Diffeomorphism Groups of Reducible 3-Manifolds

Allen Hatcher

In a 1979 announcement by César de Sá and Rourke [CR] there is a sketch of an intuitively appealing approach to measuring the difference between the homotopy type of the diffeomorphism group of a reducible 3-manifold and the homotopy types of the diffeomorphism groups of its prime factors. A more detailed write-up of this announcement did not appear, but the program was subsequently pushed through to completion by Hendriks and Laudenbach [HL], with some changes in the technical details to make everything work out correctly. In the present paper we provide a somewhat different implementation of the same general idea, in a way which we hope will lead more easily to further progress on this problem. The construction we give is also closer in spirit to a number of other well-known constructions, such as Harvey's curve complexes of surfaces and Culler-Vogtmann's "Outer Space".

Let M be a compact orientable 3-manifold which is the connected sum of irreducible manifolds $P_1, \cdots, P_n$ different from S^3. It will be convenient for us to use the term 'connected sum' in a broader sense than usual, allowing the connected sum of a manifold with itself. Thus the manifold M is allowed to have prime factors $S^1 \times S^2$ in addition to the P_i's. We will compare the group $\mathrm{Diff}(M)$ of orientation-preserving diffeomorphisms $M \to M$, with the C^∞ topology, with the corresponding groups for the summands P_i by constructing a fibration (up to homotopy) of classifying spaces

$$C(M) \longrightarrow \mathrm{BDiff}(M) \longrightarrow \mathrm{BDiff}(\textstyle\coprod_i P_i)$$

where $C(M)$ is a space whose points represent particular ways of realizing M as a connected sum of the P_i's and possibly some trivial S^3 summands. The precise definition is given in Section 1. There is also a relative version of this fibration, with $\mathrm{Diff}(M)$ replaced by the subgroup $\mathrm{Diff}(M \ \mathrm{rel}\, A)$ consisting of diffeomorphisms restricting to the identity on a compact subsurface A of ∂M, and $\mathrm{Diff}(\coprod_i P_i)$ replaced by $\mathrm{Diff}(\coprod_i P_i \ \mathrm{rel} \coprod_i A_i)$ for $A_i = A \cap \partial P_i$. The fiber in the relative case is still homotopy equivalent to the same space $C(M)$.

The group $\mathrm{Diff}(\coprod_i P_i)$ contains the product of the groups $\mathrm{Diff}(P_i)$, and equals this product if no two P_i's are diffeomorphic, but when some P_i's coincide there are also diffeomorphisms in $\mathrm{Diff}(\coprod_i P_i)$ that permute these summands.

Much of the work in constructing the main fibration will go into constructing the map $\mathrm{BDiff}(M) \to \mathrm{BDiff}(\coprod_i P_i)$ since there is no natural homomorphism $\mathrm{Diff}(M) \to \mathrm{Diff}(\coprod_i P_i)$

that could induce this map of classifying spaces. Looping the map of classifying spaces does lead to a homotopy class of maps $\mathrm{Diff}(M) \to \mathrm{Diff}(\coprod_i P_i)$, but these maps are homomorphisms only in a homotopy sense since they arise as loop maps.

It is known that if none of the P_i's are Seifert manifolds or the 3-ball then $\mathrm{BDiff}(\coprod_i P_i)$ is a $K(\pi,1)$, so the higher homotopy groups of $\mathrm{BDiff}(M)$ are the same as those of $C(M)$. These homotopy groups are generally nontrivial.

Similar constructions can be made for the operation of boundary connected sum of irreducible 3-manifolds P_i with nonempty boundary. In this case $\mathrm{BDiff}(M)$ and $\mathrm{BDiff}(\coprod_i P_i)$ are both $K(\pi,1)$'s, hence the version of $C(M)$ for this situation is also a $K(\pi,1)$.

One might expect analogous constructions to work in one lower dimension for connected sums of surfaces, but this does not seem to be the case. Applying the obvious definitions, what one gets instead is only a special case of the preceding construction for boundary connected sum of 3-manifolds, the case that M is a handlebody bounded by the given surface. The constructions do work for boundary connected sums of surfaces, although they do not really give anything new in this case. We will comment on these other situations more in the last section of the paper.

1. Connected Sum Configurations

Let us begin by defining the space $C(M)$. Let M be the connected sum of compact connected irreducible oriented 3-manifolds $P_1, \cdots, P_n$ different from S^3. We allow the connected sum of a manifold with itself, which produces $S^1 \times S^2$ summands. Let G be a finite connected graph with n of its vertices labeled $1, \cdots, n$ and such that all unlabeled vertices have valence at least three. Assume that the fundamental group of G is free of rank equal to the number of $S^1 \times S^2$ summands of M. To a vertex of G labeled i we associate the summand P_i and to each unlabeled vertex we associate a copy of S^3. Let M_0 be the disjoint union of the P_i's together with the copies of S^3 associated to the unlabeled vertices of G. We can construct a manifold diffeomorphic to M from M_0 by doing a connected sum operation for each edge of G. By specifying exactly how to perform these connected sum operations, and by letting the graph G vary, we will construct a space $C(M)$ that parametrizes the various choices made in the connected sums.

The sphere S^3_v corresponding to an unlabelel vertex v of G will be regarded as a metric object, isometric to the standard sphere of radius 1. For each edge e we perform a connected sum of the manifolds corresponding to the endpoints of e by removing balls from these manifolds and attaching a product $S^2_e \times I_e$. This product is also viewed as a metric object, the product of a sphere of radius r_e and an interval of length ℓ_e. When an end of a product $S^2_e \times I_e$ is attached to a P_i, this is done by choosing a smooth embedding of the ball B^3_e bounded by S^2_e into P_i, deleting the interior of the image of the ball, and then identifying the end of $S^2_e \times I_e$ with the resulting boundary sphere in P_i via the chosen embedding. Attaching an end of $S^2_e \times I_e$ to a sphere S^3_v is done similarly, with the added condition that the embedding of the attaching ball B^3_e in S^3_v is an isometry on S^2_e. We do these attachments for all the edges of G in such a way that all the products $S^2_e \times I_e$ in the resulting manifold M are disjoint. We assume all this is done so that the given orientations of the P_i's induce an orientation of M. The products $S^2_e \times I_e$ in M will be called *tubes* and the parts of the spheres S^3_v that remain in M will be called *nodes*. Two constructions of M as a connected sum associated to G in this way are regarded as equivalent when there is a homeomorphism between them that is an isometry on the tubes and nodes and is the identity on the complement of the union of the tubes and nodes.

Letting the various choices made in this construction of M vary, we obtain a space $C_G(M)$ parametrizing these choices. The choices are: the radii r_e, the lengths ℓ_e, and the embeddings $B^3_e \hookrightarrow P_i$ or $B^3_e \hookrightarrow S^3_v$ used to attach the tubes.

As G varies we now have a collection of disjoint spaces $C_G(M)$ which we wish to combine into a single space $C(M)$ by allowing the lengths ℓ_e of the I_e factors of certain tubes $S^2_e \times I_e$ go to zero. There are two ways in which we allow this to happen:

(a) If one end of the tube $S^2_e \times I_e$ attaches to a node and the other end to a summand P_i, then we can collapse $S^2_e \times I_e$ to S^2_e and identify the node with part of the ball in P_i bounded by the attaching sphere of the tube. The identification is done in the following way. The tube $S^2_e \times I_e$ can be viewed as part of the boundary of a 1-handle $B^3_e \times I_e$, the rest of the boundary being $B^3_e \times \partial I_e$, two 3-balls, one of which is embedded in P_i when the tube is attached, and the other of which is similarly embedded in the sphere S^3_v containing the node. When we collapse $S^2_e \times I_e$ to S^2_e we identify the node with part of the attaching ball of the tube in S^3_v by reflecting S^3_v across the attaching sphere of the tube, then we identify this ball in S^3_v with the attaching ball in P_i via the embedding used to attach the tube to P_i. The net result is that all the other tubes which attached to the node now attach to P_i, and we have a point in $C_{G/e}(M)$.

(b) If the two ends of e correspond to distinct nodes and if when we collapse the tube $S^2_e \times I_e$ to S^2_e, the union of the two nodes at its ends is, metrically, a node, then we have a point in $C_{G/e}(M)$.

Iteration of the operations in (a) and (b) corresponds to collapsing certain disjoint unions of trees in G to points, namely trees containing at most one labeled vertex.

We can enlarge the spaces $C_G(M)$ to spaces $\overline{C}_G(M)$ by allowing degenerate tubes of length $\ell_e = 0$ in the situations (a) and (b). The space $C(M)$ is then defined to be the quotient of the union of the disjoint spaces $\overline{C}_G(M)$ obtained by ignoring degenerate tubes.

In a sense the tubes are redundant. We could produce an equivalent space $C(M)$ by collapsing each tube $S_e^2 \times I_e$ to S_e^2, so the connected sums would be formed by deleting open balls and identifying the resulting boundary spheres. However it will be convenient to have the tubes later in the paper.

It would be possible to build a finite-dimensional version of $C(M)$ having the same homotopy type by choosing fixed Riemannian metrics on the factors P_i and taking the embeddings $B^3 \hookrightarrow P_i$ used to attach tubes to be scalar multiples of exponential maps. The scalar multiples could be taken to be the radii r_e of the tubes, and then the attaching would depend only on the choice of the center point of the attaching ball in P_i and a rotation parameter in $SO(3)$.

The definition of $C(M)$ works even if $n = 0$, when M is just a connected sum of copies of $S^1 \times S^2$. The condition that the graphs G have unlabeled vertices of valence at least 3 forces the number of $S^1 \times S^2$ summands to be at least 2. For the main theorem later in the paper we will assume $n \geq 1$, however.

We can think of the subspaces $C_G(M)$ in $C(M)$ as strata. In the finite-dimensional version of $C(M)$ the top-dimensional strata are those where G has the maximum number of vertices and edges, so the labeled vertices have valence 1 and the unlabeled vertices valence 3 (note that the Euler characteristic of G is equal to $1-s$ where s is the number of $S^1 \times S^2$ summands of M). The lowest dimensional strata are those where G has the fewest vertices and edges, so there are no unlabeled vertices.

The configurations in a stratum $C_G(M)$ of lowest dimension consist just of tubes joining the P_i's, and the global structure of $C_G(M)$ can be analyzed in the following way. To each configuration in $C_G(M)$ we can associate the centerpoints of the attaching balls of the tubes. These vary over a component X_G of the space of configurations of $2t$ distinct points in $\coprod_i P_i$, where t is the number of tubes, so $t = n-1+s$. More precisely, one starts with the space of configurations of $2t$ labeled points and factors out by the permutations of these points induced by symmetries of G that fix all vertices. In particular there are no such symmetries if $s = 0$. The projection $C_G(M) \to X_G$ is a fiber bundle with fiber $\mathbb{R}^{2t} \times SO(3)^t$ where the $\mathbb{R}^{2t}$ factor comes from the parameters r_e and ℓ_e and there is an $SO(3)$ factor coming from the rotation parameter in attaching each end of a tube, modulo isometries of the tube which reduce from two $SO(3)$ factors for each tube to one. The fiber bundle is a product if there are no symmetries of G fixing the vertices.

To see what the higher strata look like, consider the case that M is the sum of three irreducible manifolds P_1, P_2, P_3 with no $S^1 \times S^2$ summands. There is then one higher-dimensional stratum, consisting of configurations with one central node connected to the three P_i's by three tubes. The other three strata are obtained by collapsing each of the tubes in turn. We can see the structure of the large stratum by considering what happens to it under one of the three collapses, say the one that collapses the tube from the node to P_1. This identifies the larger stratum with triples consisting of a ball in P_1 together with two tubes from P_2 and P_3 to P_1 whose ends in P_1 lie in the specified ball in P_1. The larger stratum is attached to the smaller one by forgetting the ball.

General strata $C_G(M)$ can be described by iterating this idea, collapsing a collection of disjoint trees in G containing all the unlabeled vertices, each tree containing exactly one labeled vertex, producing configurations of nested balls in the P_i's, together with t twist parameters in $SO(3)$. If G has no symmetries fixing all the labeled vertices, these twist parameters are globally defined, and otherwise one factors out the group of symmetries.

2. Sphere Systems

Inside a fixed copy of the manifold M, consider a nonempty collection S of finitely many disjoint embedded spheres S_j such that $M - S$ consists of the manifolds P_i with disjoint balls removed, together with possibly some copies of S^3 with disjoint balls removed, at least three balls being removed from each S^3. Associated to such a *full sphere system* S is a finite graph G whoses vertices are the components of $M - S$ and whose edges are the components of S. The vertices of G not corresponding to summands P_i have valence at least three. The space of all such sphere systems S with associated graph G, and with weights $t_j > 0$ assigned to their spheres S_j, we denote $S_G(M)$. We form a space $S(M)$ from the union of these $S_G(M)$'s by letting some weights t_j go to 0 and deleting the corresponding spheres S_j, provided the remaining spheres still form a full sphere system.

We assume $S(M)$ is nonempty, so M is neither S^3 nor $S^1 \times S^2$.

Theorem. *$S(M)$ is contractible.*

Proof: We will show that the homotopy groups $\pi_k S(M)$ are trivial for all k. Let the family $S_t \in S(M)$, $t \in S^k$, represent an element of $\pi_k S(M)$. Choose a fixed system $\Sigma \in S(M)$ and thicken Σ to a family $\Sigma \times [-1,1]$ of parallel systems. Sard's theorem implies that for each $t \in S^k$ there is a slice $\Sigma \times \{s\}$ in this thickening that is transverse to S_t. This slice will remain transverse to S_t for all nearby t as well. By a compactness argument this means we can choose a finite cover of S^k by open sets U_i so that S_t is transverse to a slice Σ_i for all $t \in U_i$.

For a fixed $t \in U_i$ consider the standard procedure for surgering S_t to make it disjoint from Σ_i. The procedure starts with a circle of $S_t \cap \Sigma_i$ that cuts off a disk D in Σ_i that contains no other circles of $S_t \cap \Sigma_i$. Using D we then surger S_t to eliminate the given circle of $S_t \cap \Sigma_i$. The process is then repeated until all circles have been eliminated. Each surgery produces a system of spheres that would define a point of $S(M)$ after deleting trivial spheres that bound balls and redundant isotopic copies of the same sphere. (This follows from the standard reasoning in the proof of the existence and uniqueness of prime decompositions.) For the moment, however, we are not going to delete trivial or redundant spheres.

A convenient way to specify the order in which to perform the sequence of surgeries is to imagine the surgeries as taking place during a time interval, and then surgering a circle at the time given by the area of the disk it cuts off in Σ_i, normalized by dividing by the area of Σ_i itself, where area is computed using some choice of a Riemannian metric on M. The only ambiguity in this prescription occurs if one is surgering the last remaining circle and this circle splits Σ_i into two disks of equal area. Then one would have to make an arbitrary choice of one of these disks as the surgery disk.

We refine this procedure so that it works more smoothly in our situation. Thicken S_t to a family $S_t \times [-1,1]$ of nearby parallel systems, all still transverse to Σ_i for $t \in U_i$. Call this family of parallel systems $\mathbf{S}_t$. For $t \in U_i$, with i fixed for the moment, we perform surgery on all the disks of $\mathbf{S}_t$ family by gradually cutting through the family in a neighborhood of Σ_i. Thus we are producing a family $\mathbf{S}_{tu}$ for $u \in [0,1]$, where again we use the areas of the surgery disks in Σ_i to tell when to perform the surgeries. We allow $\mathbf{S}_{tu}$ to contain finitely many pairs of spheres that touch along a common subsurface at the instant when these spheres or disks are being surgered. To specify the surgeries more completely we choose a small neighborhood $\Sigma_i \times (-\varepsilon_i, \varepsilon_i)$ of Σ_i in $\Sigma \times [-1,1]$, which we rewrite as $\Sigma_i \times \mathbb{R}$, and we let the surgery on a sphere of $\mathbf{S}_{tu}$ produce two parallel copies of the surgery disk in the slices $\Sigma_i \times \{\pm 1/u\}$ of $\Sigma_i \times \mathbb{R}$.

To convert the thickened family $\mathbf{S}_{tu}$ back into an ordinary family S_{tu} consisting of finitely many spheres for each (t,u) we replace each family of parallel spheres in $\mathbf{S}_{tu}$ of nonzero thickness by the central sphere in this family. Thus this central disk belongs to S_{tu} for an open set of values of (t,u). Observe that this prescription for constructing S_{tu} eliminates the ambiguity in choosing one of the two equal-area surgery disks mentioned earlier.

As t varies over U_i we now have a family S_{tu}, depending on i. To combine these families for different values of i, letting t range over all of S^k rather than just over U_i, we proceed in the following way. For each i choose a continuous function $\varphi_i : U_i \to [0,1]$ that takes the value 1 near ∂U_i and the value 0 on an open set V_i inside U_i such that the different V_i's still cover S^k. Then construct S_{tu} by delaying the time when each surgery along Σ_i is performed by the value $\varphi(t)$. We may assume all the spheres Σ_i are distinct and their thickenings $\Sigma_i \times (-\varepsilon_i, \varepsilon_i)$ are disjoint, so the surgeries along different Σ_i's are completely independent of each other.

We have constructed the family S_{tu} for $(t,u) \in S^k \times [0,1]$ such that all the curves of $S_t \cap \Sigma_i$ are surgered away as u goes from 0 to $1/2$ for $t \in V_i$. We can then adjoin Σ_i to S_{tu} for $(t,u) \in V_i \times (1/2, 1)$, deleting the surgered spheres of S_{tu} for $u \geq 3/4$. We may assume all the thickenings $\Sigma_i \times (-\varepsilon_i, \varepsilon_i)$ are disjoint from the original separating system Σ. Then we adjoin Σ to S_{tu} for $u > 3/4$, so that for $u = 1$ only Σ remains in S_{tu}.

Components of S_{tu} that bound balls in M can be deleted from S_{tu}, so we will assume this has been done. All that prevents S_{tu} from being a contraction of S_t in $S(M)$ is then the possible presence of pairs of spheres in S_{tu} that are isotopic, or in other words, complementary components of S_{tu} that are products $S^2 \times I$ both of whose boundary spheres lie in S_{tu}. We will show how to replace each such collection of isotopic spheres in S_{tu} by a single sphere, in a way that varies continuously with the parameters t and u.

As a preliminary step, we will arrange that the weighting coefficients of the various spheres in S_{tu} vary piecewise linearly with t and u. For each (t,u) there is a neighborhood in $S^k \times I$ such that the spheres of S_{tu} have positive weights throughout this neighborhood. Choose piecewise linear weight functions for these spheres that are zero outside this neighborhood and positive in a smaller neighborhood of (t,u). A finite number of these smaller neighborhoods cover $S^k \times I$ by compactness. For each (t,u) let the new weight function for each sphere of S_{tu} be the sum of the piecewise linear weight functions for this sphere for the finitely many neighborhoods containing (t,u). Letting the weights go linearly from their old values to their new values gives a homotopy of S_{t0} in $S(M)$, so we can now use the new piecewise linear weight functions for S_{tu} over $S^k \times I$.

Choose a triangulation of $S^k \times I$ such that all the weight functions are linear on simplices. As one passes from a simplex to its boundary certain weight go to zero, and the corresponding spheres are deleted from S_{tu}. To give ourselves a little room to maneuver, let us expand the triangulation to a handle structure on $S^k \times I$ in the usual way, with one i-handle $h(\sigma)$ for each i-simplex σ. Thus there is a map $p : S^k \times I \to S^k \times I$ with $p^{-1}(\sigma) = h(\sigma)$ for each simplex σ. We can modify the family S_{tu} by changing the weights of S_{tu} linearly to those of $S_{p(t,u)}$. Again this gives a homotopy of S_{t0}, so we can assume this has been done and we use the notation S_{tu} for the new family. Passing from an i-handle $D^i \times D^j$ to its attaching locus $\partial D^i \times D^j$, some weights of spheres of S_{tu} go to zero, or in other words, S_{tu} is replaced by a subcollection of spheres. We want to foliate the product regions between isotopic spheres of S_{tu} by product foliations that vary continuously with (t,u), so that when we pass from a handle to its attaching locus the foliations over the attaching locus are the restrictions of the foliations over the whole handle. This can be done by downward induction on the indices of the handles. For the induction step of choosing foliations over an i-handle $D^i \times D^j$, one has foliations for $(t,u) \in D^i \times \partial D^j$ coming from the foliations over handles of larger index, and these extend to foliations over the full handle since the space of product foliations on $S^2 \times I$ is contractible (this is an equivalent form of the Smale conjecture $\mathrm{Diff}(S^3) \simeq O(4)$).

We can also arrange inductively that the foliations have compatible continuously varying transverse affine structures since the space of such structures is contractible as well. Using these transverse affine structures one can form nonnegative linear combinations of leaves in the foliations. Thus we can replace each collection of isotopic spheres in each S_{tu} by the single sphere which is their weighted average. This gives a new continuously varying family S_{tu} that lies in $S(M)$, agreeing with the previous family S_{tu} for $u = 0, 1$. Thus we have shown that the original map $S^k \to S(M)$ is homotopic to a constant map, finishing the proof. $\square$

It will be convenient in the next section to have a version of $S(M)$ in which sphere systems $S \subset M$ are thickened to products $S \times I \subset M$, regarded as laminations with sphere leaves. The weights t_j on the components S_j of S are viewed as thicknesses of the products $S_j \times I$ in such a lamination. As weights go to zero, the corresponding laminated products $S_j \times I$ shrink to single leaves S_j and disappear. The space of all such weighted laminations we denote by $T(M)$, thinking of such laminations as *tubings* of M, in the spirit of the preceding section. It is clear that $T(M)$ has the same homotopy type as $S(M)$ since the space of collar neighborhoods of a codimension one submanifold is contractible.

3. Classifying Spaces

Let us recall a classical construction for a classifying space $\mathrm{BDiff}(M)$ for a smooth manifold M. The connectivity of the space $\mathrm{Emb}(M, \mathbb{R}^k)$ of smooth embeddings $M \hookrightarrow \mathbb{R}^k$ increases with k, by Whitney's embedding theorem applied to parametrized families, so the space $\mathrm{Emb}(M, \mathbb{R}^\infty) = \bigcup_k \mathrm{Emb}(M, \mathbb{R}^k)$ has trivial homotopy groups, hence is contractible. The action of $\mathrm{Diff}(M)$ on this space by composition in the domain defines a principal fiber bundle

$$\mathrm{Diff}(M) \longrightarrow \mathrm{Emb}(M,\mathbb{R}^\infty) \longrightarrow \mathrm{Emb}(M,\mathbb{R}^\infty)/\mathrm{Diff}(M)$$

with contractible total space, so the base space $\mathrm{Emb}(M,\mathbb{R}^\infty)/\mathrm{Diff}(M)$ is a classifying space $\mathrm{BDiff}(M)$. This can be regarded as the space of smooth submanifolds of $\mathbb{R}^\infty$ diffeomorphic to M, and we will denote it by $I(M)$, the images of M in $\mathbb{R}^\infty$. When M is oriented and we take $\mathrm{Diff}(M)$ to mean orientation-preserving diffeomorphisms, then $I(M)$ becomes the space of oriented submanifolds of $\mathbb{R}^\infty$ diffeomorphic to M.

In a similar way, by replacing $\mathrm{Emb}(M,\mathbb{R}^\infty)$ by the subspace of embeddings restricting to a given embedding on a submanifold $A \subset M$ we obtain a classifying space $\mathrm{BDiff}(M \text{ rel } A)$.

Specializing to the situation of this paper, consider the product $\mathrm{Emb}(M,\mathbb{R}^\infty)\times T(M)$ where $T(M)$ is the space of tubings of M defined at the end of the previous section. We can think of points in $\mathrm{Emb}(M,\mathbb{R}^\infty)\times T(M)$ as pairs consisting of an embedding $f\colon M \to \mathbb{R}^\infty$ together with a tubing of the image $f(M)$. Again $\mathrm{Diff}(M)$ acts by composition with f, and the orbit space, consisting of tubed submanifolds of $\mathbb{R}^\infty$ diffeomorphic to M, is homotopy equivalent to $I(M)$ since $T(M)$ is contractible. We denote this space by $TI(M)$, the tubed images of M in $\mathbb{R}^\infty$. Like $I(M)$, it is a classifying space $\mathrm{BDiff}(M)$.

Let $P = \coprod_i P_i$ and let $TI(M\cup P)$ be the subspace of $TI(M)\times I(P)$ consisting of pairs (x,y) such that $x\cap y$ consists of the closure of the complement of the tubes and nodes of x. Thus x is obtained from $x\cap y$ by attaching the tubes and nodes, while y is obtained from $x\cap y$ by filling in balls where the tubes attach to $x\cap y$.

Proposition. *The projection $p\colon TI(M\cup P)\to TI(M)$, $(x,y)\mapsto x$, is a homotopy equivalence.*

Proof: As a first approximation to the proof, and as motivation, note that the fibers of p are contractible since they consist of all the ways of filling in a collection of disjoint 2-spheres in $\mathbb{R}^\infty$ with disjoint 3-balls whose interiors are also disjoint from a submanifold of $\mathbb{R}^\infty$. The disjointness conditions can be achieved by general position, and the space of balls with a fixed boundary is contractible since it is a classifying space for the diffeomorphism group of a ball fixing the boundary, which is contractible by the Smale conjecture. If the projection p were a fibration, this would finish the proof. However, verifying some form of the homotopy lifting property does not look like it would be easy, particularly since the number of 3-balls being filled in varies from fiber to fiber.

Composing p with the map $TI(M)\to I(M)$ that forgets the tubings gives a fibration $TI(M\cup P)\to I(M)$, and it will suffice to prove that the fiber F of this fibration is contractible. There is a natural projection $q\colon F\to T(M)$ which ignores the filling balls.

Since $T(M)$ is contractible, it will suffice to show q is a homotopy equivalence. We will describe how to construct a section $s:T(M) \to F$ of q and a homotopy from sq to the identity.

The section s can be constructed inductively over strata of $T(M)$. Over a given stratum one has a system of disjoint spheres in a fixed copy of M in $\mathbb{R}^\infty$, varying by isotopy over the stratum, that one is trying to fill in with balls. The restriction map from the space of ball systems to the space of sphere systems is a fibration, and its fiber is contractible. This implies (how exactly?) the existence of a section s over the stratum. For the inductive step one wants to extend a given section from the boundary of the stratum to the rest of the stratum. The spheres one wants to fill in with balls over the interior are already filled in with balls over the boundary, although these balls may intersect M in more than their boundary spheres. By contractibility of the space of balls with fixed boundary, we can fill in the balls over the whole stratum, and then use general position to make them meet M only in their boundary spheres over the interior of the stratum.

In this fashion one can construct a section s. A homotopy from sq to the identity can be constructed by a similar inductive procedure. In each fiber of q one has the space of all filling balls with fixed boundary spheres, and this space is contractible to the ball given by the section. $\square$

Forgetting the first coordinate of points (x, y) in $TI(M \cup P)$ gives a fibration

$$F \longrightarrow TI(M \cup P) \longrightarrow I(P)$$

Proposition. *The fiber F is homotopy equivalent to $C(M)$.*

Proof: The idea is to show that corresponding strata of F and $C(M)$ are homotopy equivalent by equivalences that respect the way the strata fit together. Points of both F and $C(M)$ consist of a fixed copy of P to which tubes and nodes are attached, where in F these attachments take place in $\mathbb{R}^\infty$ while for $C(M)$ the attachments are done abstractly but with extra metrical information.

Consider first a lowest stratum, where there are no nodes, only tubes. For both F and $C(M)$ there is a collection of disjoint balls in P where the tubes attach. For $C(M)$ the additional data determining each tube attachment is a twist parameter in $SO(3)$. For F the additional data is an embedding of the tube in $\mathbb{R}^\infty$. The claim is that the set of tubes in $\mathbb{R}^\infty$ with fixed ends has the homotopy type of $SO(3)$. This set of tubes is a classifying space for the group of diffeomorphisms of $S^2 \times I$ that restrict to the identity on $S^2 \times \partial I$. It follows from the Smale conjecture that this group has the homotopy type of the loopspace $\Omega SO(3)$, so its classifying space has the homotopy type of $SO(3)$.

To treat higher strata this argument needs to be extended ....

□

Combining the preceding propositions, we obtain the desired homotopy fibration

$$C(M) \longrightarrow \mathrm{BDiff}(M) \longrightarrow \mathrm{BDiff}(P)$$

The relative version mentioned in the introduction is derived in the same way, the only difference being that throughout the whole process one restricts to embeddings of P_i in $\mathbb{R}^\infty$ that are fixed on the given subsurface $A_i \subset \partial P_i$.