuts can add polyhedra which look quite similar to j-degenerate tetrahedra, $j = 1, 2, 3$, but have a fewer number of biangle faces (namely, one less). (Alternatively, these polyhedra can be obtained from j-degenerate tetrahedra by collapsing some of their biangle faces.) Contractions that are made after the cuts enable us to conclude that polyhedra of the type shown in Fig. 13(b1) and Fig. 13(c) will disappear, and so the final collection will consist of tetrahedra, 1-degenerate tetrahedra with 3 vertices, 2 triangle faces and several (1 or 2) biangle faces, and 2-degenerate tetrahedra which have two vertices and several (3 or 4) biangle faces.

Finally, making reductions to get property (A3) can only decrease the number of biangle faces in polyhedra of $\mathcal{F}^3$.

Henceforth, a 1-*degenerate tetrahedron* G^3 is defined to be a polyhedron with 3 vertices, 2 triangle faces and $\ell_1(G^3) \geq 0$ biangle faces. A 2-*degenerate tetrahedron* G^3 is defined to be a polyhedron with 2 vertices and $\ell_2(G^3) \geq 1$ biangle faces.

Suppose that M^3_{1,Ω_1} is a 3-manifold whose cell decomposition Ω_1 has properties (A1)–(A3) and polyhedra in $\mathcal{F}^3_1$ are tetrahedra and j-degenerate tetrahedra, $j = 1, 2$.

Let G^3 be a 2-degenerate tetrahedron in $\mathcal{F}^3_1$ and suppose there is a face G^2 of G^3 such that $\alpha_1(G^2) = \alpha_1(F^2)$, where F^2 is a face of a polyhedron F^3 different from G^3. Then we can eliminate G^3 by attaching it to F^3 without changing the type of F^3. Therefore, a 2-degenerate tetrahedron G^3 can be eliminated unless for every face G^2 in ∂G^3 there is another face, F^2, in ∂G^3 such that $\alpha_1(G^2) = \alpha_1(F^2)$. In the latter case, G^3 is the only polyhedron in $\mathcal{F}^3_1$ (for M^3_{1,Ω_1} is connected). Furthermore, it follows from property (A3) that M^3_{1,Ω_1} contains a single 1-cell. Hence, the link $\operatorname{Link} E^0_1$ contains 2 vertices and, obviously, 2 faces. Since $\operatorname{Link} E_1^0$ is a normal 2-sphere, it follows from the Euler formula

$$v - e + f = 2$$

for $\operatorname{Link} E_1^0$, where v, e and f denote the numbers of vertices, edges, and faces, respectively, in $\operatorname{Link} E_1^0$, and from the relation $v = f = 2$ that $e = 2$. This means that there are two edges and two faces in G^3. In particular, Ω is a Q-triangulation and M^3_{1,Ω_1} is the real projective 3-space.

The above argument enables us to 'fix' cell decompositions

$$\Omega_1, \dots, \Omega_\ell$$

of 3-manifolds

$$M^3_{1,\Omega_1}, \dots, M^3_{\ell,\Omega_\ell}, \quad \ell \geq 0,$$

that have properties (A1)–(A3) and are obtained from the original triangulated 3-manifold M^3_Ω and assume that every Ω_i is a Q-triangulation. It is clear that these cell decompositions $\Omega_1, \dots, \Omega_\ell$ can be constructed from M^3_Ω in polynomial time (of size of Ω).

Therefore, we can now assume that M^3_Ω is given by a Q-triangulation Ω (which also has property (A3)). In particular, $\operatorname{Link} E^0$ is a normal 2-sphere which, as before, is called a *trivial* normal 2-sphere.

Consider a system $\mathcal{V}^2$ of pairwise disjoint, pairwise normally nonparallel, normal 2-spheres

$$V_1^2, \dots, V_{k_V}^2$$

in M^3_Ω that is maximal with respect to these properties; that is, if V_0^2 is a normal 2-sphere disjoint from $\mathcal{V}^2$ then V_0^2 is normally parallel to one of the 2-spheres in $\mathcal{V}^2$. Such a system $\mathcal{V}^2$ will be called a *maximal system of nonparallel normal 2-spheres* in M^3_Ω. The existence of such a maximal system is well known (see [He76] and the estimate (2.3) below). Let us give an upper bound for $k_V = |\mathcal{V}^2|$.

Let $\mathcal{V}^2_{NS}$ be a maximal subset in $\mathcal{V}^2$ such that

$$M^3_\Omega - V^2_{NS}$$

is connected (so $\mathcal{V}^2_{NS}$ is a maximal subsystem of nonseparating 2-spheres in $\mathcal{V}^2$) and let $\mathcal{V}^2_{NS}$ contain ℓ_V 2-spheres. Then $M^3_\Omega - \mathcal{V}^2$ consists of

$$(k_V - \ell_V) + 1$$

connected components. Each of these connected components contains either the 0-cell E^0 or, otherwise, one of at most two pieces of $E^3 - \mathcal{V}^2$, where E^3 is a 3-cell of M^3_Ω, which does not sit between two normal disks of the same isotopy type in $\mathcal{V}^2 \cap \operatorname{Cls} E^3$. (If a connected component of $M^3_\Omega - \mathcal{V}^2$ contains no E^0 and no such pieces then it sits between two normally parallel 2-spheres of $\mathcal{V}^2$ which is impossible.)

Hence,

$$(k_V - \ell_V) + 1 \le 2r_3 + 1, \tag{2.1}$$

where $r_j = |\mathcal{E}^j|$, $j = 0, 1, 2, 3$.

Since each 2-sphere in $\mathcal{V}^2_{NS}$ gives rise to a free infinite cyclic factor in $\pi_1(M^3_\Omega)$, and the group $\pi_1(M^3_\Omega)$ has r_1 generators, it follows that

$$\ell_V \le r_1.$$

Now we have from inequality (2.1) that

$$k_V \le r_1 + 2r_3 = 2r_2 - r_1 + 2. \tag{2.2}$$

(The last equality follows from the Euler formula $r_0 - r_1 + r_2 - r_3 = 0$ for M^3_Ω and from $r_0 = 1$.)

Such a maximal system $\mathcal{V}^2$ of nonparallel 2-spheres can be effectively constructed using the Haken theory of normal surfaces. We refer the reader for more details to [Hn92], [JT95], [JR89], [HLP99] and give here only a brief outline in order to introduce the notation and definitions needed for proofs of Theorems 2–4.

For every polyhedron G^3 in $\mathcal{F}^3$ we consider unknowns which correspond to all isotopy types of normal disks in G^3. For example, if G^3 is a tetrahedron then we have 7 unknowns (corresponding to 4 types of normal triangles and 3 types of normal quadrangles), and if G^3 is a degenerate tetrahedron then we have 3 unknowns (which correspond to 3 types of simple normal disks in G^3). Hence, the total number of unknowns is

$$N(\Omega) = 7N_T + 3N_D, \tag{2.3}$$

where N_T is the number of tetrahedra in $\mathcal{F}^3$ and N_D is the number of degenerate tetrahedra in $\mathcal{F}^3$.

Let us recall how to write down matching equations in the unknowns introduced above.

For every pair of distinct edges $p = p(G^2)$ of a face $G^2 \subset \partial G^3$, G^3 in $\mathcal{F}^3$, let $x_1(p)$, $x_2(p)$ denote the unknowns that correspond to normal disks of G^3 whose boundaries contain an arc which connects points on edges in p. Note that if G^3 is a tetrahedron then there are indeed two unknowns $x_1(p)$, $x_2(p)$ associated with every pair p of edges of every face of G^3. (One of $x_1(p)$, $x_2(p)$ corresponds to a normal triangle and the other one to a normal quadrangle in G^3.) If G^2 is a biangle of a degenerate tetrahedron then there are also two unknowns $x_1(p)$, $x_2(p)$ associated with the pair of edges of ∂G^2. However, if G^2 is a triangle of a degenerate tetrahedron G^3, then there is only one unknown $x_1(p)$ associated with every pair p of edges in ∂G^2 (so $x_2(p)$ may actually be missing in the original definition).

Let $p = p(G^2)$ be a pair of distinct edges of ∂G^2, where G^2 is a face of ∂G^3 and G^3 is in $\mathcal{F}^3$, and let a^1 be an arc properly embedded in G^2 which connects the edges in p. Then there is another arc b^1 properly embedded in another face F^2 of $\mathcal{F}^3$ such that $\alpha(a^1) = \alpha(b^1)$ in M^3_Ω. Clearly, b^1 connects distinct edges in ∂F^2 and we can consider the pair q of these edges. Let $x_1(p)$, $x_2(p)$ (resp. $x_1(q)$, $x_2(q)$) denote the unknowns which are associated with p (resp. q). (One or both of $x_2(p)$ and $x_2(q)$ may be missing, as pointed out above.)

Now for every such (unordered) pair $\{p, q\}$ (over all faces G^2 of $\mathcal{F}^3$) we write down a matching equation

$$x_1(p) + x_2(p) = x_1(q) + x_2(q).$$

The set of all matching equations is called the *system of matching equations* for M^3_Ω and is denoted by

$$\mathrm{SME}(M^3_\Omega).$$

A nontrivial solution ν to $\mathrm{SME}(M^3_\Omega)$ with nonnegative integer entries will be called a *natural solution* to $\mathrm{SME}(M^3_\Omega)$. It is well known (see [Ha61], [Hn92]) that natural solutions to $\mathrm{SME}(M^3_\Omega)$ form a finitely generated additive semigroup. A natural solution to $\mathrm{SME}(M^3_\Omega)$ is called *fundamental* if it cannot be written as a sum of two other natural solutions. Clearly, the property of being finitely generated for the semigroup of natural solutions is equivalent to the finiteness of the set of fundamental solutions. Therefore, the fact that the semigroup of natural solutions is finitely generated also follows from the following result.

LEMMA 2.1. *Suppose that ν is a fundamental solution to* $\mathrm{SME}(M^3_\Omega)$. *Then every component of ν is at most*

$$N(\Omega) \cdot 2^{N(\Omega)-1},$$

where $N(\Omega)$ is the number of unknowns in $\mathrm{SME}(M^3_\Omega)$*; see also* (2.3).

Proof. To prove this, we repeat the proof of Lemma 6.1 in [HLP99]. As in [HLP99], we first use Hadamard's inequality to show that if ν is a vertex solution to $\mathrm{SME}(M^3_\Omega)$ then each entry in ν is at most $2^{N(\Omega)-1}$. (Note that the square of the Euclidean length of rows in the matrix $\mathbf{M}$ is still bounded by 4.) Then we proceed as in [HLP99] to remove the restriction of being vertex solution and obtain a weaker upper bound

$$N(\Omega) \cdot 2^{N(\Omega)-1}$$

for entries of a fundamental solution to $\mathrm{SME}(M^3_\Omega)$. □

Let U^2 be a normal surface in M^3_Ω. Then the *normal vector* $\nu(U^2)$ of U^2 is defined to be an $N(\Omega)$-tuple

$$\nu(U^2) = (\nu_1, \ldots, \nu_{N(\Omega)}),$$

where ν_i is the number of normal disks in $\alpha^{-1}(U^2)$ whose isotopy type corresponds to the ith unknown x_i of $\mathrm{SME}(M^3_\Omega)$. (As in Lemma 2.1, it is implied that unknowns of $\mathrm{SME}(M^3_\Omega)$ are ordered.) It is easy to see that $\nu(U^2)$ is a natural solution to $\mathrm{SME}(M^3_\Omega)$. Moreover, this natural solution $\nu(U^2)$ satisfies, for each tetrahedron T^3 in $\mathcal{F}^3$, the so-called *admissibility condition* which states that if x_{i_1}, x_{i_2}, x_{i_3} are the unknowns of $\mathrm{SME}(M^3_\Omega)$ which correspond to 3 normal quadrangles of T^3 then at most one of numbers ν_{i_1}, ν_{i_2}, ν_{i_3} is positive. (This is because two quadrangles of different isotopy types in T^3 would have to intersect.)

Conversely, if

$$\nu = (\nu_1, \dots, \nu_{N(\Omega)})$$

is a natural solution to $\mathrm{SME}(M^3_\Omega)$ which satisfies the admissibility condition for every tetrahedron T^3 in $\mathcal{F}^3$ (recall that Ω is a Q-triangulation) then it is not difficult to show that there is a normal surface U^2 (perhaps, not connected) in M^3_Ω whose normal vector $\nu(U^2)$ is ν and that U^2 is unique, in the sense that if $\widetilde{U}^2$ is a normal surface such that

$$\nu(\widetilde{U}^2) = \nu(U^2) = \nu$$

then $\widetilde{U}^2$ is normally parallel to U^2.

Furthermore, if

$$\nu = \nu^1 + \nu^2,$$

where ν^1, ν^2 are also natural solutions to $\mathrm{SME}(M^3_\Omega)$, then ν^1, ν^2 also satisfy the admissibility condition. Hence there are normal surfaces U^2_1, U^2_2 with

$$\nu(U^2_1) = \nu^1, \quad \nu(U^2_2) = \nu^2,$$

and it can be shown that the surface U^2 can be geometrically constructed from U^2_1, U^2_2 as a Haken sum of surfaces U^2_1, U^2_2 by means of some surgery called regular exchanges (along double curves in $U^2_1 \cap U^2_2$; for more details see [JT95], [JR89], [Hn92]).

Recall that a normal surface whose normal vector is a fundamental solution to $\mathrm{SME}(M^3_\Omega)$ is called *fundamental*. Clearly, a fundamental surface is connected.

More generally, let ν be a natural solution to $\mathrm{SME}(M^3_\Omega)$ which satisfies the admissibility condition,

$$\nu = \nu^1 + \cdots + \nu^k,$$

where $\nu^1, \dots, \nu^k$ are fundamental solutions, and let $U^2_1, \dots, U^2_k$ be fundamental surfaces with

$$\nu(U^2_1) = \nu^1, \ \dots, \ \nu(U^2_k) = \nu^k.$$

Then U^2 can also be geometrically constructed from fundamental surfaces $U^2_1, \dots, U^2_k$ by means of regular exchanges.

The first step in the construction of a maximal system $\mathcal{V}^2$ of nonparallel normal 2-spheres in M^3_Ω is given by the following result.

LEMMA 2.2. *Suppose that there is a nontrivial normal 2-sphere in M^3_Ω. Then there is also a nontrivial normal 2-sphere in M^3_Ω whose normal vector is $\delta_0\nu$, where ν is a fundamental solution to* $\mathrm{SME}(M^3_\Omega)$ *and $\delta_0 \in \{1, 2\}$.*

Proof. Let U^2 be a nontrivial 2-sphere in M^3_Ω. Suppose that U^2 is not fundamental (otherwise, there is nothing to prove). Then we can write

$$\nu(U^2) = \nu^1 + \cdots + \nu^k,$$

where $k > 1$ and $\nu^1, \ldots, \nu^k$ are fundamental solutions to SME(M^3_Ω). Let

$$U^2_1, \ldots, U^2_k$$

be fundamental surfaces whose normal vectors are, respectively, $\nu^1, \ldots, \nu^k$. Then

$$U^2 = U^2_1 + \cdots + U^2_k$$

is a Haken sum of surfaces $U^2_1, \ldots, U^2_k$. Recall that the Euler characteristic is additive under the Haken sum. Since U^2 is connected and is a nontrivial 2-sphere, it follows that none of $U^2_1, \ldots, U^2_k$ is a trivial normal 2-sphere. Since the Euler characteristic $\chi(U^2)$ is 2, it follows from the equality

$$2 = \chi(U^2) = \chi(U^2_1) + \cdots + \chi(U^2_k)$$

that there is an i such that $\chi(U^2_i) > 0$. Note that U^2_i is a connected surface (for U^2_i is fundamental). Hence $\chi(U^2_i) = 2$ or $\chi(U^2_i) = 1$. If $\chi(U^2_i) = 2$ then U^2_i is a normal 2-sphere which is nontrivial (otherwise, U^2_i would be a connected component of U^2, contradicting the connectedness of U^2), that is, U^2_i is a desired 2-sphere. If $\chi(U^2_i) = 1$ then U^2_i is a projective plane and we consider a closed regular neighborhood

$$\bar{\mathcal{N}}^3_{M^3_\Omega}(U^2_i).$$

Since M^3_Ω is orientable, it follows that $\partial\bar{\mathcal{N}}^3_{M^3_\Omega}(U^2_i)$ is a normal 2-sphere with

$$\nu(\partial\bar{\mathcal{N}}^3_{M^3_\Omega}(U^2_i)) = 2\nu(U^2_i).$$

It remains to observe that $\partial\bar{\mathcal{N}}^3_{M^3_\Omega}(U^2_i)$ is not a trivial normal 2-sphere because its normal vector consists of even components (otherwise, $\mathcal{F}^3$ would contain a single polyhedron with 2 vertices and 2 faces and no normal nontrivial 2-sphere would exist in M^3_Ω). □

Using Lemmas 2.1–2.2, we can find the first nontrivial 2-sphere V^2_1 in $\mathcal{V}^2$ (or show that there is none).

Arguing by induction on $j \geq 1$, assume that a system $\mathcal{V}^2$ of j pairwise disjoint normal 2-spheres $V^2_1, \ldots, V^2_j$ (which are pairwise normally nonparallel) has already been constructed. Consider a 3-manifold

$$M^3_\Omega - \mathcal{V}^2$$

with boundary ($M^3_\Omega - \mathcal{V}^2$ need not be connected). This 3-manifold $M^3_\Omega - \mathcal{V}^2$ has a natural cell decomposition $\Omega(\mathcal{V}^2)$ which is defined by Ω and $\mathcal{V}^2$, that is, the corresponding set $\mathcal{F}^3(\mathcal{V}^2)$ of polyhedra for $M^3_\Omega - \mathcal{V}^2$ consists of connected components of $\mathcal{F}^3 - \alpha^{-1}(\mathcal{V}^2)$. The restriction of α on

$$\mathcal{F}^3(\mathcal{V}^2) \to M^3_\Omega - \mathcal{V}^2$$

will still be denoted by α.

Let E^i be an i-cell of $M^3_\Omega - \mathcal{V}^2$, $i = 0, 1, 2, 3$. We will call E^i *outer* if

$$E^i \subset \partial(M^3_\Omega - \mathcal{V}^2)$$

and *inner* otherwise. Analogously, a closed i-cell G^i, $i = 0, 1, 2$, of $\mathcal{F}^3(\mathcal{V}^2)$ is *outer* if $\alpha(\operatorname{Int} G^i)$ is outer in $M^3_\Omega - \mathcal{V}^2$ and *inner* otherwise. As before, closed 0-, 1-, and 2-cells of $\mathcal{F}^3(\mathcal{V}^2)$ are also called vertices, edges, and faces, respectively. In particular, outer faces of $\mathcal{F}^3(\mathcal{V}^2)$ are α^{-1}-images of normal disks of $\mathcal{V}^2$.

Observe that every polyhedron G^3 in $\mathcal{F}^3(\mathcal{V}^2)$ is of one of the following 4 basic types.

(F1) G^3 is a connected component of $\mathcal{F}^3 - \alpha^{-1}(\mathcal{V}^2)$ which contains a single vertex o of $\mathcal{F}^3$ and a single outer face (whose boundary is the link of o on $\partial\mathcal{F}^3$); see Fig. 14(a)–(b).

(F2) G^3 is a connected component of $\mathcal{F}^3 - \alpha^{-1}(\mathcal{V}^2)$ which contains precisely two outer faces which are parallel normal disks in $\mathcal{F}^3$; see Fig. 14(a)–(b).

(F3) G^3 is a truncated tetrahedron or a degenerate truncated tetrahedron, that is, G^3 is obtained from a (degenerate) tetrahedron F^3 in $\mathcal{F}^3$ by cutting off some of vertices of F^3 by means of simple normal disks; see Fig. 14(a).

(F4) G^3 is one of two 'halves' obtained from a truncated tetrahedron T^3_0 (constructed as in (F3) from a tetrahedron $T^3 \in \mathcal{F}^3$) by cutting T^3_0 along a normal quadrangle of T^3; see Fig. 14(b).

A disk d^2 properly embedded in $\mathcal{F}^3(\mathcal{V}^2)$ (and its α-image $\alpha(d^2)$ in $M^3_\Omega - \mathcal{V}^2$) is called *normal* if the natural image of d^2 in $\mathcal{F}^3$ is normal in $\mathcal{F}^3$. Observe that if $G^3 \in \mathcal{F}^3(\mathcal{V}^2)$ has type (F1)–(F2) then there is a unique isotopy type of normal disks in G^3 (normally parallel in $\mathcal{F}^3$ to an outer face of G^3). If $G^3 \in \mathcal{F}^3(\mathcal{V}^2)$ has type (F3) then there are 7 isotopy types of normal disks in G^3 if G^3 is a truncated tetrahedron and there are 3 isotopy types of normal disks in G^3 if G^3 is a truncated degenerate tetrahedron. Finally, if G^3 has type (F4) then there are also 3 isotopy types of normal disks in G^3.

As before, with every isotopy type of normal disks in $\mathcal{F}^3(\mathcal{V}^2)$ we associate an unknown x. Analogously, for every pair p of distinct inner edges of an inner face G^2 of $\mathcal{F}^3(\mathcal{V}^2)$, let $x_1(p)$, $x_2(p)$ denote the unknowns that correspond to normal disks of $\mathcal{F}^3(\mathcal{V}^2)$ whose boundaries contain an arc which connects points on edges in p (as before, $x_1(p)$ can actually be missing). Again, for every such pair p there is a unique 'matching' pair q and for every $\{p, q\}$ we write down

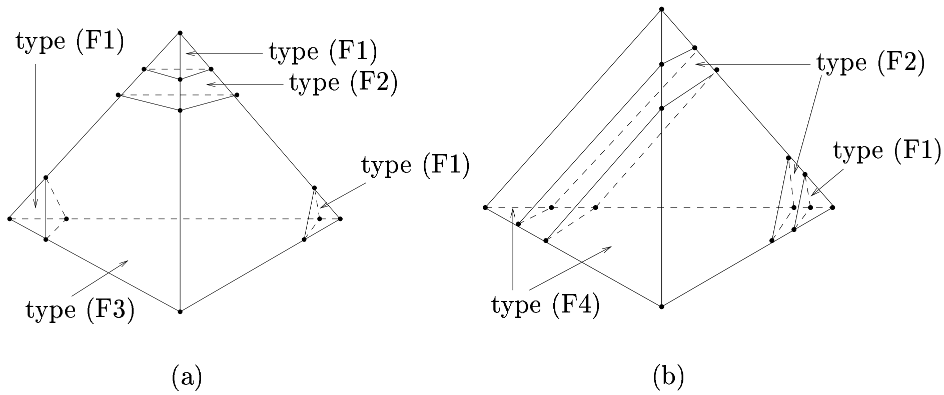


FIGURE 14

a matching equation

$$x_1(p) + x_2(p) = x_1(q) + x_2(q). \tag{2.4}$$

As above, the system of all such equations is called the *system of matching equations* for $M^3_\Omega - \mathcal{V}^2$ and denoted by

$$\mathrm{SME}'(M^3_\Omega - \mathcal{V}^2).$$

Let $G^3 \in \mathcal{F}^3(\mathcal{V}^2)$ have type (F1)–(F2) and let x be the unknown corresponding to the isotopy type of normal disks in G^3. Observe that if x occurs in the left (or right) part of any equation (2.4) then this part is simply x. This observation enables us to delete all occurrences of x in $\mathrm{SME}'(M^3_\Omega - \mathcal{V}^2)$. Repeating such deletions for all unknowns x associated with polyhedra of types (F1)–(F2) in $\mathcal{F}^3(\mathcal{V}^2)$, we obtain a new *reduced* system

$$\mathrm{SME}(M^3_\Omega - \mathcal{V}^2)$$

all of whose equations still look like (2.4).

Note that the number $N(\Omega, \mathcal{V}^2)$ of unknowns in $\mathrm{SME}(M^3_\Omega - \mathcal{V}^2)$ does not exceed $N(\Omega)$. In fact, $N(\Omega, \mathcal{V}^2)$ is less than $N(\Omega)$ by the number of tetrahedra in $\mathcal{F}^3$ that contain quadrangle normal disks of $\alpha^{-1}(\mathcal{V}^2)$ (because each such tetrahedron contributes 7 to the number $N(\Omega)$ of unknowns in $\mathrm{SME}(M^3_\Omega)$ and 6 to $N(\Omega, \mathcal{V}^2)$). Therefore, we can state the following analog of Lemma 2.1.

LEMMA 2.3. *Suppose that ν is a fundamental solution to* $\mathrm{SME}(M^3_\Omega - \mathcal{V}^2)$. *Then every component of ν is at most*

$$N(\Omega, \mathcal{V}^2) \cdot 2^{N(\Omega, \mathcal{V}^2)-1},$$

where $N(\Omega, \mathcal{V}^2)$ is the number of unknowns in $\mathrm{SME}(M^3_\Omega - \mathcal{V}^2)$.

Proof. This is proved in the same way as Lemma 2.1. □

A surface U^2 embedded in $M^3_\Omega - \mathcal{V}^2$ is called *normal* if the intersection

$$\mathcal{F}^3(\mathcal{V}^2) \cap \alpha^{-1}(U^2)$$

consists of finitely many connected components which are normal disks of $\mathcal{F}^3(\mathcal{V}^2)$.

The normal vector $\nu(U^2)$ of a normal surface U^2 in $M^3_\Omega - \mathcal{V}^2$ is defined to be

$$\nu(U^2) = (\nu_1, \dots, \nu_{N(\Omega,\mathcal{V}^2)}), \tag{2.5}$$

where ν_i is the number of normal disks in $\mathcal{F}^3(\mathcal{V}^2) \cap \alpha^{-1}(U^2)$ whose isotopy type corresponds to the ith unknown of the system

$$\mathrm{SME}(M^3_\Omega - \mathcal{V}^2).$$

(As in Lemma 2.3, the unknowns of $\mathrm{SME}(M^3_\Omega - \mathcal{V}^2)$ are assumed to be ordered.)

A normal 2-sphere U^2 in $M^3_\Omega - \mathcal{V}^2$ is called *trivial* if U^2 is normally parallel in M^3_Ω either to a 2-sphere in $\mathcal{V}^2$ or to the link of the 0-cell E^0 of M^3_Ω.

LEMMA 2.4. *Suppose that there is a nontrivial normal 2-sphere in $M^3_\Omega - \mathcal{V}^2$. Then there is also a nontrivial normal 2-sphere in $M^3_\Omega - \mathcal{V}^2$ whose normal vector is $\delta_0\nu$, where ν is a fundamental solution to* $\mathrm{SME}(M^3_\Omega - \mathcal{V}^2)$ *and $\delta_0 \in \{1, 2\}$.*

Proof. Arguing exactly as in the proof of Lemma 2.2 and keeping the notation introduced there, we additionally have to explain why the normal 2-sphere $H^2 = \partial \bar{\mathcal{N}}^3_{M^3_\Omega}(U^2_i)$ (in the case when $\chi(U^2_i) = 1$) is not trivial, that is, why H^2 is not normally parallel to a 2-sphere in $\mathcal{V}^2$. (As before, by considering the parity of components in $\nu(H^2) = 2\nu(U^2_i)$ we see that H^2 is not the link of E^0.)

Observe that the normal 2-sphere H^2 is separating in M^3_Ω. Assuming that H^2 is normally parallel to a 2-sphere V^2 in $\mathcal{V}^2$, we have that V^2, like H^2, is separating in M^3_Ω. Let V^3 be the connected component of $M^3_\Omega - V^2$ that contains H^2 and U^2_i. Since U^2 is connected and disjoint from $\mathcal{V}^2$, it follows that all fundamental surfaces $U^2_1, \dots, U^2_k$ (which are also disjoint from $\mathcal{V}^2$) are in V^3. However, V^3, like $\mathcal{N}^3_{M^3_\Omega}(U^2_i)$, contains only fundamental surfaces which are normally parallel to U^2_i. Hence, $U^2 = U^2_i + U^2_i$ and U^2 is normally parallel to $V^2 \in \mathcal{V}^2$. This contradiction completes the proof of Lemma 2.4. □

A trivial observation, which we will use implicitly in the arguments below, is that there is a natural bijective correspondence between normal surfaces in M^3_Ω disjoint from $\mathcal{V}^2$ and normal surfaces in $M^3_\Omega - \mathcal{V}^2$.

Recall that $\mathcal{V}^2$ consists of pairwise disjoint, pairwise normally nonparallel, nontrivial normal 2-spheres $V^2_1, \dots, V^2_j$, $j \geq 1$. Using Lemmas 2.3–2.4, we can either extend the system $\mathcal{V}^2$ to a new system

$$\mathcal{V}^2 = V_1^2 + \cdots + V_j^2 + V_{j+1}^2$$

or show that this is impossible, that is, $M_\Omega^3 - \mathcal{V}^2$ contains no nontrivial normal 2-spheres. By inequality (2.2), after at most $2r_2 - r_1 + 1$ similar steps we obtain a final system (in the sense that $M_\Omega^3 - \mathcal{V}^2$ contains no nontrivial normal 2-spheres)

$$\mathcal{V}^2 = V_1^2 + \cdots + V_{k_V - 1}^2$$

of pairwise disjoint, pairwise normally nonparallel, nontrivial normal 2-spheres $V_1^2, \ldots, V_{k_V-1}^2$ in M_Ω^3. Adding a trivial normal 2-sphere $V_{k_V}^2$ of M_Ω^3 (which is taken in a regular neighborhood $\mathcal{N}_{M_\Omega^3}(E^0)$ and so is disjoint form $\mathcal{V}^2$) to $\mathcal{V}^2$ yields a maximal collection $\mathcal{V}^2$ of nonparallel normal 2-spheres in M_Ω^3.

LEMMA 2.5. *Suppose that $\mathcal{V}^2$ is a maximal collection of nonparallel normal 2-spheres in M_Ω^3 and K^3 is a connected component of $M_\Omega^3 - \mathcal{V}^2$. Then K^3 contains no nonseparating 2-spheres.*

Proof. Assume, on the contrary, that U^2 is a nonseparating 2-sphere in K^3. Since $\mathcal{V}^2$ consists of normal 2-spheres, it follows that U^2 can be normalized inside K^3. Applying a normalization process to U^2, we get a collection $\mathcal{B}^2$ of 2-spheres each of which either sits in a 3-cell of M_Ω^3 or is normal in $M_\Omega^3 - \mathcal{V}^2$. Since at least one of them is also nonseparating, it follows that there is a nonseparating normal 2-sphere in K^3. However, the maximality of $\mathcal{V}^2$ means that every normal 2-sphere in K^3 is normally parallel to a connected component of ∂K^3 and so is separating. This contradiction proves Lemma 2.5. □

The following lemma is analogous to corresponding results in [R97], [T94].

LEMMA 2.6. *Suppose that $\mathcal{V}^2$ is a maximal collection of nonparallel normal 2-spheres in M_Ω^3 and K^3 is a connected component of $M_\Omega^3 - \mathcal{V}^2$ whose boundary ∂K^3 is not connected. Then K^3 is a punctured 3-ball.*

Proof. Let $K^3(1)$ and $K^3(2)$ denote

$$M_\Omega^3(1) \cap K^3, \quad M_\Omega^3(2) \cap K^3,$$

respectively. Suppose that the set $K^3(1) + \partial K^3$ is not connected and L is a connected component of $K^3(1) + \partial K^3$. Consider a closed regular neighborhood $\bar{\mathcal{N}}_{K^3}^3(L)$ and let

$$L^2 = \partial \bar{\mathcal{N}}_{K^3}(L) - \partial K^3.$$

Then $L^2 \subset \operatorname{Int} K^3$, the orientable surface L^2 intersects $K^3(2)$ in finitely many double curves and L^2 is separating in K^3. Furthermore, both connected components of $K^3 - L^2$ contain some connected components of $K^3(1) + \partial K^3$.

Using the standard innermost argument, we compress L^2 (along disks of $K^3(2)$ bounded by innermost on $K^3(2)$ curves in $K^3(2) \cap L^2$). As a result, we get an orientable surface disjoint from $K^3(2)$ one of whose connected component, say L_0^2, separates connected components of $K^3(1) + \partial K^3$. However, L_0^2 sits in a 3-cell of $M_\Omega^3 - \mathcal{V}^2$ and hence cannot separate connected components of $K^3(1) + \partial K^3$. This contradiction proves that the set $K^3(1) + \partial K^3$ is connected.

Consider a subset U^1 of connected components of $K^3(1)$ such that $U^1 + \partial K^3$ is connected and $|U^1|$ is minimal. Let $\bar{\mathcal{N}}_{K^3}^3(U^1 + \partial K^3)$ be a closed regular neighborhood and let U^2 denote

$$\partial \bar{\mathcal{N}}_{K^3}^3(U^1 + \partial K^3) - \partial K^3.$$

It follows from the minimality of U^1 that U^2 is a 2-sphere. Observe that the question whether K^3 is a punctured 3-ball is equivalent to the question whether the connected component U^3 of $K^3 - U^2$ that contains no ∂K^3 is a 3-ball. Therefore, it suffices to show that U^3 is a 3-ball. To do this, we apply a normalization process to U^2. In view of the normality of 2-spheres in $\mathcal{V}^2$, this process can be carried out in $\operatorname{Int} K^3$. As in the proof of Lemma 1.1, we note that this process does not create new points in $U^2 \cap M_\Omega^3(1)$ (it can only eliminate those). In particular, U^1 will be disjoint from a collection $\mathcal{B}^2$ of 2-spheres obtained at the end of the normalization process. Note that if $B^2 \in \mathcal{B}^2$ is a 2-sphere then either B^2 sits in a 3-cell of M_Ω^3 or B^2 is normal in M_Ω^3. Let B^2 be normal. It follows from the maximality of $\mathcal{V}^2$ that B^2 is normally parallel to a 2-sphere of ∂K^3. Since $U^1 + \partial K^3$ is connected, it follows that B^2 crosses some arcs of U^1. This, however, is impossible since U^1 is disjoint from U^2 and so from B^2. Therefore, all 2-spheres in $\mathcal{B}^2$ sit in 3-cells of M_Ω^3. As in the proof of Lemma 1.2, we consider the following proposition for a finite system $\mathcal{B}_2^2$ of pairwise disjoint 2-spheres in K^3 (which are separating by Lemma 2.5).

(B2) For every $B^2 \in \mathcal{B}_2^2$ the connected component of $K^3 - B^2$ that contains no $U^1 + \partial K^3$ is a 3-ball.

Recall that all 2-spheres obtained in the process of normalization of U^2 are disjoint from $U^1 + \partial K^3$. Hence, as before, we see that the truth value of proposition (B2) does not change during the process. Since proposition (B2) is true at the end of the process, it is true in the beginning, and Lemma 2.6 is proven. □

Let

$$C_1^3, \dots, C_m^3$$

be the connected components of $M_\Omega^3 - \mathcal{V}^2$ that have connected boundary and contain no 0-cell E^0 of M_Ω^3. Consider 3-balls

$$B_1^3, \dots, B_m^3$$

such that every B_i^3 is equipped with a cell decomposition constructed as follows. There is a bijective cellular continuous map

$$\beta_i\colon \partial C_i^3 \to \partial B_i^3,$$

that is, β_i identifies cell decompositions of ∂C_i^3 and ∂B_i^3. Next, there is a single 0-cell E_i^0 in $\operatorname{Int} B_i^3$ so that if

$$B_i^2(t), \quad t \in [0,1],$$

is a singular foliation of B_i^3, where $B_i^2(t)$, $t \in (0,1]$, is a 2-sphere, $B_i^2(1) = \partial B_i^3$ and $B_i^2(0) = E_i^0$, then every j-cell, $j = 1,2$, of B_i^3 which is not in ∂B_i^3 has the form $A^{j-1} \times (0,1)$, where A^{j-1} is a $(j-1)$-cell in ∂B_i^3.

Let us construct a closed 3-manifold

$$M_i^3$$

by attaching the 3-ball B_i^3 to C_i^3 by means of the homeomorphism β_i, $i = 1,\dots,m$. A cell decomposition for M_i^3 is obtained from those of C_i^3, B_i^3 by 'forgetting' about cells of $\partial B_i^3 = \partial C_i^3$; that is, we merge j-cells E_1^j, E_2^j, $j = 1,2,3$, of B_i^3 and C_i^3, respectively, provided that they are separated by a $(j-1)$-cell E_0^j of $\partial B_i^3 = \partial C_i^3$ by adding E_0^{j-1} to $E_1^j + E_2^j$. As a result, we get a closed 3-manifold

$$M^3_{i,\widetilde{\Omega}_i}$$

whose cell decomposition $\widetilde{\Omega}_i$ contains a single 0-cell E_i^0, $i = 1,\dots,m$. By

$$\widetilde{\alpha}_i\colon \widetilde{\mathcal{F}}_i^3 \to M^3_{i,\widetilde{\Omega}_i}$$

denote a cellular map, where $\widetilde{\mathcal{F}}_i^3 = \widetilde{\mathcal{F}}_i^3(\widetilde{\Omega}_i)$ is a suitable collection of nonsingular polyhedra. Note that every polyhedron G^3 in $\widetilde{\mathcal{F}}_i^3$ consists of several parts $G^3 - \widetilde{\alpha}_i^{-1}(\partial C_i^3)$ exactly one of which, denoted by G_C^3 and called the *central part* of G^3, has the property that

$$\widetilde{\alpha}_i(G_C^3) \subset C_i^3.$$

Clearly, G^3 is uniquely determined by its central part G_C^3 which has one of the types (F1)–(F4).

Specifically, if G_C^3 has type (F1)–(F2) then G^3 has 2 vertices and several (3 or 4) biangle faces. If G_C^3 has type (F3) then G^3 is a (degenerate) tetrahedron. Finally, if G_C^3 has type (F4) then G^3 is a degenerate tetrahedron.

Note that the 2-sphere ∂C_i^3 is normally parallel to the link of the 0-cell E_i^0 of $M^3_{i,\widetilde{\Omega}_i}$ and that any normal surface in $M^3_{i,\widetilde{\Omega}_i}$ has a natural image in C_i^3. Hence, it follows from the maximality of $\mathcal{V}^2$ that any normal 2-sphere in $M^3_{i,\widetilde{\Omega}_i}$ is normally parallel to the link of E_i^0.

Now we can eliminate those polyhedra in $\widetilde{\mathcal{F}}_i^3$ that have only two vertices by using same reductions as those described in the beginning of this section. It is not difficult to see that applying an elementary reduction (either attaching a polyhedron G^3 with two vertices to another one along a face of G^3 or closing out a face of G^3) does not change the set of normal surfaces in $M^3_{i,\widetilde{\Omega}_i}$ (in the sense that there is a natural bijection between normal surfaces with respect to cell decompositions of M^3_i before and after an elementary reduction). As before, after making all possible reductions to delete as many polyhedra in $\widetilde{\mathcal{F}}^3_i$ with two vertices as possible, we obtain a Q-triangulation Ω_i for

$$M^3_i = M^3_{i,\Omega_i}.$$

By the above remark, there is a natural bijective correspondence between normal surfaces in M^3_{i,Ω_i} and in $C^3_i \subset M^3_\Omega - \mathcal{V}^2$ (which can also be established by using normal vectors of normal surfaces in $C^3_i \subset M^3_\Omega - \mathcal{V}^2$, as defined by equality (2.5), and normal vectors of normal surfaces in M^3_{i,Ω_i}). Therefore, by the maximality of $\mathcal{V}^2$, M^3_{i,Ω_i} contains no nontrivial normal 2-spheres, and hence Ω_i is an irreducible Q-triangulation for M^3_{i,Ω_i}. It follows from the definitions and Lemma 2.6 that M^3_Ω is the connected sum of 3-manifolds

$$M^3_{1,\Omega_1}, \dots, M^3_{m,\Omega_m}$$

and finitely many 2-sphere bundles over a circle (their number is $|\mathcal{V}^2_{NS}|$; see (2.1)).

It remains to observe that each tetrahedron in $\mathcal{F}^3$ contributes a tetrahedron to

$$\mathcal{F}^3_1 + \cdots + \mathcal{F}^3_m$$

if the intersection $T^3 \cap \alpha^{-1}(\mathcal{V}^2)$ contains no normal quadrangles and, otherwise, T^3 contributes two degenerate tetrahedra to $\mathcal{F}^3_1 + \cdots + \mathcal{F}^3_m$. Similarly, each degenerate tetrahedron D^3 of $\mathcal{F}^3$ contributes one degenerate tetrahedron to $\mathcal{F}^3_1 + \cdots + \mathcal{F}^3_m$. Therefore, the difference

$$N(\Omega) - (N(\Omega_1) + \cdots + N(\Omega_m))$$

is nonnegative and equal to the number of tetrahedra in $\mathcal{F}^3$ which contain a normal quadrangle of $\alpha^{-1}(\mathcal{V}^2) \subset \mathcal{F}^3$. This completes the proof of Theorem 2.

3. Proofs of Theorems 3 and 4

We first prove Theorem 4. Keeping the notation of Sections 0–2, suppose that M^3_Ω is a 3-manifold given by a Q-triangulation Ω and

$$\nu = (\nu_1, \dots, \nu_{N(\Omega)})$$

is a natural solution to the system SME(M^3_Ω) of matching equations for M^3_Ω. Consider the following properties of ν.

(G1) ν satisfies the admissibility condition for every tetrahedron T^3 in $\mathcal{F}^3$, that is, if ν_{i_1}, ν_{i_2}, ν_{i_3} are the components of ν which correspond to three isotopy types of normal quadrangles of T^3 then at most one of ν_{i_1}, ν_{i_2}, ν_{i_3} is positive.

(G2) There is at least one negative entry in $\nu - \nu(\operatorname{Link} E^0)$.
(G3) The Euler characteristic $\chi(V^2)$ of a normal surface V^2 whose normal vector $\nu(V^2)$ is ν is positive.

LEMMA 3.1. *Let a 3-manifold M^3_Ω be given by a Q-triangulation Ω. Then Ω is reducible if and only if there is a natural solution ν to* $\mathrm{SME}(M^3_\Omega)$ *with properties* (G1)–(G3) *all of whose components are at most $N(\Omega) \cdot 2^{N(\Omega)}$.*

Proof. First suppose that Ω is reducible, that is, M^3_Ω contains a nontrivial normal 2-sphere. Then it follows from Lemmas 2.1–2.2 that M^3_Ω also contains a nontrivial normal 2-sphere U^2 whose normal vector $\nu(U^2)$ satisfies properties (G1)–(G3) and has components at most $N(\Omega) \cdot 2^{N(\Omega)}$.

Conversely, let ν be a natural solution to $\mathrm{SME}(M^3_\Omega)$ with properties (G1)–(G3) and let V^2 be a normal surface on M^3_Ω whose normal vector $\nu(V^2)$ is ν. Without loss of generality, we can assume that V^2 is connected (otherwise, we could pick a connected component of V^2 whose Euler characteristic is positive). Since $\chi(V^2) > 0$, V^2 is a 2-sphere or a projective plane. In the first case, V^2 is a nontrivial normal 2-sphere by property (G2). In the second case, $\partial(\bar{\mathcal{N}}^3_{M^3_\Omega}(V^2))$ is a nontrivial normal 2-sphere because all components in

$$\nu(\partial(\bar{\mathcal{N}}^3_{M^3_\Omega}(V^2)) = 2\nu(V^2)$$

are even. This proves Lemma 3.1. □

To prove Theorem 4, we note that, given an $N(\Omega)$-tuple ν of nonnegative integers which do not exceed $N(\Omega)\cdot 2^{N(\Omega)}$ (and written in, say, decimal form so that their size is linear in $N(\Omega)$), we can check whether ν is a natural solution to $\mathrm{SME}(M^3_\Omega)$ and verify properties (G1)–(G2) in a time that is polynomial in $N(\Omega)$. Property (G3) can also be verified in polynomial time because $\chi(V^2)$ is a linear combination of

$$\nu_1, \dots, \nu_{N(\Omega)}$$

which can be written out explicitly (see [JT95], [HLP99]) and computed in polynomial time. Hence, by Lemma 3.1, the property of being reducible for a Q-triangulation Ω is in **NP**. This completes the proof of Theorem 4 . □

Let us turn to Theorem 3.

LEMMA 3.2. *Let G^3 be a (degenerate) tetrahedron and a^2 be an $A1$-normal disk in G^3. Then G^3 is a tetrahedron, a^2 is an octagon and ∂a^2 on ∂G^3 is as depicted in Fig. 15.*

Proof. Let the edges of G^3 that are crossed twice by the curve $a^1 = \partial a^2$ be split into two nonempty groups $e_1, \dots, e_{k_e}$ and $f_1, \dots, f_{k_f}$ (as in the definition of an $A1$-normal disk in a polyhedron G^3; see Fig. 1) so that the vertices ∂e_i

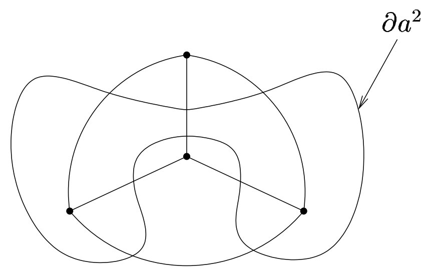


FIGURE 15

lie in distinct connected components of $a^1 - f_j$ for all i, j. This implies that the vertices

$$\partial(e_1 + \cdots + e_{k_e})$$

lie in one connected component of $\partial G^3 - a^1$ while the vertices

$$\partial(f_1 + \cdots + f_{k_f})$$

lie in the other connected component of $\partial G^3 - a^1$. In particular, the intersection

$$\partial(e_1 + \cdots + e_{k_e}) \cap \partial(f_1 + \cdots + f_{k_f})$$

is empty. Observe that if G^3 is a degenerate tetrahedron then any two edges of G^3 have a common vertex. Therefore, G^3 must be a tetrahedron in which case there are 3 potential pairs of disjoint edges to be crossed twice by a^1. Take such a pair e, f and denote by e_1^2, e_2^2 the triangles of G^3 that contain e in their boundary, and by f_1^2, f_2^2 the triangles of G^3 that contain f in their boundary; see Fig. 16.

Then the edges e, f are crossed by a^1 exactly twice and any other edge of G^3 is crossed by a^1 at most once. This means that the intersections

$$e_1^2 \cap a^1, \quad e_2^2 \cap a^1$$

consist of two arcs which connect two points $c^1 \cap e$ in e to two other sides of e_1^2 and e_2^2, respectively; see Fig. 16. Arguing similarly with the intersections

$$f_1^2 \cap a^1, \quad f_2^2 \cap a^1,$$

we have that a^1 crosses every edge of G^3 different from e, f exactly once and a^2 is an octagon as required. □

Let Ω be a Q-triangulation of a 3-manifold M^3_Ω. Consider the following property of Ω.

(A4) If T^3 is a tetrahedron in $\mathcal{F}^3 = \mathcal{F}^3(\Omega)$ then the only singularities of $\alpha(T^3)$ are vertices of T^3 (whose α-images are E^0; that is, the restriction $\alpha|_{T^3 \setminus T^3(0)}$ of α on T^3, except for the set $T^3(0)$ of vertices of T^3, is an embedding).

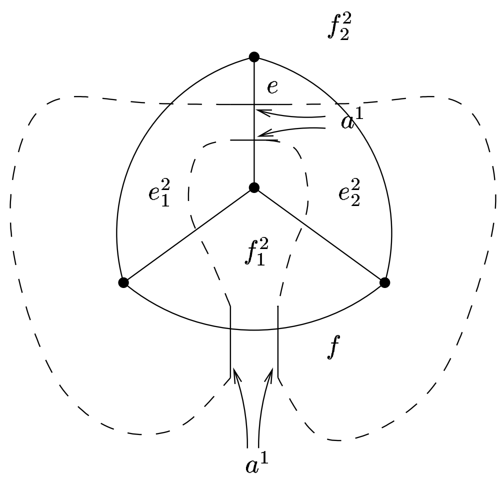


FIGURE 16

When proving Theorem 3, it will be convenient to assume that an irreducible Q-triangulation of M^3_Ω has property (A4). This is not restrictive because we can split every tetrahedron $T^3 \in \mathcal{F}^3$ into five polyhedra T^3_0, $T^3_1, \ldots, T^3_4$ such that their vertices are those of T^3, T^3_0 is a tetrahedron, $T^3_1, \ldots, T^3_4$ are degenerate tetrahedra, each T^3_i, $i = 1, \ldots, 4$, has three biangle faces and two triangle faces of T^3_i are those of T^3 and T^3_0. Therefore, making a suitable subdivision of Ω, without loss of generality, we can suppose that Ω has property (A4).

Consider the following property of a natural solution ν to the system $\mathrm{SME}(M^3_\Omega)$.

(H1) The admissibility condition is satisfied for all tetrahedra in $\mathcal{F}^3$, except for one, denoted by G^3_{A1}, and if ν_{i_1}, ν_{i_2}, ν_{i_3} are the components in ν which correspond to the isotopy types of normal quadrangles in G^3_{A1} then two of ν_{i_1}, ν_{i_2}, ν_{i_3} are equal to 1 and the other one is zero.

Just like natural solutions to $\mathrm{SME}(M^3_\Omega)$ satisfying the admissibility condition (G2) can be used to uniquely describe normal surfaces in M^3_Ω, natural solutions to $\mathrm{SME}(M^3_\Omega)$ with property (H1) can be used to uniquely describe $A1$-normal surfaces in M^3_Ω as follows.

Let U^2 be an $A1$-normal surface in M^3_Ω. By Lemma 3.2, there is a tetrahedron G^3_{A1} in $\mathcal{F}^3$ which contains an $A1$-normal octagon a^2. Let $\{e_j, f_j\}$, $j = 1, 2, 3$, be the 3 pairs of disjoint edges of G^3_{A1} and suppose that ∂a^2 crosses edges e_1, f_1 twice. Also, denote by q^2_j, $j = 1, 2, 3$, a normal quadrangle of G^3_{A1} that is disjoint from the pair $\{e_j, f_j\}$ and let q^2_1, q^2_2, and q^2_3 correspond to unknowns x_{i_1}, x_{i_2}, and x_{i_3}, respectively, of the system $\mathrm{SME}(M^3_\Omega)$. Then the *normal vector* $\nu(U^2)$ of the $A1$-normal surface U^2 is defined to be

$$\nu(U^2) = (\nu_1, \dots, \nu_{N(\Omega)}),$$

where ν_i is the number of normal disks in $\alpha^{-1}(U^2)$ that have the ith isotopy type (corresponding to the unknown x_i of $\mathrm{SME}(M^3_\Omega)$), provided that $i \notin \{i_1, i_2, i_3\}$. Otherwise, we set

$$\nu_{i_1} = 0, \quad \nu_{i_2} = \nu_{i_3} = 1.$$

Observe that the isotopy types of 8 normal arcs in $\partial(q_2^2 + q_3^2)$ are exactly the same as those in ∂a^2. (Curiously, the curve ∂a^2 can be obtained from curves ∂q_2^2, ∂q_3^2 by some surgery which is reminiscent of regular exchange surgery; see Fig. 17.) Therefore, the normal vector $\nu(U^2)$ is a natural solution to $\mathrm{SME}(M^3_\Omega)$ which, obviously, has property (H1).

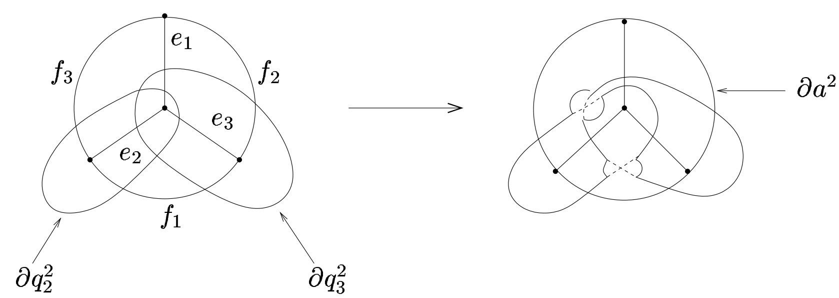


FIGURE 17

The converse is given by the following result:

LEMMA 3.3. *Suppose that ν is a natural solution to* $\mathrm{SME}(M^3_\Omega)$ *with property* (H1). *Then there is an $A1$-normal surface U^2 with $\nu(U^2) = \nu$. Furthermore, U^2 is uniquely defined by ν, that is, if $\widetilde{U}^2$ is another $A1$-normal surface with $\nu(\widetilde{U}^2) = \nu$ then $\widetilde{U}^2$ is normally parallel to U^2.*

Proof. In view of the above observation (see Fig. 17), the existence and uniqueness of an $A1$-normal surface U^2 with $\nu(U^2) = \nu$ can be proved in the same way as the existence and uniqueness of a normal surface V^2 with $\nu(V^2) = \nu'$, where ν' is a natural solution to $\mathrm{SME}(M^3_\Omega)$ satisfying the admissibility condition. □

We introduce two further properties of a natural solution ν to $\mathrm{SME}(M^3_\Omega)$ with property (H1).

(H2) There is at least one negative entry in $\nu - \nu(\mathrm{Link}\, E^0)$.

(H3) The Euler characteristic of an $A1$-normal surface V^2 whose normal vector $\nu(V^2)$ is ν equals 2.

For example, note that if V^2 is an $A1$-normal 2-sphere in M^3_Ω then its normal vector has properties (H1)–(H3).

LEMMA 3.4. *Let a 3-manifold M^3_Ω be given by an irreducible Q-triangulation Ω with property* (A4). *Then M^3_Ω is the 3-sphere if and only if there is a natural solution to* $\mathrm{SME}(M^3_\Omega)$ *with properties* (H1)–(H3) *all of whose components are at most $N(\Omega)\cdot 2^{N(\Omega)}$.*

Proof. Suppose that M^3_Ω is the 3-sphere. By Theorem 1, M^3_Ω contains an $A1$-normal 2-sphere V^2. If the normal vector $\nu(V^2)$ is fundamental or is the sum of two fundamental solutions to $\mathrm{SME}(M^3_\Omega)$ then, by Lemma 2.1, every component in $\nu(V^2)$ is at most $N(\Omega)\cdot 2^{N(\Omega)}$ and so $\nu(V^2)$ is a solution of the required form. Otherwise, by Lemma 3.2, we can write $\nu(V^2)$ in the form

$$\nu(V^2) = \nu^1 + \nu^2, \tag{3.1}$$

where ν^1, ν^2 are natural solutions to $\mathrm{SME}(M^3_\Omega)$, ν^1 has property (H1) and ν^2 satisfies the admissibility condition. Choose a representation (3.1) so that the sum of components in ν^1 is minimal. Let V_1^2 and V_2^2 be $A1$-normal and normal surfaces with

$$\nu(V_1^2) = \nu^1, \quad \nu(V_2^2) = \nu^2.$$

It follows from the choice of ν^1 that ν^1 is either a fundamental solution or a sum of two fundamental solutions and that the surface V_1^2 is connected.

By property (A4) we can suppose that double arcs of the intersection $V_1^2 \cap V_2^2$ are disjoint from the α-image of any tetrahedron of $\mathcal{F}^3$. In particular, in view of Lemma 3.2, the $A1$-normal disk of V_1^2 is disjoint from V_2^2. Then, exactly as in the case of normal surfaces, one can show that the surface V^2 is a Haken sum $V_1^2 + V_2^2$ of V_1^2 and V_2^2. In particular,

$$2 = \chi(V^2) = \chi(V_1^2) + \chi(V_2^2) \tag{3.2}$$

and so one of $\chi(V_1^2)$, $\chi(V_2^2)$ is positive. If $\chi(V_2^2) > 0$ then we can argue as in the proof of Lemma 3.1 to obtain the existence of a nontrivial normal 2-sphere in M^3_Ω. This contradiction shows that $\chi(V_2^2) \le 0$. Hence, it follows from the connectedness of V_1^2 and equation (3.2) that $\chi(V_1^2) = 2$; that is, ν^1 is a solution as required.

Conversely, let ν be a natural solution to $\mathrm{SME}(M^3_\Omega)$ with properties (H1)–(H3) and let V^2 be an $A1$-normal surface with $\nu(V^2) = \nu$. Let

$$V_1^2, \dots, V_k^2$$

be connected components of V^2. Then one of them, say V_1^2, is $A1$-normal and the others, $V_2^2, \dots, V_k^2$, are normal. Moreover,

$$\nu(V^2) = \nu(V_1^2) + \nu(V_2^2) + \cdots + \nu(V_k^2)$$

and

$$2 = \chi(V^2) = \chi(V_1^2) + \chi(V_2^2) + \cdots + \chi(V_k^2).$$

Suppose that there is an $i \geq 2$ such that $\chi(V_i^2) > 0$. If $\chi(V_i^2) = 2$ then V_i^2 is a normal 2-sphere which is nontrivial by property (H2). If $\chi(V_i^2) = 1$ then $\partial \bar{\mathcal{N}}_{M_\Omega^3}(V_i^2)$ is a normal 2-sphere which is nontrivial because, by property (H1), there is a tetrahedron in $\mathcal{F}^3$ and all components in

$$\nu(\partial \bar{\mathcal{N}}_{M_\Omega^3(2)}(V_i^2)) = 2\nu(V_i^2)$$

are even. These contradictions prove that $\chi(V_i^2) \leq 0$ for all $i \geq 2$. Therefore, $\chi(V_1^2) = 2$ whence V_1^2 is an $A1$-normal 2-sphere. An application of Theorem 1 completes the proof of Lemma 3.4. □

In view of Lemma 3.4, the proof of Theorem 3 is completely analogous to the proof of Theorem 4. As in the latter proof, given an $N(\Omega)$-tuple

$$\nu = (\nu_1, \ldots, \nu_{N(\Omega)})$$

of nonnegative integers not greater than $N(\Omega) \cdot 2^{N(\Omega)}$, we can verify in polynomial time whether ν is a solution to $\mathrm{SME}(M_\Omega^3)$ and whether properties (H1)–(H3) hold. If this is the case, Lemma 3.4 enables us to conclude that M_Ω^3 is indeed the 3-sphere. This means that the recognition problem for 3-sphere in the class of irreducible Q-triangulations is in **NP**, and the proof of Theorem 3 is complete.

References

[B59] R.H. Bing, *An alternative proof that 3-manifolds can be triangulated*, Ann. of Math. **69** (1959), 37–65.

[BH73] J. Birman and H. Hilden, *The homeomorphism problem for S^3*, Bull. Amer. Math. Soc. **79** (1973), 1006–1009.

[G87] D. Gabai, *Foliations and the topology of 3-manifolds, III*, J. Diff. Geometry **26** (1987), 479–536.

[Ha61] W. Haken, *Theorie der Normalflächen*, Acta Math. **105** (1961), 245–375.

[Ha68] ———, *Some results on surfaces in 3-manifolds*, Studies in Modern Topology, MAA Studies, vol. 5, Math. Assoc. Amer. (distributed by Prentice Hall, Englewood Cliffs, NJ), 1968, pp. 39–98.

[Ha87] ———, *Recognizing the 3-dimensional cell and the 3-sphere I*, manuscript, 1987.

[HOT80] T. Homma, M. Ochiai and M. Takahashi, *An algorithm for recognizing S^3 in 3-manifolds with Heegaard splittings of genus two*, Osaka J. Math. **17** (1980), 625–648.

[HLP99] J. Hass, J.C. Lagarias and N. Pippenger, *The computational complexity of knot and link problems*, J. ACM **46** (1999), 185–211.

[He76] J. Hempel, *3-Manifolds*, Ann. Math. Studies, vol. 86, Princeton Univ. Press, Princeton, 1976.

[Hn92] G. Hemion, *The classification of knots and 3-dimensional spaces*, Oxford Univ. Press, Oxford, 1992.

[JO84] W.H. Jaco and U. Ortel, *An algorithm to decide if a 3-manifold is a Haken manifold*, Topology **23** (1984), 195–209.

[JR89] W.H. Jaco and J.H. Rubinstein, *PL equivariant surgery and invariant decompositions of 3-manifolds*, Adv. in Math. **73** (1989), 149–191.

[JT95] W.H. Jaco and J.L. Tollefson, *Algorithms for the complete decomposition of a closed 3-manifold*, Illinois J. Math. **39** (1995), 358–406.

[K29] H. Kneser, *Geschlossene Flächen in dreidimensionalen Mannigfaltigkeiten*, Jber. D.M.V. **67** (1929), 248–260.

[Ma95] S.V. Matveev, *Algorithms for the recognition of three-dimensional sphere (after A. Thompson)*, Mat. Sbornik **186** (1995), 69–84.

[Ma99] ———, *Algorithmic classification of three-dimensional manifolds: problems and results*, Proc. Steklov Inst. Math. **225** (1999), 250–260.

[Mo52] E.E. Moise, *Affine structures in 3-manifolds. V. The triangulation theorem and Hauptvermutung*, Ann. of Math. **56** (1952), 96–114.

[R92] J.H. Rubinstein, *The solution to the recognition problem for S^3*, Lectures, Haifa, Israel, May 1992.

[R95] ———, *An algorithm to recognize the 3-sphere*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zurich, 1994), Birkhauser, Basel, 1995, pp. 601–611.

[R97] ———, *Polyhedral minimal surfaces, Heegaard splittings and decision problems for 3-dimensional manifolds*, Geometric topology (Athens, GA, 1993), AMS/IP Stud. Adv. Math., 2.1, Amer. Math. Soc., Providence, RI, 1997, pp. 1–20.

[S00] M. Stocking, *Almost normal surfaces in 3-manifolds*, Trans. Amer. Math. Soc. **352** (2000), 171–207.

[T94] A. Thompson, *Thin position and the recognition problem for S^3*, Math. Research Letters **1** (1994), 613–630.

Department of Mathematics, University of Illinois, Urbana, IL 61801, USA
E-mail address: ivanov@math.uiuc.edu