Algorithms to disprove the Poincaré Conjecture

Colin Rourke

The status of the Poincaré Conjecture has recently changed. It is now possible to describe an effective algorithm for finding a counterexample to the Poincaré Conjecture. This is a consequence of two algorithms :

(1) The Rêgo–Rourke algorithm which lists all Heegaard diagrams of homotopy 3–spheres [2].

(2) The Rubenstein–Thompson algorithm [3,4] which will decide if a given triangulated 3–manifold is homeomorphic to S^3.

Now a Heegaard diagram can be converted into a triangulation in several algorithmic ways and combining all three algorithms we have the following result :

Main Theorem *There is an algorithm with no input data which lists all non-trivial homotopy 3–spheres. Thus, if the Poincaré Conjecture is false, the algorithm will produce a proven counter-example in finite time. (If it is true, it will continue for ever with no result.)*

In this paper I shall briefly sketch the two main algorithms (of Rêgo–Rourke and Rubenstein–Thompson). The existence of an algorithm to convert a Heegaard diagram into a triangulation will be left as an exercise for the reader and I shall finish the paper with some comments on the practicality of the combined algorithm.

1. The Rêgo–Rourke algorithm

An H–**diagram** for an oriented 3–manifold is an oriented surface F of genus t together with two complete systems $\mathbf{x} = (x_1, x_2, \ldots, x_t)$, $\mathbf{y} = (y_1, y_2, \ldots, y_t)$ of simple closed curves on F. (Here a **complete system** means a set of t disjoint simple closed curves on F whose union does not separate F.)

Now let H_t donote the standard solid handlebody of genus t with standard complete systems $\mathbf{a}$, $\mathbf{b}$ as illustrated in figure 1.

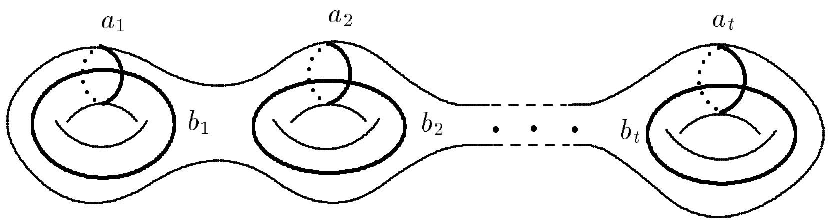

Figure 1

Given a complete system $\mathbf{x}$ on F, we can identify F with ∂H_t so that x_i is identified with a_i for each i. Let the resulting homeomorphism be denoted $h_{\mathbf{x}}$. Then the 3–manifold defined by the H–diagram $(F, \mathbf{x}, \mathbf{y})$ is defined to be

$$H_t \cup_{h_{\mathbf{x}} h_{\mathbf{y}}^{-1}} H_t$$

and will be denoted $M(\mathbf{x}, \mathbf{y})$. It is well known that any closed oriented 3–manifold can be expressed in this way.

By a **system** (not necessarily complete) of curves on F we mean a set of disjoint simple closed curves. Let $\mathbf{x} = (x_1, x_2, \ldots, x_u)$, $\mathbf{y} = (y_1, y_2, \ldots, y_s)$ be systems and suppose F and each x_i, y_i are oriented, then we can define unreduced (cyclic) words $w(x_i, \mathbf{y})$ for $i = 1, 2, \ldots, u$ in the symbols $\{y_1, y_2, \ldots, y_s\}$ by choosing a point on x_i and reading round x_i in order, noting crossings with the curves $\mathbf{y}$. Write y_j for a positive crossing with y_j and y_j^{-1} for a negative crossing.

Now consider the standard handlebody H_{s+t} of genus $s+t$ and suppose that we have a complete system

$$\{x_1, x_2, \ldots, x_t, y_1, y_2, \ldots, y_s\}$$

on ∂H_{s+t} such that

(1) $w(y_i, \mathbf{a}) = e$, $i = 1, 2, \ldots, s$,

(2) $w(b_i, \mathbf{x}) = e$, $i = 1, 2, \ldots, t+s$.

Then we call the system $\{\mathbf{x}, \mathbf{y}\}$ a P–**system** on ∂H_{s+t}.

Conditions (1) and (2) are to be understood in the appropriate free group. In other words (1) says that the word in the a_j determined by y_i cancels to the empty word, while (2) says that the word in the x_j determined by b_i cancels to the empty word.

Now condition (1) implies that the curves $\{y_1, y_2, \ldots, y_s\}$ are nulhomotopic in H_{s+t} and hence bound disjoint discs $\{D_1, D_2, \ldots, D_s\}$ in H. We can cut H along the discs $\{D_1, D_2, \ldots, D_s\}$ to obtain a solid handlebody T of genus t embedded (not necessarily standardly) in S^3. Now if we extend $\{y_1, y_2, \ldots, y_s\}$ to a complete system $\{z_1, z_2, \ldots, z_t, y_1, y_2, \ldots, y_s\}$ for ∂H such that the $\{z_i\}$ also bound discs in H, then we will have two complete systems $\mathbf{x}$ and $\mathbf{z}$ for ∂T, in other words $\{\partial T, \mathbf{x}, \mathbf{z}\}$ is an H–diagram.

Theorem 1 [2] *The 3–manifold $M(\mathbf{x}, \mathbf{z})$ determined by the H–diagram $\{\partial T, \mathbf{x}, \mathbf{z}\}$ is a homotopy 3–sphere and all H–diagrams of homotopy 3–spheres arise in this way.*

Below I shall sketch the proof of theorem 1, but first some comments about the resulting algorithm. Conditions (1) and (2) are easily checked for any given complete system $\{\mathbf{x}, \mathbf{y}\} = \{x_1, x_2, \ldots, x_t, y_1, y_2, \ldots, y_s\}$. Thus there is clearly an algorithm to list all P–systems. Moreover there is also an algorothm to list all extensions of $\{y_1, y_2, \ldots, y_s\}$ to a complete system $\{z_1, z_2, \ldots, z_t, y_1, y_2, \ldots, y_s\}$ for ∂H such that the $\{z_i\}$ also bound discs in H, since the relevant condition (namely condition (1) with z_i in place of y_i)

is readily checked. However, whichever extension is chosen, the resulting 3–manifold is always the same. Thus to list all homotopy 3–spheres (rather than all H–diagrams of homotopy 3–sphers) it suffices to find one such extension. This can be done by algorithmically deforming $(y_1, y_2, \ldots, y_s)$ to a subset of the a_i and then reversing the process on the other subset (see [2; lemmas 1 and 2] here). Thus theorem 1 implies the following :

Corollary *There is an algorithm whigh will list all H–diagrams of all homotopy 3–spheres. There is a more efficient algorithm which will list at least one H–diagram for each homotopy 3–sphere.*

Sketch of proof of theorem 1 (for full details see [2])

Let $(\mathbf{x}, \mathbf{y})$ be a P–system. Note that condition (2) implies that the curves $\mathbf{x} = (x_1, x_2, \ldots, x_t)$ bound disjoint surfaces in $\overline{S^3 - T}$. We can then define a degree 1 map of S^3 onto M by collapsing each each of these surfaces to a disc and collapsing the complementary region to a 3–ball. Thus M is a homotopy 3–sphere.

Conversely let $M = K \cup_\partial K'$ be a homotopy 3–sphere with given Heegaard decomposition. Think of this as a handle decomposition of M with K being the 0–handle and the 1–handles and K' being the 2–handles and the 3–handle. Since M is a homotopy 3–sphere there is a degree 1 map $f : S^3 \to M$. Make f transverse to the cores of the 1–handles and the cores of the 2–handles (properly embedded 2–discs D_i in K'). By a trading argument it can be assumed that $f|f^{-1}(K) \to K$ is a homeomorphism. Let $T = f^{-1}(K)$. Then T can be unknotted and identified with the standard handlebody H by adding 1–handles to T in S^3. Moreover, with care, these 1–handles can be assumed to be disjoint from the surfaces $f^{-1}(D_i)$. Now let $\mathbf{x} = (x_1, x_2, \ldots, x_t)$ be the attaching curves for the D_i and let $\mathbf{y} = (y_1, y_2, \ldots, y_s)$ be the belt spheres for the new 1–handles, then $(\mathbf{x}, \mathbf{y})$ is the required P–system. □

2. The Rubenstein–Thompsom algorithm

Some time ago, Wolfgang Haken announced that he had obtained an algorithm to decide if a given 3–manifold is S^3, however details of Haken's algorithm have never appeared. Recently Hyam Rubenstein gave a series of lectures containing such an algorithm, based on PL minimal surface theory. Subsequently Abigail Thompson has written a complete proof of a similar algorithm using "thin position" in place of minimal surface therory. In this sketch, I shall follow Thompson's treatment [4], except that I shall present alternative (shorter) proofs of lemma 2 and the first half of lemma 3 of [4] using the process of "normalisation". I believe these proofs are close to Rubenstein's treatment [3], though I have not seen details of this. The hardest part of the proof is the second half of lemma 3, and here I shall sketch Thompson's proof and refer to [4] for full details.

Let M be a closed 3–manifold with triangulation T. A **normal surface** in M is an embedded surface which meets each tetrahedron R of T in a collection in a collection of triangles and/or quadrilaterals (not necessarily planar) like figure 2.

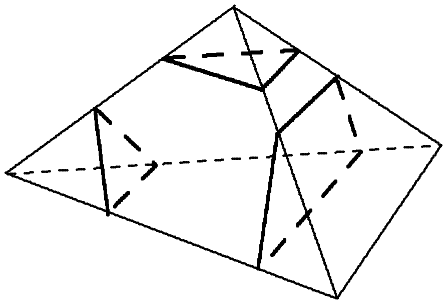

Figure 2

An **almost normal** (AN) surface is the same as a normal surface except that it has exactly one more complicated disc of intersection with one tetrahedron. This is a **saddle-disc** with boundary a curve of length eight, illustrated in figure 3.

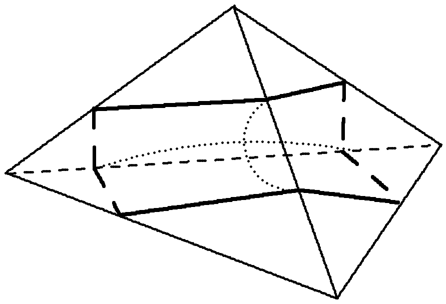

Figure 3

In figure 3, the saddle-disc is indicated by dotted curves. Note that the AN surface may well intersect the tetrahedron again in triangles near the vertices.

There is an obvious combinatorial notion of parallel for normal surfaces : two normal surfaces are **parallel** is they consist of parallel pairs of triangles and quadrilaterals in each tetrahedron which they meet.

Now let Σ be a maximal collection of disjoint non-parallel normal 2–spheres in T. Consider the components of $M - \Sigma$. There are three types :

Type I *Pineapples* Small 3–balls containing exactly one vertex and bounded by triangles near the vertex in incident tetrahedra.

Type II More than one boundary component.

Type III One boundary component but not of type I.

Lemma 2 *A component of type II is a punctured 3–ball.*

Lemma 3 *A component of type III is a 3–ball if and only if it contains an AN 2–sphere.*

The algorithm can now be described. Find a maximal collection of disjoint non-parallel normal 2–spheres in T (the algorithms of Haken *et al* can readily be adapted to find such a collection).

Check that each 2–sphere in the collection is separating.

Check that each component of type III contains an AN 2–sphere. (Again the standard algorithms can be adapted to do this, for details see [4; section 4].)

Then M is S^3 if and only if it passes both these tests.

Normalisation

Let T be a triangulated 3–manifold and let $F \subset T$ be an embedded surface transverse to T. Normalisation is a process which replaces T by a normal surface in the same homology class which has simpler intersections with the 1–skeleton of T (indeed with every 1–simplex).

Step 1 Suppose that F meets a 2–simplex q of T in a closed curve (and possibly other curves). Choose an innermost such curve and cut F along this curve gluing in the obvious discs (parallel to q) to close F up again. This removes one closed curve of intersection of F with the 2–skeleton of T. Repeat until all such curves have been eliminated.

Step 2 Suppose that F meets a 2–simplex q of T in a "D–curve" curve (and possibly other curves). By a "D–curve" I mean an arc with both end-points on the same edge of q (figure 4).

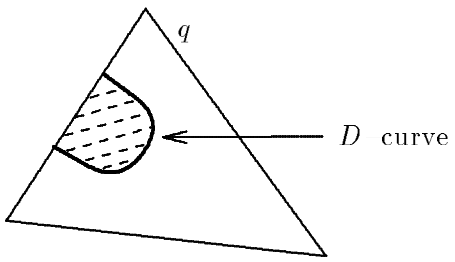

Figure 4

Choose an outermost such curve and isotope F across the D–disc (shaded in figure 4) formed by the D–curve and the edge of q. This removes two points of intersection of F with this edge. Now repeat step 1 if necessary and then repeat the process to remove all D–curves.

Step 3 After step 2, F meets the boundary of each tetrahedron R of T in a collection of **normal curves** (curves which meet each 2–simplex in arcs joining diferent edges). Now these curves bound a set of disjoint discs in R. Replace $F \cap R$ by this set of discs, if necessary. Repeat for each tetrahedron of T. Note that this step can change

F drastically, in particular any components of $F \cap R$ which do not meet ∂R will be discarded.

Step 4 Suppose that for some tetrahedron R of T, $F \cap \partial R$ contains a normal curve of length > 4. Then there will be an internal "D–disc" formed by the surface and an edge of R which the curve crosses twice (cf. figure 3). Choose an outermost such situation and isotope F across this internal D–disc as in step 2. Repeat steps 1 to 3 if necessary and repeat the process until all curves of $F \cap \partial R$ have length 3 or 4. At this point F is a normal surface.

Remark If we start the normalisation process with F an embedded S^2, then the result is a collection (which may be empty) of normal S^2's. Moreover if each of the resulting S^2's bounds a 3–ball (and in particular if the resulting collection is empty) then so does the original S^2 — this is seen by induction on the steps of the process, at each step (in reverse) we either isotope the collection of S^2's or add a trivial S^2 to the collection (step 3 reversed) or pipe two of the collection together (steps 1 and 3 reversed). In any case they always all bound 3–balls.

Proof of lemma 2

Let M be a component of type II with boundary components $S_1, S_2, \ldots, S_t$. Consider the graph formed by regarding each S_i as a vertex and the segments of the 1-simplexes of T which lie in M as 1-cells. Then this graph is connected because it is the 1–skeleton of the cell decomposition of M rel boundary obtained by taking as cells the portions of the simplexes of T which lie in M. Choose a maximal tree Γ in this graph, which will comprises the boundary components $S_1, S_2, \ldots, S_t$ and connecting arcs $\alpha_1, \alpha_2, \ldots, \alpha_{t-1}$ and let the 2-sphere S be the boundary of a regular neighbourhood of Γ, see figure 5 in which $t = 4$.

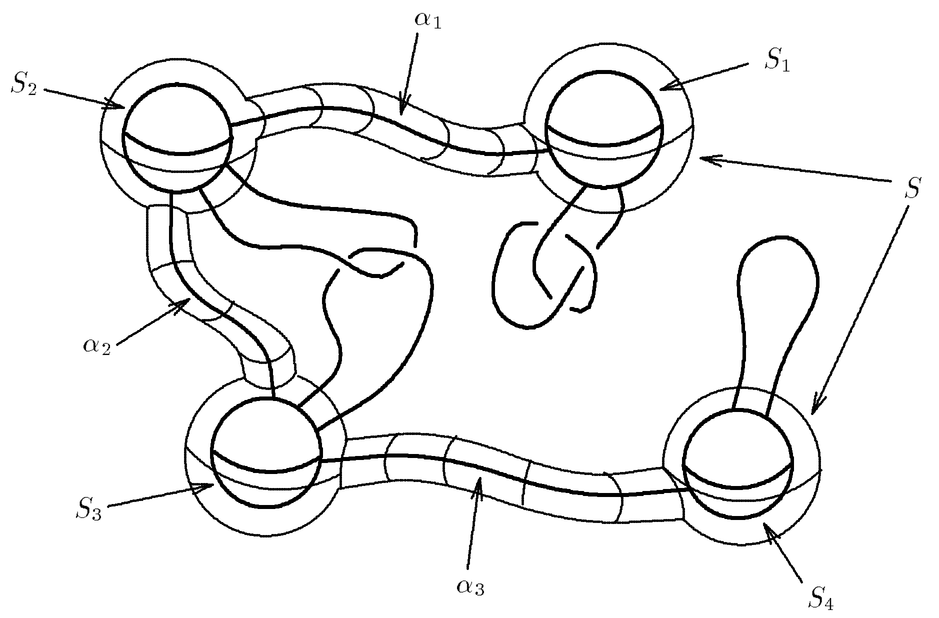

Figure 5

Normalise S. The resulting collection of normal 2–spheres does not meet the arcs $\alpha_1, \alpha_2, \dots, \alpha_{t-1}$, since normalisation simplifies intersections with the 1–skeleton of T and by construction S misses $\alpha_1, \alpha_2, \dots, \alpha_{t-1}$. Hence the collection does not contain a parallel copy of any of the boundary components of M and by maximality of Σ must be empty. Hence S bounds a 3–ball by the remark before the proof. The lemma now follows easily. $\square$

Proof of lemma 3

Let M be component of type III. The lemma breaks into two parts :

Part 1 *If M contains an AN 2–sphere then M is a 3–ball.*

Part 2 *If M is a 3–ball then M contains an AN 2–sphere.*

Proof of part 1 Let S be an AN 2–sphere. We observe that S separates M because, if not, then normalising S will result in at least one normal 2–sphere inside M which also does not separate M contradicting the maximality of Σ. Call the component of $M - S$ containing ∂M the *outside* and the other component the *inside.*

Notice that there are two ways to start to normalise S, namely the two D–type moves (step 4 moves) at the saddle–disc, one of which starts to move S inside and the other outside.

Claim 1 *Both normalisation processes continue to move S in the same direction.* In other words, after each step of the inward moving normalisation process, the resulting surface is inside the previous one (and similarly for the outward moving process). To see this it helps to imagine the inside coloured black and the outside white and to sketch the posibilities for black regions on faces of tetrahedra in T. The fact that S is AN is essential here — we never again get normal curves of length > 4 occuring during the process. The inward moving process results in black regions (and changes) of the types illustrated in figure 6. The claim can now be seen.

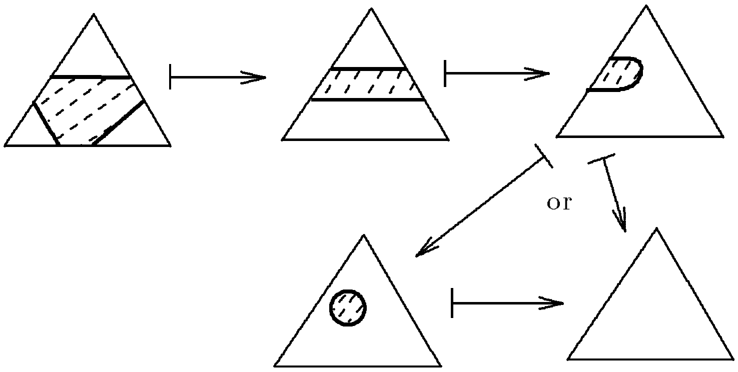

Figure 6

Since there are no normal 2–spheres inside S, it follows that the inward moving process results in the empty normal surface and hence S bounds a 3–ball inside by the remark before the proof of lemma 2. Moreover the outward moving process must result in a number of parallel copies of ∂M. If we now plug ∂M by a 3-ball then S also bounds a 3-ball outside by the same remark. Part 1 of lemma 3 follows. □

Sketch of proof of part 2 (for full details see [4])

We shall consider the interesection of $M \cong B^3$ with the 1–skeleton of T to be a link L (of arcs) in M and shall picture M (with a point on ∂M removed) as the bottom half of $\mathbb{R}^3$ with L hanging down from the top (figure 7).

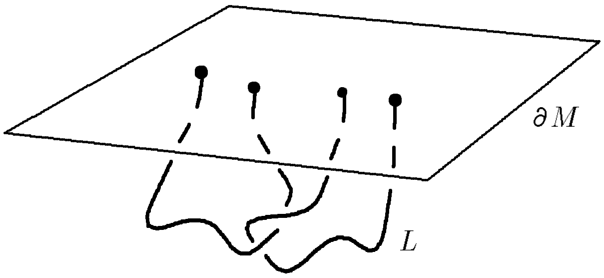

Figure 7

By general position we can assume that all but a finite number of levels (in $\mathbb{R}^3$) are **regular**, that is to say that they meet L transversally, and that the **exceptional** levels (the remaining levels) meet L in precisely one maximum or minimum.

Define the **width** of such a picture to be the sum of the number of points of L lying on one regular level between each pair of successive exceptional levels. Suppose the the picture is chosen so as to minimise width (this is called **thin position**).

Thin position implies that we do not see pictures such as those illustrated in figure 8, where the plane shown is a particular regular level. In each case in figure 8 there is an obvious isotopy which reduces width.

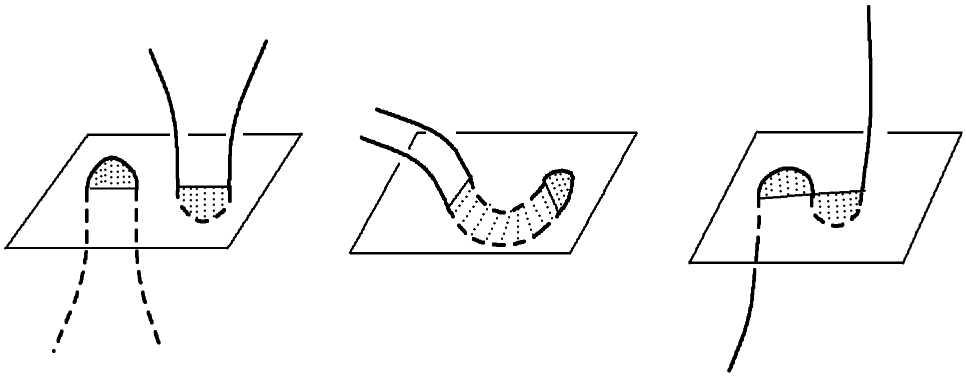

Figure 8.1 *Figure 8.2* *Figure 8.3*

Note that the shaded surfaces in figure 8 (used for the isotopies) do not need to be disjoint from the regular level, but they must not meet L except as shown.

Now consider the levels, starting at the top and dropping down the picture.

Claim 2 *The first exceptional level we meet is a maximum.* For suppose that the first exceptional level is a minimum, then there is a disc E which compresses a component of L into ∂M, figure 9.

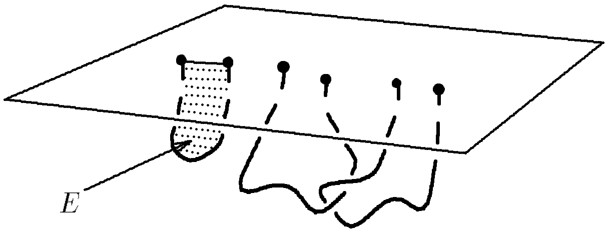

Figure 9

But a straightforward innermost curve/outermost arc argument replaces E by a disc in one tetrahedron of T and then normality of ∂M is contradicted.

By claim 2 the first exceptional level we meet is a maximum. Now we must eventually meet a minimum. Consider the regular levels between the first minimum (m) and the preceeding maximum (M).

Claim 3 *One of these regular levels has no D–curves of intersection with any 2–simplex of T.*

The argument used to prove this claim (and the concept of thin position) are due to Gabai [1]. Call the disc to the left of figure 8.1 an **up**–disc and the disc to the right a **down**–disc. Thin position implies that we never have simultaneous disjoint up and down–discs at any level. Now a D–disc (arising from a D–curve as shaded in figure 4) is either an up or a down–disc and moreover any up D–disc is disjoint from any down D–disc (except perhaps at one common boundary point as in figure 8.3). Now near the maximum M there is a small up–disc, so there cannot be any down D–discs near M and similarly there cannot be any up D–discs near m. Since we never see simultaneous up and down D–discs, there must be a level between M and m where there are no D–discs and hence no D–curves, as required.

Let S be the 2–sphere determined by a level with no D–curves given by claim 3. We shall see that S is essentially the required AN 2–sphere. We first observe that, by choice, S meets each ∂R, where R is a tetrahedron of T, in a collection of normal curves and closed curves interior to faces. Ignore the closed curves and concentrate on the normal curves. Regard S as comprising the discs which span these normal curves (forming a number of 2–spheres) piped together by thin tubes.

An easy exercise (best done by doodling in private) implies that a normal curve either has length 3, 4 or 8 (as illustrated in figures 2 and 3) or length ≥ 12. Moreover one of length ≥ 12 meets at least one edge of R in 3 adjacent intersections of alternating sign.

Claim 4 *There are no normal curves in S of length ≥ 12.* This is because the three alternating intersections give rise to the outlawed situation of figure 8.3. The required shaded surfaces in figure 8.3 can be readily found inside the tetrahedron containing the normal curve.

Claim 5 *No tetrahedron of T contains two or more normal curves of length 8.* To see this, choose an innermost pair (necessarily parallel) and consider an edge which meets each curve twice. This edge gives rise to the outlawed situation of figure 8.2. Again The required shaded surfaces are found inside the relevant tetrahedron.

Claim 6 *There are not two or more normal curves in S of length 8 (necessarily in different tetrahedra by claim 5).* To see this notice that each curve gives rise to both an up and a down–disc (cf. figure 3), thus if there are two curves of length 8 then there will be a disjoint pair of up and down discs, which is the outlawed situation of figure 8.1.

Claims 4 to 6 imply that S is the piped union of normal 2–spheres and ≤ 1 AN 2–sphere. We shall prove (cases 1 to ≥ 3 below) that S is not the piped union of normal 2–spheres, and thus discarding the normal 2–spheres, S becomes the required AN 2–sphere, completing the proof.

Case 1 *S is not just one normal 2–sphere.* For if it were then it would be a parallel copy of ∂M by maximality of Σ. But S has both an up and a down–disc given by the maximum and miminum above and below the chosen level, and we then have the same contradiction as in claim 1.

Case 2 *S is not the piped union of two normal 2–spheres.* For if it were then it would consist of two parallel copies of ∂M piped together. Consider a disc D spanning the pipe and use an innermost curve/outermost arc argument to make the up–disc for S disjoint from D and we then have the same contradiction as in case 1.

Case ≥ 3 *S is not the piped union of ≥ 3 normal 2–spheres.* Again the normal 2–spheres are all parallel to ∂M. Plug ∂M by a 3–ball B which is thought of as a (small) light-bulb and collect the portions of the 1–simplexes of T piercing the 3 outermost copies of ∂M in S together into a thick flex. Drawing S flat, we then have the situation shown in figure 10.

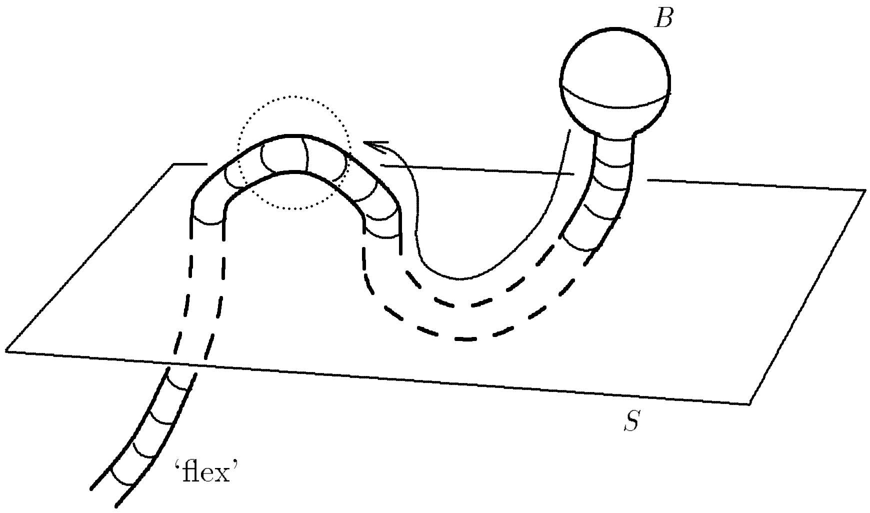

Figure 10

Consider the homeomorphism of this situation given by shortening the flex (as indicated by the arrow in figure 10) so that the bulb B is moved to the dotted position. After this homeomorphism we rechoose the levels above S to correspond to the obvious (round) 2–spheres between S and ∂B, then the new picture has considerably smaller width, contradicting thin position.

Note that figure 10 is somewhat misleading, because the pipes joining the parallel copies of ∂M get tangled in the flex, however it accurately shows the situation near the starting and finishing positions of the bulb, and the intersections of the constituent wires in the flex with S. Thus the proof is not affected by this inaccuracy of the figure. □

3. The practicality of searching for homotopy 3–spheres

The existence of an algorithm to search effectively for homotopy 3–spheres is of considerable theoretical interest, however neither the Rêgo–Rourke, nor the Rubenstein–Thompson algorithms are of much practical use as they stand. Both entail the kind of exhaustive search which would absorb a huge amount of computer time in order to accomplish very little, if implemented without further creative work. Thus we must seek ways to improve both algorithms.

Looking first at the Rêgo–Rourke algorithm, there is a simple way to make it more effective at finding suitable examples, provided one is prepared to sacrifice exhaustiveness. Since the purpose is to find a counter-example, it is sensible to implement the algorithm in a way which provides a good chance of finding an example in reasonable time rather than to make an exhaustive search which may take all the time available to produce nothing of interest.

Now the Rêgo–Rourke algorithm also uses two measures, denoted t and s in section 1, where t is the genus of the resulting H–digram and s might be called the **co-genus**.

Since the Poincaré Conjecture is true for genus two 3–namifolds it might make sense to start searching at say genus four and co-genus 2. Now the simplest way to find an complete system of curves on a surface is probably to write down an automorphism of the surface (a word in a standard set of generators) and apply it to a standard complete system. So one way to implement the algorithm (non-exhaustively) is to generate random words in the generators of the automorphism group of a surface of genus six, apply them to a standard complete system and then check whether the resulting system is a P–system (this check is of course very quick). Once a P–system is located in this way, then to write down an H–diagram of the homotopy sphere determined is fairly quick following the comments above the corollary in section 1.

Such an implementation (if designed in detail with some intelligence) might stand a good chance of locating a genuine homotopy sphere in a reasonable time, if one exists. But then the problem arises of testing the result for homeomorhism with S^3, ie using the Rubenstein–Thompson algorithm. Here there is probably a very serious problem, because normal surface theory is notoriously expensive (in terms of computer time) to implement for even quite simple manifolds and we are suggesting using it for a manifold obtained from a complicated H–diagram of genus 4. Thus to yield a practical programme to disprove the Poincaré Conjecture, the Rubenstein–Thompson algorithm will need to be developed to apply easily to say cell–complexes (rather than triangulations) or preferably to H–diagrams. This might be done by developing an appropriate theory of combinatorial minimal surfaces.

Added, November 1995 : Sergei Matveev has recast the Rubenstein–Thompson algorithm in terms of handle decompositions. This reformulation yields a more economical algorithm which dovetails well with the Rêgo–Rourke algorithm. However it is not claimed that the algorithms are practical even with this improvement.

References

[1] **David Gabai**, *Foliations and the topology of 3–manifolds III*, J. Diff. Geometry, **26** (1987)

[2] **Eduardo Rêgo and Colin Rourke**, *Heegaard diagrams and homotopy 3–spheres*, Topology, **27** (1988) 137–143

[3] **Hyam Rubenstien**, (to appear)

[4] **Abigail Thompson**, *Thin position and the recognition problem for S^3*, (to appear)

Colin Rourke
Mathematics Institute
University of warwick
Coventry CV4 7AL UK