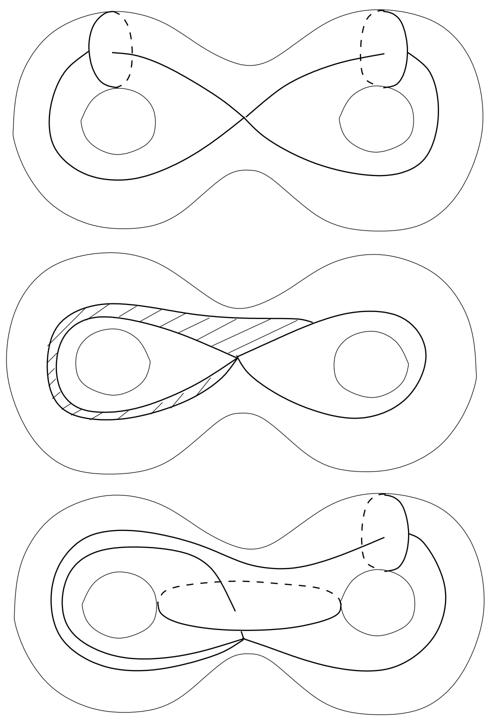
Notes on Heegaard Splittings

Jesse Johnson



E-mail address: JesseE.Johnson@yale.edu
URL: http://pantheon.yale.edu/~jj327/

Preface

The purpose of these notes is to introduce the reader to some of the tools and ideas from the study of Heegaard splittings of 3-manifolds. The discussion is motivated by three fundamental theorems that are proved in the fifth chapter: the existence of common stabilizations (the Reidemeister-Singer Theorem), the classification of Heegaard splittings of the 3-sphere (Waldhausen's Theorem) and the connection between weakly reducible splittings and incompressible surfaces (Casson and Gordon's Theorem). The presentation is based on modern methods rather than the classical proofs.

The discussion is aimed at the level of a beginning graduate student or perhaps even an advanced undergraduate. It will be helpful for the reader to have some familiarity with one or more of the following books: Rolfsen's *Knots and Links*, Hatcher's *Algebraic Topology*, Hatcher's unpublished notes on 3-manifolds (available on his web page) and Hempel's *3-manifolds*. I have tried to assume as little prior knowledge as possible.

This draft of the notes is still fairly rough in parts, particularly in the fifth chapter. Please send me any comments and suggestions you think of. I would especially appreciate feedback on which parts are unclear or could use further discussion and where more Figures would be useful.

Contents

CHAPTER 1

Piecewise Linear Topology

1. Simplicial Complexes

Before we begin to study Heegaard splittings, we will need to consider a certain structure called a triangulation, which we will define below. The class of manifolds which allow triangulations is often called the piecewise-linear category and a careful discussion of this approach to topology can be found in Rourke and Sanderson's book [**11**].

We would like to construct topological spaces by gluing together simple pieces called simplices. An n-dimensional *simplex* is a convex hull of $n+1$ linearly independent points in euclidean n-space. For $n \geq 1$, an n-simplex is homeomorphic to an n-ball and its boundary consists of simplices of lower dimension. Each simplex in the boundary of a simplex σ is called a *face* of σ. Figure 1 shows simplices of dimension 0, 1, 2 and 3.

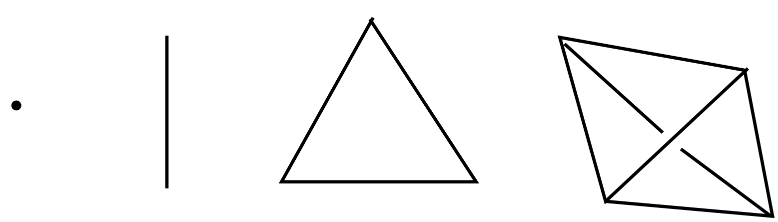


FIGURE 1. The simplices of dimension 0, 1, 2 and 3 are a point, an arc, a triangle and a tetrahedron, respectively.

We would like to construct topological spaces by "gluing" together simplices along their faces. In order to do this, we first need to define a formal structure that encodes the way in which faces of simplices are glued.

1.1. DEFINITION. A *simplicial complex* K is an ordered pair (V, F) where V is a set and F is a set of finite subsets of V. The set V is the set of *vertices* of K. An element of F is a *simplex* of K. The simplices of F must have the following properties:

(1) If v is an element of V then the set $\{v\}$ is an element of F.

(2) If A is an element of F and B is a subset of A then B is an element of F.

The first condition implies that any vertex of K is also a simplex of K (a 0-simplex). The second condition states that any face of a simplex of K is also a simplex of K.

A *subcomplex* of a simplicial complex $K = (V, F)$ is a simplicial complex $K' = (V', F')$ such that $V' \subset V$ and $F' \subset F$. In other words, every vertex of K' is a vertex of K and every simplex of K' is a simplex of K.

A simplicial complex is not a topological space, but we can construct a topological space from a simplicial complex. We stated above that an n-simplex should be topologically a convex hull of $n+1$ points in euclidean space. In $\mathbf{R}^{n+1}$, we can think of this as the set of points $(x_0, \ldots, x_n)$ such that $x_i \geq 0$ for each i and $\sum x_i = 1$.

We would like to simultaneously realize all the simplices in K so that their faces are identified in the way encoded in F. In the case when V is finite, let n be the number of vertices in V and assume that the points of F are labeled $a_1, \ldots, a_n$. The *support* of a point $(x_1, \ldots, x_n)$ is the set points $\{a_i | x_i \neq 0\}$. Define the *realization* $|K|$ of K to be the set of points $(x_1, \ldots, x_n)$ such that $x_i \geq 0$ for each i, $\sum x_i = 1$ and the support of $(x_1, \ldots, x_n)$ is a simplex in F.

The topology on $|K|$ will be the subset topology inherited from $\mathbf{R}^n$. The reader should examine how the structure of $|K|$ is encoded by K. Whenever it is clear from the context, we will use K and $|K|$ interchangeably.

Given a simplex $\sigma \in F$ containing $n+1$ vertices, the *dimension* of σ is n. The subcomplex of K consisting of all the simplices of dimension less than or equal to n is called the *n-skeleton* of K and written K^n. The dimension of a simplicial complex is largest dimension of any simplex. It is a topological space and in many cases a simplicial complex will even be a manifold. We will say that a manifold M is *piecewise-linear* if M is homeomorphic to (the realization of) a simplicial complex.

1.2. Theorem (Moise [**9**], 1952). *Every compact 3-manifold is homeomorphic to a 3-dimensional simplicial complex.*

The proof is beyond the scope of the discussion here. The main implication of the theorem is that we can narrow our focus to triangulated manifold without actually throwing out any manifolds. Given a 3-manifold M, a simplicial complex K such that $|K|$ is homeomorphic to M is often called a *triangulation* of M. From now on, we will implicitly assume that every manifold we discuss is homeomorphic to a simplicial complex.

2. Regular Neighborhoods

A closed subset A of a manfold M is piecewise-linear if for some simplicial complex K and some homeomorphism $\phi : |K| \to M$, the set A is the image in M of a sub-complex of K. Throughout the rest of these notes, we will assume that every closed set is piecewise-linear and the closure of every open set is piecewise-linear. For a given manifold, if we would like to consider more complicated subsets, we will need to define finer triangulations.

The idea will be to place a vertex at the center of each simplex of K and fill in faces in such a way that the realization is homeomorphic to $|K|$. Once again, we must translate this into a formal language.

1.3. Definition. For a given simplicial complex $K = (V, F)$ the first *barycentric subdivision* $K' = (V', F')$ of K is the simplicial complex defined as follows:

(1) The vertices of K' are the simplices of K, i.e. $V' = F$.

(2) A set $\sigma \subset V'$ is a simplex of K' if and only if for some labeling $\sigma = \{a_1, \ldots, a_n\}$ (recall each element a_i is a simplex of K) we have that a_j is a face of a_i whenever $j > i$.

The first condition fits the idea of placing a new vertex in the center of each face of K. The second condition is far less intuitive, but an example should make it clearer.

Consider a 2-simplex T, i.e. a triangle. It has three one-dimensional faces and three zero-dimensional faces. To subdivide T, we can place a vertex at the center of T, one vertex in the center of each edge, and one vertex at the center of each corner (the 0-dimensional faces). If we connect these edges in the obvious way, we will split T into six smaller triangles, as in Figure 2. Each of these triangles has one vertex at the center of T, a second vertex at the midpoint of an edge, and its third vertex at one of the endpoints of this edge. This is precisely the condition that each successive vertex is a face of the previous vertex.

The edges of the smaller triangles are of three types: Three of the edges go from the center of T to an edge of T, three edges connect the center of T to a vertex of T and six edges connect the midpoint of an edge to one of its endpoints. These all meet condition (2), and the reader can check that these are all the faces that meet this criteria.

Once we have constructed the first barycentric subdivision K' of K, we can take the second subdivision K'' of K to be the first barycentric

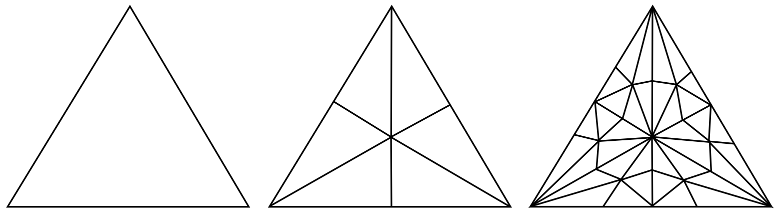


FIGURE 2. The first and second barycentric subdivisions of a triangle (a 2-simplex).

subdivision of K', as in Figure 2. In general, any simplicial complex constructed by repeatedly subdividing K will be called a barycentric subdivision of K.

There is a canonical homeomorphism from $|K'|$ to $|K|$. (It is left to the reader to work it out explicitly.) By induction, there is a canonical homeomorphism from the any barycentric subdivision of K to $|K|$. Given a subcomplex L of K, there is a unique subcomplex L' of any subdivision K' such that the canonical homeomorphism sends $|L'|$ onto L. In other words, barycentric subdivision is compatible with subcomplexes.

Our first application of the piecewise-linear method will be to define the notion of a regular neighborhood. Given a subset C of a manifold, we would like to find an open neighborhood of C which retains as much of the topology of C as possible. One can think of this as taking the closed set and dipping it into a bucket of paint, so that the result object is a "thickened" version of C.

Let $C \subset M$ be a closed, piecewise-linear subset of a manifold M. Let K be a simplicial complex, $L \subset K$ a subcomplex and $\phi : |K| \to M$ a map such that $\phi(L) = C$. Let $K' = (V', F')$ be the second barycentric subdivision of K, $\psi : K' \to K$ the canonical map and let $L' \subset L$ be such that $\psi(L') = L$.

1.4. DEFINITION. The *regular neighborhood* N of L is the set of simplices $\{\sigma \in F' | \sigma \cap L' \neq \emptyset\}$. The set $\phi(\psi(N))$ is a *closed regular neighborhood* of C and the interior of N is an *open regular neighborhood* of C.

A simplex σ of K' is in N if and only if some face of σ is also a simplex in L. This guarantees that the interior of N will be an open set containing L. Taking the second barycentric subdivision ensures that the neighborhood is "small". Figure 3 shows a regular neighborhood in a second barycentric subdivision of a surface.

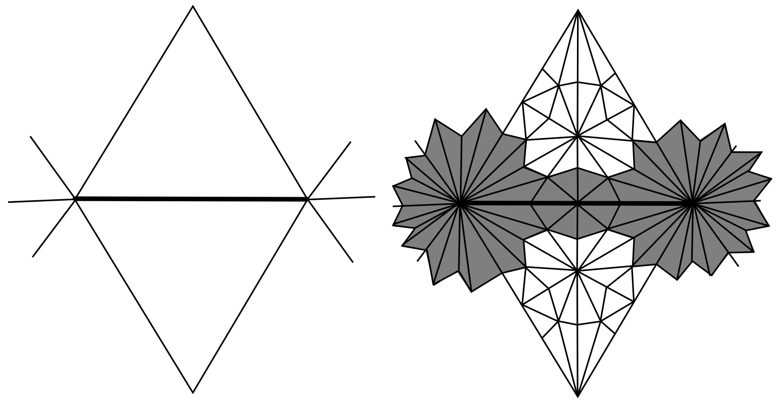


FIGURE 3. The shaded area is a regular neighborhood of an edge in a 2-complex.

Note that the image in M of N is *a* regular neighborhood, rather than *the* regular neighborhood. The construction depended on the choice of the triangulation and a different triangulation would have produced a different N. However, all is not lost, thanks to the following theorem:

1.5. THEOREM. [**11**] *Let N and N' be regular neighborhoods of a closed, piecewise-linear subset C of M. Then there is an isotopy of M which fixes C and takes N onto N'.*

Thus although a regular neighborhood is not unique as a subset of M, it is topologically unique. This is the main concept from piecewise-linear topology that we will use in our discussion of handlebodies and Heegaard splittings.

3. Gluing

A major theme in the study of Heegaard splittings is the construction of complicated manifolds by putting together simple pieces. In order to do this, we use a method called *gluing* . This idea is the same as when one glues back together a piece of plastic or wood (or more appropriately, rubber) which has broken:

Let X and Y be closed subsets of the boundaries of manifolds A, B of the same dimension or let X and Y be disjoint, closed subsets of ∂A. Let $\phi : A \to B$ be a homeomorphism. We would like to apply glue to the section A of the boundary of X, then stick this to the section B, matching them up by the identification defined by ϕ.

To do this abstractly if X and Y are in separate manifolds, we consider the disjoint union $A \sqcup B$ of our manifolds. Identify each point

$x \in X$ to $\phi(x) \in Y$ by considering the equivalence relation $x \sim y$ if $y = \phi(x)$ where $x \in X$ and $y \in Y$. The quotient of $A \sqcup B$ by this relation, endowed with the quotient topology, is the result of gluing A to B along ϕ. This set is often denoted by $A \cup_\phi B$.

If X and Y are (disjoint) subsets of the boundary of the same manifold A, then to glue X to Y, we take the quotient of A by the equivalence relation $x \sim \phi(x)$. Once again, the topology of the new manifold, denoted $A\cup_\phi$ is given by the quotient topology.

If A and B are piecewise-linear manifolds and the subsets X and Y of their respective bondaries are subcomplexes of the triangulations, we can ask if the map ϕ "respects" these triangulations.

A simplex is defined as a subset of euclidean space, so we can say that a map between two simplices (prehaps of different dimensions) is *linear* if it is the restriction of a linear map of Euclidean space. A map ϕ from a simplicial complex $|K|$ to a simplicial complex $|K'|$ is *piecewise linear* if ϕ sends each simplex of $|K|$ onto a simplex of $|K'|$ by a linear map. Finally, a map ϕ between two piecewise-linear manifolds X, Y is *piecewise-linear* of for some triangulation of X and some triangulation of Y, ϕ is piecewise-linear for those triangulations.

Just as we assumed that all manifolds will be piecewise-linear, we will assume in the future that all maps are piecewise-linear. In particular, all gluing maps will be piecewise-linear. This is a useful assumption because of the following Lemma:

1.6. Lemma. *If A and B are piecewise-linear manifolds of the same dimension and ϕ is a piecewise-linear homeomorphism from a subset of ∂A to a subset of ∂B then $A \cup_\phi B$ is a manifold with boundary and has the same dimension as A and B.*

The proof of this Lemma will be left as an exercise for the reader.

4. Handlebodies

In this section we will give the definition of a handlebody and describe its basic properties. In the next section we will describe how handlebodies combine to form Heegaard splittings. All manifolds and subsets of those manifolds will be assumed to be piecewise-linear.

1.7. Definition. Let $B_1, \dots, B_n$ be a collection of closed 3-balls and let $D_1, \dots, D_m, D'_1, \dots, D'_m$ be a collection of pairwise disjoint disks in $\bigcup \partial B_i$. For each $i \leq m$, let $\phi_i : D_i \to D'_i$ be a homeomorphism. Let H be the result of gluing along ϕ_1, then gluing along ϕ_2, and so on. After the final gluing, if H is connected then H is a *handlebody*.

A Handlebody is shown in Figure 4. We will use the symbol $H^{n,m}$ for a handlebody constructed from n balls glued along m pairs of disks. The requirement that a handlebody be connected implies that $m \geq n-1$.

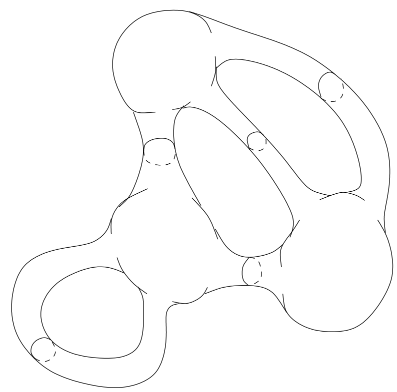


FIGURE 4. A handlebody is a collection of 3-balls, glued together along disks. This handlebody would be denoted $H^{3,5}$.

A handlebody is a manifold with a connected boundary and we can compute the genus of $\partial H^{n,m}$ from n and m. The boundary of each ball is a sphere, with Euler characteristic 2. To get $\partial H^{n,m}$, we remove the interiors of the disks $D_1, \ldots, D_m, D'_1, \ldots, D'_m$ and glue together the punctured spheres along the boundaries of the disks.

A loop has Euler characteristic zero so gluing together the boundary loops has no effect on the Euler characteristic of the union. Each disk contributes a 1 so removing a disk reduces the Euler characteristic by one. Thus if there are n spheres and we remove $2m$ disks, then glue together the boundary components, the resulting surface will have Euler characteristic $\chi(\partial H^{n,m}) = 2n - 2m$. The genus of a surface S is equal to $1 - (\chi(S)/2)$. Thus the genus of the boundary is $m + 1 - n$. Recall that two closed surfaces are homeomorphic if and only if they have the same genus.

1.8. LEMMA. *Two handlebodies are homeomorphic if and only if their boundaries have the same genus.*

We will prove this via a sequence of simpler lemmas.

1.9. LEMMA. *A handlebody of the form $H^{n,m}$, where $n > 1$, is always homeomorphic to a handleody of the form $H^{n-1,m-1}$.*

In order to prove this, we will need the following Lemma, which will be left as an exercise for the reader:

1.10. Lemma. *Let B and B' be closed (piecewise-linear) balls in a piecewise-linear 3-manifold. If $B \cap B'$ is a disk then $B \cup B$ is a ball.*

Proof of Lemma 1.9. We have a collection of balls $B_1, \ldots, B_n$ and a collection of disks $D_1, \ldots, D_m, D'_1, \ldots, D'_m$ embedded in their boundaries. Because the quotient of the balls is connected, there must be a pair of disks D_i, D'_i such that $D_i \subset B_j$ and $D'_i \subset B_k$ where $j \neq k$. Let ϕ_i be the gluing map from D_i to D'_i.

Let B be the quotient of $B_j \cup B_k$ by the identity $x \sim y$ if $\phi_i(x) = y$. The images of B_j and B_k are balls in B and the intersection $B_j \cap B_k$ is the image of D_i and of D'_i. Because the image is a disk, Lemma 1.10 implies that B is a closed 3-ball. The disks in the boundaries of B_j and B_k become disks in ∂B. The quotient of $B, B_1, \ldots, \hat{B}_j, \hat{B}_k, \ldots, B_n$ along the remaining disks is still homeomorphic to $H^{n,m}$, but it now has the form $H^{n-1,m-1}$. □

1.11. Corollary. *A handlebody of the form $H^{n,m}$ is always homeomorphic to a handlebody of the form $H^{1,m+1-n}$.*

Proof. Let $H^{n,m}$ be a handlebody and let n' be the smallest integer such that $H^{n,m}$ is homeomorphic to a handlebody constructed from n' balls. By Lemma 1.9, if $n' > 1$ then n' is not minimal. Thus n' must be 1. If $H^{n,m}$ is homeomorphic to a handlebody constructed from a single ball, then the boundaries of the two handlebodies must have the same genus, so this second handlebody must be of the form $H^{1,m+1-n}$. □

The conclusion of the proof will rely on a second result from piecewise-linear topology which, again, will be stated without proof.

1.12. Lemma. [**11**] *Let F be a compact surface, let $D_1, \ldots, D_k$ be a collection of pairwise-disjoint disks embedded in F and let $D'_1, \ldots, D'_k$ be a second collection of disjoint embedded disks. There is a homeomorphism $\phi : F \to F$, isotopic to the identity, such that ϕ sends each disk D_i onto the disk D'_i.*

This is the last ingredient in the proof of Lemma 1.8.

Proof of Lemma 1.8. By Corollary 1.11, we can reduce the problem to showing that two handlebodies of the form $H^{1,m}$ and $H^{1,m'}$ are homeomorphic if and only if their boundaries have the same genus. If the two handlebodies are homeomorphic, then their boundaries are homeomorphic and thus have the same genus. We will now prove the converse.

The genus of the boundary of a handlebody $H^{1,m}$ is m, so if the boundaries have the same genus then $m = m'$. Assume $H^{1,m}$ is constructed from a ball B and disks $D_1, \dots, D_{2m}$ and $H^{1,m'}$ is constructed from a ball B' and disks $D'_1, \dots, D'_{2m}$.

Because B and B' are balls, there is a homeomorphism $\psi : B \to B'$. Consider the collection of disks $\psi(D_1), \dots, \psi(D_{2m})$ in $\partial B'$. By Lemma 1.12, there is an automorphism $\phi : \partial B' \to \partial B'$ taking each disk $\psi(D_i)$ onto D'_i. Because B' is a ball, this map extends to a map $\hat{\phi} : B' \to B'$ of the whole ball.

The homeomorphism $\hat{\phi} \circ \psi : B \to B'$ takes each disk D_i to the disk D'_j. When we take quotients of the two balls, this map extends to a homeomorphism of the two handlebodies, completing the proof. □

Because the genus of the boundary defines the homeomorphism type of a handlebody, we are justified in defining the *genus* of a handlebody to be the genus of its boundary. We will now consider a simple way to prove that a given manifold is a handlebody:

1.13. Lemma. *Let M be a manifold with boundary and let $D_1, \dots, D_m$ be a collection of properly embedded disks. If N is the interior of M and $N \setminus \bigcup D_i$ is a collection of n open balls then M is a handlebody with genus $m + 1 - n$.*

Proof. Let $C_1, \dots, C_n$ be the components of $N \setminus \bigcup D_i$. For each $j \leq n$, let B_j be a closed ball and let ϕ_j be a homeomorphism from the interior of B_j onto C_j. Because the closure $\bar{N}$ is compact and each B_j is compact, we can extend each map ϕ_j to a map from B_j (the closure of the original set) to $\bar{N}$ by taking sequences of points which limit to the boundary.

Thus we have a collection of maps $\phi_1, \dots, \phi_n$ such that each $\phi_j : B_j \to \bar{N}$ is one-to-one in the interior of B_j. Let $\Phi : \bigcup B_j \to \bar{N}$ be the union of the maps $\phi_1, \dots, \phi_n$. This map is two-to-one along the disks $D_1, \dots, D_m$ and one-to-one exerywhere else.

The preimage of each D_i is a pair of disks in the boundaries of the balls $B_1, \dots, B_n$. Let $H^{n,m}$ be the quotient of $\bigcup B_j$ by the identity $x \sim y$ if $\Phi(x) = \Phi(y)$. This quotient is a handlebody and the map $\Phi : \bigcup B_j \to \bar{N}$ extends to a homeomorphism $\Phi' : H^{n,m} \to \bar{N}$. □

5. Heegaard Splittings

We now have a good understanding of the structure of handlebodies. The next step is to put these "simple" pieces together to form more complicated manifolds. In particular, we would like to take a pair of handlebodies H_1, H_2, of the same genus and glue them together by a homeomorphism $\phi : \partial H_1 \to \partial H_2$.

There is a natural inclusion of H_1 and of H_2 into the glued manifold $M = H_1 \cup_\phi H_2$. The images of the boundaries, ∂H_1 and ∂H_2 are surfaces in M which coincide completely. In other words, there is a surface Σ which is the boundary of H_1 in M and of H_2 in M. If we were to start with two handlebodies in M and a surface along which their boundaries coincide, we could infer that M was constructed by gluing together handlebodies. Thus in order to understand the ways a manifold M can be constructed by gluing handlebodies, we should examine the ways handlebodies can be embedded in M such that their boundaries coincide.

1.14. Definition. A *Heegaard splitting* for a 3-manifold M is an ordered triple (Σ, H_1, H_2) where Σ is a closed surface embedded in M and H_1 and H_2 are handlebodies embedded in M such that $\partial H_1 = \Sigma = \partial H_2 = H_1 \cap H_2$ and $H_1 \cup H_2 = M$. The surface Σ is called a *Heegaard surface*. The *Heegaard genus* of M is the smallest possible genus of a Heegaard splitting of M.

Because the boundaries of H_1 and H_2 are homeomorphic, by Lemma 1.8 H_1 and H_2 must by homeomorphic. Given a Heegaard surface Σ, the handlebodies are simply the closures of the components of the complement $M \setminus \Sigma$. Thus the Heegaard surface determines the two handlebodies, up to labeling.

1.15. Theorem. *Every compact, closed, connected, orientable 3-manifold allows a Heegaard splitting.*

Recall that thanks to Theorem 1.2, we can assume that a given 3-manifold homeomorphic to a simplicial complex. Thus our strategy for proving this theorem will be to construct a Heegaard splitting from a triangulation.

1.16. Lemma. *Let M be an orientable 3-manifold and let K be a (piecewise linear) graph embedded in M. If K has n vertices and m edges then the closure of a regular neighborhood of K is homeomorphic to a handlebody with genus $m + 1 - n$.*

A regular neighborhood of a graph is shown in Figure 5.

Proof. Let K be a graph embedded in a manifold M and let N be a regular neighborhood of K. Because K is piecewise linear, we can take K to be the image in M of a one dimensional subcomplex K' of a triangulation C for M. By Theorem 1.5, we can take N to be the

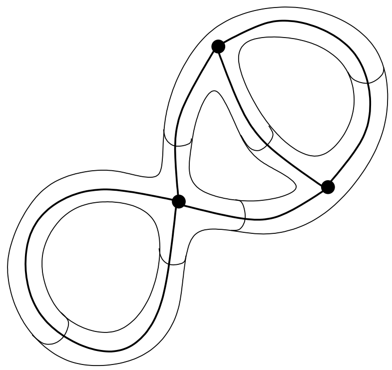


FIGURE 5. A regular neighborhood of a graph embedded in a 3-manifold is a handlebody.

image in N of the regular neighborhood of K' in the second barycentric subdivision C' of C.

Let n' and m' be the number of vertices and edges, respectively, in K'. For each edge e, let v be the vertex of C' in the center of e. Let σ be a 3-simplex of C which contains e. In the first barycentric subdivision of C, let T_1, T_2 be the triangles whose vertices consist of v, the center of a 2-face of σ which contains e and the center of σ, as shown in Figure 6.

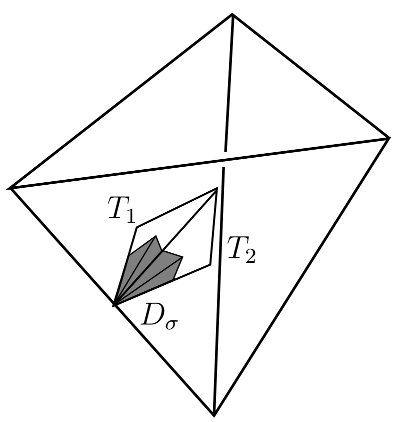


FIGURE 6. The triangles around an edge in the second barycentric subdivision of a 3-complex can be glued together to form a meridian disk for the edge.

In the second barycentric subdivision C' of C, there are two triangles in the subdivision of T_1 that contain the vertex v, and two such triangles in T_2. These four triangles form a disk D_σ. Each 3-simplex which contains e contains a similar disk, and the union of these disks is a disk D_e which is properly embedded in N and contains v.

Such a disk exists for each edge of K'. Each component of the complement of these disks contains a single vertex of K. These components have a very simple structure and the reader can check that the interior of each component is a ball. Thus Lemma 1.13 implies that N is a genus $m' + 1 - 1$ handlebody.

We would like to show that N has genus $m+1-n$ to see this, note that the graph K' is homeomorphis to K, so the $m - n = m' - n'$ (in fact, $m-n$ is the Euler characteristic of K. Thus $m+1-n = m'+1-n$, and both equal the genus of N. □

Proof of Theorem 1.15. Let M be a compact, closed, connected, orientable manifold. By Theorem 1.2, there is a simplicial complex K homeomorphic to M. Let N be a regular neighborhood of the 1-skeleton K^1 as in Figure 7. By Corollary 1.16, the closure, H_1, of N is a handlebody.

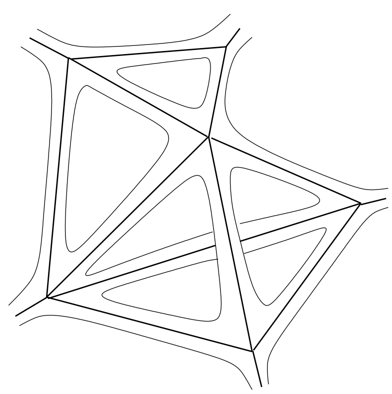


Figure 7. A triangulation of a manifold M suggests a Heegaard splitting of M. The Heegaard surface is the boundary of a regular neighborhood of the 1-skeleton.

Let H_2 be the complement of N. This set is closed because N is open. Let $\Sigma = \partial H_2 = \partial H_1$. To show that the triple (Σ, H_1, H_2) is a Heegaard splitting, all that remains is to show that H_2 is a handlebody.

The 1-skeleton of K is contained in N so the only simplices which intersect H_2 are triangles and tetrahedra. Let T be a triangle in K. Its boundary is in K^1 so $N \cap T$ is an annular neighborhood of the three edges around T. The complement of N is a disk in the interior of T, properly embedded in H_2.

Let B be a tetrahedron in K. Again, $B \cap N$ is an open neighborhood of the edges attached to B. The complement $B \setminus N$ is a ball. Thus the faces of K form a collection of disks properly embedded in H_2 such that their complement is a collection of open balls. By Lemma 1.13, H_2 is a handlebody and we're done. □

6. The 3-sphere

For each positive integer n, the unit n-sphere in $\mathbf{R}^{n+1}$ is the set of points of unit distance from the origin, i.e. $\{(x_1, \dots, x_n) : \sum |x_i| = 1\}$. An n-sphere is any n-dimensional manifold that is homeomorphic to the unit n-sphere.

We can visualize the 3-sphere without having to worry about four dimensions by "projecting" the 3-sphere into $\mathbf{R}^3$: The unit sphere in $\mathbf{R}^3$ is $S^3 = \{(x_1, x_2, x_3, x_4) : |x_1|+|x_2|+|x_3|+|x_4| = 1\}$. For every point $p \in S^3$ other than $(1, 0, 0, 0)$, there is a unique line ℓ in $\mathbf{R}^4$ that passes through both p and $(1, 0, 0, 0)$. The line ℓ intersects the 3-plane defined by the equation $x_1 = 0$ in a single point. Let $i : (S^3 \setminus (1, 0, 0, 0)) \to \mathbf{R^3}$ be the map which sends each p to the point where ℓ intersects $\mathbf{R}^3$.

For any sequence of points in S^3 that approaches $(1, 0, 0, 0)$, the images of the points in $\mathbf{R}^3$ become arbitrarily far from the origin. In other words, the sequence diverges to infinity. Thus we can think of S^3 as being $\mathbf{R}^3$ with a point added at infinity.

Let B_1 be the set of points in S^3 whose first coordinate is negative or zero, i.e. the set of points "below" the 3-plane defined by $x_1 = 0$. The map i sends this set to the unit ball in $\mathbf{R}^3$, so B_1 is a 3-ball. The set $B_1 \subset S^3$ consisting of points whose first coordinate is positive or zero is homeomorphic to B_0, so B_1 is a second ball. The intersection $S = B_1 \cap B_2$ is the set of points whose first coordinate is precisely zero. This set is a sphere and is the boundary of both B_1 and B_2. Thus by definition, the triple (S, B_1, B_2) is a genus zero Heegaard splitting for S^3. We have proved the following:

1.17. Lemma. *The Heegaard genus of S^3 is zero.*

The picture in $\mathbf{R}^3$ has B_1 as the unit ball and B_2 as everything else. The complement of the unit ball in $\mathbf{R}^3$ is not a ball. In fact, it is homeomorphic to $S^2 \times [1, \infty)$. However, adding in the point at infinity has the effect of capping of the open end of $S^2 \times [1, \infty)$ and turning it into a ball.

Next we will construct a genus one Heegaard splitting for S^3 from the genus zero splitting. Let H_1 be the set $\{(x_1, x_2, x_3) \in \mathbf{R}^3 : x_1^2 + x_2^2 + x_3^2 \leq 1 \text{ and } x_1^2 + x_2^2 \geq \frac{1}{2}\}$. In other words, we construct H_1 by taking the unit ball in $\mathbf{R}^3$ and cutting out the cylinder defined by the equation $x_1^2 + x_2^2 < \frac{1}{2}\}$.

1.18. Claim. *The set H_1 is a genus one handlebody.*

The proof of this will be left as an exercise for the reader. Note that the set of points with x_3 non-negative is a closed ball, as is the set of points with x_3 non-positive.

Let H_2 be the complement of the interior of H_1. This set is the union of B_2 with the cylinder $x_1^2 + x_2^2 \leq \frac{1}{2}\}$ that we cut out of B_1. The reader can check that $H_2 \setminus B_2$ is a ball, implying the following:

1.19. Claim. *The set H_2 is a genus one handlebody.*

The intersection $\Sigma = H_1 \cap H_2$ is a surface which is the boundary of each handlebody, so the triple (Σ, H_1, H_2) is a genus one Heegaard splitting for S^3.

This genus one Heegaard splitting was constructed from the genus zero splitting by attaching a "handle" to B_2, which was drilled out of B_1. The roles of the handlebodies could have been reversed, attaching a handle to B_1 as in Figure 8. The process can be repeated arbitrarily many times, implying the following:

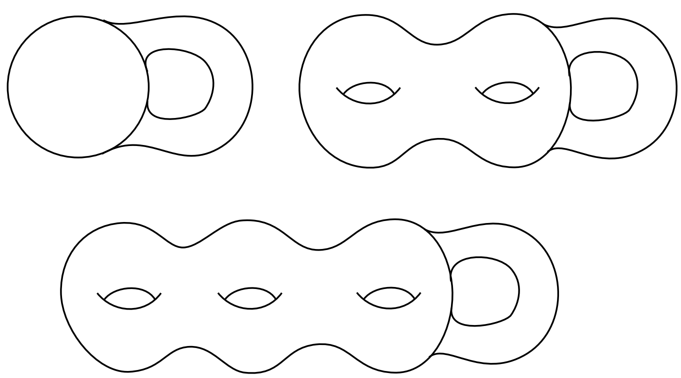


Figure 8. Higher genus Heegaard splittings for the 3-sphere can be constructed by gluing "trivial" handles on to lower genus splittings.

1.20. Lemma. *For every non-negative integer g, there is a genus g Heegaard splitting of S^3.*

The details of the proof are once again left to the reader.

7. Lens Spaces

Let H_1 be a genus one handlebody and let Σ be the boundary of H_1. From the definition of a handlebody, there is a simple closed curve $\mu \subset \Sigma$ which bounds a properly embedded disk $D \subset H_1$. Let $\lambda \subset \Sigma$ be any other simple closed curve, as in Figure 9.

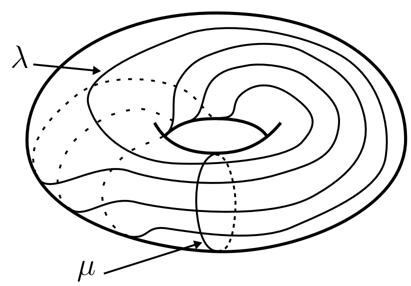


FIGURE 9. The loops μ and λ define a lens space.

Let $A \subset \Sigma$ be a regular neighborhood of λ. This set is an annulus. Let B_1 be the set $D \times [0, 1]$ where D is a disk. The set B_1 is homeomorphic to a ball. The subset of the boundary $\partial_v B_1 = \partial D \times [0, 1]$ is an annulus so there is a homeomorphism $\phi : \partial_v B_1 \to A$.

If we glue along ϕ, the result $M_0 = B_1 \cup_\phi H_1$ is a manifold with boundary. The boundary of M_0 consists of the disks $D \times \{0, 1\}$ and the annulur complement of A in ∂H_1. This is a sphere so there is a map $\psi : \partial M_0 \to B_2$ where B_2 is a ball. Let $M = M(\mu, \lambda)$ be the result of gluing along ψ.

Let Σ be the image in M of ∂H_1. The complement of H_1 consists of the balls B_1 and B_2. The map ψ sends the disks $D \times \{0\}$ and $D \times \{1\}$ in ∂B_1 to disks in ∂B_2. Thus $B_1 \cup B_2$ is the result of gluing two balls along two pairs of disks. By definition, this is a handlebody, H_2, so (Σ, H_1, H_2) is a genus-one Heegaard splitting of M.

1.21. DEFINITION. A *lens space* is a manifold which allows a genus-one Heegaard splitting.

We will see later that the above construction can produce any Lens space, i.e. any manifold with a genus-one Heegaard splitting. We also saw in Example 6 that there is a genus-one Heegaard splitting of S^3, so S^3 is also a lens space. (The manifolds S^3 and $S^1 \times S^2$ are sometimes excluded from the definition, but we will leave them in.) It is reasonable, then, to ask when the above construction will produce S^3.

1.22. LEMMA. *If $\mu \cap \lambda$ is a single point then $M(\lambda, \mu)$ is homeomorphic to S^3.*

PROOF. We will construct a map $\phi : L(\mu, \lambda) \to S^3$ in pieces, using the genus-one Heegaard splitting of S^3 defined above. Recall that we constructed a Heegaard splitting (Σ, H_1, H_2) of S^3 where H_1 is the complement of a cylinder in the unit ball.

Let $D_1' \subset S^3$ be the disk $\{(x, y, z) \in H_1 : x = 0, y > 0\}$. This is the intersection of H_1 with a half plane in $\mathbf{R^3}$. Let $\mu' = \partial D_1 \subset \Sigma$. Let $D_2' = D \setminus N = D \cap H_2$. and let $\lambda' = \partial D_2$. Notice that $\lambda' \cap \mu'$ is a single point in Σ'. Let v' be this point.

Let v be the point in $\lambda \cap \mu \subset \Sigma$ and define $\phi(v) = v'$. The set $e = \mu \setminus v$ is an open interval in Σ and the set $e' = \mu' \setminus v'$ is an open interval in Σ'. Both ends of e are on v and both ends of e' are on v' so the map ϕ can be extended to e so that it sends all of μ to μ'. The same argument applied to λ implies that ϕ can be extended to send $\mu \cup \lambda$ to $\mu' \cup \lambda'$.

There are three disks in M to which we can apply the same argument, one dimension higher. First, $\Sigma \setminus (\lambda \cup \mu)$ is an open disk, as is $\Sigma' \setminus (\lambda' \cup \mu')$ so ϕ can be extended to send Σ to Σ'. Next, there is a disk in $D_1 \subset H_1$ bounded by μ which can be sent to the disk $D_1' \subset H_1'$ bounded by μ', and a disk in $D_2 \subset H_2$ bounded by λ which can be sent to the disk D_2' bounded by λ'.

All that is left now is the complement in M of $\Sigma \cup D_1 \cup D_2$. This consists of two open balls, one in H_1 and one in H_2. The complement in S^3 of $\Sigma' \cup D_1' \cup D_2'$ is also a pair of open balls, so once again ϕ can be extended to send all of M to S^3. It is left to the reader to check that this map is a homeomorphism. $\square$

In fact, we can do slightly better:

1.23. LEMMA. *The manifold $M(\mu, \lambda)$ is homeomorphic to S^3 if and only if λ can be isotoped to intersect μ in a single point.*

The proof of this lemma will be left for later.

8. The 3-torus

The 3-dimensional analogue of the 2-dimensional torus, $S^1 \times S^1$, is the 3-torus, $T^3 = S^1 \times S^1 \times S^1$. If we parametrize S^1 as the interval $[0, 1]$ with its endpoints identified, then T^3 is the result of identifying opposite faces of the cube $[0, 1] \times [0, 1] \times [0, 1]$.

Let C be a cube and let $\phi : C \to T^3$ be the map induced by the inclusion $[0, 1] \to S^1$. The eight vertices of the cube map to a single vertex $v \in T^3$. Parallel edges in C map to the same edge in T^3 so the

12 edges of the cube map to three edges in T^3. The three edges and the vertex form a graph $K \subset T^3$.

Let N be a regular neighborhood of K in T^3. Let H_2 be the complement of N. The pre-image $B = \phi^{-1}(H_2)$ is a ball in the cube, shown in Figure 10. The faces of the cube intersect H_2 in disks which chop H_2 into this ball B. Thus by Lemma 1.13, H_2 is a handlebody.

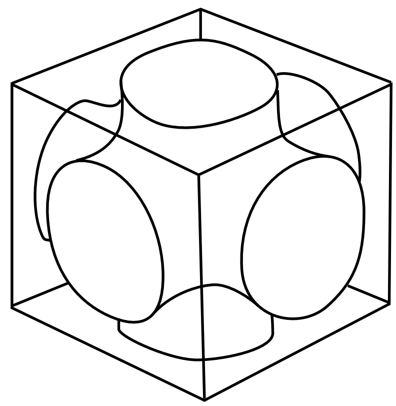


FIGURE 10. The 3-torus, $T^3 = S^1 \times S^1 \times S^1$ can be constructed by gluing together opposite faces of the cube.

Because N is a regular neighborhood of a graph, its closure H_1 is a handlebody. If Σ is the surface $H_1 \cap H_2$ then (Σ, H_1, H_2) is a Heegaard splitting for T^3. The graph K has three edges and one vertex so the splitting has genus 3.

9. Notions of Equivalence

Two Heegaard splittings, (Σ, H_1, H_2) and (Σ', H_1', H_2'), are equal if $\Sigma = \Sigma'$, $H_1 = H_1'$ and $H_2 = H_2'$. In other words, corresponding elements of the triples are equal as subsets of M. In practice, this is an impractical relationship to study because perturbing Σ' slightly will make it no longer equal to Σ.

Since we are working with topological objects, we need a definition of equivalence for Heegaard splittings which takes into account the possibility of homeomorphisms and isotopies. We will introduce two such definitions.

1.24. DEFINITION. Heegaard splittings (Σ, H_1, H_2) and (Σ', H_1', H_2') for M are *isotopic* if there is a map $i : M \times [0, 1] \to M$ such that $i|_{M\times\{0\}}$ is the identity, $i|_{M\times\{t\}}$ is a homeomorphism for each t and $i|_{M\times\{1\}}$ sends Σ to Σ', H_1 to H_1' and H_2 to H_2'.

An isotopy of the surface Σ is a map $i : \Sigma \times [0,1] \to M$ such that $i|_{\Sigma\times\{0\}}$ is the identity and $i|_{\Sigma\times\{t\}}$ is one-to-one for each t. If $i|_{\Sigma\times\{1\}}$ sends Σ to Σ' then we say Σ and Σ' are isotopic.

Isotopies of compact submanifolds in the interior of a manifold extend to the entire manifold (See Theorem 1.3 in Chapter 8 of [**6**]), so the surface Σ is isotopic to Σ' if and only if (Σ, H_1, H_2) is (ambient) isotopic to either (Σ', H_1', H_2') or (Σ', H_2', H_1').

1.25. Definition. Two Heegaard splittings are *homeomorphic* if there is a homeomorphism $h : M \to M$ that takes Σ to Σ', H_1 to H_1' and H_2 to H_2'.

If two splittings are isotopic then they are homeomorphic. (The homeomorphism is given by $i|_{M\times\{1\}}$.) However, if they are homeomorphic, they are not necessarily isotopic since the map h may not be isotopic to the identity.

One problem we will consider in upcoming chapters is to classify Heegaard splittings up to isotopy or homeomorphism. In other words, the question is: given a manifold, what are all of the isotopy classes and/or homeomorphism classes of Heegaard splittings, and how are they related? In the following sections we will begin to develop techniques that can be used to answer this question.

10. Exercises

1. Show that if K is a simplicial complex then $|K|$ is closed.

2. Prove that if K is a simplicial complex with finitely many vertices then $|K|$ is compact.

3. Prove Lemma 1.6.

4. Prove Lemma 1.10.

5. What manifold results from gluing two solid tori along their boundary by the identity map? What about higher genus handlbodies?

6. Show that if F is a surface with boundary then $F \times [0,1]$ is a handlebody.

7. Given a compact, connected, closed, orientable surface F, find a Heegaard splitting for $F \times S^1$. (Hint: Generalize the construction of the Heegaard splitting for T^3.)

CHAPTER 2

Handlebodies

1. Systems of Disks

We saw in the previous chapter that what makes a manifold H a handlebody is the existence of a collection of properly embedded disks that cut H into balls. In this chapter, we will focus on the different possible collections of disks with this property. A properly embedded disk $D \subset H$ is *essential* if its boundary does not bound a disk in ∂H.

2.1. Definition. A collection $\{D_1, \ldots, D_m\}$ of properly embedded, essential disks is called a *system of disks* for H if the complement of a regular neighborhood of $\bigcup D_i$ is a collection of balls.

2.2. Lemma. *Given a handlebody, H, there is a system of disks for H.*

Proof. Let $B_1, \ldots, B_n$ be a collection of balls and let $E_1, \ldots, E_m$, $E'_1, \ldots, E'_m$ be disks in their boundaries such that H is the result of gluing the balls together along pairs of disks. There is a natural inclusion map ϕ from the disjoint union $\sqcup B_j$ into H induced by the gluing.

For each pair of disks $E_i,$, E'_i, the image $\phi(E_i)$ coincides exactly with $\phi(E'_i)$. For each i, let $D_i \subset H$ be the disk $\phi(E_i) = \phi(E'_i)$. Let N be a regular neighborhood of the disks $\bigcup D_i$. The preimage in ϕ of N is a regular neighborhood of the disks $\bigcup(E_i \cup E'_i)$.

Given a ball B and a disk D in its boundary, the complement in B of a regular (open) neighborhood of D is a closed ball. By induction on the disks $E_i, \ldots, E_k$ and $E'_i, \ldots, E'_k$, this implies that the complement of the preimage $\phi^{-1}(N)$ in $\sqcup B_j$ is a collection of closed balls. The map ϕ is one-to-one in the complement of N so the image of these closed balls is also a collection of closed balls, which are precisely the complement in H of N.

We see that the disks $D_1, \ldots, D_n$ have the property that the complement of a regular neighborhood is a collection of balls. Each disk is properly embedded, so it only remains to show that each disks is essential. Let D_i be a disk such that the loop ∂D_i bounds a disk F in the boundary of H. If F contains the boundary of a second disk in the system, we can replace D_i with this disk and so on until we have disks D_i and F such that F is disjoint from the rest of the disks in the system.

The disk F is in the boundary of a ball B_j in the complement of N. Because the interior of F is disjoint from the system of disks, there is exactly one disk E_i of $\mathbf{E}$ in the boundary of B_j, and this is associated to D_i. Let B_k be the ball containing the second disk E_i' associated to D_i. Removing D_i from the system of disks has the effect on $B_i, \ldots, B_m$ of gluing B_j and B_k together along E_i, E_i'. This results in a new ball, so the complement of the new set of disks is still a collection of balls. Thus we can remove any inessential disks from the collection until the remaining disks are all essential, and form a system of disks. □

In the proof of Lemma 2.2, it was necessary to eliminate disks that were not essential in the handlebody. There is a similar situation which, while it is allowed in the definition of a system of disks, it is often useful to exclude. Two essential, properly embedded disks D, D' are called *parallel* if the complement of a regular neighborhood of $D \cup D'$ in the habdlebody H has a component which is a ball. Equivalently, the complement in ∂H of the boundaries of the disks contains an annulus component.

If two disks in a system are parallel then some ball B_j has exactly two disks of $\mathbf{E}$ in its boundary. The disk D' can be isotoped across B_j onto D. Because the two disks are topologically the same, it is often unnecessary to keep both of them. In particular, removing D' from the system creates a smaller system of disks so we can always assume that we have a system of disks in which no two disks are parallel.

2.3. Definition. A system of disks is *maximal* if each component of $H \setminus \bigcup D_i$ has three disks in its boundary.

A maximal system of disks, as in Figure 1, is sometimes called a *pair-of-pants decomposition* because the boundary of each ball looks like a pair of pants (a 3-punctured sphere.)

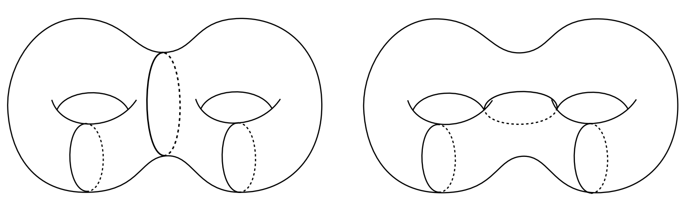


Figure 1. Two different maximal systems of disks for a genus two handlebody.

2.4. Lemma. *Let H be a handlebody and $\mathbf{D} = \{D_1, \ldots, D_m\}$ a system of disks for H. If H has genus greater than 1 and no two disks in $\mathbf{D}$ are parallel in H then there is a maximal system of disks for H containing $\mathbf{D}$.*

Proof. Because H is not a ball, the system of disks is not empty. Let N be a regular (open) neighborhood of the disks $\mathbf{D} = \{D_1, \ldots, D_n\}$. Let $\bar{N}$ be the closure of N in H. By definition, the complement $H \setminus N$ is a collection of balls $B_1, \ldots, B_m$. The intersection $\mathbf{E} = (H \setminus N) \cap \bar{N}$ is a collection of disks in the boundaries of the balls such that each disk D_i corresponds to a pair of disks E_i, E'_i in $\mathbf{E}$. In fact, the disk D_i is parallel to E_i and E'_i.

For each ball B_j, let k be the number of disks of $\mathbf{E}$ contained in ∂B_k. If $k = 2$ then the complement in ∂B_j of $\mathbf{E}$ is an annulus and E_i and E_h are parallel in H. If the disks E_i, E_h in ∂B_j come from different disks D_i, D_h in $\mathbf{D}$ then D_i and D_h are parallel in H. If E_i, E_h come from the same disk D_i then the complement in ∂H of $\partial E_i \cup \partial E_h \cup \partial D_i$ is a collection of annuli so H is a solid torus. We assumed H has genus at least two and no two disks of $\mathbf{D}$ are parallel so for each ball B_j, $k \geq 3$.

If k is strictly greater than three, the complement in ∂B_j of $\mathbf{E}$ is a sphere with four or more punctures. There is a simple closed curve l_{n+1} in ∂B_j that separates two of the disks from the rest. This loop is not parallel to the bounday of any disk in $\mathbf{E}$ because there are at least two disks on either side of l_{n+1} in ∂B_j. By the Jordan curve theorem, the loop l_{n+1} bounds a disk in the sphere. Pushing this disk into the interior of the ball B_j produces a disk D_{n+1} in the handlebody whose boundary is l_{n+1}. Add D_{n+1} into the system of disks $\mathbf{D}$.

Adding D_{n+1} into the system splitting B_j into two balls, one of which has exactly three disks in its boundary. The second has $k - 1$ disks in its boundary. Continue adding disks in this fashion until each ball has exactly three disks of $\mathbf{E}$ in its boundary. The resulting system of disks contains the original system and is maximal. □

The term "maximal" refers to the fact that no disk can be added to the system without violating the condition that no two disks are parallel.

2.5. Lemma. *If $\mathbf{D}$ is a maximal collection of disks for H and D' is a properly embedded disk which is disjoint from each disk in $\mathbf{D}$ then D' is parallel to a disk in $\mathbf{D}$.*

Proof. Let N, $B_1, \ldots, B_m$ and $\mathbf{E}$ be as in the proof of Lemma 2.4. Because D' is disjoint from the disks in $\mathbf{D}$, it is properly embedded in a ball B_i in the complement of N. The boundary of B_i contains exactly three disks of $\mathbf{E}$.

The boundary of D' is an essential loop ℓ in the complement $\partial B \setminus E$. By the Jordan curve theorem, ℓ separates ∂B into two disks. Because ℓ is disjoint from the disks $E \cap \partial B$ and does not bound a disk in their complement, one of the disks in $\partial B \setminus \ell$ contains one of the disks $E \subset \partial B \setminus E$ and the other contains two of the disks. The loop ℓ is parallel to the boundary of E so D' is parallel to E, and therefore parallel to the disk of $\mathbf{D}$ associated with E. □

In contrast to a maximal collection of disks, in which no disks can be added to the system, it will also be useful to consider systems of disks in which none of the disks can be removed. In the case of a miniml system of disks, removing a disk violates the condition that the complement of the disks is a collection of balls.

2.6. Definition. A collection of disks is *minimal* if its complement is connected.

2.7. Lemma. *Given a handlebody H and a system $\mathbf{D} = \{D_1, \dots, D_m\}$ of disks for H, there is a subset of the disks $D_1, \dots, D_m$ which form a minimal system of disks for H.*

Proof. Let N be a regular neighborhood of the system of disks. If the complement of N is connected then the system of disks is minimal. Otherwise, because H is connected, there is a disk D_j that connects two of the components of $H \setminus N$.

Each component is a ball so the union of the regular neighborhood of D_j and the two components it connects form a ball. Thus the complement in H of the disks other than D_j is a collection of balls. In other words, the disks $\mathbf{D} \setminus \{D_j\}$ form a system of disks for H. If this new system is not minimal then we can remove a second disk and form a third system. The process continues until we are left with a minimal system of disks. □

2.8. Proposition. *Let $\mathbf{D} = \{D_i\}$ be a system of disks for a handlebody H. Then the following are equivalent:*

(1) The system $\mathbf{D}$ is minimal.
(2) No proper subset of $\mathbf{D}$ is a system for H.
(3) The number of disks in $\mathbf{D}$ is equal to the genus of H.

Proof. We saw in the proof of Lemma 2.7 that if a system of disks is not minimal then there is a proper subset that is also a system of disks. The contrapositive of this statement is that (2) implies (1).

Recall that we defined the genus of a handlebody to be the genus of its boundary and calculated this genus to be $g = m + 1 - n$ where n is the number of balls and m is the number of disks used to construct H. The system of disks is minimal if and only if $n = 1$ so $g = m+1-1 = m$, implying (1) is equivalent to (3).

Finally, let $\mathbf{D}$ be a minimal system of disks for H with g disks. Thus the genus of H is g. Assume for contradiction there is a proper subset of the disks that is also a system of disks $\mathbf{D}'$ for H. Let $h < g$ be the number of disks in $\mathbf{D}'$. The complement of $\mathbf{D}'$ is connected, so it must be a single ball. This implies that the genus of H is h, a contradiction. Thus (1) implies (2). □

Given a system of disks for each handlebody in a Heegaard splitting (Σ, H_1, H_2), the boundaries of the disks lie in the Heegaard surface Σ. The ways in which these loops intersect determines the topology of the Heegaard splitting and of the ambient manifold. In Section 7, a pair of loops in a torus were used to construct a lens space. Each loop became the boundary of a disk, forming a system of disks for a genus one handlebody of a Heegaard splitting. Similarly, the boundaries of a system of disks for a higher genus handlebody completely determine its topology.

2.9. Lemma. *Let H and H' be handlebodies and let $\{D_1, \dots, D_m\}$ and $\{D'_1, \dots, D'_m\}$ be systems of disks for H and H', respectively. Assume there is homeomorphism $\phi : \partial H \to \partial H'$ such that for each i, $\phi(\partial D_i) = \partial D'_i$. Then there is a homeomorphism $\psi : H \to H'$ such that $\psi|_{\partial H} = \phi$.*

Proof. We will define the map ψ in pieces. On the boundary of H, define $\psi(x) = \phi(x)$ for each $x \in \partial H$. Each disk D_i has its boundary in ∂H so ψ sends ∂D into $\partial H'$. In fact, because ψ agrees with ϕ, it sends ∂D_i onto $\partial D'_i$.

Any map from the boundary of one disk to the boundary of a second can be extended to the interiors of the disks. Thus we can extend the map ψ to send D_i onto D'_i for each i.

All that remains of H is the complement of the system of disks. This is a collection of balls. The map ψ sends the boundary of each ball to the boundary of a ball in H. Once again, any map from the boundary of a ball to the boundary of a second ball can be extended to their interiors, so ψ can be extended to all of H. □

2. Sliding Disks

Let $\mathbf{D} = \{D_1, D_2, \ldots, D_m\}$ be a system of disks for H. Let α be an embedded arc in ∂H with one endpoint on ∂D_1, and one endpoint on ∂D_2. Assume the interior of α is disjoint from ∂D_i for each i. Let N be the closure of a regular neighborhood in H of $D_1 \cup D_2 \cup \alpha$.

The set N is a closed ball which intersects ∂H in a three punctured sphere. The complement in ∂N of this three punctured sphere consists of three disks; one is parallel to D_1, one parallel to D_2 and the third is a new disk we will call D'. (See Figure 2.) We can think of D' as the result of sliding a piece of D_1 along α and then across D_2. Write $D' = D_1 *_\alpha D_2$.

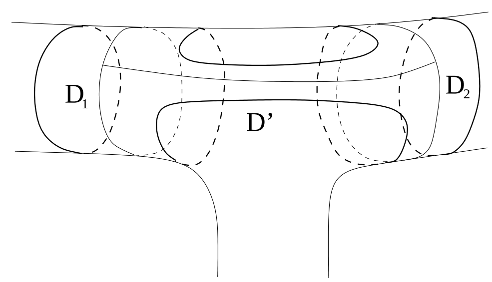


FIGURE 2. Disks D_1, D_2 and an arc α suggest a new disk D'.

2.10. LEMMA. *If* $\mathbf{D} = \{D_1, \ldots, D_m\}$ *is a system of disks for* H *and* $D' = D_i *_\alpha D_j$ $(i \neq j)$ *then* $(\mathbf{D} \cup \{D'\}) \setminus \{D_i\}$ *is a system of disks for* H.

PROOF. The neighborhood N of $D_i \cup \alpha \cup D_j$ is a ball so adding D' to the collection yields a system of disks. Now D_j separates N from some other ball in the complement of $\mathbf{D}$ so removing D_j yields a system of disks for H. □

In other words, if we add the disk D' to and remove D_j, the new set of disks is a system for H. This preserves the number of disks in the system (and the genus of H) so by Proposition 2.8, we also have the following:

2.11. COROLLARY. *If* $\mathbf{D}$ *is a minimal system of disks then adding* D' *and removing* D_i *yields a new minimal system of disks.*

Two systems of disks are *isotopic* if there is an isotopy of H (not fixing the boundary) which takes one system of disks onto the other. We will call any system of disks that is isotopic to $(\mathbf{D} \cup \{D'\}) \setminus \{D_i\}$, a *disk slide* of $\mathbf{D}$.

Systems $\mathbf{D}$ and $\mathbf{D'}$ will be called *slide equivalent* if there is a sequence of systems $\mathbf{D}_1, \ldots, \mathbf{D}_l$ such that $\mathbf{D} = \mathbf{D_1}$, $\mathbf{D}_l = \mathbf{D'}$ and for each i, $\mathbf{D}_{i+1}$ is a disk slide of $\mathbf{D}_i$. It is left as an exercise for the reader to check that this relation is an equivalence relation.

Every handlebody of genus greater than one allows an infinite number of non-isotopic systems of disks. In 1933, Reidemeister [**10**] and Singer [**17**] showed that any two minimal systems of disks are in fact slide equivalent. The rest of this section will be devoted to proving this.

2.12. Lemma. *If two minimal systems of disks are disjoint then they are slide equivalent.*

Proof. Let $\mathbf{D} = \{D_1, \ldots, D_g\}$ and $\mathbf{D'} = \{D'_1, \ldots, D'_g\}$ be minimal systems of disks for a handlebody H. Because the system $\mathbf{D}$ is minimal, the complement of a neighborhood of $\mathbf{D}$ is a ball B. For each disk D_i, the closure of the regular neighborhood of D_i intersects B in two disks E_i, E'_i in the boundary of B. Each disk D'_i is properly embedded in B such that its boundary is disjoint from each E_i and each E'_i.

The disks of $\mathbf{D'}$ cut B into $g+1$ components. If one of these components does not contain some E_i or E'_i then a disk of $\mathbf{D'}$ is boundary parallel. Because $\mathbf{D'}$ is a system of disks, this is impossible so each component contains at least one of the disks coming from $\mathbf{D}$. Because there are $m+1$ components and $2m$ disks in the boundary, there is at least one component, B', with exactly one E_i or E'_i in its boundary, say E_1.

If B' is cut off by a single disk, D'_k of $\mathbf{D'}$ then D'_k is isotopic to D_1. Otherwise, without loss of generality, assume $D'_1, \ldots, D'_k \subset \partial B'$ are the disks of $\mathbf{D'}$ which cut off B'. Because $\partial B' \setminus (D_1 \cup D'_1 \cup \cdots \cup D'_k)$ is connected there is an arc α from $\partial D'_1$ to $\partial D'_k$ that is disjoint from the other disks.

Replace D'_1 with the disk $D'_1 *_\beta D'_k$. This new disk and the disks $D'_2, \ldots, D'_{k-1}$ now cut off a component contining only D_1. Continuing in this fashion, we see that by a sequence of slides, we can replace D'_1 with a disk cutting off a component of B containing only D_1. This new disk is isotopic to D_1 so $\mathbf{D'}$ is slide equivalent to a system of disks consisting of D_1 and the disks $D'_2, \ldots, D'_m$.

The disks $D'_2, \ldots, D'_m$ cut B into m components and there are $2(m-1)$ disks $E_2, \ldots, E_m$, $E'_2, \ldots, E'_m$ in the boundary. Thus there is a collection of disks cutting off a single disk, say E_2 in the boundary.

The component containing E_2 may also contain E_1 or E'_1. However, D_1 is part of the new system of disks, $D_1, D'_2, \ldots, D'_k$ so we can slide any of the disks $D'_2, \ldots, D'_k$ over E_1 or E'_1. Thus there is a sequence of disk slides which replace D_2' with a disk isotopic to D_2. By repeating the process for the disks $D_3', \dots, D_k'$ and so, one finds a sequence of disk slides starting at $\mathbf{D}'$ and ending at $\mathbf{D}$, so the two systems are slide equivalent. □

Thus to show that any two disk systems are slide equivalent, we need only show that any two systems can be made disjoint by disk slides. This is the method used in the following proof.

2.13. Theorem. *Any two minimal systems of disks for a handlebody H are slide equivalent.*

Proof. Let $\mathbf{D} = \{D_1, \dots, D_m\}$ and $\mathbf{D}' = \{D_1', \dots, D_m'\}$ be minimal systems of disks. Assume the disks are transverse, i.e. that $D_i \cap D_j'$ is a (possibly empty) collection of properly embedded arcs and simple closed curves for any i and j.

If a component of $D_i \cap D_j'$ is a closed loop then this loop bounds a disk in D_j'. An *innermost* loop in D_j' is a loop ℓ in $D_i \cap D_j'$ such that the interior of the disk bounded by ℓ is disjoint from the disks of $\mathbf{D}$. If D_j' intersects a disk of $\mathbf{D}$ in a closed loop then D_j' contains an innermost loop $\ell \subset (D_i \cap D_j')$ for some i.

Let ℓ be an innermost loop in D_i' and let $E \subset D_i'$ be the disk bounded by ℓ. The complement in H of $\mathbf{D}$ is a ball B and ℓ is a simple closed curve in the bondary of B. The disk E is properly embedded in B so the complement $B \setminus E$ is a pair of balls. One these balls has a boundary consisting of E and a disk in D_i. Isotoping E across this ball into D_i induces an isotopy of D_j' that removes the loop of intersection ℓ. By continuing this process, we can remove all the simple closed curves of intersection.

Assume $D_i \cap D_j'$ consists of properly embedded arcs for each i and j. Define $I(\mathbf{D}, \mathbf{D}')$ to be the number of arcs of intersection over all the disks in $\mathbf{D}$ and $\mathbf{D}'$. By Lemma 2.12, we need only show that $\mathbf{D}$ is slide equivalent to a system of disks $\mathbf{D}''$ such that $I(\mathbf{D}', \mathbf{D}'') = 0$.

Let $\mathbf{D}'' = \{D_1'', \dots, D_m''\}$ be a system of disks that is slide equivalent to $\mathbf{D}$ and such that $I(\mathbf{D}', \mathbf{D}'')$ is minimal over all systems slide equivalent to $\mathbf{D}$. We will show that the minimality assumption implies $I(\mathbf{D}', \mathbf{D}'') = 0$.

If $I(\mathbf{D}', \mathbf{D}'') \neq 0$ then there is a disk D_j' such that $D_j' \cap \bigcup D_i''$ is not empty. Each arc of $D_j' \cap D_i''$ (for some i) cuts D_j' into two disks. An arc α is *outermost* if the interior of one of these disks is disjoint from $\mathbf{D}''$. If $D_j' \cap \bigcup D_i''$ contains an arc then it contains an outermost arc α. Let $E \subset D_j'$ be the subdisk disjoint from $\mathbf{D}''$ and let D_k'' be the disk such that $\alpha \subset D_j' \cap D_k''$.

Because the system $\mathbf{D}''$ is minimal, the complement in H of a regular neighborhood of $\mathbf{D}$ is a closed ball B. Each disk D_i'' is parallel to two closed disks F_i. F_i' in ∂B. The disk $E \cap B$ is a properly embedded disk whose boundary consists an arc in a disk F_k and an arc disjoint from all the disks F_i, F_i', as in Figure 3.

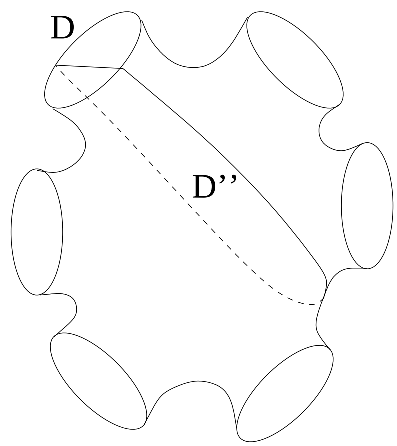


FIGURE 3. The disks $D_1, \dots, D_m$ cut H into a ball B, in which D'' is properly embedded.

Let N be a regular neighborhood in B of $E \cup F_i$. The set $N \setminus \partial B$ consists of two disks, E_1 and E_2. Every arc of intersection in $E_i \cap \mathbf{D}'$ will be an arc parallel to $D_k'' \cap \mathbf{D}'$. Since there is no arc of intersection parallel to α, the number of arcs in E_1 is strictly less than the number in D_k'', as is the number of arcs in E_2. We will reduce the number $I(\mathbf{D}', \mathbf{D}'')$ by finding a sequence of disk slides that replaces D_k'' with either E_1 or E_2.

The complement $B \setminus E_1 \cup E_2$ consists of three balls. One ball contains F_k in its boundary and one ball contains F_k'. Let $B'' \subset B$ be complement of this second ball. Either E_1 or E_2 is in the boundary of B''. (The other disk is in its interior.) Without loss of generality, assume E_1 is in the boundary of B''.

The boundary of B'' contains E_1, F_i and a collection of disks which are parallel to the disks $D_1'', \dots, D_m''$. The complement $\partial B'' \setminus (E_1 \cup D \cup F_i \cup \cdots \cup F_k)$ is connected so there is an arc β_1 from ∂D to a disk F_i in $\partial B''$. Define $G_1 = D *_{\beta_1} F_i$.

The boundary of G_1 separates $\partial B''$ into a component containing F_k and F_i and a component containing the rest of the disks. Let β_2 be an arc in $\partial B''$ from ∂G_1 to a second disk F_j in B'' and let $G_2 = G_1 *_{\beta_2} F_2$. This disk G_2 separates D, F_i and F_j from the rest of the disks. Continuing in this fashion, we eventually construct a disk G_k which separates E_1 from the rest of the disks. Thus G_k cuts off a ball containing itself and E_1. This implies that G_k is isotopic to E_1.

Recall that the disk F_k is parallel to D''_k in H. Thus there is a sequence of edge slides which replaces the disk D''_k with the disk E_1. Since E_1 intersects the disks of $\mathbf{D}'$ in fewer arcs, this sequence of edge slides reduces $I(\mathbf{D}_l, \mathbf{D}')$.

We have shown that if $I(\mathbf{D}', \mathbf{D}'') > 0$ then it can be reduced. Since we assumed the value was minimized, we have $I(\mathbf{D}, \mathbf{D}'') = 0$ so by Lemma 2.12, $\mathbf{D}''$ is slide equivalent to $\mathbf{D}'$. Since we assumed $\mathbf{D}''$ is slide equivalent to $\mathbf{D}$, this implies that $\mathbf{D}$ is slide equivalent to $\mathbf{D}'$. □

3. Spines

The systems of disks we studied in the last two sections came from constructing a handlebody by gluing together balls. Theorem 1.16 suggests a second way to construct a handlebod, as a regular neighborhood of a graph. In this case the induced structure is a graph embedded in the handlebody.

2.14. Definition. A *spine* of a handlebody H is a (piecewise linear) graph K embedded in H so that $H \setminus K$ is homeomorphic to $\partial H \times (0, 1]$.

If K is a graph embedded in a manifold and H is the closure of a regular neighborhood of K then K is a spine of H. This implies the following:

2.15. Lemma. *If H is a handlebody then there is a spine for H.*

The proof is left as an exercise for the reader. A spine of a handlebody in a Heegaard splitting is a graph in the ambient manifold and the embedding of this graph completely determines the isotopy class of the Heegaard splitting.

2.16. Lemma. *Let (Σ, H_1, H_2) and (Σ', H'_1, H'_2) be Heegaard splittings of a manifold M. Let K and K' be spines of H_1 and H'_1, respectively. If K is isotopic to K' then Σ and Σ' are isotopic.*

Before proving this Lemma, we need the following result, whose proof will be left as an exercise.

2.17. Lemma. *If K is a graph in a handlebody H and N is a regular neighborhood of K then the complement $H \setminus N$ is homeomorpic to $\partial H \times [0, 1]$ if and only if K is a spine of H.*

Proof of Lemma 2.16. Isotope K onto K'. Assume M has been triangulated such that the graph $K = K'$ and the surfaces Σ and Σ' are subcomplexes of the triangulation. Let N be a regular neighborhood of K with respect to this triangulation.

The neighborhood N is a regular neighborhood of K in H_1 so by Lemma 2.17, $H_1 \setminus N$ is homeomorphic to $\Sigma \times [0, 1]$. This parameterization defines an isotopy of Σ onto $\partial \bar{N} = \Sigma \times \{0\}$. Likewise, N is a regular neighborhood of K' in H_2 so Σ' is isotopic to $\partial \bar{N}$. Thus Σ is isotopic to Σ'. □

As we saw in Section 1, the construction of a handlebody H as a regular neighborhood of a graph G implies a system of disks for H. The graph G, which becomes a spine for H, intersects the disks in a very simple way. We will see that just as the disks were constructed from the graph, a graph of this type can also be constructed from the system of disks.

2.18. Definition. A spine K is *dual* to a system of disks $\mathbf{D}$ if each edge of K intersects a single disk of $\mathbf{D}$ exactly once, each disk intersects exactly one edge and each ball of $H \setminus \mathbf{D}$ contains exactly one vertex of V.

A system of disks $\mathbf{D}$ dual to a spine G will be minimal if and only if G has a single vertex. Similarly, $\mathbf{D}$ will be maximal if and only if every vertex of G is trivalent.

A properly embedded disk in H that intersects an edge e of G in a single point is often called a *meridian disk* for e. This duality between spines and systems of disks is exactly the relationship between a graph G and the induced system of disks for a regular neighborhood of G. The key is that the system of disks is (up to isotopy) unique.

2.19. Proposition. *Given a spine K of a handlebody H, there is a system of disks dual to K. If $\mathbf{D}$ and $\mathbf{D}'$ are systems of disks dual to K then the systems of disks are isotopic.*

Proof. Let H be a handlebody, K a spine for H and let $\Sigma = \partial H$. Let N be a regular neighborhood of K and let $\bar{N}$ be its closure. By Lemma 2.17, the complement in H of N is homeomorphic to $\Sigma \times [0, 1]$ such that $\Sigma \times \{1\} = \partial \bar{N}$.

As we saw in the proof of Lemma 1.16, for each edge e_i of G, there is a disk D_i properly embedded in N such that $e_i \cap D_i$ is a single point (a meridian disk for D_i). The boundary of D_i is a loop in $\Sigma \times \{1\}$. Let D'_i be the disk $D_i \cup (\partial D_i \times [0, 1])$, where $\partial D_i \times [0, 1]$ is an annulus in $H \setminus N$.

Each disk D_i intersects exactly one edge of K, exactly once. The graph is disjoint from $H \setminus N$ so each D'_i intersects one edge of K exactly once, and $D'_1, \ldots, D'_m$ is a system of disks dual to K. For uniqueness, we need the following Lemma:

2.20. LEMMA. *Let A be an annulus embedded in $\Sigma \times [0, 1]$ so that one boundary component is in $\Sigma \times \{0\}$ and the other boundary is in $\Sigma \times \{1\}$. Then A can be isotoped to intersect each level surface $\Sigma \times \{x\}$ in a single essentle loop.*

We will not prove this lemma here. To complete the proof of Lemma 2.19, let D and D' be properly embedded disks in H such that $D \cap K$ is a single point in an edge e of K and $D' \cap K$ is a different point in the same edge.

Let N' be a regular neighborhood of K. The intersection $D \cap N'$ and $D' \cap N'$ are disks in N'. These disks separate a ball from N' so they are parallel. Isotope D' so that $D \cap N' = D' \cap N'$. Because N' is a regular neighborhood of K, $H \setminus N'$ is homeomorphic to $H \setminus N = \Sigma \times [0, 1]$. Thus the annulus $D \setminus N$ can be isotoped so that $D \cap (\Sigma \times \{x\})$ is a loop $\alpha_x \subset \Sigma$ for each $x \in [0, 1]$. Likewise, we can assume $D' \cap (\Sigma \times \{x\})$ is a loop $\alpha'_x \subset \Sigma$ for each $x \in [0, 1]$.

For each $x_0 \in [0, 1]$, the family of loops $\{\alpha_x : x \leq x_0\}$ defines an isotopy from α_{x_0} to α_0. These isotopies lift into $\Sigma \times [0, 1]$ to define an isotopy sending $D \setminus N$ to $\alpha_0 \times [0, 1]$. There is a similar isotopy for D'. After both isotopies, D and D' agree in N and in the complement of N, so $D = D'$.

Let $D_1, \ldots, D_m$ and $D'_1, \ldots, D'_m$ be systems of disks dual to the same spine K. For each edge e_i of K, we have seen that D_i is isotopic to D'_i, so the two systems of disks are isotopic. $\square$

2.21. PROPOSITION. *Given a system of disks $\mathbf{D}$ for a handlebody H, there is a spine dual to $\mathbf{D}$. If spines K and K' are dual to $\mathbf{D}$ then there is an isotopy of H taking K onto K'.*

PROOF. Let $\mathbf{D} = \{D_1, \ldots, D_m\}$ be a system of disks for a handlebody H. Let B be a component of $H \setminus N$ where N is a regular neighborhood of $\bigcup D_i$. Let $F_1, \ldots, F_k$ be the disks in ∂B that are properly embedded in H and parallel to disks in $\mathbf{D}$.

There is a graph G in $\partial B \setminus \bigcup F_i$ such that G is a tree with one vertex v in the interior of $\partial B \setminus \bigcup (F_i \cup F'_i)$ and one vertex v_i in ∂F_i for each i. Let G' be the result of isotoping v and all the edges of G into the interior of B and isotoping each vertex v_i into the interior of F_i, then into the interior of the disk D_j parallel to F_i.

Let N be a regular neighborhood of G. The set $B \setminus N$ is homeomorphic to $(\partial B \setminus \bigcup F_i) \times [0,1]$. Construct a similar graph for each component of $H \setminus N$. For each disk D_i, there two parallel disks in the boundaries of the components of N. An arc is extened from each of these disks into the interior of D. We can assume that these arcs are extended to the same point in D_i, so that the two arcs form a single edge that intersects D_i in a single point.

Let K be the union of all the graphs for the components, with the two arcs to each disk in $\mathbf{D}$ combined into a single edge. This graph is a spine dual to the system of disks.

Now consider spines K and K' dual to the system $\mathbf{D}$. Let C be a component of $H \setminus N$ where N is a regular neighborhood of $\bigcup D_i$. Let $F_1, \dots, F_k \subset \partial B$ be parallel to disks in $\mathbf{D}$. Let S be the complement in ∂B of these disks.

The graph $K \cap B$ consists of a single vertex v in the interior of B, a vertex v_i in each disk F_i and an edge e_i from v to v_i. The graph $K' \cap B$ is a similar graph with vertices $v', v'_1, \dots, v'_k$ and edges $e'_1, \dots, e'_k$.

Let N be a regular neighborhood of $K \cap B$ in B and let N' be a regular neighborhood of $K' \cap B$. Because $D_i \setminus N$ is a vertical annulus for each disk D_i (by Lemma 2.20), $B \setminus N$ is homeomorphic to $S \times [0,1]$, as is $B \setminus N'$.

The homeomorphisms $B \setminus N \to S \times [0,1] \leftarrow B \setminus N'$ suggest a homeomorphism $B \setminus N \to B \setminus N'$. Because K is a tree, N is a ball, as is N'. Thus the map extends to an automorphism $\phi : B \to B$ which sends N to N'. Because N and N' are regular neighborhoods, this map can be chosen to send K to K'.

Because B is a ball, this automorphism is isotopic to the identity. The isotopy of B sends $K \cap B$ to $K' \cap B$. By constructing a similar isotopy for every component of $H \setminus N$, we can construct an isotopy taking K onto K'. □

4. Edge Slides

In Section 2, we defined a move called a disk slide, which can turn any system of disks for a handlebody into any other system of disks. We will now use the notion of duality from the last section to define an analogous move for spines of handlebodies.

Let e_1 and e_2 be edges in a spine K of a handlebody H, each parametrized as the closed interval $[0,1]$. Let D be a disk in H, as in Figure 4, with interior disjoint from K and whose boundary is divided into the following three sections: One arc, α_1, is the segment of e_1 from

0 to $\frac{1}{3}$. One arc, α_2 is e_2. The final arc, α_3 is an arc in the interior of H.

Let e' be the arc formed by removing α_1 from e_1 and replacing it with α_3. We can think of this as taking the subarc of e_1 between 0 and $\frac{1}{3}$ and sliding it across the disk D so that the 0 endpoint follows the edge e_2. We will write $e' = e_1 *_D e_2$. Any graph isotopic to the resulting graph K' is called an *edge slide* of K.

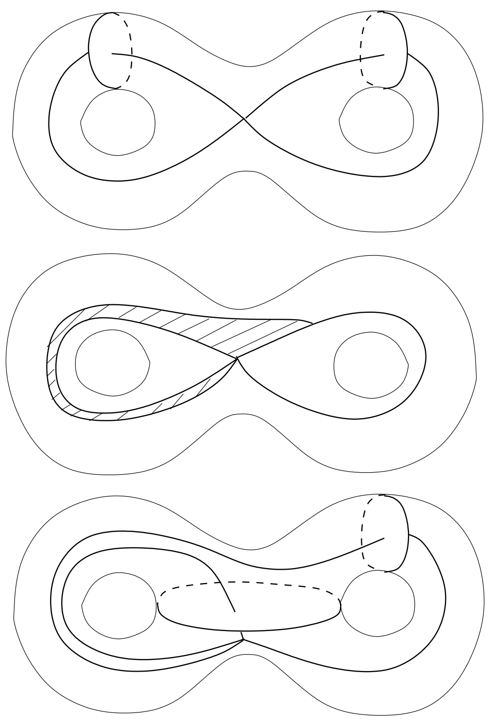


FIGURE 4. The disks before and after a disk slide suggest a corresponding edge slide for spines.

2.22. PROPOSITION. *Let K be a spine of a handlebody H and let K' be an edge slide of K. Then K' is a spine of H.*

PROOF. Let K' be an edge slide of a spine K, defined by a disk D whose boundary consists of arcs $\alpha_1 \subset e_1$, $\alpha_2 = e_2$ and $\alpha_2 \subset H$. Let N be a regular neighborhood of K and let N'' be a regular neighborhood of $K'' = K \cup D$.

The complement in N'' of N is a regular neighborhood of $K'' \setminus N$. The closure of the set $K'' \setminus N$ is a disk which intersects N in a single arc. Thus $N'' \setminus N$ is a ball which intersect N in a disk. This ball can be isotoped into N, inducing an isotopy of H that sends N'' onto N. Thus the complement in H of N'' is homeomorhic to $\partial H \times [0, 1]$.

Let K' be the result of replacing the arc $\alpha_1 \subset e_1$ with α_3' and let N' be a regular neighborhood of K'. The complement $N'' \setminus N'$ is again a ball that intersects N' in a single disk. Thus N'' is ambient isotopic to N' so $H \setminus N'$ is homeomorphic to $\partial H \times [0,1]$. By Lemma 2.17, this implies that K' is a spine of H. □

Edge slides allow us to form a large family of different spines for a handlebody. We must now show that edge slides are "dual" to disk slides.

2.23. Lemma. *Let* $\mathbf{D}$ *be a minimal system of disks for* H *and let* $\mathbf{D}'$ *be a disk slide of* $\mathbf{D}$*. If* K *is a spine dual to* $\mathbf{D}$ *then there is an edge slide* K' *of* K *that is dual to* $\mathbf{D}'$*.*

Proof. Figure 4 shows the construction that we will describe. Let K be a spine dual to a disk system $\mathbf{D} = \{D_1, \ldots, D_m\}$ and let $\mathbf{D}'$ be a disk slide of $\mathbf{D}$. Without loss of generality, assume $\mathbf{D}' = \{(D_1 *_\alpha D_2), D_2, \ldots, D_m\}$.

Let e_1 be the edge of K intersecting D_1 and let e_2 be the edge intersecting D_2. Let $B = H \setminus N$ be the complement of a regular neighborhood of $\mathbf{D}$. Because the system of disks is minimal, B is a ball.

The graph $K \cap B$ is a tree with a single vertex, v, in the interior of B and a pair of edges corresponding to each edge of K. Each disk D_i is parallel to a pair of disks F_i, F_i' in the boundary of B and α intersects B in an arc in ∂B from F_1 to F_2. (Recall, α is the arc defining the disk slide.)

Let β_1 be the edge from v to F_1 and let β_2 be the edge from v to F_2. The complement of K is homeomorphic to $S \times [0,1]$ where S is the complement in ∂B of the disks. The arc $\alpha \subset S \times \{1\}$ is parallel to an arc in $S \times \{0\}$ that runs along β_1 and β_2. Thus there is a disk E embedded in B whose boundary consists of the arc β_1, an arc in F_1, the arc α, an arc in F_2 and the arc β_2.

Let $N \subset B$ be a closed regular neighborhood of $F_1 \cup F_2 \cup \alpha$ and let E' be the closure of $E \setminus N$. Then E' is a subdisk of E whose bounary consist of a subarc of β_1, an arc in ∂N and a subarc of β_2.

Let E'' be the result of extending E' in H along the edge e_1 so that $\partial E''$ consists of all of e_1, an arc in H which passes through D_1, but is disjoint from $D_1 *_\alpha D_2$ and a subarc of e_2.

This disk defines an edge slide K' of K. The edge e_1 intersects $D_1 *_\alpha D_2$ in a single point and remains disjoint from the other disks in the system. The edge e_2 is replaced by an edge which still intersects D_2

in a single point, but is disjoint from $D_1 *_\alpha D_2$ and the disks $D_3, \ldots, D_m$. Thus the spine K' is dual to the system of disks $\mathbf{D}'$. □

Notice that the system $\mathbf{D}'$ comes from sliding D_1 across D_2, but the spine K' comes from sliding e_2 along e_1. In other words, the duality relationship reverses the roles of the objects involved in the slide.

We will say that two spines K and K' of a handlebody H are *slide equivalent* if there is a sequence of edge slides starting at K and ending at K'. By Theorem 2.13 we have the immediate corollary:

2.24. Corollary. *If K and K' are one-vertex spines of H then K and K' are slide equivalent.*

Proof. By Proposition 2.19, there is a system of disks $\mathbf{D}$ dual to K and a system of disks $\mathbf{D}'$ dual to K'. As noted above, because K and K' are one-vertex spines, the disk systems are minimal. By Theorem 2.13, there is a sequence of minimal disk systems $\mathbf{D}_0, \ldots, \mathbf{D}_n$ such that $\mathbf{D} = \mathbf{D}_0$, $\mathbf{D}' = \mathbf{D}_n$ and for each $i < n$, $\mathbf{D}_{i+1}$ is a disk slide of $\mathbf{D}_i$.

The spine $K_1 = K$ is dual to $\mathbf{D}_1$ so Lemma 2.23 implies that there is a spine K_2 that is a disk slide of K_1 and is dual to $\mathbf{D}_2$. By induction, for each $i \leq n$, there is a sequence of spines $K_1, \ldots, K_i$ such that each K_j is a disk slide of K_{j-1} and is dual to $\mathbf{D}_j$. For $i = n$, we have that $K = K_1$ is slide equivalent to a spine K_n dual to $\mathbf{D}'$. By Lemma 2.21, any two spines dual to $\mathbf{D}'$ are isotopic so K' is isotopic to K_n and is thus slide isotopic to K. □

5. Heegaard Diagrams

Let H_1 and H_2 be handlebodies of the same genus and let $h : \partial H_1 \to \partial H_2$ be a homeomorphism. Let $H_1 \cup H_2$ be the disjoint union of the two handlebodies and let M be the quotient by the relation $x \sim y$ if $x \in \partial H_1$, $y \in \partial H_2$ and $h(x) = y$. This is a closed manifold and e will write $M = H_1 \cup_h H_2$.

If we start with a manifold M and a Heegaard splitting (Σ, H_1, H_2), then the composition of the inclusion maps $\partial H_2 \to \Sigma \to \partial H_1$ is a map $h : \partial H_1 \to \partial H_2$ such that M is homeomorphic to $H_1 \cup_h H_2$. Since every manifold allows a Heegaard splitting, every manifold can be constructed by gluing together handlebodies in this way.

Let $D_1, \ldots, D_m$ be a system of disks for H_1. For each disk D_i, the image $\ell_i = h(\partial D_i)$ is a simple closed curve in ∂H_2. The collection of loops $\ell_1, \ldots, \ell_m$ can be thought of as a picture of the map ϕ, motivating the following definitions:

2.25. Definition. A *Heegaard diagram* is an ordered pair $(H, \{\ell_1, \ldots, \ell_m\})$ where H is a Handlebody and $\{\ell_i\}$ is a collection of disjoint, embedded, simple closed curves in H such that $\partial H \setminus \bigcup \ell_i$ is a collection of punctured spheres. Figure 5 shows a Heegaard diagram.

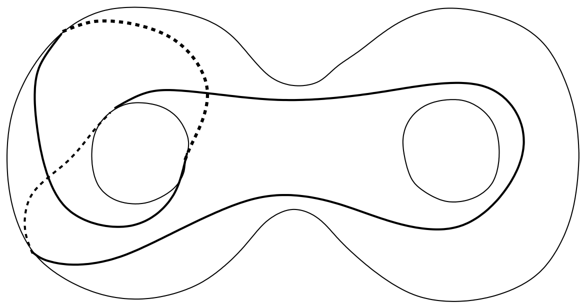


Figure 5. A Heegaard diagram for S^3 is shown here. Attaching disks along the two loops produces a manifold which is determined up to homeomorphism by the diagram.

If $\ell_1, \ldots, \ell_m$ are the images in a map $h : \partial H_1 \to \partial H_2$ of the boundaries of a system of disks for H_1 then we will say that $\{\ell_1, \ldots, \ell_m\}$ is a Heegaard diagram for h, or for M. We can now state an important Corollary of Lemma 2.9 from Section 1:

2.26. Lemma. *Let (Σ, H_1, H_2) be a Heegaard splitting of a 3-manifold M and let (Σ', H_1', H_2') be a Heegaard splitting of a manifold M'. Let $(H_2, \{\ell_1, \ldots, \ell_m\})$ and $(H_2', \{\ell_1', \ldots, \ell_m'\})$ be Heegaard diagrams for the two splittings. If there is a homeomorphism $\phi : H_2 \to H_2'$ taking each loop ℓ_i onto ℓ_i' then there is a homeomorphism $\psi : M \to M'$ taking H_1 to H_1', H_2 to H_2' and Σ to Σ'.*

Proof. We would like to extend the homeomorphism ϕ to a homeomorphism sending all of M to all of M'. Let $D_1, \ldots, D_M \subset H_1$ and $D_1', \ldots, D_m' \subset H_1'$ be the systems of disks whose boundaries define the two Heegaard diagrams. The map $\phi|_{\partial H_2}$ sends ∂H_2 to $\partial H_2'$ and sends ∂D_i to $\partial D_i'$ for each i.

By Lemma 2.9, this implies that there is a homeomorphism $\phi' : H_1 \to H_1'$ such that $\phi'|_{\partial H_1} = \phi|_{\partial H_2}$. Since ϕ and ϕ' agree on the intersection of their domains, their union is a map $\psi : M \to M'$ which has the desired properties. $\square$

This implies the following Corollary, whose proof will be left as an exercise for the reader:

2.27. COROLLARY. *Let* $(H, \{\ell_1, \ldots, \ell_n\})$ *and* $(H, \{\ell'_1, \ldots, \ell'_n\}$ *be Heegaard diagrams on the same handlebody. If the diagrams are isotopic (i.e. there is an isotopy of* ∂H *taking* $\cup \ell_i$ *onto* $\cup \ell'_j$*) then the induced Heegaard splittings are homeomorphic.*

In other words, two splittings that have the same diagram (isotopic diagrams) are homeomorphic Heegaard splittings of homeomorphic manifolds: a Heegaard diagram completely defines the topology of the Heegaard splitting.

A Heegaard splitting has an infinite number of different diagrams because a handlebody has an infinite number of systems of disks. Two Heegaard diagrams that are not homeomorphic could still come from the same Heegaard splitting. In order to find necessary and sufficient conditions for two diagrams to come from homeomorphic splittings, we would need to define a notion of loop slides and employ Theorem 2.13. The details will be left to the reader.

There is also the matter of when a diagram defines a Heegaard splitting. In other words, given a Heegaard diagram, will it be a diagram for some Heegaard splitting? The key lies in the condition that the complement of the loops be a collection of punctured spheres.

2.28. LEMMA. *Given a Heegaard diagram* $(H_2, \{\ell_1, \ldots, \ell_m\})$ *there is a map* $h : H_1 \to H_2$ *such that* $(H_2, \{\ell_1, \ldots, \ell_m)$ *is a Heegaard diagram for the induced Heegaard splitting of* $M = H_1 \cup_h H_2$*.*

The proof is very similar to the proof of Lemma 2.9, so we will only give an outline. Let H_2 be a handlebody and let $\ell_1, \ldots, \ell_m$ be simple closed curves in ∂H_2 forming a Heegaard diagram.

For each $i \leq m$, let D_i be a disk and let N_i be a regular neighborhood of ℓ_i in ∂H_2. Let $\phi_i : \partial(D_i \times [0, 1]) \to N_i$ a homeomorphism. Let M_1 be the result of gluing $D_i \times [0, 1]$ to ∂H by the map ϕ_i for each i.

Because the loops form a Heegaard diagram, the complement $\partial H \setminus \bigcup \ell_i$ is a collection of punctured spheres. Thus the boundary of M_1 is a collection $S_1, \ldots, S_n$ of spheres. For each $j \leq n$, let B_j be a closed ball and $\psi_j : \partial B_j \to S_j$ a homeomorphism and let M be the result of gluing each S_j to M_1 by the map ψ_j.

The manifold M is closed and there is a canonical embedding of H into M. From the construction, it follows that the complement of H in M is a handlebody.

6. Compression Bodies

We have so far considered only closed 3-manifolds. This limitation was imposed by the fact that a handlebody has a single boundary component which, in a Heegaard splitting, is glued to a second handlebody, leaving no boundary. There are a number of potential ways to generalize a Heegaard splitting to manifolds with boundary. The method that has proved most useful is to replace the handlebody with a similar object that has more boundary components.

Let Σ be a compact, connected closed, orientable surface and let $\ell_1, \ldots, \ell_m$ be a collection of disjoint, simple closed curves embedded in Σ. For each $i \leq m$, let N_i be a regular neighborhood in Σ of ℓ_i. Let D_i be a disk and let $\phi_i : (\partial D_i \times [0,1]) \to N_i$ be a homeomorphism. Let H' be the result of gluing $D_i \times [0,1]$ to $\Sigma \times [0,1]$ by the map ϕ_i for each i. Let H be the result of gluing balls into any sphere components of $\partial H'$.

2.29. DEFINITION. A *compression body* is a manifold resulting from the above construction.

A compression body H has a number of boundary components. We will write $\partial_+ H = \Sigma \times \{0\}$ and $\partial_- H = \partial H \setminus \partial_+ H$. (To remember which is which, notice that the genus of $\partial_+ H$ is higher than the genus of any component of $\partial_- H$.) If $\partial_- H$ is empty then the compression body H is also a handlebody.

2.30. DEFINITION. A *Heegaard splitting* of a compact, orientable manifold M with boundary is a triple (Σ, H_1, H_2) where $\Sigma \subset M$ is a compact, connected, closed, orientable surface and H_1 and H_2 are compression bodies such that $\partial_+ H_1 = \Sigma = \partial_+ H_2 = H_1 \cap H_2$ and $H_1 \cup H_2 = M$.

The proof of the following Lemma will be left as an exercise for the reader:

2.31. LEMMA. *If (Σ, H_1, H_2) is a Heegaard splitting of a manifold M then $\partial M = \partial_- H_1 \cup \partial_- H$.*

We would now like to develop a theory of compression bodies along the same lines as the theory of handlebodies developed above. For the following lemmas and theorems, the proofs use methods very similar to the analagous statements about handlebodies. Thus the discussion will be restricted to outlines of the proofs.

2.32. LEMMA. *Let H be a manifold with boundary and let Σ be a component of ∂H. Let $D_1, \ldots, D_m \subset H$ be a collection of properly embedded disks such that for each $i \leq m$, $\partial D_i \subset \Sigma$. Let N be a regular neighborhood of $\bigcup D_i$ and assume that each component of $H \setminus N$ is homeomorphic to either a ball or to $F \times [0,1]$ where F is a closed surface. Then H is a compression body.*

To prove this, note that the boundaries of the disks $D_1, \dots, D_m$ are loops in the surface Σ, which suggest a compression body, as in Definition 2.30. It can be shown that H is homeomorphic to this compression body.

This lemma suggests a way to construct a compression body which is closer to the definition of a handlebody. Let H' be a disjoint union of balls and of manifolds of the form $F \times [0,1]$ where F is a closed surface. Let $D_1, \dots, D_m, D'_1, \dots, D'_m$ be a collection of disks in $\partial H'$ such that each disk is either in the boundary of a ball component or in an $F \times \{1\}$ boundary component. For each $i \leq m$, let $\phi_i : D_i \to D'_i$ be a homeomorphism.

2.33. Lemma. *The result of gluing H' by the maps $\phi_1, \dots, \phi_m$ is a compression body.*

Lemma 2.32 also motivates the following definition:

2.34. Definition. A *system of disks* for a compression body H is a collection $D_1, \dots, D_n$ of pairwise disjoint disks properly embedded in H such that $\partial D_i \subset \partial_+ H$ for each i and the complement in H of a regular neighborhood of the disks is a collection of balls and a (not necessarily connected) manifold $\partial_- H \times [0,1]$.

Similarly, we can define a spine for a compression body. Let K be a graph embedded in H with some valence-one vertices possibly embedded in $\partial_- H$. Let N be an open neighborhood of $K \cup \partial_- H$.

2.35. Definition. If $H \setminus N$ is homeomorphic to $\partial_+ H \times [0,1]$ then $K \cup \partial_- H$ is a *spine* for H.

Note that when $\partial_- H = \emptyset$, this definition is equivalent to the definition of a spine of a handlebody. The theory of compression bodies can be developed further in analogy to handlebodies, and will be as a series of exercises for the reader.

7. Exercises

1. Define what it means for a system of disks in a compression body to be minimal or maximal.

2. Show that disk slides in a compression body are well defined and that a disk slide of a minimal system of disks produces a new minimal system of disks.

3. Define edge slides for spines of compression bodies and show that there is a one-to-one correspondence between spines and systems of disks, edge slides and disk slides.

4. Find a formula to calculate the genus of $\partial_+ H$ from the number of disks, the number of balls and the number and genera of the components of $\partial_- H$. Note that an answer should give the same result as for handlebodies when $\partial_- H = \emptyset$.

5. Show that slide equivalence is an equivalence relation.

6. Show that if K is a spine of a handlebody H and N is a regular neighborhood of K then the complement in H of N is homeomorphic to $\partial H \times [0, 1]$.

7. Show that if two properly embedded disks in a handlebody have isotopic boundaries then they are isotopic.

8. Prove Corollary 2.27.

9. Prove Lemma 2.31.

CHAPTER 3

Algebra

1. Combinatorial Group Theory

Given a positive integer, n, let A_n be a collection of symbols $\{x_1, \dots, x_n, x_1^{-1}, \dots, x_n^{-1}\}$. A *string* is a finite ordered sequence $s = a_1 a_2 \dots a_k$ where $a_i \in A_n$ for each $i \leq k$. We will define two strings, say s and s', to be equivalent if we can turn s into s' by adding a finite number of pairs of consecutive characters of the form $x_j x_j^{-1}$ or $x_j^{-1} x_j$, and then removing a number such pairs. In other words, we treat the symbol x_j^{-1} as the inverse of x_j so that removing or inserting $x_j x_j^{-1}$ is equivalent to multiplying by 1.

Let F^n be the set of equivalence classes of all finite strings made from the characters in A_n. Included in this set is the empty string, which we will represent by the character 1. We will define a multiplication on F^n that will give it the structure of a group. Given $s = a_1 \dots a_k \in F^n$ and $s' = a'_1 \dots a'_{k'} \in F^n$, the product ss' is the string defined by attaching the string s' to the end of s. In symbols, $ss' = a_1 \dots a_k a'_1 \dots a'_{k'}$.

3.1. Lemma. *Multiplication in F^n is well defined.*

It is left to the reader to check that if $s \sim s''$ and $s' \sim s'''$ then $ss' \sim s'' s'''$.

3.2. Lemma. *The empty string is the identity in F^n and every element of F^n has an inverse.*

The reader can check that for each string s, there is a string s^{-1} such that $ss^{-1} = 1$. Although ss^{-1} will not be the empty string, it will be equivalent to the empty string. The resulting group is called the *free group* on n elements.

3.3. Lemma. *Let G' be a group and let $y_1, \dots, y_n$ be a subset of G'. Then there is a group homomorphism $\phi : F^n \to G'$ sending x_i to y_i for each $i \leq n$.*

Proof. Let $\phi(1) = 1$. For each string $s = a_1 a_2 \dots a_n \in F^n$, define the element $\phi(s) \in G$ by replacing each x_i by y_i and each x_i^{-1} by

y_i^{-1}, then taking the product in G. This map is well defined for each element of F^n and respects multiplication in both groups, so ϕ is a homomorphism. □

Let $r_1, \ldots, r_m$ be a collection of strings in F^n. Let N be the subgroup of F^n generated by all conjugates of these strings. In other words, N consists of all strings which are products of elements of the form sr_is^{-1} where s is any string in F^n.

3.4. Lemma. *The subgroup N is normal in F^n and any normal subgroup containing the relations $r_1, \ldots, r_m$ contains N.*

The proof is left as an exercise for the reader. Because N is a normal subgroup, we can take the quotient. Let $G = F^n/N$. We will write $G = \langle x_1, \ldots, x_m : r_1, \ldots, r_m \rangle$. Because N was the "smallest" normal subgroup containing the relations, G should be the "largest" group generated by elements $x_1, \ldots, x_n$ such that all the products of the generators given by the strings $r_1, \ldots, r_m$ are equal to the identity. We will call the strings $r_1, \ldots, r_m$ *relations* for G.

3.5. Lemma. *Let $G = \langle x_1, \ldots, x_m : r_1, \ldots, r_m \rangle$. Let G' be a group with generators $y_1, \ldots, y_n$ such that substituting y_i for x_i (for each $i \leq m$) in any relation r_j produces the trivial element of G'. Then there is a homomorphism from G onto G' sending x_i to y_i for each i.*

Proof. By Lemma 3.3, there is a map $\phi : F^n \to G'$ sending x_i to y_i for each $i \leq n$. Because $y_1, \ldots, y_n$ generate G', this map is onto. Let N' be the kernel of ϕ. Substituting the generators of G' into a relation r_j produces the identity element of G' so each relation is in N'. Also, N' is normal because it is the kernel of a homomorphism.

By Lemma 3.4, the normal subgroup N generated by the relations $r_1, \ldots, r_j$ is contained in N' because N' is normal and contains all the relations. Thus every coset aN is contained in a coset aN', so there is an induced map $\phi' : F^n/N \to F^n/N'$. By the first isomorphism theorem, F^n/N' is isomorphic to G' so the homomorphism ϕ' sends G onto G'. Note that ϕ' will be an isomorphism if and only if $N = N'$. □

A group is called *finitely generated* if there is a finite subset of the group that generates the whole group. A group G is *finitely presented* if G is isomorphic to the group $\langle x_1, \ldots, x_m : r_1, \ldots, r_m \rangle$ for some set of relations $r_1, \ldots, r_m$. The set of generators and relations $\langle x_1, \ldots, x_m : r_1, \ldots, r_m \rangle$ is called a *presentation* for G.

Note that every finitely presented group is finitely generated, but a finitely generated group may not be finitely presented. A presentation is *balanced* if $m = n$, i.e. the number of relations is equal to the number of generators.

In general, a finitely presented group will have a number of different presentations. There is a set of moves, called the *Tietze transformations* which allow one to create a new group presentation from a given presentation. The moves are as follows:

(1) Replace a relation r_i with a conjugate relation, sr_is^{-1} or $sr_i^{-1}s^{-1}$ where s is any string in F^n.

(2) Replace a relation r_i with its product r_ir_j with any other relation r_j such that $i \neq j$.

(3) Replace a character x_i in the generating set with a new character x_i' and replace each occurence of x_i in the relations with $x_i'x_j$ or $(x_i')^{-1}$, where x_j is any distinct character, $i \neq j$.

(4) Add a character x_{n+1} to the generating set and for some string s of characters in A_n, add the relation $r_{m+1} = sx_{n+1}$. Or, if there is a generator x_i such that x_i appears in only one relation, $r_j = sx_i$ (and x_i does not appear in s) then remove the generator x_i and the relation r_j.

3.6. Lemma. *Let $G = \langle x_1, \ldots, x_m : r_1, \ldots, r_m \rangle$ and let G' be a group defined by a presentation resulting from one of the Tietze transformations on the presentation of G. Then G' is isomorphic to G.*

Proof. We will examine each transformation, in order:

(1) Let $N \subset F^n$ be the normal subgroup generated by the relations $r_1, \ldots, r_m$ and let N' be the normal subgroup generated by $sr_1s^{-1}, r_2, \ldots, r_m$. Because N is normal, sr_1s^{-1} is in N so by Lemma 3.4, $N' \subset N$. Similarly, $r_1 = s^{-1}(sr_1s^{-1})s$ is in N' so $N' \subset N$. It follows that $N' = N$ and F/N is isomorphic to (in fact, equal to) F/N'.

(2) The argument here is similar to the argument for (1). Note that any subgroup containing r_1 and r_2 will contain r_1r_2 and any subgroup containing r_1r_2 and r_2 will contain $r_1 = (r_1r_2)r_2^{-1}$. Again, $N = N'$ and the quotients are equal.

(3) We will first define an isomorphism $\phi : F^n \to \hat{F}^n$, where F^n is the free group generated by $x_1, \ldots, x_n$ and $\hat{F}^n$ is generated by $x_1', x_2, \ldots, x_n$. Define $\phi(x_1) = x_1'x_2$ and $\phi(x_i) = x_i$ for $i \neq 1$. This map extends to a homomorphism from F^n to $\hat{F}^n$.

Define the map $\hat{\phi} : \hat{F}^n \to F^n$ by $\hat{\phi}(x_1') = x_1x_2^{-1}$ and $\hat{\phi}(x_i) = x_i$ for $i \neq 1$. Again this extends to a homomorphism and $\hat{\phi}(x_1'x_2) = x_1x_2^{-1}x_2 = x_1$ so $\hat{\phi}$ is an inverse of ϕ. Thus ϕ is an isomorphism between free groups.

Let $N \subset F^n$ be the normal subgroup generated by the relations $r_1, \ldots, r_n$ and let $\hat{N} = \phi(N) \subset \hat{F}^n$. Each element of N is a product of conjugates sr_is^{-1} so each element of $\hat{N}$ is a product of conjugates $\phi(s)\phi(r_i)\phi(s^{-1})$. Thus $\hat{N}$ is the normal subgroup generated by the relations $\phi(r_1), \ldots, \phi(r_m)$. Because ϕ sends x_1 to y_1y_2 and x_i to y_i for $i > 1$, the quotient $\hat{F}^n/\hat{N}$ is the group with presentation resulting from a type-three Tietze move.

(4) Let F^{n+1} be the free group generated by $x_1, \ldots, x_{n+1}$ and let $\hat{F}^n$ be the free group generated by $x_1, \ldots, x_n$. Let G be the quotient of F^{n+1} by the relations $r_1, \ldots, r_m, sx_{n+1}$ and let $\hat{G}$ be the quotient of $\hat{F}^n$ by the relations $r_1, \ldots, r_m$.

By Lemma 3.5, there is a homomorphism $\phi : G \to G'$ sending x_i to x_i for $i \leq n$ and x_{n+1} to $s^{-1} \in G'$. This is because the relations $r_1, \ldots, r_m$ hold in G' and the relation $x_{n+1} = s^{-1}$ becomes $sx_{n+1} = 1$. Conversely, there is a map $\hat{\phi} : G' \to G$ sending x_i to x_i for $i \leq n$. One can check that $\hat{\phi}$ is the inverse of ϕ, implying that ϕ is an isomorphism from G to G'. □

2. The Fundamental Group

Recall that the fundamental group, of a manifold M is the group $\pi_1(M)$ of equivalence classes of closed paths $p : [0,1] \to M$ with their endpoints at a fixed point, $p(0) = p(1) = x \in M$. Homotopic paths are equivalent and multiplication is defined by concatenating the paths, i.e. by attaching one path to the end of the other.

In this section, we will show how to find a presentation for the fundamental group of a manifold from a Heegaard diagram. Our main tool will be Van Kampen's Theorem, which can be found in most algebraic topology books, including Hatcher's book [**5**, p.43]. It will be stated here without proof.

Given two finitely presented groups G, G', the free product $G * G'$ is the group generated by the disjoint union of the generators of G and G' and relations the disjoint union of the relations for G and G'. The free product $*_i G_i$ of an indexed set of groups $\{G_i\}$ is defined similarly. Free products can also be defined for groups that are not finitely presented, but that is beyond the scope of this discussion.

3.7. Theorem (Van Kampen). *If X is the union of path-connected open sets $A_1, \ldots, A_n$, each containing the basepoint $x_0 \in X$ and if each intersection $A_i \cap A_j$ is path connected then the homomorphism ϕ from the free product $*_i \pi_1(A_i)$ to $\pi_1(X)$ is onto. If in addition, each intersection $A_i \cap A_j \cap A_k$ is path connected then the kernel of ϕ is the* *normal subgroup generated by elements of the form* $\eta_{ij}\eta_{ji}^{-1}$*, where* η_{ij} *is the inclusion map* $\eta_{ij} : \pi_1(A_i \cap A_j) \to \pi_1(A_i)$.

Let H be a handlebody and let K be a spine for H with a single vertex, v, and g edges $e_1, \ldots, e_g$. (Here, g is the genus of H.) For each $i \leq g$, let $\ell_i : [0,1] \to H$ be a loop whose image is the edge e_i. We can think of this as a parametrization of the edge from 0 to 1. The endpoints $\ell_i(0)$ and $\ell_i(1)$ are both at v so if we take v as the base point for the fundamental group, each edge defines an element x_i of $\pi_1(K)$ and $\pi_1(H)$.

3.8. Lemma. *The fundamental group of* H *is the free group* F^g *generated by the elements* $x_1, \ldots, x_g$ *determined by the paths* $\ell_1, \ldots, \ell_g$.

To prove this, we will need the following claim, whose proof is left as an exercise for the reader:

3.9. Claim. *The handlebody* H *is homotopy equivalent to its spine* K.

Proof of Lemma 3.8. Because H is homotopy equivalent to K, we have $\pi_1(K) \cong \pi_1(H)$ and we can simply calculate $\pi_1(K)$. For each edge e_i of K, $e_i \cup v$ is a closed loop so $\pi_1(e_i \cup v)$ is the infinite cyclic group generated by ℓ_i. Let N_i be an open regular neighborhood of $e_i \cap v$ in K. Then $\pi_1(N_i) \cong Z$.

The set $e_i \cap e_j$ is a tree, as is $e_i \cap e_j \cap e_k$ for any edges e_i, e_j, e_k of K. In particular, these intersections are connected so by Van Kampen's Theorem, there is a homomorphism $\phi : *_{i=1}^{g}\mathbf{Z} \to \pi_1(K)$. The free product $*_{i=1}^{g}\mathbf{Z}$ of g copies of $\mathbf{Z}$ is precisely the free group F^g.

The kernel of the map ϕ is generated by the images of the inclusion maps η_{ij}. However, each intersection is a tree, which is simply connected so the inclusion maps are trivial. Thus ϕ is an isomorphism, completing the proof. $\square$

Let (Σ, H_1, H_2) be a Heegaard splitting for a manifold M. Let K be a one-vertex spine for H_1 as above and let $D_1, \ldots, D_g$ be a minimal system of disks for H_2. For each disk D_i, the loop ∂D_i is embedded in $\partial H_1 = \Sigma = \partial H_2$. Choose a point $b_i \in \partial D_i$ and let $k_i : [0,1] \to \partial D_i$ be a map such that $k_i(0) = k_i(1) = b_i$.

The image of k_i is contained in H_1. Let $j_i : [0,1] \to H_1$ be a path from v to b_i. The concatenation $j_i^{-1}k_ij_i$ is a closed curve based at v, so it defines an element r_i of $\pi_1(H_1)$. (Here, j_i^{-1} is the reverse of j_i, not the inverse map.)

3.10. Lemma. *The fundamental group of* M *is isomorphic to* $\langle x_1, \ldots, x_g : r_1, \ldots, r_g \rangle$.

PROOF. Let M_1 be the result of gluing a disk D_1 to H_1 along the loop ℓ_1. Let $A_1 \subset M_1$ be a regular neighborhood of H_1 in M_1. Let $A_2 \subset M_1$ be the interior of D_1.

The fundamental group $\pi_1(A_2)$ is the trivial group and $\pi_1(A_1) \cong \pi_1(H_1) \cong F^g$. The intersection $A_1 \cap A_2$ is an open annulus $S^1 \times (0,1) \subset D_1$.

By Van Kampen's Theorem, there is a homomorphism ϕ_1 from $F^g * \{1\}$ $(= F^g)$ to $\pi_1(M_1)$ and the kernel of ϕ_1 is generated by elements of the form $\eta_{A_1}(x)\eta_{A_2}^{-1}(x)$ where x is an element of $\pi_1(A_1 \cap A_2) = \mathbf{Z}$. Because D is simply connected, $\eta_{A_2}^{-1}(x)$ is the identity in F^g. The image $\eta_{A_1}(x)$ is a power of the relation r_1 defined by l_1, so the kernel of ϕ_1 is the normal subgroup generated by r_1.

Let M_2 be the result of gluing a disk D_2 to ∂M_1 along the loop l_2. By Van Kampen's Theorem, there is a homomorphism $\phi_2 : \pi_1(M_1) \to \pi_1(M_2)$ whose kernel is the normal subgroup generated by the relation r_2. The map $\phi_2 \circ \phi_1$ is a map from F^n to $\pi_1(M_2)$ whose kernel is the normal subgroup generated by r_1 and r_2.

Continue in this fashion, constructing a manifold M_i for each $i \leq g$ by gluing the disk D_i to ∂M_{i-1}. Let ϕ be the homomorphism $\phi_g \circ \cdots \circ \phi_1 : F^n \to \pi_1(M_g)$. The kernel of ϕ is the normal subgroup generated by the relations $r_1, \ldots, r_g$. Because ϕ is onto, the first homomorphism theorem implies $\pi_1(M_g) \cong \langle x_1, \ldots, x_g : r_1, \ldots, r_g \rangle$.

Because M_g came from gluing disks into a Heegaard diagram for M, there is an embedding $h : M_g \to M$ such that the complement of $h(M_g)$ is an open ball in M. Let A_1 be this open ball and let A_2 be a regular neighborhood of the image of M_g.

The open set A_1 is simply connected so Van Kampen's Theorem implies a homomorphism $\pi_1(M_g) \to \pi_1(M)$. The intersection $A_1 \cap A_2$ is homeomorphic to $S^2 \times (0,1)$. This set is simply connected so the inclusion maps $\eta : \pi_1(A_1 \cap A_2) \to \pi_1(A_i)$ are trivial. Thus the kernel of the homomorphism is trivial, and η is an isomorphism. Thus $\pi_1(M) \cong \pi_1(M_g) \cong \langle x_1, \ldots, x_g : r_1, \ldots, r_g \rangle$. □

3.11. COROLLARY. *If M is a compact, closed 3-manifold then $\pi_1(M)$ is finitely presented and has a balanced presentation.*

Notice that the relation r_i depends on the path j_i from v to b_i. If we choose a different path, j_i' then j_i and j_i' form a closed loop in H_1. Let s be the element of $\pi_1(H_1)$ defined by this closed loop. The relation r_i' defined by taking j_i' instead of j_i is precisely the conjugation of r_i by s. Thus the presentation is defined up to the first Tietze transformation. (See Section 1.) These moves do not affect the isomorphism class of the resulting group. As a result, we can ignore the path j_i and find a nice way to calculate the relations.

Let $D'_1, \ldots, D'_g$ be a system of disks for H_1 dual to K such that $\partial D'_j$ is transverse to ∂D_i for each i, j. Choose an orientation for H_1. The map ℓ_i from $[0, 1]$ to e_i defines an orientation for the edge. This along with the orientation of H_1 defines an orientation for D_i, which determines an orientation of the boundary $\partial D'_i$. The orientation for H_1 determines an orientation for Σ and we can arbitrarily choose an orientation for each loop ∂D_j.

For a loop l_i, let $p_1, \ldots, p_n$ be the points of intersection between ∂D_i and all the loops $\partial D'_1, \ldots, \partial D'_g$, in the order defined by l_i. At each point, the orientations of ∂D_i and $\partial D'_j$ define an orientation for Σ.

The string r_i will be defined as follows: For each p_k, if the orientation defined by ∂D_i and $\partial D'_j$ agrees with the orientation of Σ then append the generator x_j to the left of r_i. If the orientations disagree, append the element x_j^{-1} to r_i. An example is shown in Figure 1.

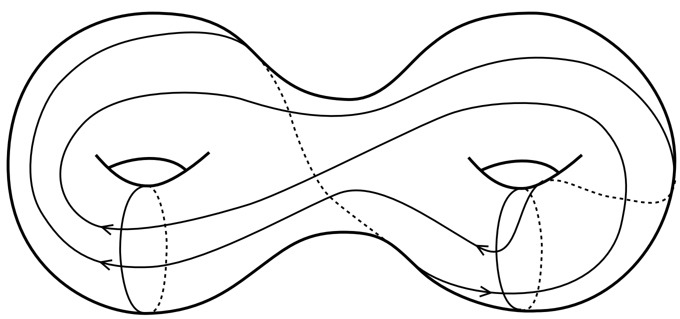


FIGURE 1. This loop suggests the relation $ab^{-1}ab = 1$.

3.12. LEMMA. *Given a loop ℓ_i and a path j_i from v to $\ell_i(0)$, the element of $\pi_1(H_1)$ defined by ℓ_i and j_i is conjugate to the element defined by the string r_i described above.*

PROOF. Let $B \subset H$ be the complement of a regular neighborhood N of the disks $D_1, \ldots, D_n$. The intersection of B with the closure of N is a collection of disks $E_1, E'_1, \ldots, E_n, E'_n$ in ∂B such that each D_j is parallel to E_j and E'_j. The intersection of ℓ_i with B is a collection of arcs in ∂B, properly embedded in the complement of the disks $E_1, E'_1, \ldots, E_n, E'_n$.

The intersection of the spine K with B is a tree with a single vertex v in the center of B and a vertex v_j, v'_j in the interior of each disk E_j,

E'_j, respectively. For each D_j, let N_j be the closure of the component of N that contains D_j. This set is homeomorphic to $D_j \times [0,1]$.

The spine K intersects N_j in an arc of the form $\{x\} \times [0,1]$, where x is in the interior of D_j. The loop ℓ_i intersects N_i in some number of arcs, each of which is of the form $\{b\} \times [0,1]$, with b a point of ∂D_j. Let α be an arc in D_j from b to x. We can homotope the arc $\{b\} \times [0,1]$ to the arc $(\alpha \times \{0\}) \cup (\{x\} \times [0,1]) \cup (\alpha \times \{1\})$. The homotopy of $\{b\} \times [0,1]$ induces a homotopy of ℓ_i in H. Do the same for each arc of $\ell \cap N_j$ for each j.

The intersection of the spine K with B is a graph with a single vertex v in the center of B, a vertex v_j or v'_j in the center of each disk D_j or D'_j, respectively, and an edge from v to each v_j or v'_j. After the homotopy of ℓ_i, the intersection $\ell_i \cap B$ consists of a collection of arcs between the vertices of $K \cap B$ in ∂B.

Let β be an arc of $\ell_i \cap B$. Because B is simply connected, there is a homotopy of β taking it into $K \cap B$ so that its image consists of the edge from one endpoint of β to v and the edge from v to the second endpoint. This homotopy lifts to a homotopy of l_i in H.

Carrying out this homotopy for each arc of $\ell_i \cap B$ turns ℓ_i into an edge path in K. By the construction, each edge e_j of K traversed by ℓ_i corresponds to an intersection with the corresponding disk D_j and the direction across the edge is determined by the orientation of ℓ_i with respect to D_j. Thus the element of $\pi_1(K)$ defined by ℓ_i is precisely the string calculated above.

During the homotopy of ℓ_i, the path j_i can be extended, becoming a loop from v to v. Replacing the arc j_i with the trivial path corresponds to a conjugating the element of $\pi_1(H)$. Thus up to conjugation, the element of $\pi_1(H)$ defined by $j_i \ell_i j_i^{-1}$ is the string r_i. □

Keep in mind that we made an arbitrary choice, taking the generators of $\pi_1(M)$ from H_1 and the relations from H_2. If we were to switch the roles, the result would be a different presentation which is, in some sense, dual to the original. These dual roles will be important in the next few sections.

We also made an arbitrary choice of orientation for the disks in each system. If we choose a different orientation for a disk in H_1, it changes the presentation by replacing a generator by its negative in each relation. Switching the orientation of a disk in H_2 corresponds to replacing a relation with its negative. These are both allowable Tietze moves so, as we would hope, switching an orientation does not change the isomorphism class of the group, even though it changes the presentation.

3.13. EXAMPLE. *The 3-torus*

Recall from Chapter 1 the construction of the 3-torus. To form T^3, we glue together pairs of faces of the cube. The vertices and edges of the cube form a graph K with a single vertex, v, and three edges which we will label a, b and c.

In Figure 2, the cube is shown with the edges labeled according to their images in ϕ. Arrows indicate the orientations. Because the graph has a single vertex, the edges define the three generators for the free group $\pi_1(H_1)$ where H_1 is the closure of a regular neighborhood of K.

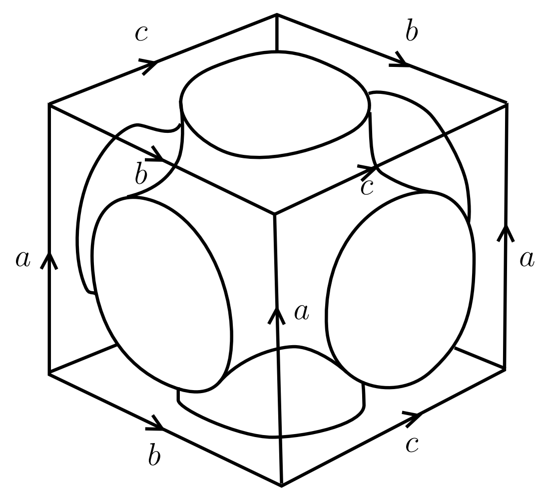


FIGURE 2. The edges of the cube are identified in groups of four to form edges of a spine of a Heegaard splitting..

The faces of the cube intersect the complementary handlebody H_2 in three disks, which cut H_2 into a ball. The boundaries of these disks define elements of $\pi_1(H_1)$ which can be read from the figure. The front left face defines $aba^{-1}b^{-1}$, the front right defines $aca^{-1}c^{-1}$ and the top face defines $bcb^{-1}c^{-1}$.

By Lemma 3.10, the fundamental group of the 3-torus has presentation $\pi_1(T^3) = \langle a, b, c : aba^{-1}b^{-1}, aca^{-1}c^{-1}, bcb^{-1}c^{-1}\rangle$. The relation $aba^{-1}b^{-1} = 1$ can be rewritten as $ab = ba$. In other words, the relation makes a and b commute. The other relations make a, c and b, c commute. Thus the presentation determines the abelian group $\mathbf{Z}^3$.

This group has rank three, meaning that no set of two or fewer elements can generate $\mathbf{Z}^3$. Thus T^3 does not allow a genus-two Heegaard splitting because such a splitting would imply a presentation for $\mathbf{Z}^3$ with only two generators. We have proved the following:

3.14. LEMMA. *The 3-torus has Heegaard genus three.*

3. A Second Look at Lens Spaces

In this section we will calculate the fundamental groups of lens spaces. Recall from Chapter 1 that a lens space $M(\mu, \lambda)$ is defined by simple closed curves μ and λ in a torus Σ. The loop μ is the boundary of a disk embedded in a genus-one handlebody H such that $\partial H = \Sigma$. The loop $\lambda \subset \partial H$ forms a Heegaard diagram for a genus-one Heegaard splitting of $M(\mu, \lambda)$.

Let λ and λ' be closed loops in Σ and assume there is an ambient isotopy $\phi : \Sigma \times [0, 1] \to \Sigma$ taking λ to λ'. By Corollary 2.27, the Heegaard splittings induced by λ and λ' are homeomorphic. Thus we can isotope λ in Σ without changing $M(\lambda, \mu)$. We will isotope λ so as to make its intersection with μ as "simple" as possible.

The complement in Σ of μ is an open annulus, homeomorphic to $S^1 \times (0, 1)$. The homeomorphism extends to a map $h : S^1 \times [0, 1] \to \Sigma$ sending $S^1 \times \{0\}$ and $S^1 \times \{1\}$ onto μ. Define $A = S^1 \times [0, 1]$. If λ is not disjont from μ then the preimage in A of λ is a collection of properly embedded arcs. These arcs can have their endpoints on the same boundary component of A, or on opposite components.

3.15. Lemma. *There is an isotopy of λ after which $h^{-1}(\lambda)$ is either an essential closed loop or a collection of parallel arcs with endpoints on boundary components of A.*

Proof. Isotope λ so as to minimize the number of points of intersection with μ. If $\lambda \cap \mu$ is empty then λ is a closed loop in the complement. Otherwise, let α be an arc of $h^{-1}(\lambda)$. Let ℓ_1 and ℓ_2 be the boundary loops of A. The arc α has endpoints $v, v' \in \ell_1 \cup \ell_2$. The proof of the following claim is left as an exercise for the reader:

3.16. Claim. *If both vertices are in the same boundary component, say ℓ_1, then there is a disk $D \subset A$ such that ∂A consists the arc α and an arc $\beta \subset \ell_1$.*

Any arc of $h^{-1}(\lambda)$ in D also has both boundary components in A. By taking an innermost disk, we can assume that there are no arcs of $h^{-1}(\lambda)$ in the interior of D. Isotoping α across D reduces the number points of intersection between λ and μ. Since we assumed this number was minimized, this is not possible, so α must have one endpoint in ℓ_1 and one in ℓ_2.

Because $h^{-1}(\lambda)$ is properly embedded in A, any two arcs in $h^{-1}(\lambda)$ are disjoint and have their endpoints in opposite boundary components. Because A is an annulus, this implies that any two arcs are parallel. Thus if $h^{-1}(\lambda)$ is not an essential simple closed curve, then it consists of a collection of parallel arcs. $\square$

If $h^{-1}(\lambda)$ is a single arc then $\lambda \cap \mu$ is a single point, so by Lemma 1.22, $M(\mu, \lambda)$ is S^3. If $h^{-1}(\lambda)$ is a closed loop, then λ is isotopic to μ. It is left as an exercise for the reader to figure out what manifold results in this case.

Otherwise, let p be the number of arcs of $h^{-1}(\lambda)$ and let α be one of the arcs with endpoint y in $S^1 \times \{0\}$. Then $h(y)$ is a point $x \in \lambda$ and $h^{-1}(x)$ consists of two points, y and a second point, $x_0 \in S^1 \times \{1\}$. Label the rest of the endpoints in $S^1 \times \{1\}$ $x_1, \ldots, x_{p-1}$ in order from x_0 as in Figure 3.

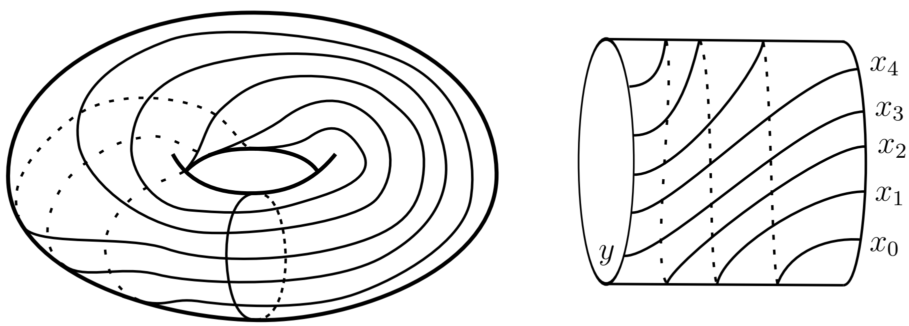


FIGURE 3. The loop shown here suggests the integers $p = 5$ and $q = 3$. The resulting lens space is $L(5, 3)$.

Let q be the integer such that x_q is the endpoint of α in $S^1 \times \{1\}$. The integers p and q with $q < p$ define the arcs $h^{-1}(\lambda)$ up to isotopy. The image of the arcs in Σ will be connected if and only if the greatest common divisor of p and q is 1. Thus we have the following:

3.17. LEMMA. *Given positive integers p and q such that the greatest common divisor of p and q is 1 and $q < p$, there is a loop $\lambda \subset \Sigma$ which defines the ordered pair (p, q) with respect to μ.*

3.18. LEMMA. *The isotopy class of λ uniquely determines a pair of integers (p, q) such that $q < p$ and the greatest common divisor of p and q is 1.*

PROOF. The fundamental group of the torus Σ is $\mathbf{Z} \times \mathbf{Z}$, generated by the loop μ and a loop ν such that $\mu \cap \nu$ is a single point, v and the pre-image $h^{-1}(\nu)$ is a single arc in A from y to x_0. We will think of $\pi_1(\Sigma)$ as ordered pairs (a, b) where a is the number of ν components and b is the number of μ components.

If λ defines the ordered pair (p, q) then $h^{-1}(\lambda)$ consists of p arcs, each of which traverses ν exactly once. Thus $a = p$. Each arc traverses μ a number of times (possibly zero) and then goes q/p of the way along μ. Since there are p arcs, $b = p(q/p + k) = q + kp$.

Thus we have $a = p$ and $b \equiv q \mod p$. The integers a and b are uniquely determined by the isotopy class of λ. Thus p is uniquely determined. The requirement that q be positive and no greater than p implies $q = b$ so q is uniquely determined. □

From now on, instead of writing $M(\mu, \lambda)$, we will write $M(p, q)$ where p and q are the integers uniquely defined by λ with respect to μ. This may seem foolish because the integers p and q do not uniquely determine the isotopy class of λ in ∂H. However, we will see that they do determine the isotopy class up to automorphisms of H. The following theorem implies that the integers p, q completely determine the homeomorphism class of the manifold $M(p, q)$.

3.19. THEOREM. *Let $\lambda \subset \partial H$ and $\lambda' \subset \partial H'$ be loops corresponding to the ordered pairs (p, q) and (p', q'), respectively. Then there is a homeomorphism $H' \to H$ taking λ' to λ if and only if $p = p'$ and $q = q'$.*

Before proving this theorem, a lemma is required:

3.20. LEMMA. *If $\mu, \mu' \subset \partial H$ are essential loops that bound disks in a genus one handlebody H then μ' is isotopic to μ.*

PROOF. Lemma 3.16 implies that we can isotope μ' so that the intersection of μ' with the complement $A = \Sigma \setminus \mu$ is either a closed loop or a collection of arcs with their endpoints on opposite components of ∂A. If μ and μ' are disjoint then μ' is a closed essential loop in the complement of μ, an annulus, so μ' is isotopic to μ. Otherwise, Let D and D' be transverse disks bounded by μ and μ', respectively.

Any point of intersection in $\mu \cap \mu'$ is an endpoint of an arc in $D \cap D'$. Let α be an outermost arc of intersection in D', i.e. assume there is a dist $D'' \subset D$ whose boundary consists of α and an arc $\beta \subset \partial D'$ such that the interior of D'' does not contain any arcs of $D \cap D'$.

The arc β is a subarc of μ' with its endpoints in μ and its interior disjoint from μ. The arc α is contained in D and there is an embedding of $[0, 1] \times [0, 1]$ into a neighborhood of α in D' sending $\{\frac{1}{2}\} \times [0, 1]$ onto α. The arc β intersects this neighborhood in the arcs $[0, \frac{1}{2}] \times \{0\}$ and $[0, \frac{1}{2}] \times \{1\}$. In the complement of μ, these arcs are on the same boundary component of A, so β has both endpoints on the same boundary component. This contradicts the assumption at the beginning of the proof, implying that $D \cap D'$ must be empty. □

PROOF OF THEOREM 3.19. First we will show that if $(p, q) = (p', q')$ then there is a homeomorphism taking λ' to λ. Let A and

A' be annuli and let $h : A \to \partial H$ and $g : A' \to \partial H'$ be maps sending the interiors of A and A' to the complements of μ and μ'.

Because $p = p'$, the number of arcs in $h^{-1}(\lambda)$ and $g^{-1}(\lambda')$ is the same so there is a homeomorphism $\phi : A' \to A$ taking $g^{-1}(\lambda')$ to $h^{-1}(\lambda)$. Because $q = q'$, the image in ϕ of the point y is identified to the image of the point x_0 by the map sending A' onto $\partial H'$. These two points are also identified by the map sending A onto ∂H so ϕ induces a continuous map from ∂H to $\partial H'$. This map takes λ to λ' so Lemma 2.9 implies that ϕ extends to a homeomorphism from H' to H which takes λ' to λ.

Conversely, let λ' and λ be loops and let $\phi : H' \to H$ be a map that sends λ' onto λ. The image $\phi(\mu')$ bounds a disk in H so there is an isotopy of ∂H taking $\phi(\mu')$ onto μ.

After applying this isotopy to $\phi(\lambda')$, the loop will intersect μ in the same pattern as it intersects μ'. Since this patter determines the ordered pair, $\phi(\lambda')$ must define the ordered pair (p', q') with respect to μ. By Lemma 3.18, since $\phi(\mu') = \mu$ we must have $(p', q') = (p, q)$. □

Using the techniques of the previous section, we can calculate the fundamental group of $M(p, q)$. A spine for H_1 has a single edge and μ is the boundary of meridian disk for this edge. The loop λ defined by the integers p and q intersects μ in p points, each with the same orientation. Thus a presentation for the fundamental group is $\pi_1(M(p, q)) = \langle x : x^p\rangle = \mathbf{Z}_p$. We have proved the following:

3.21. Lemma. *The fundamental group of the manifold $M(p, q)$ is the finite cyclic group $\mathbf{Z}_p$.*

Although the homeomorphism class of $M(p, q)$ is determined by the coefficients p, q, it is possible that two lens spaces with different coefficients may be homeomorphic. The fundamental group allows us to differentiate many lens spaces, in particular leading to the following:

3.22. Lemma. *If $M(p, q)$ is homeomorhic to $M(p', q')$ then $p = p'$.*

The fundamental group of S^3 is the trivial group, with presentation $\langle x : x\rangle$. This will be the resulting presentation when $p = 1$. Thus we have proved Lemma 1.23, that $M(p, q)$ is S^3 if and only if $\lambda \cap \mu$ is a single point. For higher values of p, things are not so simple. In fact, there are many lens spaces with different coefficients that are homeomorphic.

3.23. Lemma. *If $q \equiv \pm q' \mod p$ or $qq' \equiv \pm 1 \mod p$ then $M(p, q)$ is homeomorphic to $M(p, q')$.*

Proof. If $q = q'$ then by Theorem 3.19, the loops λ and λ' defined by (p, q) and (p, q') define homeomorphic Heegaard diagrams, so the resulting manifolds are homeomorphic. If $q \equiv -q' \mod p$ then there is an orientation-reversing automorphism of H_1 taking λ to a loop defined by the pair (p, q') and again the manifolds are homeomorphic.

Assume $qq' \equiv 1 \mod p$. We will construct a homeomorphism by "flipping" the two handlebodies, sending μ to λ' and λ to μ'. The loop μ' in $\partial H_2'$ defines a pair of integers (p'', q'') with respect to the loop λ'. The integer p'' is the number of arcs of μ' in the complement of λ'. This is equal to the number of points in $\lambda' \cap \mu'$, which is precisely the number of arcs of λ' in the complement of μ'. Thus $p'' = p$.

Let α be the union of q'' consecutive arcs of λ' in the complement of μ'. The endpoint of α are connected by a single arc in μ'. Since each arc of λ' advances by q' arcs along μ, this implies that $q'q'' \equiv 1 \mod p$. Since q' is relatively prime to p, there is a unique integer with this property. We assumed $q'q \equiv 1 \mod p$ so we must have $q'' = q$.

The endpoints of an arc in $\mu' \setminus \lambda'$ are $q'' = q$ arcs apart along λ'. Thus μ' and λ' (with reversed roles) define the integers (p, q). We showed above that this implies $M(\mu', \lambda') = M(p', q')$ is homeomorphic to $M(p, q)$. A similar analysis works for the case when $q'q \equiv -1$. □

As an example, consider the lens space $L(5, 3)$ shown in Figure 3. The first arc of λ goes from the point labeled x_0 on the left to the point labeled x_3 on the right. The next arc goes from x_3 on the left to the point labeled x_1. Thus the one arc of μ goes from the first point on λ to the third point. Switching the roles of λ and μ defines the lens space $L(5, 2)$, which is what we would expect because $2 * 3 \equiv 1 \mod 5$.

Notice that when $q' \equiv -q \mod p$, the homeomorphism defined in the proof is orientation reversing. When $qq' \equiv 1 \mod p$, the homeomorphism is orientation preserving.

4. Nielsen Equivalence

In a genus-one handlebody, any two essential, properly embedded disks are isotopic. Thus a genus-one Heegaard splitting of a lens space suggests a unique presentation for the fundamental group. In a higher genus handlebody, however, there are an infinite number of non-isotopic systems of disks, and therefore a potentially infinite number of group presentations. We will now see that these presentations are all related by the first three Tietze transformations.

Let (Σ, H_1, H_2) be a genus $g > 1$ Heegaard splitting of a manifold M. Let $K \subset H_1$ be a one-vertex spine for H_1 with edges $e_1, \dots, e_g$ and let $\mathbf{D} = \{D_1, \dots, D_g\}$ be a minimal system of disks for H_2. We have seen that this defines a presentation $\pi_1(M) = \langle x_1, \dots, x_g : r_1, \dots, r_g \rangle$

for the fundamental group of M and this presentation is defined up to type-one Titeze moves.

3.24. LEMMA. *Let* $\mathbf{D}'$ *be a disk slide of* $\mathbf{D}$. *Then the presentation for* $\pi_1(M)$ *coming from* K *and* $\mathbf{D}'$ *is related to the original presentation by a type-two Tietze move and a number of type-one moves.*

PROOF. Without loss of generality, let D_1 and D_2 be the disks involved in a disk slide, defined by an arc $\alpha \subset \Sigma$. Let v_1 and v_2 be the endpoints of α, with $v_1 \in \partial D_1$ and $v_1 \in \partial D_2$.

Let $D'_1, \dots, D'_2$ be the system of disks dual to the spine K of H_1 and let r_1 be the relation defined by the intersections of ∂D_1 with $\partial D'_1, \dots, \partial D'_n$. To produce r_1, we will begin and end at $v_1 \in D_1$. By Lemma 3.12, this relation is conjugate to the relation defined by D_1, so after a type-one Tietze move we can assume r_1 is in the presentation.

Let r_2 be the relation defined by ∂D_2, starting and ending at the point v_2. Let s be the string read from the intersections of α with $\partial D'_1, \dots, \partial D'_n$ in the same way as the relations were read from ∂D_1 and ∂D_2.

The disk $D^* = D_1 *_\alpha D_2$ is defined by a regular neighborhood of $D_1 \cup \alpha \cup D_2$. Its boundary is defined by a regular neighborhood in Σ of $\partial D_1 \cup \alpha \cup \partial D_2$. This loop is parallel to ∂D_1, then the arc α from v_1 to v_2, then ∂D_2 and finally the arc α from v_2 to v_1.

The intersections with $\partial D'_1, \dots, \partial D'_2$ define the string $s^{-1}r_1sr_2$. This relation comes from a type-one move followed by a type-two move. By Lemma 3.12, this string is conjugate to the relation defined by ∂D^*, so removing D_1 and replacing it with D^* corresponds to a type-two Tietze move and two type-one moves. Any presentation defined by the new set of disks is related to this presentation by more type-one moves. □

3.25. LEMMA. *Let* K' *be an edge slide of* K. *Then the presentation for* $\pi_1(M)$ *coming from* K' *and* $\mathbf{D}$ *is related to the original presentation by a type-three Tietze move.*

PROOF. Without loss of generality, assume K' is the result of sliding an edge e_1 along an edge e_2. In $\pi_1(H_1)$, the resulting set of generators is $x_1x_2, x_2, \dots, x_n$ where x_1 and x_2 are the generators defined by the edges e_1 and e_2, respectively. To rewrite the relations in terms of the new generator $x' = x_1x_2$, we simply replace each x_1 with $x'x_2^{-1}$. This is a type-three Tietze move. □

We have seen that any two spines or systems of disks are related by a sequence of edge slides or disk slides, respectively. Applying this to these two lemmas, we have the following Theorem:

3.26. Theorem. *Any two presentations coming from (Σ, H_1, H_2) taking the generators from H_1 and the relators from H_2 are related by a sequence of the first three Tietze moves.*

Proof. Recall that a presentation for $\pi_1(M)$ is defined, up to type-one Tietze moves by a spine for H_1 and a system of disks for H_2. Let K and $\mathbf{D}$ be a spine and a system of disks, respectively, suggesting one presentation and let K', $\mathbf{D}'$ suggest a second presentation.

By Corollary 2.24, any two spines for H_1 are related by a sequence of edge slides. Thus by Lemma 3.25 there is a sequence of type-three Tietze moves turning the presentation coming from K' and $\mathbf{D}'$ into a presentation coming from K and $\mathbf{D}'$.

By Theorem 2.13, the systems of disks $\mathbf{D}$ and $\mathbf{D}'$ are related by a sequence of disk slides, so there is a sequence of type-two Tietze moves turning the presentation into a presentation defined by K and $\mathbf{D}$. This presentation is unique up to type-one Tietze moves. □

Later on, we will define a geometric construction corresponding to the fourth Tietze move. Such a construction will have to increase or decrease the number of generators, and therefore the genus, which means producing a new Heegaard splitting. For now, however, we will consider the different sets of generators that can be induced by spines of a single Heegaard splitting.

The question is: for all possible sets of g generators for $\pi_1(M)$, which of these sets can be induced by a spine for a handlebody of a given Heegaard splitting. We have seen that edge slides correspond to type three Tietze moves. Conversely, for any type three Tietze move there is an edge slide that will induce it. Thus the generating sets induced by the different spines of a handlebody are precisely those that can be produced by these moves. This is the motivation for the following definition:

3.27. Definition. Let $x_1, \ldots, x_g$ and $y_1, \ldots, y_g$ be subsets of a group G such that each subset generates G. The two sets are *Nielsen equivalent* if there is a sequence of type-three Tietze moves that turn the first set into the second set.

Nielsen equivalence is a relation on certain subsets of a given group G, namely subsets containing g elements that generate G. By the nature of its definition, it is an equivalence relation. The importance for Heegaard splittings was first pointed out by Lustig and Moriah [**8**].

3.28. Lemma. *Let (Σ, H_1, H_2) and (Σ', H_1', H_2') be genus g Heegaard splittings for a manifold M. Let K, K' be spines for H_1, H_1', respectively. If Σ is isotopic to Σ' by an isotopy taking H_1 to H_1' then the* *systems of generators for* $\pi_1(M)$ *defined by* K *and* K' *are Nielsen equivalent.*

PROOF. Assume there is an isotopy of M taking H_1' onto H_1. The isotopy does not change the homotopy type of the edges of K' so the Nielsen class defined by the spine is the same after the isotopy.

Thus we can take $H_1' = H_1$ and the graph K' is a spine of H_1. By Corollary 2.24, there is a sequence of edge slides which turn the spine K' of H_1 into the spine K. By Lemma 3.25 and the definition of Nielsen equivalence, edge slides do not change the Nielsen class of the generators. Thus the Nielsen classes defined by K and K' are the same. □

We now have a tool for differentiating two Heegaard splittings. If we can show that the generating sets induced by the handlebodies of one splitting are not Nielsen equivalent to the generating sets defined by the other handlebodies, then the splitting surfaces cannot be isotopic. We will use this at the end of the chapter to find our first example of a 3-manifold with non-isotopic Heegaard splittings of the same genus. But first, we will consider Nielsen class in a familiar example.

Let (Σ, H_1, H_2) be the genus-one Heegaard splitting constructed above for a lens space $M(p,q)$. Let $x \in \pi_1(M)$ be the element defined by the edge of a spine K_1 of H_1. Then $\pi_1(M) = \{1, x, x^2, \ldots, x^{p-1}\}$. Let e_2 be the edge of a spine K_2 of H_2.

3.29. LEMMA. *The edge* e_2 *defines the element* $x^a \in \pi_1(M)$ *where* a *is the unique element of* $\mathbf{Z}_p$ *such that* $aq \equiv 1 \mod p$.

PROOF. Let a be the unique element of $\mathbf{Z}_p$ such that $aq \equiv 1 \mod p$. Assume λ has been isotoped to form p essential arcs in the complement of μ. Let α be the path consisting of a consecutive arcs in λ. The endpoints of α are points in $\lambda \cap \mu$ which are joined by an arc in $\mu \setminus \lambda$. Let ℓ be the loop formed by α and this arc.

By pushing ℓ away from λ along the arc α, we can isotope ℓ to intersect λ in a single point. Thus in H_2, ℓ is a loop in the boundary which intersects the meridian disk in a single point. Pushing ℓ into H_2 produces a spine of H_2 dual to the disk whose boundary is λ.

We can easily calculate the Nielsen class of this spine. The original l was created from a arcs of λ so, after a small isotopy, l can be made to intersect μ in a points, all with the same orientation. By Lemma 3.12, this implies that l defines the element $x^a \in \mathbf{Z}_p = \pi_1(M)$. □

5. The First Homology Group

Recall that the first homology group $\mathscr{H}_1(M)$ of a manifold M is a group of equivalence classes of finite sets of closed, oriented curves in M. Two sets L_1 and L_2 of oriented loops are equivalent if there is a map from an orientable surface S into M sending the boundary of S into $L_1 \cup L_2$ such that the orientations of the loops in L_1 and L_2 induce the opposite orientations for S. The product of the elements of $\mathscr{H}_1(M)$ determined be L_1 and L_2 is simply the union of the two sets. This multiplication makes $\mathscr{H}_1(M)$ a group. See [**5**] for details.

A loop ℓ in M determines the trivial element of $\mathscr{H}_1(M)$ if and only if ℓ bounds a one punctured, orientable surface in M (not necessarily embedded). Every loop that is trivial in $\pi_1(M)$ is trivial in $\mathscr{H}_1(M)$, but a trivial loop in $\mathscr{H}_1(M)$ may not be trivial in $\pi_1(M)$. This suggests that the first homology should be a "smaller" group then the fundamental group. In fact, $\mathscr{H}_1(M)$ can be written as a quotient of $\pi_1(M)$.

Let ℓ be a closed curve in M that bounds an orientable surface S in M. We can construct S from a polygon by removing a disk from the interior and identifying the edges in pairs. The pairs of edges forms closed loops in M that determine elements $a_1, \ldots, a_{2n}$ of $\pi_1(M)$. The loop ℓ is isotopic in the polygon to the boundary of the polygon. In M, this isotopy determines a homotopy from ℓ onto the loop $a_1 a_2 a_1^{-1} a_2^{-1} \ldots a_{2n-1} a_{2n} a_{2n-1}^{-1} a_{2n}^{-1}$.

The *commutator subgroup* of a group G is the subgroup generated by elements of the form $aba^{-1}b^{-1}$ for all $a, b \in G$. (Elements of the form $aba^{-1}b^{-1}$ are called *commutators*.) The construction above shows that every loop in M that is trivial in $\mathscr{H}_1(M)$ is a product of commutators in $\pi_1(M)$. In fact, there is a canonical homomorphism from $\pi_1(M)$ into $\mathscr{H}_1(M)$ whose kernel is precisely the commutator subgroup of $\pi_1(M)$. (See Hatcher's book [**5**] for a proof.) By the first isomorphism theorem, this implies that $\mathscr{H}_1(M)$ is isomorphic to the quotient of $\pi_1(M)$ by its commutator subgroup.

Taking the quotient of a group G by its commutator subgroup is equivalent to adding the relation $aba^{-1}b^{-1} = 1$, or equivalently $ab = ba$ for each $a, b \in G$. In other words, we form a group with the same generators and relations as G, but with the added condition that any two elements commute. The quotient of a group G by its commutator subgroup is, of course, abelian and is called the *abelianization* of G. We will use the expression $[x_1, \ldots, x_n : r_1, \ldots, r_m]$ to denote the abelianization of the group determined by the presentation $\langle x_1, \ldots, x_n : r_1, \ldots, r_m \rangle$.

Because the generators of the group $[x_1, \ldots, x_n : r_1, \ldots, r_m]$ commute with each other, the elements in each relation can be reordered without changing the isomorphism type. In particular, they can be reordered so that each relation is of the form $r_j = x_1^{k_1^j} \ldots x_n^{k_n^j}$ where k_i is an integer (possibly zero or negative) for each i.

We saw that for a closed 3-manifold M, the fundamental group has a balanced presentation $\langle x_1, \ldots, x_g : r_1, \ldots, r_g \rangle$ determined by a Heegaard splitting for M. Because $\mathscr{H}_1(M)$ is the abelianization of $\pi_1(M)$, it has a balanced presentation $[x_1, \ldots, x_g : r_1, \ldots, r_g]$. Define A to be the matrix whose entries are the coeffecients $a_{i,j}$ of the relations $r_j = x_1^{a_{j,1}} \ldots x_g^{a_{j,g}}$.

The matrix A encodes all of the information necessary to reconstruct a presentation for $\mathscr{H}_1(M)$. Like the presentation for $\pi_1(M)$, the entries of matrix depend on the choice of systems of disks for the two handlebodies in the Heegaard splitting. However, as with the presentation for $\pi_1(M)$ there is a simple correspondence between changes to the systems of disks and changes to A.

3.30. LEMMA. *Disk slides in the two handlebodies correspond to elementary row and column operations in A: adding a row or column to a different row or column, or replacing a row or column with its negative.*

PROOF. Given a Heegaard splittings (Σ, H_1, H_2) and minimal systems of disks $\mathbf{D}_1$ and $\mathbf{D}_2$ for H_1 and H_2, respectively, we saw in Section 4 that disk slides in H_1 and H_2 correspond to combinations of the first three tietze transformations on a presentation $\langle x_1, \ldots, x_g : r_1, \ldots, r_g \rangle$ for $\pi_1(M)$. To prove this Lemma, we need only to translate these three moves into moves on the matrix A.

The first Tietze move is to replace a relation r_j with a conjugate $s r_j s^{-1}$ or $s r_j^{-1} s^{-1}$. The entry $a_{j,i}$ of A is determined by adding the powers of the x_is in r_j. For every occurence of x_i in s, it occurs with the opposite power in s^{-1} so these cancel. Thus replacing r_j with $s r_j s^{-1}$ does not affect A. If we replace r_j with $s r_j^{-1} s^{-1}$ then the powers of x_i in s and s^{-1} cancel out, but the powers of x_i in r_j^{-1} are precisely the negatives of the powers in r_j. This transformtion thus corresponds to replacing row j of A with its negative.

For the second Tietze move, we replace a relation r_j with $r_j r_{j'}$ $(j \neq j')$. The sum of the powers of x_i in $r_j r_{j'}$ is the sum of the powers in r_j plus the sum of the powers in $r_{j'}$. Thus the second Tietze move corresponds to adding the j'th row of A to the jth row.

Finally, for the third move we replace a generator x_i with x'_i and in each relation, we replace x_i with $x'_i x_{i'}$ or $(x'_i)^{-1}$. In the first case, for every occurence of x_i, we add an occurence of $x_{i'}$, so the sum of the powers of $x_{i'}$ is the sum of the powers of x_i plus the sum of $x_{i'}$. The sum of the powers of x'_i is the same as that of x_i, so in A this move corresponds to adding the ith column to the i'th column. In the second case, the sum of the powers of x'_i is the negative of the sum of the powers of x_i, so it corresponds to replacing column i with its negative. □

Row and column operations on a matrix preserve the determinant of a matrix. Because any two pairs of systems of disks for H_1 and H_2 are slide equivalent, this means that the determinant of A is determined by the Heegaard splitting, independent of the systems of disks. In fact, the determinate has a lot to do with the topology of the ambient manifold.

3.31. Lemma. *The first homology group of M is finite if and only if the determinant of A is non-zero. Moreover, if the first homology is finite then the order of $\mathscr{H}_1(M)$ is equal to the absolute value of the determinant of A.*

Proof. There is a sequence of row and column operations that turn A into an upper triangular matrix. By Lemma 3.30 there is a sequence of disk slides in H_1 and H_2 that induce the necessary matrix operations, so we can assume that A is upper triangular. In this situation the determinant of A is precisely the product of the diagonal entries.

Let $b = (b_1, \ldots, b_g)$ represent an arbitrary element of $\mathscr{H}_1(M)$. If $a_{1,1}$ is not zero then by adding or subtracting the first relation from b, we can ensure that $0 \leq b_1 < a_{1,1}$. Because the matrix is upper triangular, adding and subtracting the remaining relations will not affect b_1. If $a_{2,2}$ is not zero then by adding and subtracting the second relation, we can ensure that $0 \leq b_2 < a_{2,2}$ and so on for each non-zero diagonal of A.

Each element of $\mathscr{H}_1(M)$ can be written as $(b_1, \ldots, b_g)$ where b_i is bounded by $a_{i,i}$ if $a_{i,i} \neq 0$ and b_i is arbitrary otherwise. Moreover, because A is upper triangular, this form is unique so $\mathscr{H}_1(M)$ is in one-to-one correspondence with g-tuples of this form.

If the determinant of A is zero then some diagonal $a_{i,i}$ is zero and the entry b_i can be arbitrary. This implies $\mathscr{H}_1(M)$ is infinite. If the determinant is non-zero then each diagonal is non-zero, so $\mathscr{H}_1(M)$ is in one-to-one correspondence with the set of g-tuples such that $0 \leq b_i < a_{i,i}$. The number of such g-tuples is the product of the diagonal entries, precisely the determinant of A. □

Before ending this section, we will use this connection between Heegaard splittings and the first homology group to prove the following Lemma:

3.32. Lemma. *If M is a compact, connected, closed, orientable 3-manifold and $\mathscr{H}_1(M)$ is infinite then M contains a two sided, non-separating, embedded surface.*

The converse of this Lemma will be left as an exercise for the reader.

Proof. Let M be a compact, connected, closed, orientable 3-manifold with $\mathscr{H}_1(M)$ infinite. Let (Σ, H_1, H_2) be a Heegaard splitting for M and let A be the matrix for the homology determined by a pair of systems of disks for H_1, H_2. Because $\mathscr{H}_1(M)$ is infinite, Lemma 3.31 implies that A is a singular matrix, so there is a sequence of row and column operations after which the bottom row of A consists entirely of zeros. Let $\mathbf{D}_1$, $\mathbf{D}_2$ be systems of disks for H_1, H_2, respectively that determine the transformed A with empty bottom row.

Recall that the relations are determined by orienting Σ, then orienting the boundaries of each disk in $\mathbf{D}_1$ and $\mathbf{D}_2$. Each point of intersection between a disk of $\mathbf{D}_1$ and a disk of $\mathbf{D}_2$ is called positive if the orientation of Σ induced by the orientations of the loops at that point agrees with the chosen orientation. Otherwise, the intersection is called negative. The entry $a_{i,j}$ is equal to the number of positive intersections between the ith disk in $\mathbf{D}_1$ and the jth disk in $\mathbf{D}_2$ minus the number of negative intersections.

Let $D_g \in \mathbf{D}_2$ be the disk corresponding to the last generator in the presentation. Because the bottom row of A consists of zeros, for each disk $D \in \mathbf{D}_1$, the number of positive and negative intersections between ∂D_g and ∂D are equal.

If $\partial D_g \cap D$ is not empty then there are at least two points in the intersection. Moreover, there must be a positive point of intersection adjacent to a negative point in ∂D. Let $\alpha \subset \partial D$ be an arc with one endpoint at a positive intersection, one endpoint at a negative intersection and interior disjoint from ∂D_g. Let N be the closure of a regular neighborhood of α in H_2.

The set N is a ball and $\partial N \setminus \Sigma$ is a disk E that intersects D_g in two arcs. Let $Q \subset E$ be the square bounded by these two arcs and two arcs in $\partial E (\subset \Sigma)$. Let F_1 be the surface consisting of the complement $D_g \setminus N$ and the square Q. In other words, we construct F_1 by attaching a band to D_g along the arc α.

The boundary of F_1 coincides with the boundary of D_g except at the arc α where rather than crossing ∂D at the endpoints of α, it runs along α. Thus $\partial F_1 \cap \partial D$ contains exactly two fewer points than $\partial D_g \cap F_1$. The following two steps are exercises for the reader:

3.33. Claim. *Because D_g is non-separating in H_1 (it is part of a minimal system of disks), F_1 is also non-separating.*

3.34. Claim. *Because D_g is two sided, one endpoint of α is at a positive intersection and the other is at a negative intersection, F_1 is also two sided.*

For each disk $D \in \mathbf{D}_1$, the number of positive intersections in $\partial F_1 \cap \partial D$ is again zero, so we can repeat the above construction to find a two sided, non-separating surface F_2 whose boundary intersects D in strictly fewer points. Continue the process for each disk in $\mathbf{D}_1$ until we have a surfac F_n whose boundary is disjoint from every disk in $\mathbf{D}_1$.

We saw in the proof of Lemma 2.4 that if a loop in the boundary of a handlebody is disjoint from a system of disks for the handlebody then it bounds a disk in that handlebody. Each component of ∂F_n is a simple closed curve in ∂H_1 disjoint from $\mathbf{D}_1$ so each boundary component of F_n bounds a disk in H_1. These loops are disjoint so they bound disjoint disks. Because F_n is two sided and non separating, the union of F_n and the disks in H_1 is a two sided, non-separating surface embedded in M. □

6. Connect Sums of Lens Spaces

In this section we will construct a manifold admitting two Heegaard splittings of the same genus and use Lemma 3.28 to show that the splittings are not isotopic. Birman [**1**] first showed that this manifold has distinct Heegaard splittings, using a slightly different method than is presented here.

Let M and M' be manifolds, each homeomorphic to $M(7,3)$. Let (Σ, H_1, H_2) and (Σ', H_1', H_2') be Heegaard splittings for M and M', respectively, homeomorphic to the genus-one splitting constructed above for $M(7,3)$. Let $B \subset M$ be a ball embedded so that $B \cap \Sigma$ is a disk in the interior of B. Let $B' \subset M'$ be a similar ball with respect to Σ'. Let $\phi_1 : \partial B \to \partial B'$ be an orientation reversing homeomorphism that sends the loop $\Sigma \cap \partial B$ to $\Sigma' \cap \partial B'$ and the disk $H_1 \cap \partial B$ to $H_1' \cap \partial B'$. Let ϕ_2 be a similar map, but sending $H_1 \cap \partial B$ to $H_2' \cap \partial B'$ and $H_2 \cap \partial B$ to $H_1' \cap \partial B'$.

Define M^1 to be the connect sum $M \# M'$, the result of removing the interiors of B and B' and gluing M to M' along ∂B and $\partial B'$. We will define two distinct Heegaard splittings M^1. In the manifold M^1, let $\Sigma^1 = \Sigma \cup \Sigma' \subset M^1$ and $H_i^1 = H_i \cup H_i'$ for $i = 1, 2$. Similarly, construct

M^2 by gluing along the map ϕ_2, defining $\Sigma^2 = \Sigma \cup \Sigma'$, $H_1^2 = H_1 \cup H_2'$ and $H_2^2 = H_2 \cup H_1'$.

Each of the sets H_j^i is constructed by gluing together handlebodies along disks on their boundaries. The result of such a construction is a new handlebody, so we have the following:

3.35. LEMMA. *The triple* (Σ^1, H_1^1, H_2^1) *is a Heegaard splitting for* M^1 *and* (Σ^2, H_1^2, H_2^2) *is a Heegaard splitting for* M^2*.*

The maps ϕ_1 and ϕ_2 are orientation reversing maps between a pair of 2-spheres, so the maps are isotopic. Gluing along an isotopic map does not change the homeomorphism class. In fact, we can construct a very specific homeomorphism between M^2 and M^1.

3.36. LEMMA. *There is a homeomorphism* $h : M^2 \to M^1$ *that restricts to the identity on* M *and* M' *outside a neighborhood of* $\partial B = \partial B'$*.*

PROOF. Because the maps ϕ_1 and ϕ_2 are both orientation reversing automorphisms between spheres, the composition $\phi = \phi_1^{-1} \circ \phi_2$ is an orientation preserving automorphism of S^2. Such a map is always isotopic to the identity on S^2 so there is an isotopy $\{\Phi_t : S^2 \to S^2\}$ such that Φ_0 is the identity and $\Phi_1 = \phi_1^{-1} \circ \phi_2$.

Let N_1 be the closure of a regular neighborhood in M_1 of ∂B (the sphere along which M and M' are glued.) Let N_2 be a similar neighborhood in M_2. The sets N_1 and N_2 are each homeomorphic to $S^2 \times [0,1]$. The complement of each set is homeomorphic to $(M \setminus B) \cup (M' \setminus B')$ so there is a homeomorphism $h : M_1 \setminus N_1 \to M_2 \setminus N_2$ which is the identity on each component.

Extend h to $N_1 = S^2 \times [0,1]$ and $N_2 = S^2 \times [0,1]$ by taking $h|_{S^1 \times \{x\}} = \phi_1^{-1} \circ \Phi_x$. This map is the identity on $S^2 \times \{0\}$ so h is continuous on $\partial(M \setminus B)$. The map h on $S^2 \times \{1\}$ is $\phi_1^{-1} \circ \phi_2$. This sphere $S^2 \times \{1\}$, N_1 is glued to M' by the map ϕ_1, so its image in h is glued to M' by the map $\phi_1 \circ \phi_1^{-1} \circ \phi_2 = \phi_2$. Thus h is continuous at both ends of $S^2 \times [0,1]$. This completes the proof. □

The image in h of (Σ^2, H_1^2, H_2^2) is a Heegaard splitting for M^1, which we will also denote (Σ^2, H_1^2, H_2^2).

Let K be a spine for H_1 with a single vertex v sitting in ∂B and let K' be a spine for H_1' with a single vertex v' in $\partial B'$. In fact, we can take $v' = \phi_1(v)$ so that $K \cup K' \subset M^1$ is a spine K_1^1 for H_1^1. By a similar construction, we can construct spines K_2^1, K_1^2, K_2^2 for H_2^1, H_1^2, H_2^2, respectively.

The meridian disks for H_2 and H_2' define a minimal system of disks for H_2^1, giving us a presentation, $\pi_1(M^1) = \langle x_1, x_2 : x_1^7, x_2^7 \rangle = \mathbf{Z}_7 * \mathbf{Z}_7$.

This is the fundamental group for M^1 we would expect from Van Kampen's Theorem. The first homology group $\mathscr{H}_1(M)$ is the abelianization $[x_1, x_2 : x_1^7, x_2^7]$ of $\pi_1(M^1)$. This is the abelian group $\mathbf{Z}_7 \times \mathbf{Z}_7$.

Because $\mathscr{H}_1(M)$ is a quotient of $\pi_1(M)$, there is an induced map $\mathbf{Z}_7 * \mathbf{Z}_7 \to \mathbf{Z}_7 \times \mathbf{Z}_7$ sending an element $y \in \pi_1$ to the ordered pair (a, b) where $a, b \in \mathbf{Z}_7$ are the sums of the powers of the x_1s and x_2s, respectively, in y. The pair of edges in each spine suggests a pair of elements, (a, b) and (c, d), which we will represent as a matrix

$$\begin{bmatrix} \mathrm{a} & \mathrm{b} \\ \mathrm{c} & \mathrm{d} \end{bmatrix}.$$

Note that this matrix is different than the matrix A constructed in Section 5. The entries in the matrix A compare the generators in one handlebody to the relations in the other handlebody. The matrix constructed here compares the generators in a handlebody to a fixed set of generators for $\mathscr{H}_1(M)$. In particular, the determinant of this matrix will not, in general equal the order of $\mathscr{H}_1(M)$.

Like the matrix A, row slides in the matrix here correspond to edge slides in the spine and changing the orientation of an edge corresponds to replacing a row with its negative. Thus the determinant of the matrix is an invariant of the handlebody, but rather than containing information about the ambient manifold, it has to do with the Nielsen class of the spine. In particular, we have the following:

3.37. Lemma. *If $\{y_1, y_2\}$ and $\{y_1'.y_2'\}$ are Nielsen equivalent generating sets for $\pi_1(M^1)$ then the determinants of the correspondong matrices (taken over $\mathbf{Z}_7$) are equal.*

The spines of H_1 and H_2 define the elements x_1 and x_1^5, respectively, by Lemma 3.29. Similarly, the spines of H_1' and H_2' define the elements x_2 and x_2^5, respectively. Thus the spines K_1^1, K_2^1, K_1^2, K_2^2 define the generator sets corresponding to the following matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}.$$

The determinants of these matrices (modulo 7) are 1, 3, 5 and 5, respectively.

By Lemma 3.37, we have that the Nielsen class defined by the handlebody H_1^1 is different from both the Nielsen classes defined by H_1^2 and H_2^2. By Lemma 3.28, this implies the following Theorem:

3.38. Theorem. *There is no isotopy of M^1 taking the genus-two Heegaard surface Σ^1 onto the genus-two Heegaard surface Σ^2*

This is our first example of a manifold with two non-isotopic Heegaard splittings of the same genus. This example is not ideal because it is a connect sum. More sophisticated examples of irreducible manifolds with multiple splittings of the same genus are known, but the proofs are much more difficult.

7. Exercises

1. Prove Lemma 3.4.

2. Generalize the method used to prove Lemma 3.32 to show that given a compact manifold M with non-empty boundary, there is a two sided, non-separating surface properly embedded in M.

3. Show that if M contains a two sided, non-separating, incompressible surface then $\mathscr{H}_1(M)$ is infinite. (This is the converse of Lemma 3.32.

4. Prove Claim 3.33.

5. Prove Claim 3.34.

6. Show that the fundamental group of a connect sum of manifolds is the free product of their fundamental groups. What is its first homology group?

7. Prove Claim 3.9.

8. What is the manifold $M(\mu, \lambda)$, when λ is isotopic to μ?

9. Prove Claim 3.16.

11. Prove Lemma 3.37.

12. Generalize the methods used to prove Theorem 3.38 to as many connect sums of lens spaces as possible.

CHAPTER 4

Differential Topology

1. Smooth Manifolds

Up to this point we have been working in the piecewise linear category. Not only did our proof of the existence of Heegaard splittings rely on our 3-manifold having a triangulation, but we have made extensive use of regular neighborhoods, which are defined in terms of a second barycentric subdivision.

At this point, it will be necessary to switch to the smooth category. By introducing a notion of derivatives in our manifolds, we can define Morse functions. For a more in depth discussion of smooth structures on manifolds, the reader should consult Hirsch's *Differential Topology* [**6**], which will be referred to throughout the chapter.

Given a d-dimensional manifold M and a point $p \in M$, there is by definition a neighborhood $N \subset M$ of p homeomorphic to a ball. In other words there is an open ball $B \subset \mathbf{R}^d$ and a homeomorphism $\phi : N \to B$. We can think of the homeomorphism as an embedding of N into $\mathbf{R}^d$.

Given a function $f : M \to \mathbf{R}$, we can define a map $\hat{f} : B \to \mathbf{R}$ by taking $\hat{f}(x) = f(\phi^{-1}(x))$. Because B is an open subset of $\mathbf{R}^d$, partial derivatives of $\hat{f}$ are well defined. This gives us a way to think about partial derivatives of f.

The problem is that the function $\hat{f}$ depends on the map ϕ, which is not unique. In order to define derivatives across the entire manifold, we must fix a collection of compatible maps from subsets of M into $\mathbf{R}^d$.

4.1. Definition. A *chart* is an open ball $N \subset M$ and a one-to-one map $\phi_N : N \to \mathbf{R}^d$.

Because M is a manifold, every point is contained in an open ball and there is always an embedding of such an open ball into $\mathbf{R}^d$. In fact, there are an infinite number of possible charts at any point, given by all embeddings of N into $\mathbf{R}^d$. We need to restrict our attention to a specific collection of charts that "agree" in the areas of overlap.

Let (B, ϕ_B) and (C, ϕ_C) be charts in M, as in Figure 1. Then $\phi_B(B \cap C)$ and $\phi_C(B \cap C)$ are subsets of $\mathbf{R}^3$. Let $A = \phi_B(B \cap C)$. The map $(\phi_C \circ \phi_B^{-1})$ sends A to the image $\phi_C(B \cap C)$.

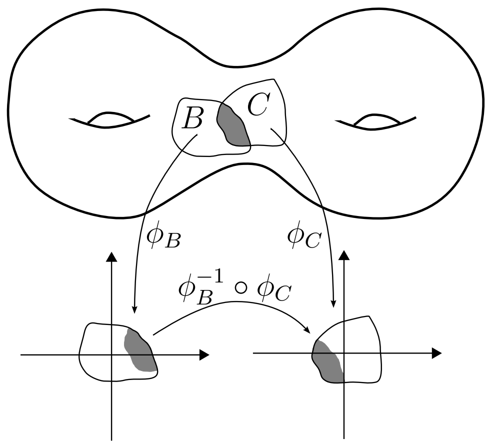


FIGURE 1. A pair of charts are compatible if the composition of the two maps is smooth on the intersection of the open sets.

4.2. DEFINITION. The charts (B, ϕ_B), (C, ϕ_C) are *compatible* if $(\phi_C \circ \phi_B^{-1})$ is smooth (has infinitely many derivatives) as a map from $\mathbf{R}^d$ to $\mathbf{R}^d$.

4.3. DEFINITION. An *atlas* for a manifold M is a collection of charts such that every point in M is contained in a chart and any two charts are compatible. A manifold endowed with an atlas is called a *smooth manifold*.

There is a subtle point which it is important to note. When we say "manifold", we mean a *topological* manifold: a topological space with the property that every point has a manifold-like neighborhood. A *smooth* manifold is a topological manifold with an additional structure added on top: a collection of compatible charts. Keep this subtlety in mind when you read the following theorem:

4.4. THEOREM. *Every compact 2-manifold or 3-manifold is homeomorphic to a smooth manifold.*

We will not prove this theorem here. Note that the statement is not true for higher dimensional manifolds. That is, there are higher dimensional topological manifolds for which any attempt to define a smooth atlas is doomed to failure. From now on, we will assume that any manifold we are dealing with is endowed with a smooth atlas.

Let M be a compact manifold and $f : M \to \mathbf{R}$ a continuous function on M. For a chart (B, ϕ_B), define the function $\hat{f}$ on $\phi_B(B)$ by $\hat{f} = f \circ \phi_B^{-1}$. For a second chart (C, ϕ_C), let $\bar{f}$ be the function $\bar{f} = f \circ \phi_C^{-1}$. Assume there is a point $p \in B \cap C$.

By composing the maps $\hat{f}$ with ϕ_B, we get $\hat{f} \circ \phi_B = f$ on the open set $B \cap C$. By definition, $\bar{f} = f \circ \phi_C^{-1} = \hat{f} \circ \phi_B \circ \phi_C^{-1}$. The map $\phi_B \circ \phi_C^{-1} : \mathbf{R}^d \to \mathbf{R}^d$ is smooth so $\bar{f}$ is the composition of $\hat{f}$ with a smooth function. Thus if $\bar{f}$ has n derivatives at p then so will $\hat{f}$. We have proved the following:

4.5. Lemma. *If $\hat{f}$ has n derivatives at $\phi_B(p)$ then $\bar{f}$ will have n derivatives at $\phi_C(p)$. If $\hat{f}$ is smooth (i.e. has derivatives of all orders) at $\phi_B(p)$, then $\bar{f}$ will be smooth at $\phi_C(p)$.*

This allows us to make the following definition:

4.6. Definition. Given a smooth manifold M, a function $f : M \to \mathbf{R}$ is *smooth* if $\hat{f}$ is smooth for every chart in the atlas.

For a differentiable map $\hat{f} : \mathbf{R}^n \to \mathbf{R}$ and a point $p \in \mathbf{R}^d$, the *gradient* of $\hat{f}$ at p is the vector

$$\nabla_x \hat{f} = \left(\frac{\partial \hat{f}}{\partial x_1}, \ldots, \frac{\partial \hat{f}}{\partial x_n} \right)$$

Where $(x_1, \ldots, x_n)$ are the coordinates in $\mathbf{R}^d$ and the derivatives are taken at p. This vector points in the direction in which $\hat{f}$ increases most rapidly. For a map $f : M \to \mathbf{R}$, it is possible to define the gradient in a way that does not depend on a chart. However, for our purposes we can use a simpler definition.

4.7. Definition. Given a function $f : M \to \mathbf{R}$ and a point $p \in M$, we will write $\nabla_p f = 0$ if $\nabla_p \hat{f} = 0$ in some chart, and $\nabla_p f \neq 0$ otherwise.

We need to check that this definition is useful. Let (B, ϕ_B) and (C, ϕ_C) be charts and $p \in B \cap C$ a point. Let $f : M \to \mathbf{R}$ be a smooth function and let $\hat{f}$ and $\bar{f}$ be the functions on $\mathbf{R}^3$ coming from the charts B and C. Let $x = \phi_B(p)$ and $x' = \phi_C(p)$.

4.8. Lemma. *The gradient $\nabla_x \hat{f}$ will be the zero vector if and only if $\nabla_{x'} \bar{f}$ is also the zero vector.*

Proof. As in the proof of Lemma 4.5, we can write $\bar{f} = \hat{f} \circ g$ where $g = \phi_B \circ \phi_C^{-1}$ is a map from $\mathbf{R}^d$ to $\mathbf{R}^d$. The chain rule from multi-variable calculus implies that $\nabla \bar{f} = (\nabla \hat{f} \cdot A)$ where A is the matrix of partial derivatives of g. That is, entry i, j of A is the partial derivative $\partial g_i / \partial x_j$.

The matrix A is non-singular because g is one-to-one. (This is a non-trivial fact which we will not prove here.) Since the kernel of A is trivial, this implies $\nabla \bar{f} = 0$ if and only if $\nabla \hat{f} = 0$. □

So, $\nabla_p f$ is zero if and only if $\nabla_p \hat{f}$ is zero for every chart.

The *Hessian* $H_x(\hat{f})$ of a function $\hat{f} : \mathbf{R}^n \to \mathbf{R}$ at a point x is a matrix of the second derivatives of $\hat{f}$. For $n = 3$, the matrix is:

$$\begin{bmatrix} \frac{\partial^2 \hat{f}}{\partial x^2} & \frac{\partial^2 \hat{f}}{\partial x \partial y} & \frac{\partial^2 \hat{f}}{\partial x \partial z} \\ \frac{\partial^2 \hat{f}}{\partial y \partial x} & \frac{\partial^2 \hat{f}}{\partial y^2} & \frac{\partial^2 \hat{f}}{\partial y \partial z} \\ \frac{\partial^2 \hat{f}}{\partial z \partial x} & \frac{\partial^2 \hat{f}}{\partial z \partial y} & \frac{\partial^2 \hat{f}}{\partial z^2} \end{bmatrix}$$

Once again, for a function $f : M \to \mathbf{R}$, it is possible to define the Hessian in a way that does not depend on the chart. For our purposes, however, the following definition will suffice:

4.9. Definition. For a function $f : M \to \mathbf{R}$ and a point $p \in M$, we will say that the Hessian is *singular* if in some chart, the determinant of the Hessian is zero.

Once again, there is a Lemma to justify this definition.

4.10. Lemma. *If $\nabla f = 0$ at p and the Hessian is singular in one chart then it will be singular in every chart.*

The proof is very similar to that of Lemma 4.8. The chain rule suggests a transition matrix that gives the Hessian of one chart in terms of the other. Because the transition matrix is non-singular, the composition will be singular if and only if the original Hessian is singular. It is left as an ecercise for the reader to fill in the details of the proof. Note that if the condition $\nabla f = 0$ is removed, the statement is no longer true.

2. Morse Functions

Now that we have defined a way of talking about derivatives on smooth manifolds, we can define Morse functions, the main concept which we will study throughout the rest of the chapter. Let M be a smooth d-dimensional manifold and $f : M \to \mathbf{R}$ a smooth function.

4.11. Definition. A point $p \in M$ is a *critical point* of f if $\nabla_p f = 0$. The value $f(p) \in \mathbf{R}$ is called a *critical value*. If $f^{-1}(a)$ does not contain any critical points then a is a *regular value*.

By Lemma 4.10, at a critical point the expression $|H_p(f)| = 0$ is well defined. The following Theorem is the main ingredient of Morse theory:

4.12. THEOREM (Morse). *Let $p \in M$ and let $f : M \to \mathbf{R}$ be a smooth function. If p is a critical point and the Hessian, is non-singular then there is a neighborhood $N \subset M$ and a map $\phi : N \to \mathbf{R}^d$, compatible with the atlas, sending p to the origin, such that if $\hat{f} = f \circ \phi^{-1}$ then*

$$\hat{f}(x_1, \ldots, x_d) = \pm x_1^2 \pm \cdots \pm x_d^2 + f(p).$$

If p is not a critical point of f then there is a chart (N, ϕ) such that

$$\hat{f}(x_1, \ldots, x_d) = x_1 + f(p).$$

We will not prove this here. This is Theorem 1.1 in Chapter 6 of Hirsch's book in [**6**].

4.13. DEFINITION. A critical point p of a function f is *degenerate* if $|H_p(f)| = 0$ and is *non-degenerate* if $|H_p(f)| \neq 0$.

For a non-degenerate critical point, we can rearrange the variables so that $\hat{f}$ is of the form

$$\hat{f} = x_1^2 + \ldots, x_k^2 - x_{k-1}^2 - \cdots - x_d^2.$$

The integer $d - k$ is the *index* of the critical point and determines the behavior of f at that point.

Notice that for the function $\hat{f}(x_1, \ldots, x_d) = \pm x_1^2 \pm \cdots \pm x_d^2 + f(p)$, the origin is the only critical point. For the function $\hat{f}(x_1, \ldots, x_d) = x_d$, there are no critical points. Thus every regular point and every non-degenerate critical point has a neighborhood containing at most one critical point. This implies the following Corollary of Theorem 4.12:

4.14. COROLLARY. *If f is a function on a compact, smooth manifold M and every critical point of f is non-degenerate then there are a finite number of critical points.*

PROOF. Let M be a compact, smooth manifold and let $f : M \to \mathbf{R}$ be a function such that every critical point of f is non-degenerate. For each point $x \in M$, define an open set N_x as follows: If x is a critical point of f then x is non-degenerate so by Theorem 4.12, there is a neighborhood of x in which x is the only critical point. Let N_x be this neighborhood. If x is not a critical point, let N_x be a neighborhood of x containing no critical points. Again, this neighborhood is implied by Theorem 4.12.

The collection $\{N_x : x \in M\}$ is an open cover for M. Because M is compact, there is a finite sub-cover. Each N_x contains at most one critical point so each set in the finite sub-cover contains at most one critical point. Thus there are a finite number of critical points. □

4.15. DEFINITION. A smooth function $f : M \to \mathbf{R}$ is a *Morse function* if every critical level of f contains exactly one critical point and this critical point is non-degenerate.

By Corollary 4.14, a Morse function contains finitely many critical points, all non-degenerate, and any two critical points are on different critical levels. What makes Morse functions useful are the following two theorems, which will be stated without proof. The first can be summarized as saying that Morse functions are "dense" (every function is pretty close to a Morse function) and the second says that Morse functions are "stable" (you can perturb them without changing them significantly.) The following theorems follow from the $s = \infty$ case of Theorem 1.2 in Chapter 6 of Hirsch's book [**6**].

4.16. THEOREM. *Let M be a smooth manifold and let $g : M \to \mathbf{R}$ be any smooth function. Given $\epsilon > 0$, there is a Morse function $f : M \to \mathbf{R}$ such that for any $x \in M$, $g(x) - f(x) < \epsilon$.*

4.17. THEOREM. *Let M be a compact, smooth manifold and let $f : M \to \mathbf{R}$ be a Morse function on M. Then there is an $\epsilon > 0$ such that for any function $g : M \to \mathbf{R}$, if $f(x) - g(x) < \epsilon$ for every $x \in M$ then there is a map $\phi : M \to M$ and a map $\psi : \mathbf{R} \to \mathbf{R}$, each isotopic to the identity, such that $g = \psi \circ f \circ \phi$.*

Let f be a Morse function. A *level set* of f is a subset $f_x = f^{-1}(x) \subset M$ for some $x \in \mathbf{R}$.

4.18. LEMMA. *If x is a regular value then f_x is a closed submanifold of M of dimension $d - 1$ (where d is the dimension of M.)*

PROOF. Fix x and let $p \in M$ be a point in the level set f_x. In other words, assume $f(p) = x$. Since x is a regular level, p is a regular point and there is a neighborhood N of p and map $\phi : N \to \mathbf{R}^n$ such that $\hat{f} = f \circ \phi^{-1}$ is of the form $\hat{f}(x_1, \dots, x_d) = x_1 + f(p)$.

The map ϕ sends $N \cap f^{-1}(x)$ to a subset of the subspace $x_1 = x$ of $\mathbf{R}^d$. The map is one-to-one so $N \cap f^{-1}(x)$ is homeomorphic to an open subset of $\mathbf{R}^{d-1}$. By taking a smaller neighborhood of p in $f^{-1}(x)$, we can ensure that the neighborhood is, in fact, homeomorphic to a ball in $\mathbf{R}^{d-1}$. Such a neighborhood exists for any point in $f^{-1}(x)$ so this subset is a $(n - 1)$-dimensional manifold embedded in M. □

The subset $f^{-1}(x)$ is closed because it is the pre-image of a closed set in a continuous function. If we choose M to be a compact manifold then $f^{-1}(x)$ must also be compact. If there is an interval in $\mathbf{R}$ with no critical levels then the surfaces in the pre-image must fit together in a fairly simple way. The proof of the following Lemma can again be found in Section 6.2 of Hirsch's book [**6**].

4.19. LEMMA. *Let $a, b \in \mathbf{R}$ be regular values such that $a < b$ and let N be the submanifold $f^{-1}(a)$. If there are no critical levels in the interval $[a, b]$ then $f^{-1}([a, b])$ is homeomorphic to $N \times [a, b]$.*

A *sub-level set* is everything "below" a level set. Define $f_x^- = f^{-1}((-\infty, x])$. If f_x is a regular level then f_x^- will be an n-dimensional submanifold and if M is closed then $f_x = \partial f_x^-$.

4.20. LEMMA. *Let $a, b \in \mathbf{R}$ be regular values such that $a < b$. If there are no critical levels in the interval $[a, b]$ then there is an isotopy of M taking the sub-level set f_a^- onto f_b^-.*

PROOF. Let c be a regular value that is less than a and such that there are no critical values in the interval $[c, a]$. This value exists because there are a finite number of critical values. Then there are and no critical values in the interval $[c, b]$. By Lemma 4.19, $f^{-1}[c, b]$ is homeomorphic to $N \times [c, b]$ (where $N = f^{-1}(a)$) and $f^{-1}[c, a]$ is the subset $N \times [c, a]$ of $N \times [c, b]$.

There is an isotopy of the interval $[c, a]$ to the interval $[c, b]$. This isotopy lifts to an isotopy of $f^{-1}([c, a])$ onto $f^{-1}([c, b])$ and then to an isotopy of f_a^- onto f_b^-. □

The final step in understanding the nature of Morse functions is to define an analogous combinatorial object.

4.21. DEFINITION. In a d-dimensional manifold, a *k-handle* is a ball parameterized as $D^k \times D^{d-k}$.

Let N be a d-dimensional manifold with boundary. Let h be a k-handle and $\phi : (D^k \times \partial D^{n-k}) \to \partial N$ a one-to-one map. Let N' be the result of gluing h to N by the map ϕ. If $k = 0$ then N' is just the disjoint union of N and the closed ball h. We say that N' is the result of attaching a k-handle to N.

4.22. LEMMA. *Let $a, b \in \mathbf{R}$ be values such that $a < b$. If there is a single critical level in $[a, b]$ and this level contains an index-k critical point then f_b^- is isotopic to the result of adding a k-handle to f_a^-.*

Rather than proving this Lemma in the general case, we will illustrate it in dimensions 2 and 3 in the following sections.

3. Morse Functions on Surfaces

In dimension two there are three possible indices for critical points: 0, 1 and 2, defined by the polynomials $x^2 + y^2$, $x^2 - y^2$ and $-x^2 - y^2$, respectively. To visualize these types of points, consider a surface embedded in $\mathbf{R}^3$ so that f is a height function, i.e. $f(x, y, z) = z$. The three types of points are shown in Figure 2. These types of points will be called a *minimum*, a *saddle* and a *maximum*, respectively. Minima and maxima are also called *central singularitities*.

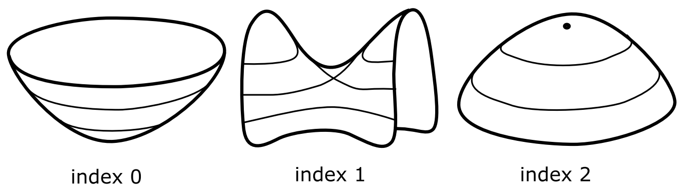


FIGURE 2. The three types of critical points for a Morse function on a surface are shown as height functions.

On a closed surface, a regular level is a collection of loops. Recall that by definition, there is at most one critical point in a critical level of a Morse function. A critical level set containing an index-zero or index-two critical point will consist of a (possible empty) collection of loops and a point. A critical level with an index-one critical point will consist of a figure eight (two circles attached at a point) and possibly one or more loops.

4.23. THEOREM. *If f is a Morse function on a compact, closed surface S then there is a cell decomposition for S such that each k-dimensional cell contains exactly one critical point and this critical point has index k.*

To prove this theorem, we will prove a slightly more general, and more technical lemma:

4.24. LEMMA. *If f is a Morse function on a compact, closed surface S then for each regular level x, there is a cell complex embedded in the sub-level set f_x^- such that:*

(1) Each k-dimensional cell contains exactly one critical point and this critical point has index k.

(2) The complement in f_x^- of the cell complex is a neighborhood of $f^{-1}(x)$ (the boundary of f_x^-) homeomorphic to $f^{-1}(x) \times [0, 1]$.

PROOF. We will prove this by induction on the number of critical points in f. Let $S' = f_x^-$. For the base step assume there are no critical values below x. The set f_x^- is closed in S and therefore compact. Thus its preimage $f(f_x^-)$ is a compact subset of $(-\infty, x]$. If this set is non-empty then there is a minimum value $y \in (-\infty, x]$. The image $f(y)$ cannot contain a regular value because a regular value cannot be a minimum. It cannot contain a critical point because we assumed there were no critical points below x. Thus there cannot be a minimum y, so $f(f_x^-)$ is empty. The sub-level set f_x^- is the empty set and the cell complex called for by the Lemma is the empty cell complex.

For the induction step, assume there are $n > 0$ critical values below x. Let c be the largest critical value below x. In other words, assume that if b is a critical value of f then $b \leq c$. Because there are a finite number of critical values, there is a regular value y such that c is the only critical value greater than y.

There is a single critical point p such that $f(p) = c$ so there are exactly $n - 1$ critical points in the sub-level set $S'' = f_y^-$. By the inductive assumption, there is a cell complex embedded in S'' meeting conditions (1) and (2). (Note that this may be the empty cell complex and S'' the empty set.) There are three cases to consider, defined by the possible indices for the critical point p.

Case 1: If p is an index-zero critical point then we see from Figure 2 that there is a disk D properly embedded in S' whose image in f is the interval $[c, x]$. The complement in S' of D contains the same $n-1$ critical points as S''. By Lemma 4.20, the sub-surface $S' \setminus D$ is isotopic to S'' in S. To construct a cell complex for S', start with the cell complex for S'' and add a vertex at p.

Case 2: If p has index one then $f^{-1}[y, x]$ is a sub-surface of S containing a single index-one critical point. By Lemma 4.19, every component of $f^{-1}[y, x]$ that does not contain a critical point is homeomorphic to an annulus (a loop cross an interval.) As seen in Figure 2, the component containing p is the result of attaching a one-handle to an annulus. This produces a a three-punctured sphere.

There is an arc that passes through p and has both endpoints in $f^{-1}(y)$. This arc cuts the component into one or two annuli which are neighborhoods of the $f^{-1}(x)$ boundary components. In S', the arc can be extended and attached to vertices of the cell complex. (Note that there are often multiple ways to extend the arc.) The cell complex resulting from extending the arc is the required cell complex for S'.

Case 3: If the index of p is two then there is a disk D containing p such that $f^{-1}(D) = [y, c]$. By the induction assumption, there is an annulus in the complement of the cell complex that is a neighborhood of ∂D. The union of D and this annulus is a disk (a 2-cell.) To get a cell complex for S', attach this cell to the cell complex for S''. □

PROOF OF THEOREM 4.23. The surface S is compact so its image in f is a compact subset of $\mathbf{R}$. Thus there is an $x \in \mathbf{R}$ such that $f(p) < x$ for every p in S. In particular, this implies $f^{-1}(x)$ is the empty set. By Lemma 4.24, there is a cell complex embedded in S whose complement is a neighborhood of $f^{-1}(x)$. Since $f^{-1}(x)$ is empty, the complement of the cell complex is empty, so it is a cell decomposition for S. □

The converse of Theorem 4.23 is also true. That is, given a cell decomposition of a surface it is possible to construct a Morse function with an index-i critical point for each i-dimensional cell of the decomposition. In order to better illustrate the construction, we will consider an example.

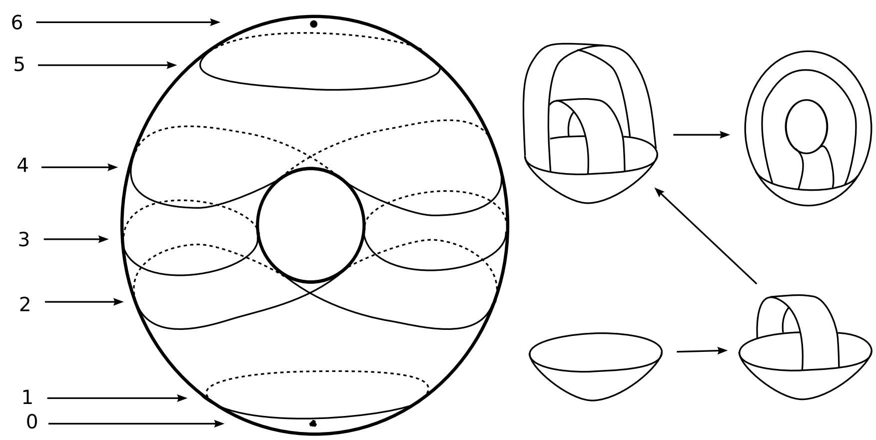


FIGURE 3. The height function on a standardly embedded torus defines a Morse function. Some of the level sets are shown here.

4.25. EXAMPLE. *The torus*

Figure 3 shows a torus T^2 embedded in $\mathbf{R}^3$. The height function $f(x, y, z) = z$ in $\mathbf{R}^3$ restricts to a Morse function on T^2. The level sets for certain values are shown.

At level 0 there is an index-zero critical point, a minimum. At level 2 is an index-two critical point, a saddle. In between, the level sets are loops. As the loops approach the figure eight at level 2, two points squeeze together, splitting the level set into two loops, between 2 and 4.

There is a second saddle point at level 4 at which the two loops are joined and become a single loop. Between 4 and 6, these loops shrink onto the maximum at level 6.

The levels 1, 3 and 5 are regular levels. The sub-level set below 1 is a disk, a 0-handle. Below 3, the level set is an annulus. This is homeomorphic to the result of attaching a 1-handle to a 0-handle along two arcs, as shown on the right side of Figure 3. The level set below 5 results from adding a second 1-handle to form a punctured torus and the level set below 7, the entire torus, results from filling in the puncture with a 2-handle.

The definition of a smooth structure on a manifold does not take into account the possibility of a boundary. Given a manifold M with boundary, the interior of M is a non-compact manifold and (at least in dimensions two and three) can be given a smooth structure. However, because the smooth structure doesn't extend to the boundary, we have to be very careful when we define Morse functions on these manifolds.

There are two ways to define Morse functions on a surface S with boundary. A *proper Morse function* f is a function that is Morse on the interior of S with the property that some neighborhood N of the boundary can be parameterized as $\partial S \times [0, 1]$ so that $f(a, x) = x$. In other words near the boundary, level sets of a proper Morse function will be boundary parallel loops.

To define the second type of Morse function on a manifold with boundary, we need to introduce some terminology. Let S be a surface with boundary. The *double* S^d of S is the surface resulting from gluing two copies of S together by the identity map on their boundaries. We can choose an atlas for S^d that is compatible with the atlases on each copy of S.

If $f : S \to \mathbf{R}$ is a function on S, then there is an induced function f^d on S^d. An *improper Morse function* f on a surface S with boundary is a function such that $f|_{\partial S}$ is a Morse function (on the 1-manifold) and the induced function on the double of S is Morse. Critical points on the boundary in an improper Morse function are shown in Figure 9.

4.26. Lemma. *If $f : S \to \mathbf{R}$ is an improper Morse function on S then a critical point x of the one-dimensional Morse function $f|_{\partial S}$ corresponds to a critical point in the two-dimensional Morse function f^d.*

Proof. Let p be a point in ∂S. By definition, there is a neighborhood N of p that is homeomorphic to the upper half plane in $\mathbf{R}^2$. In the double S^d of S, there is a neighborhood of p homeomorphic to a disk in $\mathbf{R}$ such that ∂S is sent to the horizontal axis.

The induced function $\hat{f}$ is symmetric across the horizontal axis so the vertical component of $\nabla \hat{f}$ must be zero. By assumption, p is a critical point of f restricted to the boundary. Thus the horizontal component of $\nabla \hat{f}$ is zero. This implies $\nabla \hat{f} = 0$ and p is a critical point of the double. □

The point x will be called a *half saddle* if the corresponding critical point of f^d is a saddle point. It will be called a *lower half disk* if the corresponding critical point of f^d is a minimum and an *upper half disk* if the corresponding critical point of f^d is a maximum.

4. Morse Functions on 3-Manifolds

Unlike functions on surfaces, Morse functions on 3-manifolds cannot be visualized as height functions. Instead, we will visualize the function by drawing the level sets within a ball in the 3-manifold.

There are four types of critical points in a 3-dimensional Morse function, defined by the following polynomials:

$$\textit{index } 0 : f(x, y, z) = x^2 + y^2 + z^2 + f(0)$$
$$\textit{index } 1 : f(x, y, z) = x^2 + y^2 - z^2 + f(0)$$
$$\textit{index } 2 : f(x, y, z) = x^2 - y^2 - z^2 + f(0)$$
$$\textit{index } 3 : f(x, y, z) = -x^2 - y^2 - z^2 + f(0)$$

Figure 4 shows the level sets of a Morse function near these four types of points.

4.27. Lemma. *Let $f : M \to \mathbf{R}$ be a Morse function on a closed 3-manifold M. Let a be a regular value of f and assume that all the critical points of f in the sub-level set $f^{-1}((-\infty, a])$ have index 0 or 1. Then every component of $f^{-1}((-\infty, a])$ is a handlebody.*

Proof. Let $p_1, \ldots, p_n$ be the critical points below a and let $x_1, \ldots, x_{n-1}$ be regular values such that $p_i < x_i < p_{i+1}$. Because M is closed, any sub-level set below p_1 is empty. By Lemma 4.22, $f^-_{x_i}$ is the result of attaching a 0-handle or 1-handle to $f^-_{x_{i-1}}$.

In three dimensions, attaching a 0-handle is equivalent to taking the disjoint union with a closed 3-ball. Attaching a 1-handle is equivalent to gluing a ball to the manifold along two pairs of disjoint disks. The result is orientable because M is orientable. Thus each component is a handlebody, the result of gluing together disjoint balls along pairs of disks. □

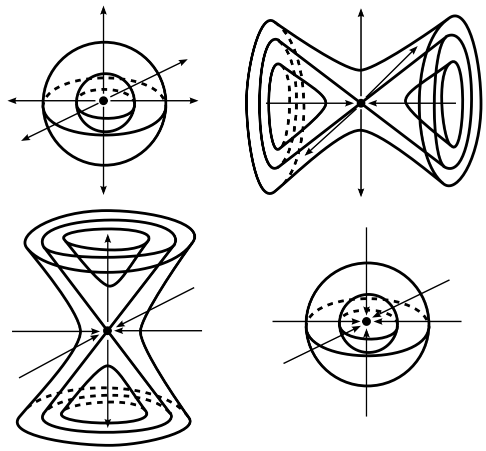


FIGURE 4. The four types of critical points for a Morse function on a 3-manifold are shown via level sets. The arrows indicate the directions in which the function is increasing.

If we replace f with the function $-f$ (the function that takes on the value $-f(x)$ at each point x) then index-0 critical points become index-3 critical points, and vice versa. Index-1 critical points switch with index-2 critical points, giving us the following immediate corollary of Lemma 4.27:

4.28. COROLLARY. *Let $f : M \to \mathbf{R}$ be a Morse function on a compact, closed, orientable 3-manifold M. Let a be a regular value of f and assume that all the critical points of f in the set $f^{-1}([a, \infty))$ have index 2 or 3. Then every component of $f^{-1}([a, -\infty))$ is a handlebody.*

Combining these two Lemmas, we get the following:

4.29. THEOREM. *Let $f : M \to \mathbf{R}$ be a Morse function on a compact, connected, closed manifold M. Assume there is a value $a \in \mathbf{R}$ such that every critical point below the level set f_a has index 0 or 1 and every critical point above f_a has index 2 or 3. Then the surface f_a is a Heegaard surface for M.*

PROOF. By Lemma 4.27, the sub-level set $H_1 = f_x^-$ is a collection of handlebodies and by Corollary 4.28, $H_2 = f^{-1}[x, \infty]$ is a collection of handlebodies. Let $\Sigma = f^{-1}(x)$. Then (Σ, H_1, H_2) will be a Heegaard splitting if and only if Σ is connected.

The ambient manifold M is connected (and path connected) so if Σ is not connected then there is a path α from one component to another. The arc α is a union of subarcs, each of which is properly embedded in H_1 or H_2. Each component of H_1 is a handlebody, so the endpoints of any arc are in the same component of Σ. The same is true for the arcs in H_2, so Σ is connected and (Σ, H_1, H_2) is a Heegaard splitting of M. □

This tells us that for a large class of Morse functions, there is an associated Heegaard splitting. There is a converse to this theorem that associates to each Heegaard splitting one or more Morse functions.

If M is a manifold with boundary, we will say that a function $f : M \to \mathbf{R}$ is a *proper Morse function* if f is constant on ∂M, a Morse function on the interior of M and near ∂M, level sets of f are surfaces parallel to the boundary.

4.30. Lemma. *Let H be a handlebody with a smooth atlas defined on its interior. There is a proper Morse function $f : H \to \mathbf{R}$ such that f is constant on ∂H and all the critical points of f have index 0 or 1.*

The function can be constructed from a system of disks for H, as a converse to the proof of Lemma 4.27. The details will be left to the reader, who may wish to consult the upcoming example in Secton 6. The proof of the following theorem will also be an exercise for the reader.

4.31. Theorem. *Let (Σ, H_1, H_2) be a Heegaard splitting for a closed manifold M. Then there is a Morse function $f : M \to \mathbf{R}$ such that at least one level surface of f is isotopic to Σ.*

There are similar lemmas and theorems for manifolds with boundary.

4.32. Lemma. *Let M be a compact, orientable manifold, possibly with boundary, and let $f : M \to \mathbf{R}$ be a (proper) Morse function. Let a and b ($a < b$) be regular levels of f and assume all the critical points in the set $f^{-1}([a, b])$ have index 0 or 1. Then each component of $H = f^{-1}([a, b])$ is a compression body with $\partial_+ H = f_b$ and $\partial_- H = f_a$.*

The proof is almost identical to the proof of Lemma 4.27, except that rather then starting from the empty set, we begin with a collection of closed surfaces, $f^{-1}(a)$. The subset $f^{-1}[a, b]$ is the result of attaching 0-handles and 1-handles. By definition, each component of $f^{-1}[a, b]$ is a compression body.

This lemma implies analogues to Lemma 4.28 and Theorem 4.29, which we will not state here.

5. Generalized Heegaard Splittings

Given a generic Morse function, there is no reason to expect the index-0 and index-1 critical points to come before the index-2 and index-3 critical points, as in the assumptions of Theorem 4.29. However, we can always split the level sets into intervals which contain either all index-0 and index-1 critical points or all index-2 and index-3 critical points.

The level sets between these regions are no longer Heegaard splittings, but by Lemma 4.32, they split M into compression bodies, suggesting a new structure closely related to Heegaard splittings. This idea was first suggested by Scharlemann and Thompson [**14**] and expanded upon by Schultens [**15**].

Given a 3-manifold, M, let $\mathbf{H_1} = \{H_1^1, \dots, H_1^k\}$ be a collection of pairwise disjoint sets such that each H_i is one or more compression bodies embedded in M. Let $\mathbf{H_2} = \{H_2^1, \dots, H_2^k\}$ be a second collection of pairwise disjoint compression bodies in M. Let $\mathbf{\Sigma} = \{\Sigma^1, \dots, \Sigma^k\}$ and $\mathbf{T} = \{T^1, \dots, T^{k-1}\}$ be collections of pairwise disjoint closed surfaces in M. (We will not require that any Σ^i or T^i be connected.)

4.33. Definition. The 4-tuple $(\mathbf{\Sigma}, \mathbf{T}, \mathbf{H}_1, \mathbf{H}_2)$ is a *generalized Heegaard splitting* if $\partial_- H_1^1 \cup \partial_- H_2^k = \partial M$ and for each $i \leq k$, $\partial_- H_1^i = T_{i-1} = \partial_- H_2^{i-1}$ and $\partial_+ H_1^i = \Sigma^i = \partial_+ H_2^i$. The interiors of all the compression bodies H_j^i must be pairwise disjoint and their union must be all of M.

A schematic diagram comparing a Heegaard splitting to a generalized Heegaard splitting is shown in Figure 5. A surface T^i is often called a *thin surface* and each Σ^i is called a *thick surface*. This is due to the fact that the genera of the components of T^i are less than the genera of Σ^i and Σ^{i+1}.

Note that a generalized Heegaard splitting with $k = 1$ is just a Heegaard splitting. For $k > 1$, the thin surfaces $\bigcup T^i$ cut the manifold in a collection of manifolds with boundary and the thick surfaces Σ^i form Heegaard splittings for these manifolds. The following Lemma implies that a generalized Heegaard splitting gives a combinatorial model of a Morse function.

4.34. Lemma. *If $f : M \to \mathbf{R}$ is a Morse function then there is a generalized Heegaard splitting $(\mathbf{\Sigma}, \mathbf{T}, \mathbf{H}_1, \mathbf{H}_2)$ of M such that each Σ^i and each T^i is a level set $f^{-1}(x)$ for some regular value x.*

Proof. Given a Morse function f on M, there are regular levels $t_1, \dots, t_k$ and $s_1, \dots, s_{k-1}$ such that $t_i < s_i < t_{i+1}$, between t_i and s_i

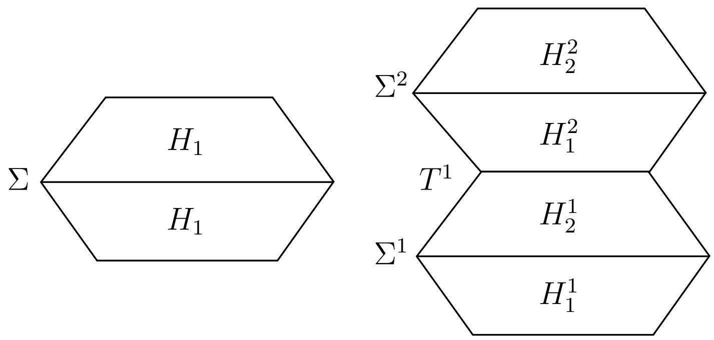


FIGURE 5. The schematic diagrams compare a Heegaard splitting to a generalized Heegaard splitting. Horizontal lines represent surfaces in the manifold. The top and bottom lines of each figure represent the boundary (if any) of the manifold. The remaining arcs represent the surfaces that cut the manifold into compression bodies. For each compression body, the longer arc represents ∂_+.

there are only index-0 and index-1 critical points, and between s_i and t_{i+1} there are only index-2 and index-3 critical points.

By Lemma 4.32, each set $f^{-1}[t_i, s_i]$ is a collection of compression bodies. By applying the same lemma to $-f$, $f^{-1}[s_i, t_{i+1}]$ is a collection of compression bodies. For each $i \leq k$, define $\Sigma^i = f^{-1}(s_i)$, $T^i = f^{-1}(t_i)$, $\mathbf{H}_1 = \{f^{-1}[s_i, t_i]\}$ and $\mathbf{H}_2 = \{f^{-1}[s_i, t_{i+1}]\}$. Then $(\mathbf{\Sigma}, \mathbf{T}, \mathbf{H}_1, \mathbf{H}_2)$ is a generalized Heegaard splitting of M. □

Given a generalized Heegaard splitting for M, we can construct a (standard) Heegaard splitting by a process called amalgamation. The reverse construction, producing a generalized Heegaard splitting from a standard splitting, will be discussed in the next chapter.

Let $(\mathbf{\Sigma}, \mathbf{T}, \mathbf{H}_1, \mathbf{H}_2)$ be a generalized Heegaard splitting and assume $k > 1$. In this case, H_2^1 is a compression body. We will think of H_1^2 as the union $H_1^2 = P^1 \cup Q^1$ where P^1 contains $\partial_- H_1^2$ and is homeomorphic to $\partial_- H_1^2 \times [0,1]$. The subset Q^1 is homeomorphic to a collection of pairwise disjoint balls which intersect P^1 in disks. In other words, we think of constructing H_1^2 by starting with P^1 and gluing on the 1-handles in Q^1. Every compression body has such a decomposition.

Let $\mathbf{D}^1$ be the collection of disks $P^1 \cap Q^1$ and let X be the set $\mathbf{D}^1 \times [0,1]$ in $P^1 = \partial_- H_1^2 \times [0,1]$. In other words, extend disks to tubes from one boundary component to the other, as in Figure 6. Then $X \cap \partial_- H_1^2$ is a collection of pairwise disjoint disks in $\partial_- H_1^2 = T_1 = \partial_- H_2^1$. The compression body H_2^1 has a similar decomposition $H_2^1 =$

$P^2 \cup Q^2$. Because $(P^2 \cap Q^2)$ is a collection of disks, we can assume that $(P^2 \cap Q^2) \times [0,1]$ is disjoint from X in T_1.

Let $\mathbf{D}^2$ be the collection of disks $X \cap \partial_- H_2^1$. We can extend X into H_2^1 by attaching $Y = \mathbf{D}^2 \times [0,1]$ in $P^2 = \partial_- H_2^1 \times [0,1]$. Notice that $Q^1 \cup X \cup Y$ is homeomorphic to Q^1 (a collection of balls) and intersects $\partial_+ H_2^1 = \Sigma_2 = H_1^1$ in a collection of disks. Thus $H_2' = H_1^1 \cup (Q^1 \cup X \cup Y)$ is a handlebody. The construction is illustrated in Figure 6.

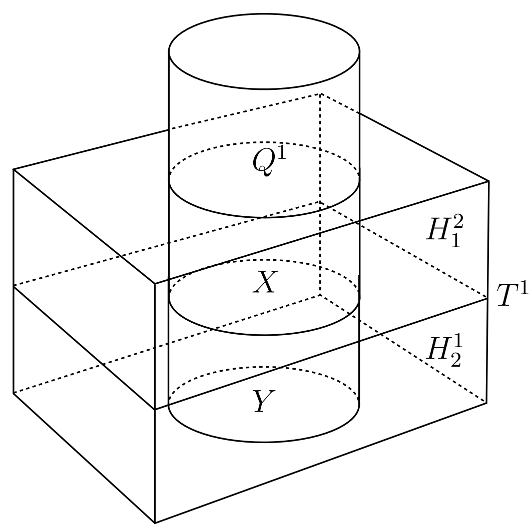


FIGURE 6. To construct an amalgamation of a generalized Heegaard splitting, we extend the one-handles (P^1) of H_2^1 through the surface cross interval parts of H_2^1 and H_1^2, adding the pieces X and Y.

We constructed the handlebody H_2' by extending the 1-handles in H_1^2 through H_1^2 and H_2^1 and attaching them to H_1^1. Similarly, we can extend the handles in H_2^1 in the opposite direction, through H_2^1 and H_1^2, and attach them to H_2^2. The resulting set H_2' is a compression body.

The complement in each of H_1^2 and H_2^1 of H_1' and H_2' is a punctured surface cross an interval. The complement of H_2^1 is parallel to the boundary of H_1' so the union H_1'' of H_1 and the complement of H_2^1 of a handlebody. Similarly, union of H_1^2 and the complement of H_1^2 is a compression body.

Replacing H_1^1 and H_1^2 by H_1'' and replacing H_2^1 and H_2^2 by H_2'' produces a new generalized Heegaard splitting in which k is smaller. If we continue the process, we will eventually produce a Heegaard splitting (Σ, H_1, H_2).

Notice the following aspects of the construction: (1) If we switch the roles of H_2^1 and H_1^2 in the construction, the resulting generalized Heegaard splitting is isotopic to the result of the original construction. (2) We chose to collapse the splitting along the thin surface T_1, but the construction can be used to collapse along any of the thin surfaces.

4.35. DEFINITION. If (Σ, H_1, H_2) is the result of the process described above then (Σ, H_1, H_2) is an *amalgamation* of $(\mathbf{\Sigma}, \mathbf{T}, \mathbf{H}_1, \mathbf{H}_2)$. The generalized Heegaard splitting is a *telescope* of (Σ, H_1, H_2).

If we collapse a generalized Heegaard splitting along one or more (but not all) of its thin surfaces, the result is a new generalized Heegaard splitting, which we will also call an *amalgamation* of the original. The original generalized splitting is a *telescope* of the new generalized splitting.

(Note: In the literature, the word "untelescope" is often used where we say "telescope". The prefix "un-" seems only to make things cumbersome and removing it will hopefully not cause any confusion.)

The process of amalgamation allows us to associate a Heegaard splitting for M to every Morse function on M. It also proves to be a more natural construction for gluing together manifolds.

4.36. PROPOSITION. *Let M_1 and M_2 be manifolds with boundary and let M be the result of gluing M_1 to M_2 along a genus-g boundary component of M_1 and a genus-g boundary component of M_2. If M_1 has Heegaard genus g_1 and M_2 has Heegaard genus g_2 then M has Heegaard genus at most $g_1 + g_2 - g$.*

PROOF. Let (Σ, H_1, H_2) be a genus-g_1 Heegaard splitting of M_1. Let (Σ', H_1', H_2') be a genus-g_2 splitting of M_2. The gluing induces an embedding of Σ, Σ' and the four compression bodies into M. Let T be the image in M of the boundary components along which M_1 and M_2 are glued.

Define $\mathbf{\Sigma} = \{\Sigma, \Sigma'\}$, $\mathbf{T} = \{T\}$, $\mathbf{H}_1 = \{H_1, H_1'\}$ and $\mathbf{H}_2 = \{H_2, H_2'\}$. Then $(\mathbf{\Sigma}, \mathbf{T}, \mathbf{H}_1, \mathbf{H}_2)$ is a generalized Heegaard splitting of M. Let (Σ'', H_1'', H_2'') be an amalgamation of this generalized Heegaard splitting. By construction, the genus of Σ'' is $g_1 + g_2 - g$. □

The converse of this proposition is not necessarily true. In other words, given manifolds with genera g_1 and g_2, it is possible that the result of gluing the manifolds along a boundary component is less than $g_1 + g_2 - g$. Understanding when the genus drops and how far it can drop is an open problem.

6. More of the 3-Torus

Recall that the 3-torus, T^3, is the result of gluing together opposite faces of a cube. In Section 1.8 we constructed a genus three Heegaard splitting (Σ, H_1, H_2) for T^3 and in Section 3.2, Lemma 3.14, we showed that this is a minimal genus Heegaard splitting.

We will construct a Morse function f for T^3 such that Σ is isotopic to a level surface of f. The sub-level sets of this Morse function are shown in Figure 7. Begin with an index-0 critical point in the center of the cube, creating a ball. The ball expands until it is almost tangent to the left and right walls of the cube. There is an index-one critical point where the cube touches the left and the right walls of the cube. In T^3 these walls are identified so the critical point turns the ball into a solid torus.

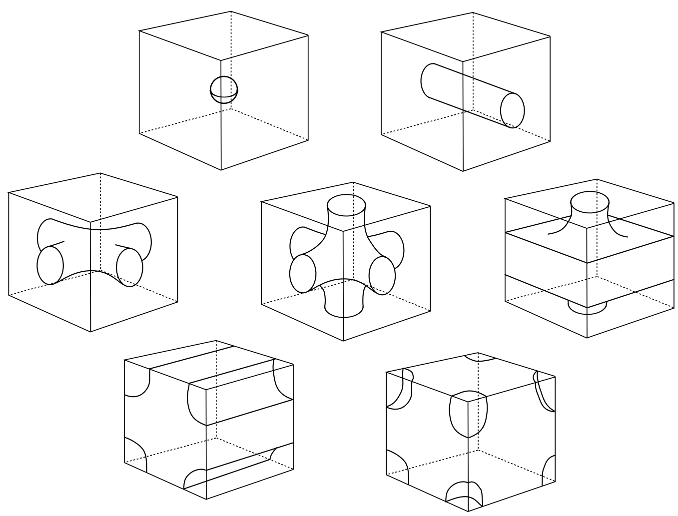


FIGURE 7. The sub-level sets of a Morse function corresponding to the standard Heegaard splitting in the 3-torus.

As the solid torus expands, there is a second index-one critical point at the front and back walls of the cube, forming a genus-two handlebody in T^3. A third critical point at the top and bottom walls forms a genus-three handlebody. At this point, any level surface is isotopic to Σ.

As the genus-three handlebody continues to expand, it begins to approach the four vertical edges in the cube. In T^3, these edges form a single edge of K and the handlebody is shrinking around this edge. As it collapses onto the edge, there is an index-two critical point which adds a two-handle. The sub-level set is no longer a handlebody.

As the sub-level set continues expanding, there is a second index-two critical point around a second edge of K, and then a third critical point around the last edge of K. The complement of the sub-level set is now a ball which contains the vertex of K. In the cube, the sub-level set looks like the result of removing the eight corners. The Morse function ends with an index-three critical point as the sub-level set swallows the vertex of K.

There is a second way to construct a Morse function on T^3 with the same critical points, but in a different order. The sub-level sets are shown in Figure 8. We once again begin with an index-zero critical point in the center of the cube. After this, there is an index-one critical point as the ball touches the left and right walls, then a second index-one critical point at the front and back walls.

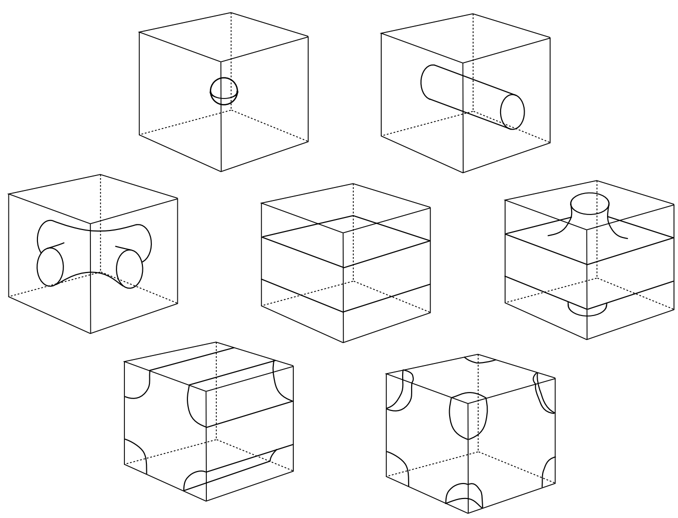


FIGURE 8. The sub-level sets of a second Morse function on the 3-torus. For this Morse function, none of the level sets are Heegaard surfaces.

After this, rather than having a third index-one critical point, the genus-two handlebody defined by the sub-level sets will pinch the edge of K defined by the four vertical edges of the cube, forming an index-two critical point. After this critical point, the sub-level set is homeomorphic to a torus cross and interval. The level set at this point is a pair of tori.

After the index-two critical point, we have the index-one critical point at the top and bottom faces of the cube. This turns the pair of tori into a single genus-two surface. This is followed by two index-two critical points at the two remaining edges of K, and then an index-three critical point at the vertex.

None of the level sets of this second Morse function are Heegaard surfaces for T^3 because the highest genus of any level surface is two and by Lemma 3.14, any Heegaard surface of T^3 must have genus at least three. The reason Theorem 4.29 fails is that the index-two critical point at the vertical edge of K comes before the index-one critical point at the upper and lower faces of the cube. However, Lemma 4.34 implies that there should be a generalized Heegaard splitting of T^3 whose surfaces come from level surfaces of this Morse function. We will now describe this splitting.

Recall that there is a spine K' for H_1 consisting of a vertex in the center of the cube and three edges, each going through a pair of faces. Let H_1^1 be the closure of a regular neighborhood of the two horizontal edges. Consider horizontal squares S_1 and S_2 in the cube, just above and below H_1^1. These form tori in T^3. Let H_2^1 be the result of thickening these tori, then attaching a 1-handle along the edge of K formed by the vertical edges of the cube. This can be done so that $\partial_+ H_2^1 = \partial H_1^1$.

Next consider the tori in T^3 formed by a pair of squares just above and below S_1 and S_2. Thicken these tori and attach them by a 1-handle through the top and bottom faces of the cube. The result is a compression body H_1^2 such that $\partial_- H_1^2 = \partial_- H_2^1$. (The boundary $\partial_- H_1^2$ is a pair of tori formed by horizontal squares in the cube.)

The complement of the three compression bodies defined so far is a regular neighborhood of the edges of K formed by the horizontal edges of the cube. This final handlebody H_2^2 completes the generalized Heegaard splitting. Notice that we can construct this generalized splitting so that all the surfaces involved are level sets of the second Morse function.

7. Sweep-outs

4.37. Definition. Let M be a compact, closed manifold. A function $f : M \to [-1, 1]$ is a *sweep-out* of M if the following axioms hold:

(1) For each $a \in (-1, 1)$, $f^{-1}(a)$ is a closed surface S embedded in M.

(2) For $a = +1$ and $a = -1$, $f^{-1}(a)$ is a graph embedded in M.

(3) For each $a, b \in (-1, 1)$ such that $a < b$, $f^{-1}([a, b])$ is homeomorphic to $S \times [a, b]$.

(4) For each $a \in (-1, 1)$, $f^{-1}([-1, a])$ is a handlebody and $f^{-1}([a, 1])$ is a handlebody.

The axioms are very restrictive, but they are necessary for our purposes. In particular, they immediately imply the following two propositions:

4.38. PROPOSITION. *If f is a sweep-out of a closed manifold M then for $a, b \in (-1, 1)$, the surfaces $f^{-1}(a)$ and $f^{-1}(b)$ are homeomorphic and are isotopic in M.*

4.39. PROPOSITION. *If f is a sweep-out of a closed manifold M then for any $a \in (-1, 1)$, the triple $(f^{-1}(a), f^{-1}([-1, a]), f^{-1}([a, 1]))$ is a Heegaard splitting for M.*

Compare the second Corollary to Theorem 4.29. The reason sweep-outs are important is the converse of this:

4.40. LEMMA. *If (Σ, H_1, H_2) is a Heegaard splitting of a closed manifold M then there is a sweep-out $f : M \to [-1, 1]$ such that $f^{-1}(0) = \Sigma$.*

PROOF. Let K_1 be a spine of H_1. The complement $H_1 \setminus K_1$ is homeomorphic to $\partial H_1 \times [0, 1)$. Let ϕ be a homeomorphism from $H_1 \setminus K_1$ to $\Sigma \times (-1, 0]$ (because ∂H_1 is homeomorphic to Σ.) Let $p : \Sigma \times (-1, 0] \to (-1, 0]$ be projection onto the $(-1, 0]$ factor. Then $f_1 = p \circ \phi : (H_1 \setminus K_1) \to (-1, 0]$ is a map such that the preimage of any point is a surface homeomorphic to Σ.

There is a similar map $f_2 : (H_2 \setminus K_2) \to [0, 1)$ where K_2 is a spine of H_2. Notice that $f_1(\Sigma) = 0 = f_2(\Sigma)$. Define $f : M \to [-1, 1]$ to send K_1 and K_2 to 1, -1, respectively and to agree with f_1 and f_2 on H_1 and H_2, respectively. This maps satisfies the axioms of a sweep-out and $f^{-1}(0) = \Sigma$. □

4.41. LEMMA. *Let $f : M \to [0, 1]$ be a smooth sweep-out and let F be a smooth surface embedded in M. Then there is an isotopy of F after which the restriction $f|_F$ is a Morse function on F.*

The proof of this Lemma is beyond the scope of our discussion, so we will take it for granted.

A sweep-out is very similar to a Morse function in the sense that away from the spines of the sweep-out and the critical points of the Morse function, both look locally like a surface cross an interval. On the other hand, the two types of functions are in some ways opposites. In fact, if we choose a sweep-out that is also a smooth function then the set of critical points will be the union of $f^{-1}(-1)$ and $f^{-1}(1)$. So, rather than a finite number of critical points, we have uncountably many, all of which are degenerate.

8. Lens Spaces, Again

In this section, we will employ the results from our discussion of Morse functions and sweep-outs to completely classify genus-one Heegaard splittings of most lens spaces. Higher genus splittings of lens spaces will be considered in the next chapter, with Theorem 5.44. The following theorem was first proved by Bonahon and Otal [**2**], but the proof presented here is based on techniques used by Schultens [**16**].

4.42. Theorem. *Let M be a lens space, other than S^3 or $S^1 \times S^2$. Let (Σ, H_1, H_2) and (Σ', H_1', H_2') be genus-one Heegaard splittings of M. Then there is an isotopy of M taking Σ' onto Σ.*

This is true for S^3 and $S^1 \times S^2$ as well, but we don't have the machinery to prove that quite yet. The main consequence is a converse to Lemma 3.23:

4.43. Corollary. *Lens spaces $L(p, q)$ and $L(p', q')$ with $p > 1$ are homeomorphic if and only of $p = p'$ and either $q' = \pm q$ or $qq' \equiv \pm 1$* mod *p.*

Proof. By Lemma 3.23, we know that if $p = p'$ and $q \equiv \pm q'$ mod p then $L(p, q)$ is homeomorphic to $L(p', q')$. To prove the converse, assume there is a homeomorphism $\phi : L(p', q') \to L(p, q)$. Let (Σ, H_1, H_2) be the standard Heegaard splitting for $L(p, q)$ and let (Σ', H_1', H_2') be the standard Heegaard splitting for $L(p', q')$.

Because ϕ is a homeomorphism, the images $\phi(\Sigma')$, $\phi(H_1')$, and $\phi(H_2')$ form a Heegaard splitting of $L(p, q)$. We will write (Σ', H_1', H_2') for this induced Heegaard splitting as well as for the splitting of $L(p', q')$.

By Theorem 4.42, the splitting surfaces Σ and Σ' are isotopic. If the isotopy takes H_1' onto H_1 then a meridian disk for H_2 becomes a meridian disk for H_2'. The boundary loops form the same loop on H_1 and H_1' so they define the same integers and $p' = p$, $q' = q$.

If the isotopy takes H_1' to H_2 then a meridian disk for H_2' comes from H_1 rather than H_2. As we saw in Section 3.3, the boundary of this disk in ∂H_2 defines integers p', q' such that $p' = p$ and $qq' \equiv \pm 1$ mod p. □

4.44. Definition. A loop ℓ embedded in the interior of a compression body H is a *core* of H if for a regular neighborhood N of ℓ, the complement $H \setminus N$ is a compression body.

Note that if H is a solid torus and ℓ is a core of H then the complement of N is a thickened torus, so ℓ is also a spine of H.

4.45. Lemma. *Let D be a disk properly embedded in a compression body H and let $\ell \subset \partial_+ H$ be a loop embedded so that $\ell \cap \partial D$ is a single point. Then ℓ is isotopic (in H) to a core of H.*

Proof. Let N be the closure of a regular neighborhood in H of $D \cup \ell$. Because $D \cap \ell$ is a single point, the set N is a solid torus and $\partial N \setminus \partial H$ is an open disk. Let D' be the closure of this disk. Then D' is properly embedded and separates H into a compression body and the solid torus N.

Let ℓ' be the result of pushing ℓ into the interior of N and let N' be a regular neighborhood of ℓ' in N. Then $N \setminus N'$ is a compression body (a torus cross an interval) and $H \setminus N'$ is the result of gluing together two compression bodies along a disk. This is a compression body so ℓ' is a core of H. □

Theorem 4.42 follows from the following Lemma, whose proof will take up most of the rest of this section.

4.46. Lemma. *Let (Σ, H_1, H_2) be a genus-one Heegaard splitting of a lense space M that is not homeomorphic to S^3 or $S^1 \times S^2$. If (Σ', H_1', H_2') is a Heegard splitting of M then a spine ℓ of H_1 is isotopic to a core of either H_1' or H_2'.*

Proof. Let D' be a meridian disk for H_2 and let ℓ be a spine for H_1. Because $p \geq 2$, the loop $\partial D'$ in ∂H_1 will intersect the boundary of a meridian disk at least twice. Let $A = S^1 \times [0,1]$ be an annulus. Because M is not S^3 or $S^1 \times S^2$, there is a map $\phi' : A \to H_1$ that is one-to-one on the interior of A, sends $S^1 \times \{0\}$ onto $\partial D'$ and sends $S^1 \times \{1\}$ onto ℓ. We can assume that the image of $S^1 \times \{1\}$ is locally one-to-one. It will wind around ℓ at least twice.

By attaching the meridian disk D' to the annulus A, we can find a map ϕ' from a disk D into M such that ϕ' is one-to-one on the interior of D and such that ∂D winds around ℓ at least twice. Let $f : M \to [-1,1]$ be a sweep-out of M induced by the Heegaard splitting (Σ', H_1', H_2') and let $\hat{f} = f \circ \phi : D \to [-1,1]$ be the push-back of f onto D.

By isotoping ℓ, we can assume that the restriction of $\hat{f}$ to ℓ is a Morse function. Isotope ℓ and D further so as to minimize the number of critical points in this Morse function.

4.47. Lemma. *By isotoping the interior of D in M, we can assume that $\hat{f}$ is an improper Morse function on D.*

Assume D and ℓ have been isotoped so that $\hat{f}$ is a Morse function and the number of critical points has been minimized. By Lemma 4.26, every critical point p of $\hat{f}|_{\partial D}$ defines a critical point of the double of $\hat{f}$

in D^d. As before, we will say that p is a *lower half disk* if the induced critical point is an index-zero critical point, a *half-saddle* if it is an index-one critical point (a saddle singularity,) or an *upper half-disk* if the critical point has index-two in D^d. These are shown in Figure 9.

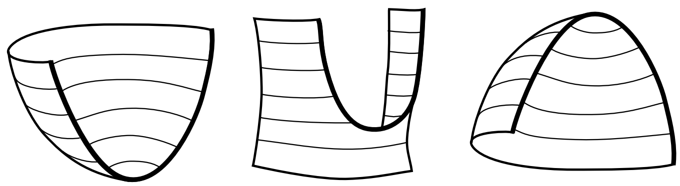


FIGURE 9. The critical points in the double D^d define, from left to right, lower half disks, half saddles and upper half disks near the boundary of D.

4.48. LEMMA. *The disk D can be isotoped so that all the critical points in ∂D are upper or lower half disks.*

PROOF. Let $p \in \partial D$ be a half-saddle and without loss of generality, assume p is a minimum of $\hat{f}$ restricted to ∂D. Near p, D lies below p as in Figure 9. If we push D up past p, as in Figure 10, we turn the half saddle into a lower half disk and create a saddle singularity in the interior of D. This construction reduces the number of half saddles so, by induction, we can eliminate all half-saddles in ∂D. □

Let s be a saddle singularity in the interior of D at level x. The level set $S = \hat{f}^{-1}(x)$ is almost a 1-manifold in the sense that the complement of s is a collection of arcs and loops. From Figure 2, we see that s forms a valence-four vertex in $f^{-1}(x)$.

The arcs that leave s can either loop back to s, or they can terminate at ∂D, as in Figure 11. Because s is the only critical point at level x,

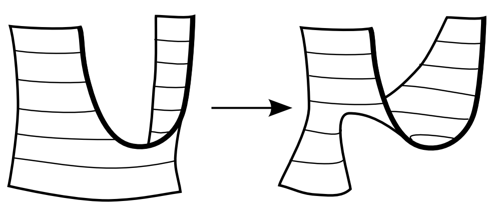


FIGURE 10. A small isotopy of D changes a half saddle into a saddle singularity in the interior and a half disk on the boundary.

the arcs cannot terminate at a different critical point. We will call s an *essential saddle* if all four arcs terminate at ∂D. Assume D has been isotoped so that $\hat{f}$ is an improper Morse function and there are no half-saddles in ∂D.

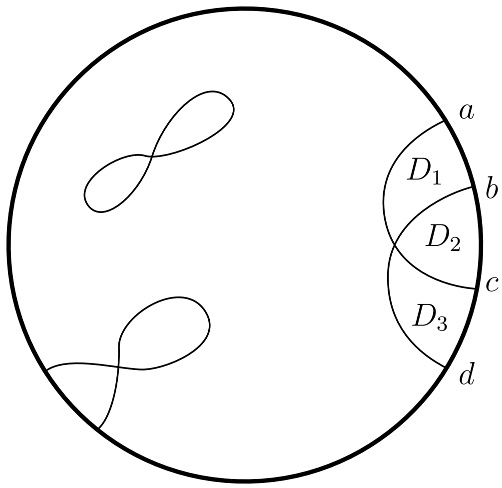


FIGURE 11. A saddle singularity sits in one of three kinds of level components. The level set shown at the far right is an essential saddle.

4.49. LEMMA. *The interior of D contains an essential saddle.*

PROOF. Assume for contradiction there are no essential saddles. Then for each saddle singularity, s, in the interior of D, one of the arcs leaving s returns to s, forming a loop α. By the Jordan Curve Theorem, the loop α bounds a disk D' in D. A proper Morse function on a disk must contain an index zero or an index two critical point. The restriction of f to D' is proper so D' contains a central singularity.

If there are no saddles in the disk D' then there is a single central singularity c in D'' and we can connect s to c by an arc. If there is a single saddle s' in D'' then the arcs comming out of s' bound two disks, each of which contains a single central singularity. If s' is connected to one of these point then we can conect s to the other. Continuing in this fashion, we can connect each saddle in the interior of D to a unique central singularity.

Let f^d be the double of $f|_D$ on the double D^d of D. This surface is closed so by Lemma 4.24, there is a cell complex for S^2 with cells corresponding to the critical points of f^d. The number of index one critical points coming from the interior of D is less than or equal to the nuber of central singularities in the interior. Because $p > 1$, there are at least two index-zero critical points coming from the lower half disks in ∂D and at least two index-three critical points from the upper half disks.

The Euler characteristic of a sphere is 2. The number of vertices and 2-cells in the cell complex coming from f^d is greater than the number of edges by at least four. This suggests an Euler characteristic of at least 4. The contradiction implies that there must be an essential saddle in D. □

This last Lemma is the final ingredient that allows us to prove Lemma 4.46. For each essential saddle s_i (and there is at least one,) let X_i be the union of s_i and the four arcs that terminate at ∂D. Each X_i cuts D into four disks. An *outermost essential saddle* is a component X_i such that three of the complementary disks do not contain any essential saddles. There must be at least one outermost essential saddle. Let X be this saddle and let D_1, D_2, D_3 be the complementary disks which do not contain essential saddles.

The arcs that leave s terminate in points a, b, c, d in ∂D as in Figure 11. Assume for contradiction the points a and c map to the distinct points a' and c' in ℓ. We will show that the number of critical points in ℓ is not minimal.

Isotope the arc $a'c'$ in ℓ along disks D_1 and D_2 to the level arc in X from a to c. Isotoping ℓ slightly further puts it into Morse position, as in Figure 12 with two fewer critical points in ℓ. Because we assumed the number of critical points was minimized, this is impossible so a and c must map to the same point in ℓ. Similarly, b and d must map to the same point in ℓ. This implies that the arc in ∂D from a to c maps onto ℓ.

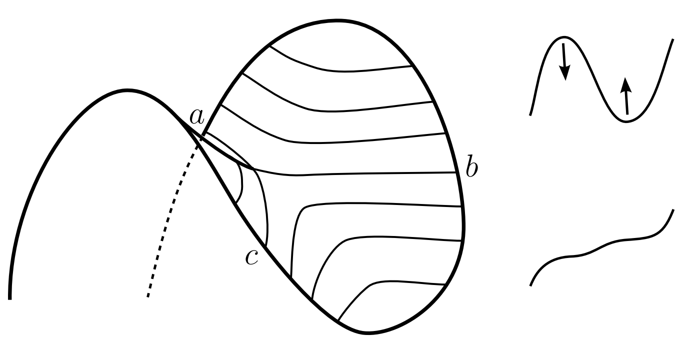


FIGURE 12. If points a and c map to distinct points in ∂D, then we can reduce the number of critical points.

The arcs from a to b and from c to d map to the same arc of ℓ so the interior of $D_1 \cup D_3$ is an open a disk D'' in M, shown in Figure 13. Define $\bar{D}''$ to be the closure of D'' in M. The set $\bar{D}'' \setminus D$ is contained in a level set of f and is homeomorphic to a figure eight (a graph with one vertex and two edges).

Sliding the loop ℓ along the disks D_1 and D_2 brings it onto one of the loops in the figure eight. Near the vertex of the figure eight, we can isotope D_1 and D_2 away from each other so that their union is a new disk D'''. This disk will intersect in an arc. If we continue to slide D_1 and D_2, we can make $\partial D'''$ transverse to ℓ.

A choice of orientation for D'' induces an orientation on D_1 and D_2. Because of the direction in which ab and cd are glued together, these orientations will disagree in D. Thus at the vertex of $\bar{D}'' \setminus D$, the two disks come together with a twist, as in Figure 13. When the loops are made transverse, there will be a single point of intersection. By Lemma 4.45, ℓ is a core of H_1 or H_2. □

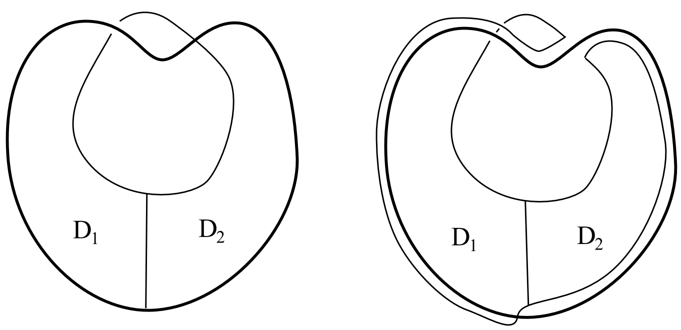


FIGURE 13. The disks D_1 and D_2 come together with opposite orientations at the vertex. This creates a twist so that when $\partial D'''$ is made transverse to ℓ, there is a single point of intersection. (In order to simplify the figure, the surface containing ℓ is not shown.)

PROOF OF THEOREM 4.42. Let (Σ, H_1, H_2) and (Σ', H_1', H_2') be genus-one Heegaard splittings of a Lens space $L(p, q)$ with $p > 1$. (The condition $p > 1$ is equivalent to requiring that $L(p, q)$ not be S^3 or $S^1 \times S^2$.) By Lemma 4.46, a spine ℓ of H_1 is a core of either H_1' or H_2'.

The handlebodies H_1' and H_2' have genus one so a loop that is a core is also a spine. Thus the loop ℓ is isotopic to both a spine of H_1 and a spine of either H_1' or H_2'. The spines are isotopic to Σ is isotopic to Σ'. □

CHAPTER 5

Reducibility

1. Stabilization

Let M be a 3-manifold, (Σ, H_1, H_2) a Heegaard splitting of M and let $\alpha \subset H_1$ be the image of an embedding $e : [0,1] \to H_1$ that sends the endpoints into ∂H_1 and sends the interior into the interior of H_1.

We will say that α is *unknotted* if there is a disk D embedded in H_1 such that ∂D consists of a single arc β in ∂H_1 and the arc α in the interior of H_1. Notice that D is not properly embedded in H_1 because its boundary is not entirely on ∂H_1. Sliding α along the disk D defines an isotopy of α onto $\beta \subset \partial H_1$. In fact, a properly embedded arc $\alpha \subset H_1$ is unknotted if and only if it is isotopic into ∂H_1.

Let $N_1 \subset H_1$ be a closed regular neighborhood of α and let N_2 be a closed regular neighborhood of $D \setminus N_1$ in $H_1 \setminus N_1$, as in Figure 1. Each of N_1, N_2 is a ball and $N_1 \cup N_2$ is a regular neighborhood of D in H_1, also a ball.

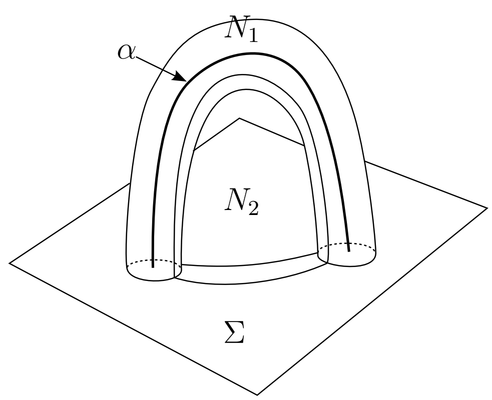


FIGURE 1. An unknotted arc in H_1 defines a stabilization of (Σ, H_1, H_2).

Define $H_2' = H_2 \cup N_1$. This set is a handlebody because H_2 is a handlebody and N_1 is a ball that intersects H_2 in two disks. Let $\Sigma' = \partial H_2'$.

Let H_1' be the closure in M of $H_1 \setminus N_1$. The set $H_1 \setminus (N_1 \cup N_2)$ is a handlebody because $N_1 \cup N_2$ is a ball and intersects H_1 in a single disk. The set N_2 is a ball that intersects this handlebody in two disks and $H_1 \setminus N_1 = (H_1 \setminus (N_1 \cup N_2) \cup N_2)$. Thus H_1' is a handlebody and (Σ', H_1', H_2') is a Heegaard splitting of M.

5.1. Definition. Any Heegaard splitting constructed by iterating the above process one or more times, possibly reversing the roles of H_1 and H_2, is a *stabilization* of (Σ, H_1, H_2).

Note that the construction depends only on the arc α, not on the disk D. Stabilization allows us to compare Heegaard splittings of different genera. If two splittings do not have the same genus, then they can never be isotopic, but we can ask if one of the splittings is a stabilization of the other. We can also ask if, given two splittings, there is a third splitting that is isotopic to a stabilization of each. In 1935, Reidemeister [**10**] and Singer [**17**] independently proved the following following Theorem:

5.2. Theorem (Reidemeister-Singer [**10**],[**17**]). *Let (Σ, H_1, H_2) and (Σ', H_1', H_2') be Heegaard splittings (not necessarily of the same genus) of a compact 3-manifold M. Then there is a Heegaard splitting surface Σ'' that is isotopic to both a stabilization of Σ and a stabilization of Σ'.*

The proof of this theorem will be presented at the end of the chapter, after we have developed a little more machinery. For now, we will consider the basic properties of stabilization.

5.3. Theorem. *If (Σ', H_1', H_2') and (Σ'', H_1'', H_2'') are stabilizations of the same Heegaard splitting (Σ, H_1, H_2) and Σ', Σ'' have the same genus then Σ' is isotopic to Σ''.*

Proof. First we will assume that each splitting Σ', Σ'' is a single stabilization of Σ, i.e. the result of carrying out the above process exactly once. The general result follows from this case by induction. Let α be the arc defining the stabilization Σ' and let D' be a disk with one boundary arc on α and the other in Σ. Let β and D'' be the arc and disk defining Σ''.

Without loss of generality we can assume that α lies in H_1. If β does not lie in H_1, we would like to find an alternative arc in H_1 that also defines Σ''. Let N_1'' be a regular neighborhood of β and let N_2'' be a regular neighborhood of $D'' \setminus N_1''$.

As we saw in the construction above, $N_1'' \cup N_2''$ is a ball B' which intersects Σ in a disk $E \subset \partial B$. Let E' be the disk $\partial B \setminus E$. Sliding E across B defines an isotopy of E onto E'. This isotopy extends to all of Σ so that after the isotopy, Σ will intersect ∂B in E' instead of E.

Before the isotopy, N_1 intersects Σ in a pair of disks and N_2 intersects $\partial N_1 \cup \Sigma$ in an annulus. After the isotopy, N_2 intersects Σ in a pair of disks and N_1 intersects $\partial N_2 \cup \Sigma$ in an annulus. In other words, the isotopy changes N_1 from a 1-handle to a 2-handle and changes N_2 from a 2-handle to a 1-handle.

Let β' be an arc in N_2 from one component of $N_2 \cap E'$ to the other such that N_2 is a regular neighborhood of β'. After the isotopy of Σ, β' is properly embedded and unknotted in H_1. We can then isotope Σ back to its original position and isotope β' with it. The splitting defined by β' before the second isotopy of Σ is precisely (Σ'', H_1'', H_2''). Thus the splitting defined by β' after the second isotopy is isotopic to (Σ'', H_1'', H_2'').

Thus, without loss of generality we assume α and β lie in H_1. As above, construct balls B', B'' containing α and β, respectively and intersecting Σ in disks $E' \subset B'$ and $E'' \subset B''$, respectively.

By Lemma 1.12, there is an isotopy of Σ that takes E' to E''. This isotopy can be extended to an isotopy of M that brings B' to B''. Another small isotopy brings N_1' to N_1'' and it follows immediatly that the composition of these isotopies take Σ' to Σ''. □

Note that there is a statement very similar to this theorem that is not true. If Σ is a stabilization of Σ' and Σ is a stabilization of Σ'' then it is not necessarily the case that Σ' and Σ'' are isotopic.

As mentioned above, we would like to restrict our attention to Heegaard splittings that are not stabilizations. We will now define criteria to determine when a Heegaard splitting is a stabilization of some other splitting. Consider the pair of disks in H_1' and H_2', shown in Figure 2, suggested by the handles N_1, N_2 that were added to H_1 and H_2. The disk in H_2' is given by $D_2' = D \cap H_2' = D \setminus N_1$. The disk $D_1' \subset H_1$ is a meridian disk of α. The boundaries of these disks intersect in a single point in Σ'.

5.4. Definition. A Heegaard splitting (Σ', H_1', H_2') is *stabilized* if there is a pair of essential, properly embedded disks $D_1' \subset H_1'$, $D_2' \subset H_2'$, such that $\partial D_1' \cap \partial D_2'$ consists of a single point in Σ'.

5.5. Lemma. *A Heegaard splitting* (Σ', H_1', H_2') *is stabilized if and only if it is a stabilization of a Heegaard splitting* (Σ, H_1, H_2).

Proof. From the above discussion, we know that if Σ' is a stabilization then Σ' is stabilized. To prove the converse, we need to construct a Heegaard splitting Σ from the disks D_1' and D_2' and show that Σ' is a stabilization of Σ.

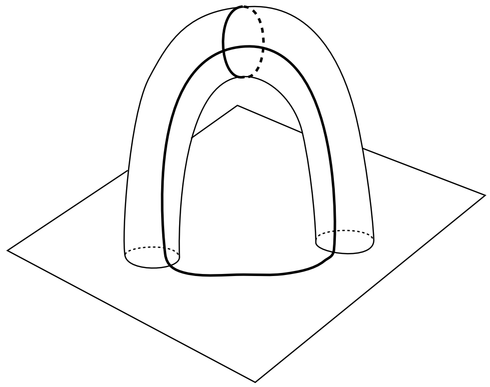


FIGURE 2. After stabilization, the resulting handlebodies contain disks whose boundaries intersect in a single point.

Let N_1 be a regular neighborhood of D_1 in H_1' and let N_2 be a regular neibhborhood of D_2 in H_2'. Let $B = N_1 \cup N_2$. Each of N_1 and N_2 is homeomorphic to an open ball and $N_1 \cap N_2$ is a single disk (since $D_1 \cap D_2$ was a single point) so B is a ball.

Define $H_2 = H_2' \setminus N_2$ and H_1 the closure of $H_1' \cup N_2 = (H_1' \setminus N_1) \cup B$. Then H_2 is a handlebody because H_2' is a handlebody and N_2 is a neighborhood of a properly embedded disk. Also, H_1 is a handlebody because $H_1' \setminus N_1$ is a handlebody and H_1 is the result of attaching the ball B along a disk.

Define $\Sigma = \partial H_1$. Then (Σ, H_1, H_2) is a Heegaard splitting. It follows from the construction that there is a properly embedded arc $\alpha \subset H_1$ that defines Σ' as a stabilization of Σ. $\square$

We can use the stabilization disks to find a presentation of $\pi_1(M)$ induced by a stabilization of (Σ, H_1, H_2): Given minimal systems of disks for H_1 and H_2, we can construct systems for H_1' and H_2' by adding the stabilization disks to the systems for H_1 and H_2. These disks intersect each other in a single point and are disjoint from the rest of the disks in the system, so they define the generator x_{g+1} and the relation $x_{g+1} = 1$. This is precisely the fourth Tietze transformation, completing the analogy started in Chapter 3 between group presentations and Heegaard splittings.

The second major theorem that we will prove in this chapter is the following, originally proved by Waldhausen in 1968 [**18**]:

5.6. THEOREM (Waldhausen). *Every positive genus Heegaard splitting of S^3 is stabilized.*

This theorem has two important consequences:

5.7. Corollary. *Every Heegaard splitting of S^3 is a stabilization of a genus-zero Heegaard splitting.*

Proof. Let (Σ', H'_1, H'_2) be a Heegaard splitting of S^3 other than a genus-zero splitting. By Theorem 5.6, Σ' is stabilized. Let g be the smallest integer such that Σ' is a stabilization of of a genus-g splitting (Σ, H_1, H_2) of S^3. If Σ is stabilized then Σ is a stabilization of a lower genus spliting. The splitting Σ is also a stabilization of this lower-genus splitting so if Σ is stabilized then g is not minimal. Because g is minimal, Σ must not be stabilized so Theorem 5.6 implies $g = 0$. □

5.8. Corollary. *Any two Heegaard splittings of S^3 of the same genus are isotopic.*

Proof. For the genus zero case, note that a spine of a genus-zero handlebody is a single point. Because S^3 is connected, any two points in S^3 are isotopic. By Lemma 2.16, this implies that any two genus-zero splittings of S^3 are isotopic.

Given two Heegaard splittings of S^3 of the same genus (not zero), Corollary 5.7 implies that both are stabilizations of genus-zero splittings. Any two genus-zero splittings are isotopic so by Theorem 5.3 any two splittings of the same genus are isotopic. □

2. Reducibility

Consider a stabilized Heegaard splitting (Σ', H'_1, H'_2). Let D'_1, D'_2 be properly embedded disks in H'_1, H'_2, respectively such that $\partial D'_1 \cap \partial D'_2$ is a single point. Let N_1 be a closed regular neighborhood of $D'_1 \cup \partial D'_2$. Because $\partial D'_1 \cap \partial D'_2$ is a single point, the set $\partial N_1 \setminus \Sigma$ is an open disk. Let $D_1 \subset H'_1$ be the closure of this open disk and let $D_2 \subset H'_2$ be a similar disk constructed from $D'_2 \cup \partial D'_1$.

The boundary of D_1, shown in Figure 3, coincides with the boundary of a neighborhood in Σ' of $\partial D'_1 \cup \partial D'_2$. The same is true for ∂D_2, so we have $\partial D_1 = \partial D_2$ (as subsets of Σ'). The union $D_1 \cup D_2$ is a sphere which bounds a ball B and $\Sigma' \cap B$ is genus-one subsurface. Define $\Sigma = (\Sigma' \setminus B) \cup D_1$. The reader can check that Σ' is a stabilization of Σ.

Conversely, if we start with disks D_1, D_2, properly embedded in the handlebodies H_1, H_2, respectively, if $\partial D_1 = \partial D_2$ then the union $D_1 \cup D_2$ is a sphere along which we can cut the Heegaard surface. We will see that if $D_1 \cup D_2$ bounds a ball B in M, the complement $\Sigma \setminus B$ leads to a new Heegaard surface for M of smaller genus. This fact motivates the following definition:

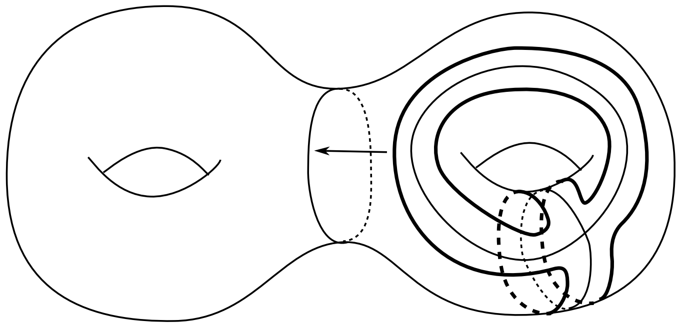


FIGURE 3. A pair of stabilizing disks imply a pair of disks in opposite handlebodies whose boundaries coincide.

5.9. DEFINITION. A Heegaard splitting (Σ, H_1, H_2) is *reducible* if there are essential, properly embedded disks D_1, D_2 in H_1, H_2, respectively, such that $\partial D_1 = \partial D_2 \subset \Sigma$. The splitting is *irreducible* if it is not reducible.

5.10. LEMMA. *If a Heegaard splitting* (Σ, H_1, H_2) *is stabilized then either* Σ *is reducible or* Σ *is a genus-one Heegaard splitting of* S^3.

PROOF. Let D_1', D_2' be essential disks whose boundaries intersect in a single point in Σ. As described above, we can construct properly embedded disks D_1, D_2 whose boundaries coincide. If D_1 and D_2 are essential then Σ is reducible.

If the disks are not essential then ∂D_1 bounds a disk in Σ. Since ∂D_1 bounds a punctured torus on the other side (by construction), Σ has genus 1. The disks D_1, D_2 form a sphere which bounds a 3-ball on either side, so the manifold is S^3. □

The converse of this theorem is not true. For example, we can take a connect sum of manifolds and consider the induced Heegaard splitting, as in Section 4.6. Recall that given be closed 3-manifolds M and M' and Heegaard splittings (Σ, H_1, H_2), (Σ', H_1', H_2') for M, M', respectively we construct the connect sum M'' and a Heegaard splitting for it by gluing $M \setminus B$ to $M' \setminus B'$ where B and B' are open balls that intersect Σ, Σ', respectively in disks.

The boundary of B is a union of two disks: $D_1 = \partial B \cap H_1$ and $D_2 = \partial B \cap H_2$. The image of the disks D_1 in M'' is essential and properly embedded in the handlebody $H_1'' = H_1 \cup H_1'$. The image of D_2 is essential and properly embedded in $H_2'' = H_2 \cup H_2'$. Moreover, $\partial D_1 = \partial D_2$ so (Σ'', H_1'', H_2'') is reducible. However, if neither M_1 nor M_2 is not S^3 then D_1 and D_2 do not correspond to a pair of stabilization disks. (Such a pair would require that M or M' was a 3-sphere with a genus-one splitting. The sphere $D_1 \cup D_2$ does not bound a ball on either side.

5.11. Definition. A 3-manifold M is *reducible* if it contains a sphere S that does not bound a ball and we say that S is *essential.* (Such a sphere need not separate M.)

Waldhausen's Theorem implies the following Lemma:

5.12. Lemma. *Let (Σ, H_1, H_2) be a Heegaard splitting of M. If Σ is reducible then Σ is stabilized or M is reducible.*

Proof. Let D_1, D_2 be disks in H_1, H_2, respectively, such that $\partial D_1 = \partial D_2$. The disks form a sphere $S = D_1 \cup D_2$. If S is essential then M is reducible and we are done. Otherwise, S bounds a ball B. We will find a pair of stabilization disks in B.

The surface $\Sigma' = \Sigma \cap B$ is properly embedded and splits B into two handlebodies. Because D_1 and D_2 are essential disks, Σ' is a punctured surface of positive genus (i.e. not a disk). It is not a Heegaard splitting because Σ' is not closed. In fact, $\partial\Sigma'$ is a loop in $S = \partial B$ that cuts S into two disks. Let B' be the result of removing B from M and gluing these disks together (the choice of gluing maps is unique up to isotopy.)

The image Σ'' of Σ' in B' is a Heegaard splitting of B'. Let $D \subset \Sigma''$ be the image in B' of the disks that were glued together. By Waldhausen's Theorem, Σ'' is stabilized because it is a Heegaard splitting surface for the 3-sphere, B' of positive genus. Let D_1'', D_2'' be a pair of stabilization disks for Σ''.

The loop ∂D is trivial in Σ so we can isotope D_1'' and D_2'' disjoint from ∂D and therefore disjoint from D. Thus D_1'' and D_2'' are the images of disks D_1' and D_2' in B and these two disks are the images of a pair of stabilization disks in M. □

3. Incompressible Surfaces

Let S be a properly embedded surface in a 3-manifold M. We will say that S is *two sided* if a regular neighborhood N of S is homeomorphic to $S \times [0, 1]$. The set $S \times [0, 1]$ is orientable if and only if S is orientable so if M is orientable then any two-sided surface is orientable. The converse holds as well:

5.13. Lemma. *Let M be an orientable 3-manifold. Then a properly embedded surface $S \subset M$ is two-sided if and only if S is orientable.*

Details of the proof will be left to the reader. The key to the proof is that given an orientation for the surface, an orientation for the ambient manifold is determined by a normal vector to the surface. If S is two sided then sliding a normal vector along a loop in S cannot switch the loop to the other side of S. Thus the induced orientation for S cannot switch.

5.14. Definition. A two sided surface S properly embedded in a 3-manifold M is *compressible* if there is a simple closed curve $\ell \subset S$ such that ℓ does not bound a disks in S but ℓ is the boundary of an embedded disk $D \subset M$ such that the interior of D is disjoint from S. If a surface S is not a sphere and not compressible then S is *incompressible*.

Given a two-sided surface S, each element of $\pi_1(S)$ is represented by a closed curve γ in S. If we choose the base point for $\pi_1(M)$ the same as that for $\pi_1(S)$ then the image in M of γ determines an element of $\pi_1(M)$. A homotopy of γ in S defines a homotopy in M, so the induced map $i : \pi_1(S) \to \pi_1(M)$ is well defined.

If S is compressible, the loop ∂D determines a non-trivial element of $\pi_1(S)$. The disk D determines a homotopy of γ down to a point, so the induced element of $\pi_1(M)$ is the identity. Thus if S is compressible then i has non-trivial kernel. In other words, if i is one-to-one then S is incompressible. The converse of this statement is also true, but the proof is non-trivial.

5.15. Theorem. *If S is a closed incompressible surface embedded in a manifold M then the inclusion map $\pi_1(S) \to \pi_1(M)$ is one-to-one.*

The proof of this theorem can be found in [**5**].

5.16. Example. *Incompressible tori in the 3-torus*

Recall that the 3-torus T^3 is formed by identifying opposite faces of the cube $[0,1] \times [0,1] \times [0,1]$. We showed in section 3.2 that the fundamental group of T^3 is $\mathbf{Z}^3$ generated by the edges of the graph formed from the edges of the cube.

The image in T^3 of the square $[0,1] \times [0,1] \times \{0\}$ is a torus. The fundamental group ($\mathbf{Z}^2$) of this torus is generated by the graph in T^3 formed by the edges of this square. The induced homomorphism $\mathbf{Z}^2 \to \mathbf{Z}^3$ is one-to-one so this torus is incompressible in T^3.

5.17. Corollary. *Every properly embedded surface in S^3 or B^3 is a disk, a sphere or is compressible.*

Proof. Let S be a closed surface embedded in S^3 or B^3. If S is not a sphere, then $\pi_1(S)$ is non-trivial. The fundamental group of S^3 is trivial, as is $\pi_1(B^3)$ so the kernel of the induced map i is $\pi_1(S)$. Thus Theorem 5.15 implies that S is compressible.

Given a properly embedded surface S in B^3, any component of ∂S bounds a disk in ∂B^3. Pushing this disk into the interior of B^3 will create a compression disk for S if ∂S is essential in S. Thus the only properly embedded, incompressible surface in B^3 is a disk. □

5.18. Proposition. *Let S be an incompressible surface, properly embedded in a handlebody H and assume that no component of S is a sphere. Let N be an open regular neighborhood of S. Then each component of $H \setminus N$ is a handlebody.*

Proof. Consider a system of disks $\mathbf{D} = \{D_1, \ldots, D_n\}$ for H. Choose the disks so as to minimize the number of components of $S \cap (\bigcup D_i)$. Assume for contradiction $S \cap D_i$ contains a loop component for some i. Let ℓ be an innermost loop in D_i. Because ℓ bounds a disk with interior disjoint from S and S is incompressible, ℓ must bound a disk in S. Compressing D_i along this disk in S produces a new disk that intersects S in fewer components. Thus each component of $D_i \cap S$ must be an arc.

Let B be the complement in H of a regular neighborhood of $\bigcup D_i$ and assume for contradiction some component of $S \cap B$ is compressible in B. Let D be a compressing disk for $S \cap B$. Because S is incompressible, ∂D must be trivial in S, bounding a disk $D' \subset S$. The intersection $D_i \cap D'$ cannot contain loops since $D_i \cap S$ contains no loops. In addition, $D_i \cap D'$ cannot contain arcs because $\partial D'$ is disjoint from each D_i. Thus D' is contained in a component of $S \cap B$, contradicting the assumption that $\partial D = \partial D'$ is essential in $S \cap B$.

Because B is a ball, an incompressible surface in B is a sphere or a disk. No component of $S \cap H$ is a sphere so each component must be a disk. This implies that $B \setminus N$ is a collection of balls. For each i, $S \cap D_i$ is a collection of arcs so $D_i \setminus N$ is a collections of disks. Thus $(\bigcup D_i) \cap (H \setminus N)$ is a collection of disks that cut $H \setminus N$ into a collection of balls. It follows that each component of $H \setminus N$ is a handlebody, □

If a surface S is compressible, the compressing disk D suggests a way to construct a new, simpler surface, as follows: Let N be an open neighborhood of D and define S' to be the result of capping off $S \setminus N$ with disks parallel to D. This construction is called a *compression*. There is a similar construction for properly embedded surfaces with non-trivial boundary. However, rather than compressing the surface in the interior of M, we can compress it against the boundary.

5.19. Definition. A surface S with boundary, properly embedded in a 3-manifold M is *boundary compressible* if there is a properly embedded arc $\alpha \subset S$ and an arc $\beta \subset \partial M$ such that α does not bound a disk in S but $\alpha \cup \beta$ is the boundary of an embedded disk $D \subset M$ such that the interior of D is disjoint from S. If a surface S is not a disk and not boundary compressible then S is *boundary incompressible.*

5.20. Lemma. *Any surface properly embedded in a handlebody is compressible, boundary compressible, or a collection of spheres and disks.*

Proof. Consider a system of disks $\mathbf{D} = \{D_1, \ldots, D_n\}$ for H. Choose the disks so as to minimize the number of components of $S \cap (\bigcup D_i)$. We saw in the previous argument that if S is not compressible then for each i, $S \cap D_i$ contains no closed loops. If S is disjoint from each D_i then S is properly embedded in the complement B of the disks. If S is not compressible then Corollary 5.17 implies that each component of S is a disk or a sphere.

For some i, $D_i \cap S$ must be non-empty so $D_i \cap S$ contains one or more arcs. Let α be an outermost arc of $D_i \cap S$. If α separates a disk from S then compressing D_i along this disk produces a new system of disks which intersects S in fewer components. Thus α must be essential in S. Thus the outermost disk in D_i defines a boundary compression of S. □

For compression bodies, we need to be a little more careful. A proof of the following Lemma can be found in [**5**]:

5.21. Lemma. *Let F be a compact, closed, orientable surface and define $M = F \times [0,1]$. Assume S is an incompressible, boundary incompressible, properly embedded surface in M. Then each component of S is a disk, a sphere, a closed surface isotopic to $F \times \{\frac{1}{2}\}$, or an annulus isotopic to $\lambda \times [0,1]$ for some simple closed curve $\lambda \subset F$.*

An argument similar to the proof of Lemma 5.20 leads to the following corollary. The details of the proof will be left to the reader.

5.22. Corollary. *Let H be a compression body and S an incompressible, boundary incompressible surface properly embedded in H. Then each component of S is a disk, a sphere, a closed surface parallel to a component of $\partial_- H$, or an annulus with one boundary component in $\partial_- H$ and the other component in $\partial_+ H$.*

4. Essential Disks and Spheres

Recall that a sphere S embedded in a manifold M is called essential if no component of its complement is an open ball. (An essential sphere may be separating or non-separating.) A manifold M is reducible if it contains an essential sphere. Lemma 5.12 makes a connection between reducible manifolds and reducible Heegaard splittings. In this section, we will prove the following partial converse to Lemma 5.12:

5.23. THEOREM (Haken [**4**]). *Let (Σ, H_1, H_2) be a Heegaard splitting of a manifold M. If M is reducible then (Σ, H_1, H_2) is reducible.*

Before we begin the proof, it is necessary to state a technical lemma:

5.24. LEMMA. *Let F be a connected, planar surface, not a disk, and let $\{\alpha_1, \dots, \alpha_n\}$ be an ordered collection of pairwise disjoint, properly embedded arcs in F such that $F \setminus \{\alpha_1, \dots, \alpha_n\}$ is a collection of disks. Assume that for each i, α_i is essential in the complement of a regular neighborhood of the arcs $\alpha_{i+1}, \dots, \alpha_n$. Then the complement in F of a regular neighborhood of $\alpha_1, \dots, \alpha_n$ has strictly fewer boundary components than F.*

PROOF. Induct on the number n or arcs. If $n = 1$ then F must be an annulus so F has two boundary components and $F \setminus \alpha_1$, a disk, has a single boundary component.

If $n > 1$, consider the arc $\alpha_n \subset F$ and let F' be the complement in F of a regular neighborhood of α_n. There are two cases to consider: the endpoints of α_n may be on the different components of F or on the same component.

First assume the endpoints are on different boundary components of F. Then F' has one fewer boundary components than F. The arcs $\alpha_1, \dots, \alpha_{n-1}$ are properly embedded in F'. Each arc α_i is essential in the complement of $\alpha_{i+1}, \dots, \alpha_n$ in F, so it is essential in the complement of $\alpha_{i+1}, \dots, \alpha_{n-1}$ in F'. Thus by the induction hypothesis, $F \setminus (\alpha_1 \cup \dots \cup \alpha_n)$ has strictly fewer boundary components than F', and therefore has fewer components than F.

For the second case, assume the endpoints of α_1 are on the same component of ∂F. Then F' has one more boundary component than F. To continue, we need following Lemma, which follows from the Jordan Curve Theorem and is left as an exercise for the reader:

5.25. LEMMA. *If α is an arc in a planar surface F and the endpoint of α are on the same component of ∂F then α separates F.*

Returning to the proof, because F is planar, the arc α_n separates F. By assumption, α_1 does not separate a disk from F so the induction assumption applies to each component of F'. The arcs cut each component into pieces with strictly fewer boundary components, so the total number is at most two less than F'. Since F' has exactly one more component than F, $F \setminus (\alpha_1 \cup \dots \cup \alpha_n)$ has strictly fewer boundary components than F. This completes the proof. $\square$

Proof of Theorem 5.23. Let (Σ, H_1, H_2) be a Heegaard splitting for M and let S be an essential sphere in M, transverse to Σ. Define $L = \Sigma \cap S$ and assume that we have chosen S from all the essential spheres in M so as to minimize the number of components of L. If L is empty then S sits in one of the handlebodies. Any sphere embedded in a handlebody bounds a ball and is not essential so the set L contains at least one loop. We will show that if L is minimized then L contains exactly one loop.

If L contains more than one loop then $S \cap H_i$ contains a non-disk component for $i = 1$ or 2. Without loss of generality, assume that $S \cap H_1$ contains a non-disk component and let $F = S \cap H_1$. Each component of F has boundary contained in $\partial_+ H$ so no component can be parallel to a component of $\partial_- H$ or an annulus with a boundary component in $\partial_- H_1$. Because F is not a collection of disks, Lemma 5.22 implies that F is either compressible or boundary compressible in M.

Assume F is compressible in H_1 and let D be a compressing disk. Then ∂D is a simple closed curve in the sphere S. Compressing S along the disk D defines two spheres, S' and S''. The following Lemma is left as an exercise for the reader:

5.26. Lemma. *If S is an essential sphere in a manifold M then one of the spheres S', S'', produced by compressing S, is essential in M.*

Each of S' and S'' intersects Σ in fewer components than S. By the Lemma, one of them is essential in M. This contradicts the assumption that L is minimal, so F must be incompressible.

Because F is not compressible in H and not a collection of disks, by Lemma 5.20, it must be boundary compressible. Let $D \subset H$ be a boundary compression disk for F. Define $\alpha = F \cap D \subset F$. This arc is properly embedded and essential in F. Isotope a neighborhood in S of α across D, sending α out of H_1. After the isotopy, S intersects H_1 in a surface F', which is homeomorphic to the complement in F of a regular neighborhod of β_1.

If F' is not a collection of disks then F' is boundary compressible and there is a boundary compressing disk D'. We can again isotope S across this disk, cutting F' along a second arc. Continuing in this fashion, we can find a collection of arcs $a_1, \dots, a_n$ properly embedded in F with α_n the first arc we removed and α_1 the last arc. By construction, each α_i is essential in the complement of $\alpha_{i+1}, \dots, \alpha_n$. The process terminates when we reach a surface that consists of only disks.

The number of boundary components for $F \setminus (\bigcup a_i)$ is precisely the number of disks in the final surface. By Lemma 5.24, each non-disk component of F is cut into pieces with strictly fewer boundary components. By construction, there is an isotopy of S after which S intersects H_1 in the final surface in the sequence of boundary compressions. This sphere intersects Σ in strictly fewer loops than S, contradicting minimality.

The contradiction implies that $S \cap \Sigma$ must be connected. This implies that $S \cap H_1$ is a single essential disk which we will call D_1 and $S \cap H_2$ is a single essential disk D_2. These disks have the property that $\partial D_1 = \partial D_2$ so (Σ, H_1, H_2) is reducible. □

One way to restate Theorem 5.23 is to say that if M contains an essential sphere then it contains an essential sphere that intersects Σ in a single loop. In the next section, we will need a similar result for disks. We have defined what it means for a properly embedded disk in a handlebody to be essential. We will now generalize this idea to arbitrary manifolds with boundary.

5.27. Definition. An *essential disk* D in a manifold M with boundary is a properly embedded disk such that ∂D does not bound a disk in ∂M. A manifold that contains an essential disk is *boundary reducible.*

5.28. Lemma. *Let* (Σ, H_1, H_2) *be a Heegaard splitting of a boundary reducible manifold* M*. Then there is an essential disk* D *such that* $D \cap \Sigma$ *is a single loop.*

Proof. The proof follows the argument used to prove Theorem 5.23. Let D be a boundary compression disk for M and assume D intersects Σ in a minimal number of components. Without loss of generality, assume ∂D is contained in a component of $\partial_- H_1$. Then exactly one component of $D \cap H_1$ has a single boundary loop in $\partial_- H$. This component cannot be a disk because ∂D is essential in ∂M, and any disk in H_1 with boundary in $\partial_- H_1$ is inessential.

Each component of $D \cap H_2$ has nontrivial boundary contained in $\partial_+ H_2$ so as in the proof of Theorem 5.23, minimality implies that every component of $D \cap H_2$ is a disk.

If $D \cap H_1$ does not consist of one annulus and some number of disks then $D \cap H_1$ is boundary compressible. Let F be a surface homeomorphic to $D \cap H_1$ and let $\alpha_1, \dots, \alpha_n$ be a set of properly embedded arcs in F that keep track of a sequence of boundary compressions turning $D \cap H_1$ into an annulus and disks.

The surface $F \setminus (\bigcup_1^n \alpha_i)$ consists of disks and one annulus. Let $\alpha_{n+1} \subset F$ be an arc in the annulus component from one boundary loop to the other which is disjoint from the arcs $\alpha_1, \dots, \alpha_n$. The arc α_{n+1} cuts the annulus component into a disk, implying that $F \setminus (\bigcup_1^{n+1} \alpha_i)$ is a collection of disks. Lemma 5.24 implies that this second surface has strictly fewer boundary components than F. It has exactly one more component than $F\backslash(\bigcup_1^n \alpha_i)$ so after the boundary compressions, $D\cap H_1$ has at most as many boundary components as before the compressions.

Thus we can choose D so that L is minimized and $D \cap H_1$ consists of one annulus and some number of disks. This implies $D \cap H_2$ consists entirely of disks so L must consist of a single loop. This completes the proof. □

5. Weakly Reducible

5.29. Definition. A Heegaard splitting (Σ, H_1, H_2) is *weakly reducible* if there is a pair of essential, properly embedded disks D_1, D_2 in H_1, H_2, respectively, such that ∂D_1 and ∂D_2 are disjoint in Σ. The splitting is *strongly irreducible* if it is not weakly reducible.

Notice that if $\partial D_1 = \partial D_2 \subset \Sigma$ then a small isotopy of D_1 makes the boundaries disjoint. Thus a reducible splitting is always weakly reducible.

In this section, we will prove the following analogue to Theorem 5.23 for weakly reducible splittings.

5.30. Theorem (Casson and Gordon). *If (Σ, H_1, H_2) is a weakly reducible Heegaard splitting of a manifold M then (Σ, H_1, H_2) is reducible or M contains a two-sided, closed incompressible surface.*

The proof here will follow a method due to Scharlemann and Thompson [**14**], employing the notion of thin position for 3-manifolds.

We will see that if we think of a Heegaard splitting as a sequence of 1-handles and 2-handles (as in a Morse function) then the condition of weakly reducible means that there is a 2-handle that can be attached to the 1-handles before all the 1-handles have been added, as in the example in Section 6.

5.31. Lemma. *If (Σ, H_1, H_2) is a weakly reducible Heegaard splitting then (Σ, H_1, H_2) is a non-trivial amalgamation of a generalized Heegaard splitting.*

In other words, a weakly reducible Heegaard splitting has a non-trivial telescope.

Proof. Let D_1, D_2, be properly embedded essential disks in H_1, H_2, respectively with disjoint boundaries. Let N_1, N_2 be disjoint open regular neighborhoods of D_1, D_2, respectively.

Define $H_1^1 = H_1 \setminus N_1$. This is a handlebody and $N_2 \cap \partial H_1^1$ is an annulus. Thus if N is a closed regular neighborhood of ∂H_1^1 in the complement of H_1^1 then the closure H_2^1 of $N \cup N_2$ is a compression body with $\partial_+ H_2^1 = \partial H_1$. Define $\Sigma^1 = \partial H_1$.

The closure of the intersection $N_1 \cap \partial_- H_2^1$ consists of two disks. Let N' be a regular neighborhood of $\partial_- H_2^1$ in the complement of H_2'. Define $N'' = N_1 \setminus H_2^1$. Then the closure H_1^2 of $N' \cup N''$ is a compression body and $\partial_- H_1^2 = \partial_- H_2^1$. Define $T^1 = \partial_- H_1^2$.

Finally, let H_2^2 be the closure of the complement in H_2 of $H_1^2 \cup H_2^1$. The later set intersects the first in the union of N_2 and a regular neighborhood of ∂H_2. Thus H_2^2 is a handlebody and $\partial H_2^2 = \partial_+ H_1^2$. Define $\Sigma^1 = \partial_+ H_1^2$. This collection of compression bodies and thin and thick surfaces defines a generalized Heegaard splitting. The reader can check that the amalgamation of this generalized splitting is (Σ, H_1, H_2). □

As noted above, the genus three Heegaard splitting of T^3 is weakly reducible and we saw in Chapter 4 that T^3 allows a non-trivial telescope, shown in Figure 4.8.

Note that the genera of the components of Σ^1, T^1 and Σ^2 are strictly lower than the genus of Σ. Thus the generalized splitting is, in some sense, simpler than the original Heegaard splitting. In order to understand weakly reducible Heegaard splittings, we will define a notion of complexity such that the complexity of a telescope of a Heegaard splitting will be strictly lower than that of the original splitting.

For a compact, closed, orientable surface F (not necessarily connected) the *complexity* of F is the sum over each non-sphere component of F of $2g - 1$, where g is the genus of the component. Given a generalized Heegaard splitting $(\mathbf{\Sigma}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{T})$, for each surface $T_i \in \mathbf{T}$ let $c_1, \dots, c_n$ be the complexities of $T_1, \dots, T_n$. However, rather than having c_i be the complexity of T_i for each i, we will choose the indices so that $c_i > c_j$ whenever $i < j$. The *complexity* of the generalized Heegaard splitting is the ordered n-tuple $(c_1, \dots, c_n)$.

We will impose on the complexities the lexicographic ordering. In other words, we define $(c_1, \dots, c_n) > (d_1, \dots, d_m)$ if for some $k \geq 0$, the first k entries are equal and $c_{k+1} > d_{k+1}$. For example, in the generalized Heegaard splittings defined for T^3 in Chapter 4, we have $(2, 2) < (3)$. As expected, the complexity of the telescope is strictly less than the complexity of the original splitting.

A telescope of a Heegaard splitting (Σ, H_1, H_2) that has minimal complexity over all telescopes for that splitting is often called a *thin position* for (Σ, H_1, H_2). For such a generalized Heegaard splitting, the thick and thin surfaces are as "thin" (low genus) as possible. We will see that if a generalized Heegaard splitting of a 3-manifold M is a thin position then the thin surfaces are incompressible in M.

5.32. Definition. A compression body H is *trivial* if H is homeomorphic to $F \times [0,1]$ for some closed surface F.

5.33. Lemma. *If (Σ, H_1, H_2) is a Heegaard splitting of a manifold M, neither H_1 nor H_2 is trivial and M is boundary reducible then (Σ, H_1, H_2) is weakly reducible.*

Proof. By Lemma 5.28, there is a boundary compressing disk for M that intersects Σ in a single loop. Without loss of generality, assume ∂D is contained in $\partial_- H_1$. Then $D \cap H_1$ is an annulus with one boundary component in $\partial_- H_1$ and the other in $\partial_+ H_1$. Because H_1 is non-trivial, there is a properly embedded essential disk D_1 in H_1 that is disjoint from this annulus.

Define $D_2 = D \cap H_2$. Then D_2 is a properly embedded essential disk in H_2 and ∂D_2 is disjoint from ∂D_1. Thus Σ is weakly reducible. □

5.34. Corollary. *If $(\mathbf{\Sigma}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{T})$ is a thin position of a Heegaard splitting (Σ, H_1, H_2) then each thin surface $T_i \in \mathbf{T}$ is either a sphere or an incompressible surface in M.*

Proof. Recall that for each i, (Σ_i, H_1^i, H_2^i) is a Heegaard splitting of the submanifold $(H_1^i \cup H_2^i)$. If H_1^i is trivial for some i then amalgamating (Σ_i, H_1^i, H_2^i) with $(\Sigma_{i-1}, H_1^{i-1}, H_2^{i-1})$ eliminates a thin surface and a thick surface with out changing the complexities of the other surfaces. This reduces the overall complexity of the splitting. We can reduce the complexity similarly if H_2^i is trivial. Thus minimal complexity implies that H_1^i and H_2^i are non-trivial for each i. This allows us to employ Lemma 5.33.

Assume for contradiction some thin surface $T_i \in \mathbf{T}$ is compressible in M. Let D be a compressing disk for T_i. For any other surface $T_j \in \mathbf{T}$, if $D \cap T_j$ is minimized then each loop of $D \cap T_j$ is essential in T_j. Thus after the isotopy, an innermost disk D' in D determines a boundary compression disk for some $T_j \in T$ which is disjoint from any other $T_k \in \mathbf{T}$.

The disk D' is a boundary compression disk for either $H_1^i \cup H_2^i$ or $H_1^{i-1} \cup H_2^{i-1}$. Thus Lemma 5.33 implies that one of these Heegaard splittings is weakly reducible. By Lemma 5.31, the weakly reducible splitting has a non-trivial telescope. Replacing this splitting with its non-trivial telescope determines a non-trivial telescope of $(\mathbf{\Sigma}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{T})$. This telescope has lower complexity, contradicting the assumption that $(\mathbf{\Sigma}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{T})$ has minimal complexity. □

PROOF OF THEOREM 5.30. Let (Σ, H_1, H_2) be a weakly reducible Heegaard splitting for a manifold M and let $(\mathbf{\Sigma}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{T})$ be a minimal complexity telescope for (Σ, H_1, H_2). By Lemma 5.31, this telescope is non-trivial. Corollary 5.34 implies that each surface in $\mathbf{T}$ is either a sphere or an incompressible surface. If some element of $\mathbf{T}$ is not a sphere then the proof is complete.

Otherwise, assume each element of $\mathbf{T}$ is a sphere. It is an exercise for the reader to check that after amalgamating the generalized splitting to (Σ, H_1, H_2) the sphere $T_1 \in \mathbf{T}$ intersects Σ in a single essential loop. Thus Σ is reducible. □

Note that Theorem 5.6 was not used in the proof of Theorem 5.30. Thus we are free to emply Theorem 5.30 to prove Theorem 5.6 in the next section.

6. The Rubinstein-Scharlemann Graphic

We will now use the sweep-outs defined in Chapter 4 to analyze a pair of Heegaard splittings. The level sets of a sweep-out form a family of parallel Heegaard surfaces. We will be interested in the intersection of one level set in the first sweep-out and a second level surface in the second sweep-out. Rubinstein and Scharlemann [**12**] introduced the notion of the graphic as a way to analyze all possible pairs simultaneously.

Let f and g be sweep-outs of M. We will denote the level set at level t by $f_t = f^{-1}(t)$. The intersection of f_t with a level surface $g^{-1}(s)$ is precisely the pre-image of s in the restriction of g to the level surface f_t. Thus to understand the sets of intersection of f_t with all the level sets of g, we should consider the restriction of g to f_t. Let $g_t = g|_{f_t}$ be the function induced on f_t.

Recall that a Morse function is defined by two conditions: that each critical point be non-degenerate and that no two critical points be on the same level. We will call a function *near-Morse* if there is a single degenerate critical point or there are two non-degenerate critical points at the same level. We will see what near Morse functions look like and how they come into play after the following definitions.

5.35. DEFINITION. We will say that f and g are in *general position* if g_t is a Morse function on f_t for all but finitely many t, and g_t is a near-Morse function for the remaining values.

The following Theorem is usually attributed to Cerf [**3**]. A proof in the context of Heegaard splittings (and in English) is given by Kobayashi [**7**, Theorem 4.2].

5.36. Theorem. *Given f and g, there are arbitrarily small isotopies of f and g after which the functions are in general position.*

This theorem allows us assume, in most situations, that any given pair of sweep-outs are in general position.

5.37. Definition. Given a pair of sweep-outs in general positon, the *Rubinstein-Scharlemann graphic* is the closure in $[-1,1] \times [-1,1]$ of the set of points (t,s) such that s is a critical level of g_t.

As noted above, the level set $g_t^{-1}(s)$ is the intersection of the surface f_t with the level surface $g^{-1}(s)$. From the classification of critical points of 2-dimensional Morse functions in Section 4.3, one sees that s will be a critical level of f_t if and only if f_t is tangent to $g^{-1}(s)$ at some point. Thus the Rubinstein-Scharlemann graphic is precisely the set of points (t,s) such that f_t is not transverse to $g^{-1}(s)$.

The graphic forms a graph in $[-1.1] \times [-1,1]$ with vertices of valence 2 and 4 in the interior. The valence-2 vertices in the interior arise when g_t is a near-Morse function with a degenerate critical point. As t passes the vertex, a central singularity and a saddle singularity in f_t come together and cancel, as in Figure 4.

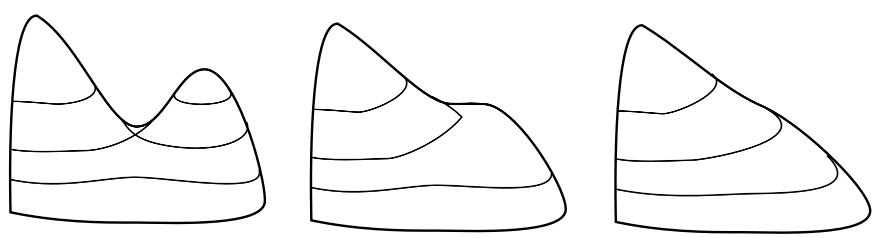


Figure 4. A cusp arises in the graphic when a central singularity and a vertex cancel. The height function on the middle surface has a degenerate critical point.

At a valence-4 vertex, (t, s_t), g_t has two critical points at the same level. As t passes the vertex, one of the critical points rises past the other, as in Figure 5. The two critical points may be on the same component of $g_t^{-1}(s_t)$ or different components.

Assume (t,s) and (t,s') are points such that the horizontal arc between them is disjoint from the graphic. Then for every point in the interval, the corresponding surfaces are disjoint. This implies an ambient isotopy of M which takes $g^{-1}(s)$ onto $g^{-1}(s')$ and takes f_t onto itself. Thus the intersection of f_t with $g^{-1}(s)$ is essentially the same as the intersection with $g^{-1}(s')$.

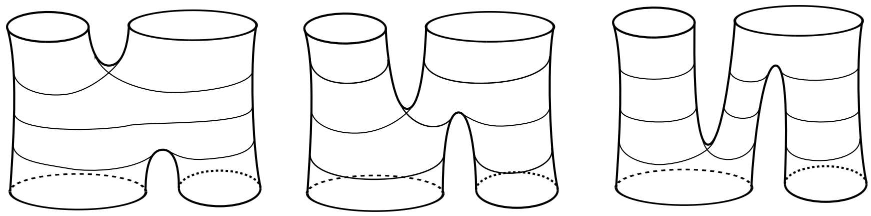


FIGURE 5. A valence four vertex arises when two critical points pass each other. The height function on the middle surface has two critical points at the same level.

Similarly, if the vertical interval between points (t, s) and (t', s) is disjoint from the graphic then there is an ambient isotopy taking f_t onto $f_{t'}$ and $g^{-1}(s)$ onto itself. If points (t, s) and (t', s') are in the same component of the complement of the graphic then there is a piecewise vertical and horizontal path from (t, s) to (t', s') that is disjoint from the graphic. Composing the ambient isotopies defined by the horizontal and vertical arcs, we find an ambient isotopy that takes f_t onto $f_{t'}$ and $g^{-1}(s)$ onto $g^{-1}(s')$.

Thus at any two points in the same component of the complement, the intersections of the corresponding surfaces are essentially the same. In the next section, we will define a method of labeling the components based on these intersections. An analysis of the labeling will allow us to prove Waldhausen's Theorem.

7. Heegaard Splittings of the 3-Sphere

Waldhausen's Theorem follows from the following lemma:

5.38. LEMMA. *Every genus $g > 1$ Heegaard splitting of S^3 is weakly reducible. Every genus $g = 1$ Heegaard splitting og S^3 is stabilized.*

Before we prove this Lemma, we need to state and prove the following:

5.39. LEMMA. *Let (Σ, H_1, H_2) be a Heegaard splitting for M and let S be a 2-sphere embedded in M such that $S \cap \Sigma$ is a non-empty collection of loops. If one or more of the loops is essential in Σ then there is a loop of $S \cap \Sigma$ that bounds a properly embedded, essential disk in either H_1 or H_2.*

PROOF. Induct on the number of components of $S \cap \Sigma$. If there is a single component ℓ of $S \cap \Sigma$ then ℓ must be essential. This ℓ splits S into two disks, each of which is a properly embedded, essential disk in one of the handlebodies.

If there are more than one components of $S \cap \Sigma$, let n be the number of loops and assume the result is true whenever there are $n-1$ loops. Let ℓ be an innermost loop of $S \cap \Sigma$ (one that bounds a disk D in S disjoint from Σ.) If ℓ is essential, then D is is properly embedded and essential and we are done.

Otherwise, D is boundary parallel (since handlebodies are irreducible) so we can isotope S to a surface S' such that $S' \cap \Sigma$ is a proper subset of $S \cap \Sigma$. Since $S' \cap \Sigma$ has $n-1$ components, one of them bounds a properly embedded, essential disk, so the corresponding component of $S \cap \Sigma$ does as well. □

Consider a Heegaard splitting (Σ, H_1, H_2) of S^3 and let f be a sweepout for Σ. Let g be a sweep-out for a genus-zero Heegaard splitting of S^3, i.e. assume $g^{-1}(s)$ is a 2-sphere in S^3 for each $s \in (-1, 1)$. By Theorem 5.36, assume f and g are in general position.

A *regular loop* in a level set of a Morse function is a component of a level set that does not contain a critical point. For each $t \in [-1, 1]$, label t with a 1 if an essential regular loop in some level set of g_t bounds a properly embedded disk in the handlebody $f^{-1}([-1, t])$. If an essential regular level loop bounds a properly embedded disk in $f^{-1}([t, 1])$, label t with a 2. (A point may have both labels or neither.)

If a level loop in $g_t^{-1}(s)$ is essential in f_t then $f_t \cap g^{-1}(s)$ contains an essential loop. The surface $g^{-1}(s)$ is a sphere, so Lemma 5.39 implies that some loop in L_p bounds a disk in either $f^{-1})([-1, t])$ or $f^{-1})([t, 1])$. Thus if t is not labeled then all the loops of L_p are trivial in f_t.

Any two distinct regular level loops in f_t are disjoint. If t has both labels 1 and 2, then there is a pair of disjoint loops that bound disks on opposite sides of f_t, or one loop which bounds disks on both sides. Because f_t is isotopic to Σ, if some value of t has both labels then (Σ, H_1, H_2) is weakly reducible.

5.40. Lemma. *Let Σ be a compact, connected, closed, orientable surface and $f : \Sigma \to \mathbf{R}$ a Morse or near-Morse function on Σ. If every regular loop of f is trivial in Σ then Σ is a torus or a sphere.*

Proof. It is left as an exercise for the reader to show that if f is a Morse function such that all regular level loops are trivial in Σ then Σ is a sphere.

If f is a near-Morse function with a degenerate critical point (such as the middle picture in Figure 4) then the level sets are isotopic to the level sets of a Morse function on Σ. If all the level sets of a Morse function are trivial then Σ is a sphere. Thus if f is a near-Morse function with a degenerate critical point then Σ is a sphere.

If f has two (non-degenerate) critical points at the same level then the level component containing these critical points is a graph G with two valence four vertices, and thus four edges. The restriction of f to the complement of G is a Morse function with no essential level sets so the complement must be a collection of disks. Thus G defines a cellular decomposition for Σ with two vertices, four edges and at least one face.

Calculating the Euler characteristic from this decomposition, we find that is must be at least $2 - 4 + 1 = -1$. Because Σ is closed and orientable, its Euler characteristic is even, so there must be at least two faces. Moreover, the Euler characteristic is non-negative so Σ is a torus or a sphere. □

With this Lemma we now understand the graphic well enough to prove that Heegaard splittings of S^3 of genus greater than one are weakly reducible.

Proof of Lemma 5.38. We will show that a strongly irreducible Heegaard splitting of S^3 must be a genus one or zero splitting. This is equivalent to the statement of the Lemma. Assume (Σ, H_1, H_2) is a strongly irreducible Heegaard splitting of S^3. Let $f : S^3 \to [-1, 1]$ be a sweep-out for Σ and assume f is in general position with respect to a sweep-out g of a genus zero splitting.

Let $t_1, \dots, t_n$ be the values of t for which f_t is near-Morse rather than Morse. The complement of this finite set is a collection of open intervals. Any two values of t in the same interval must have the same labels because the isotopy classes of the level loops can only change when g_t passes through a near Morse function.

If a near-Morse function g_{t_i} is labeled, then the essential level loops of g_{t_i} are isotopic to essential level loops of the Morse functions right before and right after it. Thus if t_i has a label then the open intervals right before and right after it have the same label.

For t near -1, f_t is the boundary of a small neighborhood of a spine of H_1. For fixed s and small enough t, the intersection of the handlebody $f^{-1}([-1, t])$ and the surface $g^{-1}(s)$ is a collection of disks. Thus the values of t near -1 are labeled with a 1. Similarly, the values of t near 1 are labeled with a 2.

The set of values of t labeled with a 1 form an open set, as do the values labeled with a 2. Because (Σ, H_1, H_2) is strongly irreducible, no value of t has both labels, so the sets are disjoint. The interval $[-1, 1]$ is connected so it is not the union of two open sets. Thus some value of t is not labeled. By Lemma 5.39, if t is unlabeled then every regular level loop of g_t is trivial in Σ, so by Lemma 5.40, Σ is a torus or a sphere. (In fact, if Σ is a torus, the unlabeled value t is a near-Morse function.) □

We can now bring Lemma 5.38 together with many of the techniques developed earlier in the chapter to prove Waldhausen's Theorem.

Proof of Theorem 5.6. We will induct on the genus g of a Heegaard splitting (Σ, H_1, H_2) of S^3. For the base case, $g = 1$, Lemma 5.38 implies that Σ is stabilized.

For $g > 1$, assume that every Heegaard splitting of genus strictly less than g is stabilized. Lemma 5.38 implies that Σ is weakly reducible. By Theorem 5.30, either Σ is reducible or S^3 contains an incompressible surface. Corollary 5.17 states that S^3 does not contain any non-trivial incompressible surfaces, so Σ must be reducible.

Let S be a reducing sphere for Σ. Because S^3 is irreducible, S bounds a ball B and $\Sigma \setminus B$ defines a Heegaard splitting for S^3 of strictly lower genus than Σ. By the induction hypothesis, this splitting is stabilized so Σ is also stabilized. □

8. Heegaard Splittings of Handlebodies

We now return to the Theorem with which we began the chapter. Theorem 5.2 states that for any two Heegaard splittings of a given manifold, there is a common stabilization. Historically, this was proved before any of the other Theorems in this chapter. Here, however, we will present a modern proof, first discovered by Scharlemann and Thompson [**13**], using the machinery we have accumulated so far. We will begin by classifying Heegaard splittings of handlebodies.

Let H be a handlebody of genus g and let $N \subset H$ be a regular neighborhood of ∂H. Then N is a (trivial) compression body, homeomorphic to $\partial H \times I$ and the closure of $H \setminus N$ is a handlebody. Thus $\partial(H \setminus N)$ is a Heegaard splitting surface for H.

5.41. Definition. A Heegaard splitting (Σ, H_1, H_2) of a handlebody H is *standard* if it is isotopic to the Heegaard splitting defined above or to a stabilization of this splitting.

In general, if a manifold has the property that any two Heegaard splittings of the same genus are isotopic, it is common to say that Heegaard splittings of this manifold are *standard*.

5.42. Lemma. *Let H be a handlebody. Every Heegaard splitting of H is standard.*

Proof. We will show that every Heegaard surface of H that is not parallel to ∂H is stabilized. Because there is a unique (up to isotopy) surface parallel to ∂H and any two stabilizations of the same Heegaard splitting are isotopic, this implies a unique (up to isotopy) Heegaard splitting for each genus.

We will induct on the genus of H. If H is a genus-zero handlebody, then the union of H with a ball is S^3. Theorem 5.6 implies that every positive genus Heegaard splitting of S^3 is stabilized, so every positive genus splitting of the 3-ball is stabilized.

For genus $g > 0$, assume the result is true for all lower genera. The boundary of H is compressible, so Lemma 5.28 implies there is a disk D such that $D \cap \Sigma$ is a single loop. Let N be a regular neighborhood of D such that $\Sigma \cap N$ is an open annulus. By capping off the ends of Σ, we can produce a Heegaard splitting for the handlebody $H \setminus N$.

This new splitting surface will be boundary parallel or stabilized if and only if the original splitting for H is boundary parallel or stabilized. (This is exercise 7.) The new handlebody has strictly lower genus, so the inductive assumption implies that the induced Heegaard splitting surface is either boundary parallel or stabilized. Thus the same is true of the original splitting. □

5.43. Corollary. *Let (Σ, H_1, H_2) and (Σ', H_1', H_2') be Heegaard splittings of a closed, compact manifold M. Let K be a spine of H_1 and let K' be a spine of H_1'. If K is a subgraph of K' then Σ' is isotopic to a stabilization of Σ.*

Proof. Let N be an open regular neighborhood of K in M such that $K' \cap N$ is an open regular neighborhood of K in K'. Because K is a spine of H_1, the complement $M \setminus N$ is a handlebody H. The complement in H of a neighborhood N' of K' is a second handlebody, and the union of N' and a neighborhood of ∂H is a compression body. Thus $K' \setminus N$ defines a Heegaard splitting of H. By Lemma 5.42, this Heegaard splitting is a stabilization of the standard splitting for H, so Σ' is a stabilization of Σ. □

Thanks to Lemma 4.45, we can apply this to lens spaces to prove the following:

5.44. Theorem. *Let M be a lens space. Then any two Heegaard surfaces in M of the same genus are isotopic.*

Proof. Let (Σ, H_1, H_2) be the standard genus-one Heegaard splitting of M. We will show that any Heegaard splitting of a lens space of genus greater than one is a stabilization of Σ. The Theorem follows immediately from this and Theorem 5.3.

Let (Σ', H_1', H_2') be a Heegaard splitting of M and let ℓ be a spine of H_1. Lemma 4.46 states that ℓ is isotopic to a core of either H_1' or H_2'.

After the isotopy, one can construct a spine of H_1' or H_2' that contains ℓ. Because (after isotopy) a spine of H_1 is contained in a spine of H_1' or H_2', Corollary 5.43 implies Σ' is isotopic to a stabilization of Σ. □

Proof of Theorem 5.2. Let (Σ, H_1, H_2) and (Σ', H_1', H_2') be Heegaard splittings of a closed manifold M. Let K be a spine of H_1 and K' a spine of H_1'. After a small isotopy, we can assume K and K' are disjoint. Let M' be the complement of an open regular neighborhood of $K \cup K'$.

By Theorem 1.15, there is a Heegaard splitting (Σ'', H_1'', H_2'') for M'. In fact, we can choose Σ'' such that H_2 is a handlebody and $\partial M = \partial_- H_1$. Let K'' be a graph in H_1'' such that $K \cup \partial_- H_1''$ is a spine for H_1''. The inclusion into M of Σ' defines a Heegaard spliting $(\Sigma''', H_1''', H_2''')$ for M.

We can extend K'' into H_1 and H_1' so that $K \cup K' \cup K''$ is a spine for H_1'''. This spine of H_1''' contains a spine of H_1 so Corollary 5.43 implies Σ''' is a stabilization of both Σ. Similarly, this spine contains a spine of H_1' so Σ''' is also a stabilization of Σ'. In other words, Σ''' is a common stabilization and the proof is complete. □

9. Exercises

1. Give examples of one sided and two-sided surfaces in orientable and non-orientable manifolds.
2. Prove Lemma 5.25.
3. Prove Lemma 5.26.
4. Find a pair of disks with disjoint boundaries in opposite handlebodies of the genus three Heegaard splitting of T^3.
5. Show that if $(\mathbf{\Sigma}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{T})$ is a telescope of (Σ, H_1, H_2) and some thin surface in $\mathbf{T}$ is a sphere then (Σ, H_1, H_2) is reducible.
6. Show that if f is a Morse function such that all regular level loops are trivial in Σ then Σ is a sphere.
7. Show that in the proof of Lemma 5.42, the Heegaard splitting surface constructed by compressing H is boundary parallel or stabilized if and only if the original splitting for H was boundary parallel or stabilized.

References

1. Joan Birman, *On the equivalences of Heegaard splittings of closed orientable 3-manifolds*, Knots, Groups and 3-Manifolds, Ann. of Math. Studies, Princeton Univ. Press, Princeton, N.J., 1975.
2. Francis Bonahon and Jean-Pierre Otal, *Scindements de Heegaard des espaces lenticulares*, C. R. Acad. Sci. Paris **294** (1982), no. 17, 585–587.
3. Jean Cerf, *Sur les diffeomorphismes de la sphere de dimension trois ($\gamma_4 = 0$)*, Lecture Notes in Mathematics, vol. 53, Springer, 1968.
4. W. Haken, *Some results on surfaces in 3-manifolds*, MAA Studies in Mathematics, vol. 5, The Mathematical Association of America, 1968.
5. Allen Hatcher, *Basic Topology of 3-Manifolds*, Unpublished, available on author's webpage, http://www.math.cornell.edu/∼hatcher/.
6. Morris W. Hirsch, *Differential topology*, Springer-Verlag, 1976.
7. Tsuyoshi Kobayashi and Osamu Saeki, *The Rubinstein-Scharlemann graphic of a 3-manifold as the discriminant set of a stable map*, Pacific Journal of Mathematics **195** (2000), no. 1, 101–156.
8. Martin Lustig and Yoav Moriah, *Nielsen equivalence in Fuchsian groups and Seifert fibered spaces*, Topology **30** (1991), no. 2, 191–204.
9. Edwin E. Moise, *Affine structures in 3-manifolds. V. The triangulation theorem and Hauptvermutung*, Ann. of Math.(2) **56** (1952), 96–114.
10. K. Reidemeister, *Zur dreidimensionalen Topologie*, Abh. Math. Sem. Univ. Hamburg **11** (1933), 189–194.
11. C. P. Rourke and B. J. Sanderson, *Introduction to piecewise-linear topology*, Springer-Verlag, 1972.

12. Hyam Rubinstein and Martin Scharlemann, *Comparing Heegaard splittings of non-Haken 3-manifolds*, Topology **35** (1996), no. 4, 1005–1026.

13. Martin Scharlemann and Abigail Thompson, *Heegaard splittings of* $(surface) \times I$ *are standard*, Mathematische Annalen **295** (1993), 549–564.

14. ———, *Thin position for 3-manifolds*, Contemporary Mathematics **164** (1994), 231–238.

15. Jennifer Schultens, *The classification of Heegaard splittings for (compact orientable surface)* $\times S^1$, Bull. Lond. Math. Soc. **15** (1993), 425–448.

16. ———, *Heegaard splittings of Seifert fibered spaces with boundary*, Trans. Amer. Math. Soc. **347** (1995), no. 7, 2533–2552.

17. J. Singer, *Three-dimensional manifolds and their Heegaard diagrams*, Trans. Amer. Math. Soc. **35** (1933), 88–111.

18. F. Waldhausen, *Heegaard-Zerlegungen der 3-Sphare*, Topology **7** (1968), 195–203.

Index