4

Applications of the Theory of Normal Surfaces

4.1 Examples of Algorithms Based on Haken's Theory

The theory of normal surfaces is used extensively in algorithmic topology. Algorithms based on it most often follow the General Scheme described below. Suppose that we wish to solve a problem about a given 3-manifold M.

General Scheme

1. Reduce the problem at hand to one of the existence in M of a surface with some specific characteristic property, which we denote by P. Let $\mathcal{P}$ be the class of all characteristic surfaces in M, i.e., the class of all surfaces that possess P.
2. Choose a triangulation of M and show that if M contains at least one characteristic surface F, then there exists a normal characteristic surface. Quite often it can be done by proving that P is stable with respect to the normalization procedure, i.e., with respect to isotopies and moves N_1–N_8 that bring surfaces in normal position, see Theorem 3.3.21. By stability we mean that if F_1 is obtained from $F \in \mathcal{P}$ by isotopies and moves N_1–N_8, then at least one connected component of F_1 is also in $\mathcal{P}$.
3. Show that if there is a normal characteristic surface, then there is a fundamental characteristic surface. One possible way to do that is to prove that if a characteristic surface F is not fundamental, then M contains a less complicated characteristic surface. Certainly, we should know how to measure the complexity of a surface in M. The edge degree $e(F)$, i.e., the total number of points in the intersection of F with the edges, may serve as a good candidate for the purpose.
4. Construct an algorithm to decide whether or not a given surface is characteristic.

Assume that all four steps of the General Scheme are carried out. Then the algorithm that solves the problem works as follows:

1. Choose a triangulation T of M.
2. Write down the corresponding matching system of linear equations.
3. Find the finite set of fundamental solutions.
4. Realize the fundamental solutions by normal surfaces.
5. Test each of the obtained fundamental surfaces for being characteristic. It follows that M contains a characteristic surface (i.e., that the problem in question has a positive answer) if and only if at least one of the fundamental surfaces is characteristic.

4.1.1 Recognition of Splittable Links

We will illustrate the above scheme by describing algorithms for recognizing splittable links in S^3. Recall that a link $L \subset S^3$ is a collection of disjoint simple closed curves in S^3. The curves are called the *components* of L. A link L is called *splittable*, if there is a 2-sphere $S \subset S^3$ such that $S \cap L = \emptyset$ and each connected component of the complement $S^3 \setminus S$ contains at least one component of L. We will call S *a splitting sphere*. For example, a splittable link of two components is nothing more than the union of two knots contained in disjoint balls. The boundary sphere of either ball can be taken as a splitting sphere.

Theorem 4.1.1. *There is an algorithm to decide if a given link $L \subset S^3$ is splittable.*

Proof. We will follow the General Scheme.

STEP 1. It is convenient to replace the noncompact 3-manifold $S^3 \setminus L$ by the compact manifold $M = S^3 \setminus \text{Int } N(L)$, where $N(L)$ is a regular neighborhood of L. The boundary of M is a collection of 2-dimensional tori. Let us formulate the following property P of closed surfaces in M: a surface $S \subset \text{Int } M$ possesses P if S is a splitting sphere for L. Clearly, P is characteristic for the problem at hand: *L is splittable $\Longleftrightarrow$ M contains a sphere $S \in \mathcal{P}$.*

Note that a sphere $S \subset M$ splits L if it is essential in M. Even more: S splits L if and only if it determines a nontrivial element $[S]$ of $H_2(M; Z_2)$. Further, a closed surface in M determines a nontrivial element of $H_2(M; Z_2)$ if and only if there exists a proper arc $a \subset M$ which crosses L transversally in an odd number of points.

STEP 2. Choose a triangulation T of M. The statement we need to prove here is that if M contains a splitting sphere, then it also contains a normal splitting sphere.

The proof is natural: Take the splitting sphere S which exists by assumption, and normalize it by a sequence of moves N_1–N_4. Note that moves N_5–N_8 (see Sect. 3.3.3) are irrelevant since S is closed. Each time we apply move N_1, we get two 2-spheres S', S'' such that $[S'] + [S''] = [S] \neq 0$ in $H_2(M; Z_2)$. Therefore at least one of them splits M. We cross out the other.

STEP 3. We wish to prove that if M contains a normal splitting 2-sphere S, then such a sphere can be found among fundamental surfaces. Our strategy is to show that if S is not fundamental, then there is another splitting sphere which is simpler than S. For measuring the complexity of S we use the edge degree $e(S)$.

Let S be presented as a geometric sum $S = F_1 + F_2$ of two surfaces. There are many ways to present S in such form. By Lemma 3.3.30, we may assume that F_1, F_2 are connected, and no component of $F_1 \cap F_2$ separates both surfaces. Taking into account that the Euler characteristic is additive with respect to geometric sums and $\chi(S) = 2$, $\chi(F_i) \leq 2$, we come naturally to two options: 2=1+1 and 2=2+0. The first option does not occur, since the only closed surface with $\chi = 1$ is the projective plane RP^2, which cannot be embedded into S^3.

In the second case we may conclude that one of the surfaces (say, F_1) is a sphere, while the other (F_2) is a torus. The Klein bottle, which also has zero Euler characteristic, does not embed into S^3 either. Since $[F_1]+[F_2] = [S] \neq 0$, at least one of elements $[F_1], [F_2]$ is not zero.

Case 1. If $[F_1] \neq 0$, then F_1 splits M. Clearly, this new sphere is simpler than S.

Case 2. Suppose $[F_1] = 0, [F_2] \neq 0$. We claim that there exists a proper arc $a \subset M$ which does not intersect F_1 and intersects F_2 transversally at an odd number of points. Indeed, consider an arc b that joins two points $x, y \in \partial M$ contained in different components of $M \setminus F_2$ and is transversal to F_1, F_2. Since $[F_1] = 0$, the intersection $F_1 \cap b$ consists of an even number of points, while the number of points in $F_2 \cap b$ is always odd. The desired arc a can be obtained by replacing the subarc of b between the first and the last points in $F_1 \cap b$ by an arc that runs near F_1 but does not intersect it.

Recall that each circle $c \subset F_1 \cap F_2$ does not separate at least one of the surfaces. Since all circles on the sphere F_1 do separate, c is a nonseparating curve on F_2 and thus is nontrivial. Note that any collection of disjoint nontrivial simple closed curves in the torus F_2 decomposes it into annuli. Since a intersects F_2 at an odd number of points, at least one of the annuli (denote it by A) contains an odd number of crossing points.

Denote by d_1, d_2 the boundary circles of A. They bound in F_1 disjoint discs D_1, D_2. Let us perform (not necessarily regular) switches along d_1, d_2 which adjoin the discs to A. The switches produce a sphere S' that corresponds to $D_1 \cup A \cup D_2$, and the remaining part R of $F_1 \cup F_2$ that corresponds to the union of $F_1 \setminus \text{Int } (D_1 \cup D_2)$ and $F_2 \setminus \text{Int } A$. Note that S' splits M, since a does not intersect D_1, D_2 and thus the number of intersection points of a and S' is odd, see Fig. 4.1.

Let us prove that either S' is simpler than S or S' is isotopic to a simpler sphere. Denote by R' the surface obtained from R by performing regular switches along all the curves in the self-intersection of R. Of course, at least one of the switches at d_1, d_2 is irregular, because otherwise S would consist of S' and R' and thus be disconnected. This irregular switch produces a return, i.e., an arc in a triangle of the triangulation having both endpoints on the same edge. If the return is contained in R', then $e(R') > 0$, and we obtain

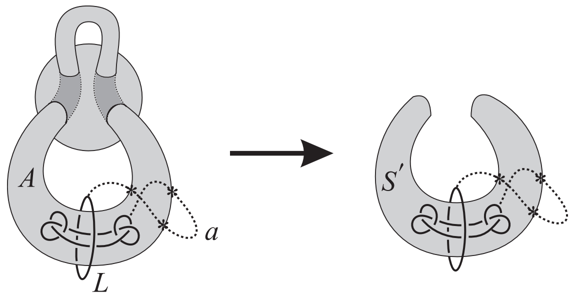

Fig. 4.1. Constructing a simpler splitting sphere

$e(S') = e(S) - e(R') < e(S)$. Let the return be in S'. Then we decrease the degree of S' by an isotopy which annihilates the return together with its endpoints. Applying the normalization procedure to the resulting surface, we get a new essential 2-sphere that has a strictly smaller edge degree.

Now we apply the simplification process as long as the sphere remains nonfundamental. Since each time we get a smaller edge degree, the process is finite and we end up with a fundamental splitting sphere. The last step of the General Scheme is easy. According to the scheme, we get a recognition algorithm for splittable links.

In conclusion we note that the proof works for any polyhedral subset of S^3. We have never used that L is the union of disjoint circles. □

4.1.2 Getting Rid of Clean Disc Patches

Later we will describe other problems of 3-manifold topology whose algorithmic solutions are based on the theory of normal surfaces. Among they are:

1. Recognizing the Unknot.
2. Calculating the genus of a given simple closed curve on the boundary of a 3-manifold.
3. Recognizing irreducibility and boundary irreducibility of a 3-manifold.
4. Testing two-sided surfaces for incompressibility and boundary incompressibility.
5. Detecting sufficiently large 3-manifolds (see Sect. 4.1.6 for the definition of a sufficiently large 3-manifold).

Solutions of all these problems follow the same General Scheme. In all cases Step 3 (if M contains a surface having a specific characteristic property, then such a surface can be found among the fundamental ones) plays a crucial role. The right strategy for transforming a given normal surface into a fundamental one consists in considering the so-called *clean disc patches*; in many cases they are responsible for the existence of characteristic surfaces which are not fundamental.

It is surprising that four technical tricks, elaborated by Jaco and Oertel, [56] work in all five cases listed above. We present the tricks in the form of four lemmas, but before doing that we describe clean disc patches and related notions.

Let a normal surface F in a triangulated 3-manifold (M, Γ) be presented as the geometric sum of two normal surfaces, i.e., have the form $F = G_1 + G_2$. Then the union $G_1 \cup G_2$ is a 2-dimensional polyhedron of a very specific type. The set $G_1 \cap G_2$ of its singular points consists of double lines (intersection lines of the surfaces). If we cut $G_1 \cup G_2$ along all double curves, we get a collection of pieces called *patches*. Each patch P is a compact surface. If $P \cap \partial M \neq \emptyset$, P is said to be a *boundary patch*. Otherwise P is an *interior patch*.

Performing the regular switches along all double curves of $G_1 \cup G_2$, i.e., decomposing $G_1 \cup G_2$ into patches and gluing the patches together in the appropriate way, we restore F. It is convenient to think of F as being decomposed into the same patches by *trace curves* (images of the boundaries of the patches under the gluing). If a double curve l of $G_1 \cup G_2$ is two-sided on both G_1 and G_2, then it contributes two trace curves $l_1, l_2 \subset F$. They are called *twins*. If l is one-sided on G_1 and hence on G_2, it contributes only one trace curve, which is the twin of itself.

Let l be a double curve contained in two distinct patches P_1, P_2 of $G_1 \cup G_2$. We say that P_1, P_2 are *opposite* at l, if they lie in the same surface $G_i, i = 1, 2$. The patches are *adjacent*, if they are in different surfaces. The terminology is motivated by the actual position of patches near l, see Fig. 4.2. It may happen that the same patch P approaches l from opposite sides. Then we say that P is *self-opposite*.

By a *clean disc patch* we mean a patch P homeomorphic to a disc such that either ∂P is a closed double curve, or ∂P consists of a double arc and a clean arc on ∂M. In the first case P is an interior clean disc patch, in the second it is a boundary clean disc patch. We emphasize that if a clean patch P intersects ∂M along two or more disjoint arcs, then P is *not* a clean disc patch, even if it is homeomorphic to a disc. A boundary patch P is called *quadrilateral*, if ∂P consists of two trace arcs and two arcs on ∂M such that these pairs correspond to opposite sides of the quadrilateral.

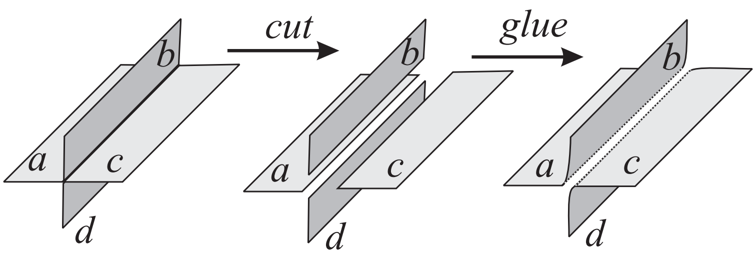

Fig. 4.2. Trace curves are shown by *dotted lines*. Pairs a, c and b, d consist of opposite patches, all other pairs consist of adjacent ones

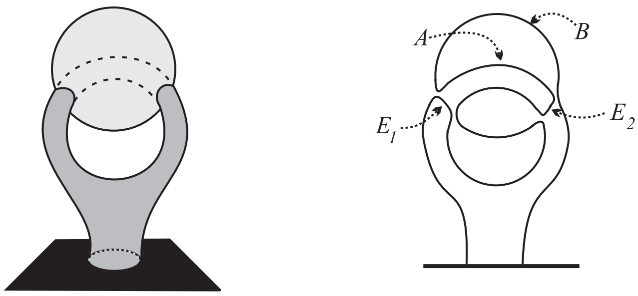

Fig. 4.3. Adjacent and opposite companions

Assume that $E, \partial E = s$, is a not self-opposite clean disc patch of F such that the twin curve s' of s cuts off a clean disc E' from F. We say that E' is *an adjacent* or *an opposite companion* of E, if the patch of E' containing s' is adjacent or opposite to E, respectively. For example, there are two clean disc patches in Fig. 4.3. Patch E_1 has the adjacent companion $A \cup E_2$, patch E_2 has the opposite companion $B \cup A \cup E_2$.

The next four lemmas show how the presence of clean disc patches helps us to simplify nonfundamental normal surfaces. Here and later on we will measure the complexity of a surface by its edge degree. So F is *simpler* than F' if $e(F) < e(F')$.

Lemma 4.1.2. *Let a normal surface F in a triangulated 3-manifold M be presented in the form $F = G_1 + G_2$. Suppose that $G_1 \cup G_2$ has a self-opposite disc patch E. Then the following holds:*

1. *M contains a normal projective plane P such that P is simpler than F.*
2. *M contains a surface F' such that $\partial F' = \partial F$, F' is homeomorphic to F, and F' is simpler than F.*

Proof. We may assume that E is contained in G_1. Then the connected component G_1' of G_1 containing E is a normal projective plane having a smaller edge degree. It intersects G_2 along a unique closed curve l, which corresponds to ∂E. This gives us the first conclusion.

To get the second conclusion, we replace the regular switch along l by the irregular one. We get a surface F' which is homeomorphic to F and has the same edge degree and the same boundary. See Fig. 4.4. Since the switch is irregular, this surface has a return. Thus we can eliminate the return and hence decrease $e(F)$ by an isotopy of F fixed on ∂F. □

Lemma 4.1.3. *Let a normal surface F in a triangulated 3-manifold (M, Γ) be presented in the form $F = G_1 + G_2$. Suppose that there is a not self-opposed clean disc patch E of $G_1 \cup G_2$ having no clean companion discs. In other words, we suppose that the twin curve s' of $s = \partial E$ is nontrivial in F. Then F is*

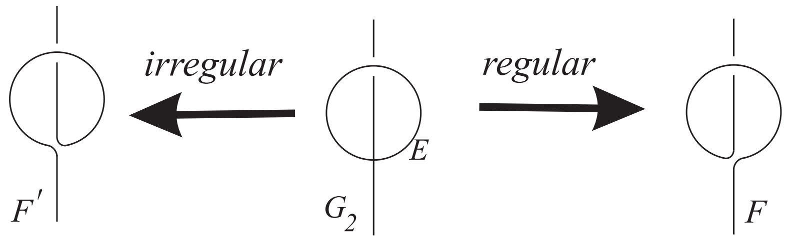

Fig. 4.4. Regular and irregular switches along l produce homeomorphic surfaces

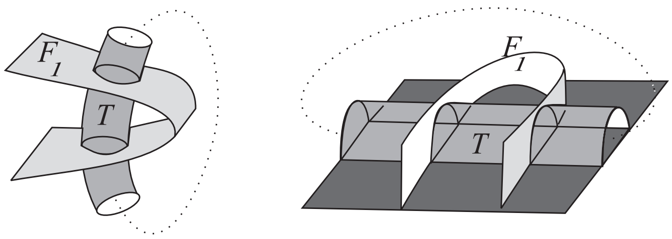

Fig. 4.5. T consists of annular or quadrilateral patches and therefore is a surface with $\chi(T) = 0$: a torus, a Klein bottle, an annulus, or a Möbius band

either compressible or boundary compressible, depending on whether E is an interior or a boundary patch.

Proof. Evident, since a parallel copy E' of E is an essential compressing or boundary compressing disc for F. □

Lemma 4.1.4. *Let a normal surface F in a 3-manifold (M, Γ) with boundary pattern be presented in the form $F = G_1 + G_2$. Assume that any clean disc patch E of $G_1 \cup G_2$ admits an adjacent companion disc E'. Then either there is a clean disc patch $E \subset F$ such that its adjacent companion disc E' is also a patch of F, or the following holds:*

1. *F can be presented in the form $F = F_1 + T$, where T is a torus, a Klein bottle, an annulus, or a Möbius band. Certainly, F_1 is simpler than F.*
2. *All patches of T are either annuli or quadrilaterals, and each double curve of $F_1 \cup T$ cuts off a clean disc patch from F_1, see Fig. 4.5. Here by a* quadrilateral *we mean a patch P whose boundary consists of two trace arcs and two arcs on ∂M such that these pairs correspond to opposite sides of P.*
3. *F_1 is homeomorphic to F as well as to the surface F' obtained from $F_1 \cup T$ by irregular switches along all the curves in $F_1 \cap T$.*
4. *If (M, Γ) is irreducible and boundary irreducible, then F is admissibly isotopic to F_1 and to F' (see Fig. 4.6).*

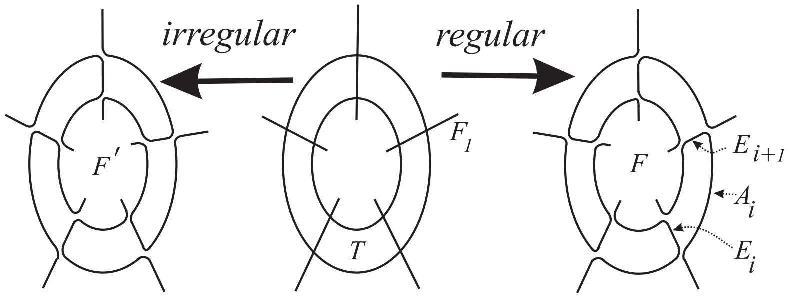

Fig. 4.6. Both regular and irregular switches convert $F_1 \cup T$ into homeomorphic surfaces

Proof. Let us construct an oriented graph Γ whose vertices are clean disc patches of F. Two patches E_1, E_2 are joined by an oriented edge $\overrightarrow{E_1E_2}$ if E_2 is contained in the adjacent companion disc E_1' of E_1. Since any clean disc patch admits an adjacent companion disc, which necessarily contains at least one clean disc patch, every vertex of Γ possesses at least one outgoing edge. It follows that Γ contains a simple cycle Z (a subgraph of Γ which consists of coherently oriented edges and is homeomorphic to a circle).

Denote by $E_0, E_1, \dots, E_{k-1}$ the successive vertices of Z. For each i, the clean disc patch E_{i+1} is contained in the adjacent companion disc E_i' of E_i (indices are taken modulo k). It may happen that $k = 2$ and $E_0' = E_1$. Then we get a pair of clean disc patches, each being a companion of the other. This corresponds to the first alternative of the conclusion of the lemma.

Assume that this situation never occurs. Then all $A_i = E_i' \setminus \text{Int } E_{i+1}, 0 \le i \le k-1$, are either annuli (if E_i are interior patches) or quadrilaterals (if they are boundary ones). Let us glue now each trace curve ∂E_i to its twin trace curve back (the same result can be obtained by making regular switches of $G_1 \cup G_2$ along all double curves except those that correspond to ∂E_i). We get a presentation $F = F_1 + T$, where T is obtained by gluing A_i while F_1 is the union of all patches of F that are not contained in T. Conditions 1, 2 are fulfilled by construction.

Let us prove that F_1, F, and F' are homeomorphic. Performing regular switches along all the curves in $F_1 \cap T$, which correspond to $\cup_{i=0}^{k-1} \partial E_i$, we actually replace each clean disc patch E_i of F_1 by the disc $A_i \cup E_{i+1}$. Similarly, irregular switches replace each disc E_i by the disc $A_{i-1} \cup E_{i-1}$. These replacements preserve the homeomorphism type of F_1. On the other hand, the result is F in the case of regular switches, and F' in the case of irregular ones. Thus F_1, F, and F' are homeomorphic.

To prove 4, consider a collection $\{S_i, 0 \le i \le k-1\}$ of surfaces in M consisting either of 2-spheres or of clean proper discs. Each S_i is composed from A_i, E_{i+1}, and a copy of E_i, $0 \le i \le k-1$. Since (M, Γ) is irreducible and boundary irreducible, each member of the collection cuts off a clean ball from M. These balls can be used for constructing an isotopy from F to F_1 and from F to F'. □

Lemma 4.1.5. *Let F be an incompressible normal surface in a triangulated 3-manifold (M, Γ). Suppose that F can be presented in the form $F = G_1 + G_2$ such that a not self-opposite clean disc patch E of $G_1 \cup G_2$ admits an opposite companion disc E'. Then the following holds:*

1. *If E is an interior disc patch, than there exists a general position surface $F' \subset M$ such that F' is homeomorphic with F, has the same boundary, and $e(F') < e(F)$. If, in addition, M is irreducible, then F' is admissibly isotopic to F.*
2. *If E is a boundary clean disc patch and (M, Γ) is irreducible and boundary irreducible, then there exists a general position surface $F' \subset M$ such that $e(F') < e(F)$ and F' is admissibly isotopic to F.*

Proof. Denote by s the double curve of $G_1 \cup G_2$ which corresponds to the twin trace curves $s_1 \subset \partial E, s_2 \subset \partial E'$. Suppose that E is an interior disc patch.

CASE 1. *E' does not contain E.* Then the sphere $S = E \cup E'$ does not decompose M, thus M is reducible. Therefore this situation never occurs for an irreducible manifold. Replacing the regular switch along s by the irregular one, we get a homeomorphic surface F_1 with a return, see Fig. 4.7a. Removing the return by an isotopy, we get a surface F' with $e(F') < e(F_1) = e(F)$. This modification of F takes place far from ∂F, so $\partial F = \partial F'$.

CASE 2. *E' contains E.* Let us replace again the regular switch along s by the irregular one. We get the disjoint union $F_1 \cup T$ of two surfaces. The first surface F_1 is homeomorphic to F and has the same boundary. The second surface T is obtained from the annulus $E' \setminus \text{Int } E$ by identifying its boundary circles and thus is either a torus or a Klein bottle. See Fig. 4.7b. Since the switch is irregular, $F_1 \cup T$ has a return. If the return is in F_1, then we can remove it by an admissible isotopy and get a surface F' with $e(F') < e(F_1) = e(F)$. If it is in T, then $e(T) \neq 0$ and $e(F_1) = e(F) - e(T) < e(F)$. Therefore, we can take $F' = F_1$.

It remains to prove that F is admissibly isotopic to F_1 (and hence to F'), provided that M is irreducible. Indeed, assuming that, attach to F a parallel copy E_1 of E. Then the sphere $S = E' \cup E_1$ bounds a ball which can be used for constructing an isotopy from F to F_1.

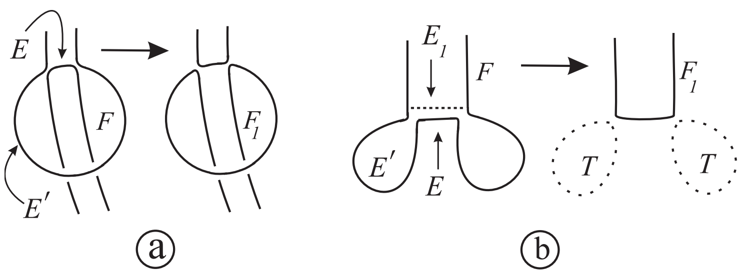

Fig. 4.7. Simplifying surfaces

If E is a boundary clean disc patch and (M, Γ) is irreducible and boundary irreducible, we apply the same tricks. The only difference is that in this case S is clean proper disc and T is either an annulus or a Möbius band. □

As a first application of Lemmas 4.1.2–4.1.5 we describe an important situation when clean disc patches are impossible. Let (M, Γ) be a triangulated 3-manifold. As before, we will measure the complexity of a proper surface $F \subset M$ by its edge degree $e(F)$.

Definition 4.1.6. *A normal surface F in (M, Γ) is called* minimal*, if $e(F)$ is the minimum for the values $e(F')$, where F' ranges over all general position surfaces in M that are admissibly isotopic to F.*

Definition 4.1.7. *Let a normal surface F in (M, Γ) be presented in the form $F = G_1 + G_2$. Then the sum $G_1 + G_2$ is in* reduced form *if F cannot be written as $F = G_1' + G_2'$, where G_1', G_2' are normal surfaces admissibly isotopic to G_1, G_2, respectively, and $G_1' \cap G_2'$ consists of fewer components than $G_1 \cap G_2$.*

It follows from conclusion 4 of Proposition 3.3.24 that if (M, Γ) is irreducible and boundary irreducible, F is incompressible and boundary incompressible, and no component of F is an essential 2-sphere, an essential clean disc, or an essential semiclean disc, then F is admissibly isotopic to a minimal normal surface. Also, if a normal surface F is written in the form $F = G_1 + G_2$, then F can be written as a sum (of isotopic surfaces) in reduced form. Recall that it makes sense to speak about incompressible and boundary incompressible patches, though they are almost never proper.

Lemma 4.1.8. *Let (M, Γ) be an irreducible boundary irreducible 3-manifold and $F \subset (M, \Gamma)$ a minimal normal surface presented in a reduced form $F = G_1 + G_2$. If F is incompressible and boundary incompressible, then $G_1 \cup G_2$ has no clean disc patches. Moreover, all patches are incompressible and boundary incompressible.*

Proof. First, we note that since F is normal and M is irreducible, no component of F is a sphere. Otherwise the sphere would be inessential and we could decrease $e(F)$ by shifting the component into the interior of a tetrahedron. Similarly, no component of F is a clean proper disc or RP^2. Let us prove that $G_1 \cup G_2$ has no clean disc patches. To the contrary, suppose that such a patch E does exist. Let us consider all possible types of E.

If E is self-opposite, then by Lemma 4.1.2 M contains a normal projective plane P such that $e(P) < e(F)$. It follows that M, being irreducible, is homeomorphic to RP^3. Up to isotopy, RP^3 contains only one closed incompressible surface without spherical components. Therefore, P is isotopic to F and we get a contradiction with the minimality of F.

If E has no companion disc, then F is either compressible or boundary compressible by Lemma 4.1.3. This contradicts our assumption. Suppose that E has an opposite clean companion disc. Then Lemma 4.1.5 tells us that

F is admissibly isotopic to a general position surface F' such that $e(F') < e(F)$, in contradiction with our assumption. Finally, suppose that every clean disc patch of $G_1 \cup G_2$ has an adjacent companion disc. Then we can apply Lemma 4.1.4. Since F is minimal, conclusion 4 of Lemma 4.1.4 does not hold. Therefore, there is a clean disc patch E such that its adjacent clean companion disc E' is also a patch. Let us merely permute E and E' by performing the regular switch of $G_1 \cup G_2$ only along the double curve $\partial E \cap \partial E'$. This gives another presentation $F = G'_1 + G'_2$ having a fewer number of double lines (since $\#(G'_1 \cap G'_2) = \#(G_1 \cap G_2) - 1$). On the other hand, $E \cup E'$ is a sphere or a clean proper disc and thus cuts off a clean ball from the irreducible boundary irreducible manifold (M, Γ). Deforming E to E' and E' to E through the ball, construct isotopies from G'_1 to G_1 and from G'_2 to G_2. But this contradicts the assumption that $G_1 + G_2$ is in reduced form.

Since all possibilities led to a contradiction, $G_1 \cup G_2$ has no clean disc patches. To prove that any patch P of F is incompressible, consider a compressing disc D for P. Our aim is to prove that D is inessential. Note that D can intersect F not only along ∂D. We will assume that the intersection is transversal.

Choose an innermost circle $c \subset F \cap D$, which bounds a disc $D' \subset D$ such that $F \cap \text{Int } D' = \emptyset$. Since F is incompressible, c bounds a disc $D'' \subset F$. It cannot happen that $D'' \supset P$; otherwise D'' would contain at least one clean disc patch of F. Therefore c can be eliminated by an isotopy of D' to the other side of F through the ball bounded by $D' \cup D''$. Doing so as long as possible, we get a new compressing disc for P (still denoted by D) such that D has the same boundary and $F \cap \text{Int } D = \emptyset$. Since F is incompressible, ∂D must bound a disc in F and hence in P (otherwise that disc would contain at least one clean disc patch). We may conclude that P is incompressible.

To see that P is boundary incompressible, consider a boundary compressing disc D for P and follow actually the same procedure for eliminating first all circles and then all arcs in $F \cap \text{Int } D$, and for proving that D must be inessential. The only difference is that we use boundary incompressibility of F, boundary irreducibility of (M, Γ), and an outermost arc argument for eliminating arcs. □

4.1.3 Recognizing the Unknot and Calculating the Genus of a Circle in the Boundary of a 3-Manifold

Let K be a knot in S^3. We would like to know whether K is trivial, i.e., bounds a disc embedded in S^3. It is well known that K is always spanned by a connected orientable surface $F \subset S^3$. Any such surface is called *a Seifert surface* for K. The minimal possible genus of Seifert surfaces for K is called the *genus* of K. Since D^2 is the only connected nonclosed surface with $\chi = 1$, K is trivial if and only if its genus is 0. Hence the recognition problem for the unknot is a partial case of the genus calculation problem.

Denote by $N(K)$ a tubular neighborhood of K in S^3, and by M_K the complement space $S^3 \setminus \mathrm{Int}\ N(K)$. Recall that a nontrivial simple closed curve $m \subset \partial N(K)$ is a *meridian* of K if it bounds a disc in $N(K)$.

Any simple closed curve $l \subset \partial N(K)$ which crosses m transversally exactly once is a *longitude* of K. A longitude l_0 is *principal* if there is an orientable surface $F \subset M_K$ such that $\partial F = l_0$. The principal longitude always exists and is unique up to isotopy. To construct it, one may take any longitude l and improve it by k negative twists along m, where $k = \mathrm{lk}(l, K)$ is the linking number. Since the homology group $H_1(M_K, Z)$ is cyclic and generated by the meridional cycle $[m]$, we get a curve l_0 such that $[l_0] = [l] - k[m] = 0 \in H_1(M_K, Z)$. It follows that l_0 bounds an orientable surface F in M_K and hence is a principal longitude.

It is important to note that there is actually no difference between Seifert surfaces for K and surfaces that bound l_0. Indeed, l_0 is isotopic to the core circle of the torus $N(K)$, i.e., to K. Therefore any surface $F \subset M_K$ that spans l_0 is isotopic to a spanning surface for K. The converse is also true, since any Seifert surface is isotopic rel K to a surface $F \subset S^3$ such that $F \cap N(K)$ is an annulus. Then the surface $F \cap M_K$ spans l_0.

Theorem 4.1.9. *There exists an algorithm that calculates the genus of a knot.*

It seems reasonable to expect that one can prove a more general statement which deals with arbitrary curves in the boundary of arbitrary 3-manifolds. Let l_0 be a simple closed curve on the boundary of a 3-manifold M. Consider the set of connected orientable surfaces in M that are bounded by l_0. The minimal genus of such surfaces is called the *genus* of l_0. If l_0 bounds no orientable surface, then the genus is ∞.

Theorem 4.1.10. *There exists an algorithm that calculates the genus of any simple closed curve l_0 in the boundary of any 3-manifold M.*

Proof. Let us triangulate M so that l_0 crosses each edge of the triangulation no more than once and is not contained inside a triangle of the triangulation. We will use the asterisk to indicate that a general position surface is connected and bounded by l_0. So F^* is *a star surface*, if F^* is connected, intersects the edges transversally, and $\partial F^* = l_0$. Let us prove that if the genus g_0 of l_0 is finite, then there is an orientable star surface of genus g_0 which is fundamental. This is quite sufficient for algorithmic calculation of the genus. All what we have to do is to construct all fundamental surfaces, select among them orientable star surfaces, and take the minimum of their genera.

By the definition of the genus, there exists an orientable star surface of genus g_0. Note that it is incompressible (otherwise we could compress it and get an orientable star surface of a smaller genus). Among all such surfaces chose a *minimal* surface, which has the smallest edge degree, and apply to it the normalization procedure described in Sect. 3.3.3. It follows from item 3 of Proposition 3.3.24 that l_0 bounds a normal homeomorphic copy F^* of the minimal surface. Since the normalization procedure does not increase the edge degree, F^* is also minimal.

Let us prove that F^* is fundamental. Assuming the contrary, choose a presentation of F^* in the form $F^* = G_1^* + G_2$. By Lemma 3.3.30 we may suppose that G_1^*, G_2 are connected, and no circle from $G_1^* \cap G_2$ separates both surfaces. One of the surfaces (G_1^*) is bounded by l_0, the other one is closed (otherwise l_0 would intersect some edge more than once). It follows that there are no boundary disc patches. Let us show that the presence of an interior disc patch leads to a contradiction.

1. Assume that there is a self-opposite interior disc patch. Then we apply conclusion 2 of Lemma 4.1.2 to construct an orientable star surface which has a return and thus can be normalized into a simpler star surface by items 3, 6 of Proposition 3.3.24. This contradicts the minimality of F^*.
2. Since F^* is incompressible, every interior disc patch of $G_1^* \cup G_2$ has a companion disc by Lemma 4.1.3.
3. Assume that an interior disc patch of $G_1^* \cup G_2$ admits an opposite companion disc. Using conclusion 1 of Lemma 4.1.5, we can replace F^* by an orientable star surface which is simpler than F^*. This contradicts the minimality of F^*.
4. Assume that every interior disc patch E admits an adjacent companion disc E'. Since the presentation $F^* = G_1^* + G_2$ had been chosen so that no double curve of $G_1^* \cup G_2$ separates both surfaces, there is no pair of companion interior disc patches. Then conclusions 1–4 of Lemma 4.1.4 hold. In particular, F^* has the form $F^* = F_1 + T$, where F_1^* is an orientable star surface homeomorphic to F^*. Obviously, F_1^* has a smaller edge degree, in contradiction with the minimality of F^*.

In Cases 1–4 we have considered all the logical possibilities for interior disc patches and found that all of them lead to a contradiction. Therefore we can suppose that $G_1^* \cup G_2$ contains no clean disc patches at all. Observe the following important fact: The no-disc-patch condition implies that G_2 is neither a sphere nor a projective plane. Therefore, $\chi(G_2) \le 0$ and $-\chi(G_1^*) \le -\chi(F^*)$.

Let us prove that G_1^* is nonorientable. Suppose, on the contrary, that it is orientable. Then its genus $g(G_1^*) = (1 - \chi(G_1^*))/2$ does not exceed $g_0 = (1 - \chi(F))/2$ and hence equals g_0 because of the minimality of g_0. On the other hand, we have $e(G_1^*) = e(F) - e(G_2) < e(F)$, in contradiction with our choice of F.

To proceed further, choose an orientation of F^* and supply all the patches of $G_1^* \cup G_2$ with the inherited orientations. Decompose the set of all double curves of $G_1^* \cup G_2$ into two subsets A, B. A double curve l is in A, if by the irregular switch along l the orientations of the uniting patches match together (by the regular switch they always do). Otherwise l is in B, see Fig. 4.8. Note that l is in A if and only if the orientations of patches that are opposite at l do not match. Since G_1^* is nonorientable, $A \ne \emptyset$.

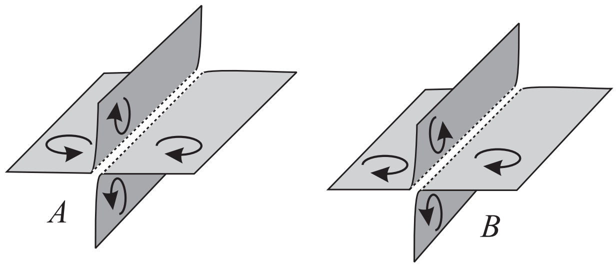

Fig. 4.8. In Case A opposite patches have opposite orientations, and the orientations of all patches match after any switch. In Case B opposite patches have coherent orientations

Now perform irregular switches at all curves in A and regular switches at all curves in B. We get an orientable star surface F'^*. Normalizing it, we decrease the edge degree (since at least one irregular switch has been made). Just as above, the no-disc-patch condition assures us that the star component of F'^* has the same or a bigger Euler characteristic. Therefore that component is simpler than F^*, in contradiction with the choice of F.

We may conclude that if the orientable star surface F^* is not fundamental, then there always exists a simpler orientable star surface. So the simplest orientable star surface must be fundamental. This completes Step 3 of the General Scheme.

In order to calculate the genus of l_0, it remains to construct all fundamental surfaces and choose the simplest orientable star surface among them. Its genus gives us the answer (if there are no orientable star surfaces at all, the genus of l_0 is ∞). □

Recall that the genus of a nonorientable surface F with boundary S^1 can be defined as $g(F) = (1 - \chi(F))/2$, i.e., by the same formula as for orientable surfaces. For example, if F is a punctured connected sum of m projective planes, then $g(F) = m/2$. The *nonorientable genus* of a simple closed curve $l_0 \subset \partial M$ can be defined as the minimum of genera of all connected surfaces which bound l_0, including nonorientable ones.

Theorem 4.1.11. *There exists an algorithm that calculates the nonorientable genus of any simple closed curve l_0 in the boundary of any 3-manifold M.*

The proof is similar to the one of Theorem 4.1.10. The only difference is that we do not need to care about orientability of surfaces and thus do not need the last step of the proof. Instead, we simply replace F^* by G_1^*.

4.1.4 Is M^3 Irreducible and Boundary Irreducible?

As we have indicated in the introduction to the book, Haken manifolds play an important role in 3-manifold topology. By definition, Haken manifolds are irreducible, boundary irreducible, and sufficiently large (see Sect. 4.1.6). But how do we know that a given 3-manifold M possesses these properties? Here we show that irreducible and boundary irreducible 3-manifolds can be recognized algorithmically (modulo algorithmic recognition of S^3, which we describe in Sect. 5).

Irreducibility. Recall that a 3-manifold M is irreducible, if every 2-sphere $S \subset M$ is inessential, i.e., bounds a 3-ball in M. Note that if M contains a one-sided projective plane P, then either M is reducible or $M = RP^3$. Indeed, the boundary S of a regular neighborhood of P is a 2-sphere (if P is normal, one can take $S = 2P$, having in mind the geometric summation). The sphere can be either essential or not. In the first case M is reducible, in the second it is RP^3.

Theorem 4.1.12. *Let M be an orientable triangulated 3-manifold. Then the following holds:*

1. *If M contains a projective plane, then at least one of the projective planes in M is fundamental.*
2. *If M is reducible, then there exists an essential sphere $S \subset M$ such that S either is fundamental or has of the form $S = 2P$, where P is a fundamental projective plane.*

Proof. Any projective plane $P \subset M$ is incompressible, since the only nontrivial simple closed curve on P is orientation-reversing, and thus cannot bound a compressing disc. By conclusion 3 of Proposition 3.3.24, P can be replaced by a normal projective plane (still denoted by P). Suppose that P is not fundamental. We claim that then M contains a normal projective plane having a smaller edge degree.

To prove this, present P as a nontrivial sum $G_1 + G_2$, where normal surfaces G_1, G_2 are connected and no double curve of $G_1 \cup G_2$ decomposes both G_1, G_2. Since the Euler characteristics of the patches of $G_1 \cup G_2$ sum up to $\chi(P) = 1$, there is at least one clean disc patch E. If E is self-opposite, then by Lemma 4.1.2 one of the surfaces G_1, G_2 is a simpler projective plane. "No companion disc" situation is impossible, since P is incompressible (see Lemma 4.1.3). If there is a clean disc patch with an opposite companion disc, or if every clean disc patch admits an adjacent companion disc, then a simpler normal projective plane in M can be found by conclusion 1 of Lemma 4.1.5 or conclusion 3 of Lemma 4.1.4, and normalization.

We have considered all logical possibilities and found that if P is not fundamental, then there is a simpler normal projective plane. It follows that any normal projective plane in M having the smallest edge degree must be fundamental.

Suppose now that M is reducible and contains no projective planes (otherwise one can take $S = 2P$, where P is a fundamental projective plane). By conclusion 3 of Proposition 3.3.24, M contains an essential normal sphere S. Assuming that S is not fundamental, present it as a nontrivial sum $G_1 + G_2$ such that normal surfaces G_1, G_2 are connected and no double curve of $G_1 \cup G_2$

decomposes both G_1, G_2. Our goal is to prove the existence of an essential normal sphere which is simpler than S, i.e., has a smaller edge degree. Indeed, since by the Jordan Theorem any circle in S bounds discs on both sides, every clean disc patch of $G_1 \cup G_2$ admits an adjacent companion disc. By Lemma 4.1.4, we can construct a presentation $S = S_1 + T$, where S_1 is a sphere and T is a torus or a Klein bottle. T consists of $k \geq 1$ annular patches $A_i, 0 \leq i \leq k-1$, while S_1 has k clean disc patches $E_i, 0 \leq i \leq k-1$, and one patch homeomorphic to a sphere with k holes.

If $k = 1$, then S_1 does not decompose M and hence is essential. Evidently, S_1 is simpler than S. Assume now that $k > 1$. We get a collection of spheres composed from $E_i \cup A_i$ and a copy of E_{i+1}, where $0 \leq i \leq k$ and indices are taken modulo k. Each of these spheres is simpler than S. If at least one of them is essential, then after normalization we get a simpler essential normal sphere (see conclusion 3 of Proposition 3.3.24). Assume that all of them bound balls. Then these balls can be used for constructing an isotopy from S to the simpler sphere S_1.

We may conclude that if S is not fundamental, then there is a simpler essential normal sphere. It follows that the simplest essential normal sphere in M must be fundamental. □

Theorem 4.1.12 is insufficient for recognition of irreducibility. It is true that if M is reducible, then an essential sphere can be found among a finite algorithmically constructible set of 2-spheres. But how can we be sure that a given 2-sphere $S \subset M$ is essential? If S does not decompose M or if it decomposes M into two parts, each containing a component of ∂M (as in the recognition of splittable links), then we are happy. In general, to make the last step of the General Scheme, we have to have a recognition algorithm for the standard 3-ball or, equivalently, a recognition algorithm for S^3. We describe both algorithms in Chap. 6.

Boundary irreducibility. It is easy to show that the connected sum of 3-manifolds is boundary irreducible if and only so are the summands. Therefore, it suffices to construct an algorithm that recognizes boundary irreducibility of irreducible manifolds.

Theorem 4.1.13. *There exists an algorithm to decide whether or not a given irreducible 3-manifold (M, Γ) is boundary irreducible. In case it is boundary reducible, the algorithm constructs an essential compressing disc.*

Proof. Let us triangulate (M, Γ). Suppose that ∂M admits an essential clean compressing disc D. By conclusion 2 of Proposition 3.3.24, after normalization we obtain a normal essential compressing disc (still denoted by D). Our goal is to show that then such a disc can be found among fundamental surfaces.

We claim that if D is not fundamental, then there is a simpler essential compressing disc. This is sufficient for proving the theorem, since then the simplest essential compressing disc must be fundamental. For proving the claim we apply the same procedure as in the proof of Theorem 4.1.12. Let D

be presented in the form $D = G_1 + G_2$, where G_1, G_2 are connected surfaces and each double curve does not separate at least one of them.

Assume that $G_1 \cup G_2$ contains an interior disc patch. Then one can construct a simpler disc D_1 with the same boundary by the same tricks as in the proof of Theorem 4.1.12. Indeed, if there exists a self-opposite disc patch E or if an interior disc patch E admits an opposite companion disc E', then it suffices to switch all curves in $G_1 \cap G_2$ regular except ∂E which we switch irregular. By conclusion 1 of Lemma 4.1.2 or conclusion 1 of Lemma 4.1.5 and normalization, we get a simpler disc with the same boundary. Since every circle in D bounds a disc, "no companion disc" situation is impossible. Thus we can use Lemma 4.1.4 to present D in the form $D = F_1 + T$ and replace D with the simpler disc $D_1 = F_1$ having the same boundary.x

Now assume that $G_1 \cup G_2$ has no interior disc patches and has a clean boundary disc patch. Observe that since every proper arc in D divides D into two discs, every boundary disc patch admits an adjacent clean companion disc. By Lemma 4.1.4, one may present D in the form $D = F_1 + T$, where F_1 is a proper disc and T is either an annulus or a Möbius band. Recall that T consists of $k \geq 1$ quadrilateral patches A_i and F_1 is the union of k boundary clean disc patches E_i plus one exceptional patch. If $k > 1$ and the collection of proper discs $E_i \cup A_i \cup E_{i+1}, 0 \leq i \leq k-1$, contains at least one essential disc, then we replace D by that disc and normalize it, thus obtaining a simpler essential compressing disc. If all the disc in the collection are inessential or if $k = 1$, then we replace D by F_1. In both cases we get a simpler essential compressing disc, since F_1 is admissibly isotopic to D for $k > 1$ and does not separate ∂M for $k = 1$. □

Corollary 4.1.14. *There is an algorithm to decide whether a given irreducible 3-manifold M is a solid torus.*

Proof. Evident, since the solid torus is the only irreducible boundary reducible 3-manifold whose boundary is homeomorphic to $S^1 \times S^1$. □

4.1.5 Is a Proper Surface Incompressible and Boundary Incompressible?

Theorem 4.1.15. *There is an algorithm to decide if a given two-sided surface in a 3-manifold M is incompressible.*

Proof. Denote by M_F the manifold obtained from M by cutting along F. Then ∂M_F contains two copies F_1, F_2 of F. Triangulate M_F such that $\partial F_1 \cup \partial F_2$ is the union of edges and supply M_F with the boundary pattern Γ composed of all edges in ∂M_F that have no common points with Int $(F_1 \cup F_2)$. It follows from the construction that (M_F, Γ) is boundary reducible if and only if F is compressible. It remains to apply Theorem 4.1.13. □

Remark 4.1.16. There is no algorithm known to decide if a one-sided surface F is incompressible. However, injectivity of F can be recognized: It suffices to test the double of F for incompressibility.

To describe a recognition algorithm for boundary incompressibility of surfaces we need some preparation.

Lemma 4.1.17. *Let an incompressible boundary incompressible connected normal surface F in an irreducible boundary irreducible triangulated 3-manifold $(M\Gamma)$ be presented in the form $F = G_1 + G_2$. Assume that there are two clean patches $P_1 \subset G_1$, $P_2 \subset G_2$ such that the following holds:*

1. *P_1, P_2 are quadrilateral discs, each having exactly two opposite sides in ∂M.*
2. *These four sides bound two clean biangles in ∂M.*
3. *$P_1 \cap P_2$ consists of the other two opposite sides of the quadrilaterals. Those sides are not in ∂M.*

Then either F is admissibly isotopic to a surface of a smaller edge degree or the sum $G_1 + G_2$ is not in reduced form.

Proof. We say that a common side l of P_1, P_2 is of type **a**, if by the regular switch at l the patches are pasted together, and of type b, if not. Then we have four possibilities: aa, ab, ba, bb, see Fig. 4.9. Since F is connected, case aa is impossible. In Cases ab and ba replace the regular switch along the b-type arc by the irregular one. We get a new surface F' consisting of a compressible annulus that cuts off a clean ball from M, and of a surface F'' that is admissibly isotopic to F. The isotopy consists in deforming a portion of F through the ball. Normalizing F'', we get a normal surface admissibly isotopic to F and having a smaller edge degree.

Consider the Case bb. Replacing G_1 by $G_1' = (G_1 \setminus P_1) \cup P_2$ and G_2 by $G_2' = (G_2 \setminus P_2) \cup P_1$, i.e., merely switching $G_1 \cup G_2$ along the two segments $P_1 \cap P_2$ irregular, we get another presentation $F = G_1' + G_2'$ such that G_i' is admissible isotopic to $G_i, i = 1, 2$, and $G_1' \cap G_2'$ consists of fewer components than $G_1 \cap G_2$. This contradicts the assumption that $G_1 \cup G_2$ is in REDUCED form. □

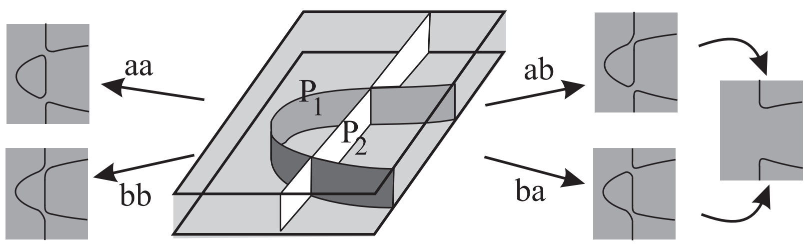

Fig. 4.9. (**aa**) F is not connected; (**ab, ba**) F is not minimal; (**bb**) F is not in reduced form

Recall that a proper disc $D \subset M$ is called *semiclean* if $\partial D \cap \Gamma$ consists of two points, see Definition 3.3.19. D is said to be *inessential*, if D is parallel rel ∂D to a disc $D' \subset \partial M$ whose intersection with Γ consists of one arc.

Lemma 4.1.18. *Let (M, Γ) be an irreducible boundary irreducible 3-manifold and L a collection of disjoint circles in Γ such that each edge of L separates two different components of $M \setminus \Gamma$. Then one can algorithmically decide, whether or not (M, Γ) contains an essential semiclean disc D such that both points of $\partial D \cap \Gamma$ lie in L.*

Proof. Denote by $\mathcal{D}$ the set of all essential semiclean normal discs in (M, Γ) whose boundaries intersect L in two points. Our goal is to construct an algorithm to decide whether or not $\mathcal{D}$ is nonempty. Let us choose a triangulation of M such that Γ is a subcomplex. We claim that $\mathcal{D} \neq \emptyset$ if and only if $\mathcal{D}$ contains a fundamental disc. Obviously, this is sufficient for proving the lemma.

To prove the claim, choose a disc $D \in \mathcal{D}$ having the minimal edge degree. Clearly, D is incompressible. Since L separates two different regions of $\partial M \setminus \Gamma$, D is boundary incompressible. Therefore we can normalize it by an admissible isotopy (see conclusion 4 of Proposition 3.3.24). Let us prove that the resulting essential semiclean normal disc D is fundamental.

Suppose, on the contrary, that there is a nontrivial presentation $D = G_1 + G_2$. By Lemma 3.3.30, we can assume that G_1, G_2 are connected and no double curve of $G_1 \cup G_2$ decomposes both surfaces. We can also assume that the presentation is in reduced form. Then by Lemma 4.1.8, $G_1 \cup G_2$ contains no clean disc patches.

Recalling that $\chi(D) = 1, \chi(G_i) \leq 2$, and the Euler characteristic is additive, we arrive at two options: 1=2+(-1) and 1=1+0. The first option is impossible, since then one of the surfaces G_i is a sphere, which must contain a disc patch. In the second case one of the surfaces (say, G_1) is a disc, the other one is either an annulus or a Möbius band. Moreover, both points in $D \cap L$ lie in G_1, since otherwise G_1 would contain a clean disc patch. It follows that G_1 is a semiclean disc, automatically inessential, since it is simpler than the minimal essential semiclean disc D.

The double curves decompose G_1 into two disc patches, each containing a point of $\partial G_1 \cap D$, and several clean quadrilaterals. Recall that by our assumption no double curve of $G_1 \cup G_2$ decomposes both surfaces. Since each proper curve on G_1 does decompose it, no double curve decompose G_2. It follows that all the patches of G_2 (which is either an annulus or a Möbius band) are clean quadrilaterals, each having two opposite sides in G_1, the other two in ∂M.

Let P_2 be a quadrilateral patch of G_2 contained in the ball B bounded by G_1 and a disc in ∂M. Suppose that P_2 is outermost with respect to B. Then the arcs $P_2 \cap G_1$ cut out a quadrilateral patch $P_1 \subset G_1$ from G_1, see Fig. 4.10. By construction, P_1, P_2 satisfy the assumption of Lemma 4.1.17. Applying it, we come to a contradiction with our assumption that $D = G_1 + G_2$ is minimal and in reduced form. This contradiction shows that D is fundamental. □

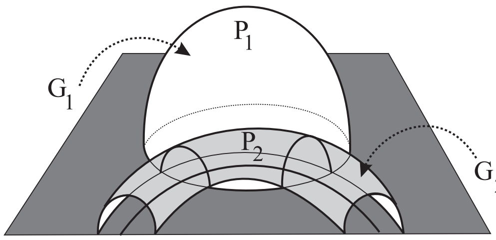

Fig. 4.10. P_1 and P_2 are two quadrilateral patches having a common pair of opposite sides

Theorem 4.1.19. *There is an algorithm to decide if a given incompressible two-sided surface in an irreducible boundary irreducible 3-manifold* (M, Γ) *is boundary incompressible.*

Proof. Denote by M_F the 3-manifold obtained from M by cutting along F. Let $F_1, F_2 \subset \partial M_F$ be two copies of F thus obtained. We supply M_F with the boundary pattern $\Gamma' = \Gamma_F \cup L'$, where $\Gamma_F \subset \partial M_F$ is obtained from Γ by cutting along ∂F, and $L' = \partial F_1 \cup \partial F_2$. Obviously, each edge of L' separates two different components of $\partial M_F \setminus \Gamma'$.

Let us show that F is boundary compressible if and only if (M_F, Γ') contains an essential semiclean disc D such that both points of $\partial D \cap \Gamma'$ lie in L'. Indeed, there is a natural map $\varphi: M_F \to M$ obtained by the reverse identification of F_1 with F_2. It is easy to see that if $D \subset M_F$ is as above, then $\varphi(D) \subset M$ is an essential boundary compressing disc for F. The converse implication is also easy, since cutting along F transforms any essential boundary compressing disc for F into an essential semiclean disc in M_F.

The desired algorithm for recognition of boundary incompressible surfaces can be now constructed as follows: we simply apply Lemma 4.1.18 to (M_G, Γ') and all the circles of $L' \subset \Gamma'$. It follows from the claim that F is boundary reducible if and only if we get at least one essential semiclean disc. □

4.1.6 Is M^3 Sufficiently Large?

Definition 4.1.20. *A compact 3-manifold* M *is* sufficiently large*, if it contains a closed connected surface which is different from* S^2, RP^2 *and is incompressible and two-sided.*

Note that the exceptional surfaces listed above are the only closed surfaces having a finite fundamental group. Since every two-sided incompressible surface is injective (see Lemma 3.3.5), the fundamental group of every sufficiently large manifold is infinite.

The class of sufficiently large 3-manifolds is "sufficiently large." For example, every irreducible 3-manifold M with nonempty boundary is either a handlebody (orientable or not), or sufficiently large. To prove that, we define a *core* of a compact 3-manifold M. Recall that if F is a proper surface in M, then M_F denotes the 3-manifold obtained from M by cutting it along F. It is convenient to think of cutting as removing an open regular neighborhood of F and thus consider M_F as a submanifold of M.

Definition 4.1.21. *Let M be an irreducible 3-manifold with nonempty boundary. Then we call a compact 3-manifold $Q \subset Int\ M$ a* core *of M, if the following holds:*

1. *Q is boundary irreducible and no connected component of Q is a 3-ball.*
2. *Q is obtained from M by successive cutting along proper discs, removing all 3-ball components that might appear under this cutting, and removing an open collar of the resulting 3-manifold to get a submanifold of $Int\ M$.*

Remark 4.1.22. It follows from condition 1 that any continuation of the cutting process brings us nothing new. Since Q is boundary irreducible, each next cut results in the appearance of a new 3-ball and the removal of it. The isotopy class of Q remains the same.

Remark 4.1.23. Sometimes the following reformulation of condition 2 is more convenient: M can be obtained from Q by collaring ∂Q, adding disjoint 3-balls, and attaching handles of index 1, see Fig. 4.11. In particular, if Q is empty, then M is a handlebody (the converse is also true).

Remark 4.1.24. Let $D_1, D_2, \ldots, D_n$ be the sequence of discs that determines Q. Then each next disc is proper in the manifold obtained by cutting along the previous discs, but it may not be proper in M. Nevertheless, the discs can be modified by an isotopy so that afterward they are disjoint and proper in M. The core remains the same (modulo isotopy). Indeed, arguing by induction, we can assume that the first k discs $D_1, \ldots, D_k$ are already disjoint and proper in M. Then we simply shift the boundary ∂D_{k+1} of the next disc from the copies of $D_1, \ldots, D_k$ in the boundary of M_{k+1}, where M_{k+1} is the manifold obtained from M by cutting along $D_1, \ldots, D_k$.

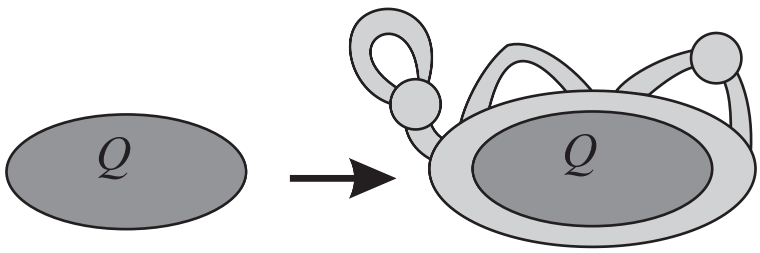

Fig. 4.11. Any 3-manifold can be obtained from its core Q by adding $\partial Q \times I$, disjoint 3-balls, and disjoint handles of index 1

Proposition 4.1.25. *Any irreducible 3-manifold has a core, which is unique up to isotopy.*

Proof. Let us construct a sequence $D_1, D_2, \ldots$ of disjoint proper discs in M. Each next disc D_k must be essential in the 3-manifold M_k obtained from $M_1 = M$ by cutting along $D_1 \cup \ldots \cup D_{k-1}$. If there appear 3-balls, we remove them at once. Let us observe that for each k we have the following:

1. $\chi(M_{k+1}) \geq \chi(M_k)$ and $\beta_1(M_{k+1}) \leq \beta_1(M_k)$, where χ is the Euler characteristic and β_1 is the first Betti number.
2. If $\chi(M_{k+1}) = \chi(M_k)$, then $\beta_1(M_{k+1}) < \beta_1(M_k)$.

Indeed, cutting M_k along D_k increases χ by one. We can get $\chi(M_{k+1}) = \chi(M_k)$ only if there appears a 3-ball, whose removal decreases χ by one. In this situation D_k, being essential, is a meridional disc of a component of M_k homeomorphic to a solid torus or to a solid Klein bottle and we have $\beta_1(M_{k+1}) = \beta_1(M_k) - 1$.

It follows that the process of constructing new essential discs and new 3-submanifolds without 3-ball components is finite and must stop with a boundary irreducible 3-manifold $M_n \subset M$, which admits no essential proper discs and contains no 3-ball components. Removing an open collar of ∂M_n in M_n, we get a core Q of M.

To prove that Q is unique up to isotopy, consider another core Q' with the defining system $D'_1, D'_2, \ldots, D'_m$ of discs which are disjoint and proper in M (see Remark 4.1.24). Let $\mathcal{D} = \cup_{i=1}^n D_i$ and $\mathcal{D}' = \cup_{j=1}^m D'_j$. Since M is irreducible, we can eliminate all circles in $\mathcal{D} \cap \mathcal{D}'$ and assume that the intersection consists of proper arcs. The arcs decompose the discs into smaller discs.

Let $M_{\mathcal{D}}, M_{\mathcal{D}'}$, and $M_{\mathcal{D}\cup\mathcal{D}'}$ be 3-manifolds obtained from M by cutting along, respectively, $\mathcal{D}, \mathcal{D}'$, and $\mathcal{D} \cup \mathcal{D}'$ (the latter can be considered as the union of those smaller discs). See Fig. 4.12. Since $M_{\mathcal{D}}$ is boundary irreducible, any disc in $\mathcal{D}' \cap M_{\mathcal{D}}$ is inessential. Therefore, cutting $M_{\mathcal{D}}$ along it preserves the isotopy class of Q. We can conclude that Q is isotopic to a submanifold $Q_0 \subset M$ obtained from $M_{\mathcal{D}\cup\mathcal{D}'}$ by throwing away 3-ball components and

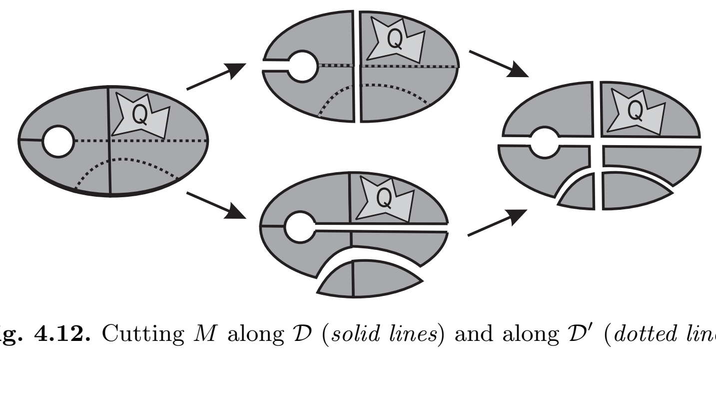

Fig. 4.12. Cutting M along $\mathcal{D}$ (*solid lines*) and along $\mathcal{D}'$ (*dotted lines*)

removing an open collar of the boundary of the resulting manifold. The same is true for Q'. Therefore, Q and Q' are isotopic. □

Lemma 4.1.26. *Let $Q \subset Int\ M$ be the core of an irreducible 3-manifold M. Then ∂Q is incompressible in M.*

Proof. Consider an arbitrary compressing disc $\Delta \subset M$ for ∂Q. We wish to prove that Δ is inessential, that is, its boundary curve bounds a disc in ∂Q. If Δ lies in Q, then it is inessential by condition 1 of Definition 4.1.21.

Suppose that Δ lies outside Q. Let $\mathcal{D} = D_1 \cup \ldots \cup D_n$ be a collection of disjoint proper discs in M which decomposes M into 3-balls and a copy M_n of Q. Since M is irreducible, we can deform Δ isotopically away from $\mathcal{D}$. Then Δ is contained in the collar $M_n \subset \text{Int } Q = \partial Q \times I$ and hence must be inessential in this case as well. □

Corollary 4.1.27. *Every irreducible 3-manifold M with nonempty boundary is either a handlebody or sufficiently large.*

Proof. Follows from Proposition 4.1.25 and Lemma 4.1.26: if M is not a handlebody, then the boundary of the core of M is an incompressible surface in M different from S^2 and RP^2. □

Definition 4.1.28. *A compact 3-manifold M is called* Haken*, if it is irreducible, boundary irreducible, and sufficiently large.*

It follows from Corollary 4.1.27 that every irreducible boundary irreducible 3-manifold M with nonempty boundary is either a 3-ball or Haken. So all interesting non-Haken 3-manifolds are closed. Examples of such manifolds can be found among Seifert manifolds fibered over S^2 with three exceptional fibers. Some closed hyperbolic 3-manifolds also possess this property, among them the manifolds M_1, M_2 described in Sect. 2.5.1. In general, if you take a closed irreducible 3-manifold by chance, then most probably it would be Haken.

Lemma 4.1.29. *Let M be a closed irreducible 3-manifold such that the group $H_1(M; Z)$ is infinite and M contains no projective planes. Then M is Haken.*

Proof. Since $H_1(M; Z)$ is infinite, there exists a map $f: M \to S^1$ such that the induced homomorphism $f_\star: \pi_1(M) \to H_1(M; Z)$ is surjective. Then the inverse image of any regular point $a \in S^1$ is a nonseparating surface. Compressing it as long as possible, we get an incompressible surface F. Since the property of a surface to be nonseparating is preserved under compressions, F is nonempty. It follows from our assumption that F is different from S^2, RP^2. □

In this section we present the following important result of Jaco and Oertel [56].

Theorem 4.1.30. *There exists an algorithm to decide if an irreducible manifold is sufficiently large.*

Since irreducibility and boundary irreducibility of a 3-manifold can be recognized algorithmically (see Theorem 4.1.13), Corollary 4.1.31 is evident.

Corollary 4.1.31. *There exists an algorithm to decide if an irreducible manifold is Haken*

The ideas used in the proof of Theorem 4.1.30 influenced significantly the development of the normal surface theory and turned out to be very important for analyzing the complexity of algorithms based on it. The proof follows the same General Scheme, but before presenting it we prove several lemmas that are useful not only for this particular proof.

Let (M, Γ) be a triangulated 3-manifold. As before, we will measure the complexity of a general position proper surface $F \subset M$ by its edge degree $e(F)$. We slightly generalize the notion of compressing disc for a surface by extending it to the case of arbitrary subpolyhedra. Let X be a compact 2-dimensional subpolyhedron of a 3-manifold (M, Γ). By *a singular graph* $S(X)$ of X we mean the union of all points in X having no disc neighborhood.

Definition 4.1.32. *A disc* $D \subset M$ *is called* a compressing disc *for* X *if* $D \cap X = \partial D$ *and the curve* ∂D *is transversal to* $S(X)$*. Similarly, a clean disc* $D \subset M$ *is* a boundary compressing disc *for* X *if* $l = D \cap X$ *is an arc in* ∂D*,* l *is transversal to* $S(X)$*, and* $D \cap \partial M$ *is the complementary arc in* ∂D*.*

Let $F = G_1 + G_2$ be a normal surface in a 3-manifold M presented as a geometric sum of two normal surfaces, and let l be a component of $G_1 \cap G_2$. Locally, in a close neighborhood of every point $x \in l$, the surfaces G_1, G_2 look like two planes (or half-planes, if $x \in \partial M$) forming four dihedral angles. A dihedral angle is called *good*, if the patches forming its sides are pasted together under the regular switch at l. Otherwise the angle is *bad*.

Any compressing disc D for $G_1 \cup G_2$ can be considered as a curvilinear polygon whose angles are labeled by letters g or b depending on whether the corresponding dihedral angles are good or bad, respectively. Analogously, any boundary compressing disc D for $G_1 \cup G_2$ is a polygon whose all but two angles are labeled by letters g or b (each of the two angles without labels is formed by an arc in $D \cap \partial M$ and an arc in $D \cap G_1$ or $D \cap G_2$). Let us describe the behavior of D with respect to the regular switch of $G_1 \cup G_2$ and to the resulting surface $F = G_1 + G_2$. Near each double line l of $G_1 \cup G_2$, consider a strip A that runs along l and spans the trace curves of l. If l is closed, then A is either an annulus (if l is two-sided in both G_1, G_2) or a Möbius band (if l is one-sided). In case l is a proper arc, A is a disc band (homeomorphic image of a rectangle) such that two opposite sides of A are in ∂M while the other two coincide with the trace curves of l.

Consider the union of F with all strips obtained in this way. Then D determines a compressing or boundary compressing disc $\tilde{D}$ for the union. Near each good angle of D, the boundary curve of $\tilde{D}$ is contained in F while

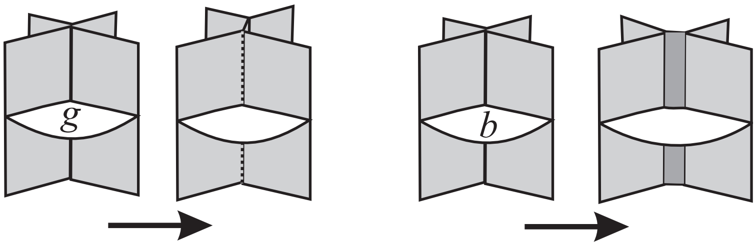

Fig. 4.13. Behavior of compressing discs near good and bad angles

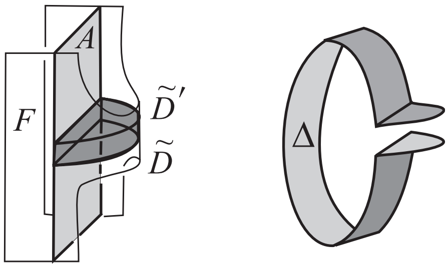

Fig. 4.14. Parallel compressing discs for $F \cup A$ and a compressing disc Δ for F

near each bad angle it crosses the corresponding strip. See Fig. 4.13. If all angles of D are good, then $\tilde{D}$ is a compressing or boundary compressing disc for F.

Lemma 4.1.33. *Let an incompressible normal surface F in an irreducible boundary irreducible manifold (M, Γ) be presented in the form $F = G_1 + G_2$ such that there are no clean disc patches. Assume that $G_1 \cup G_2$ admits a compressing disc D such that precisely one angle of D is bad. Then the edge degree of F can be decreased by an admissible isotopy.*

Proof. Denote by l the double line of $G_1 \cup G_2$ that passes through the vertex of the bad angle of D. Let A and $\tilde{D}$ be the corresponding band and compressing disc for $F \cup A$. Assume that l is closed. Denote by B' the ball bounded by $\tilde{D}$, a close parallel copy $\tilde{D}'$ of $\tilde{D}$ (see Fig. 4.14), and the portions A' of A and F' of F between them. Consider the disc $\Delta \subset M$ composed of $\tilde{D}$, $\tilde{D}'$, and the band $\mathrm{Cl}(A \setminus A')$, see Fig. 4.14 to the right. It is easy to see that Δ is a compressing disc for F. Since F is incompressible and M is irreducible, there is a ball $B \subset M$ bounded by Δ and a disc δ in F. It cannot happen that $B' \subset B$, otherwise each trace curve of l would be contained in δ and bound a disc in $F \cap \delta \subset F$, which contradicts the "no clean disc patch" assumption. Therefore $B' \cap B = \tilde{D} \cup \tilde{D}'$, and the union $B \cup B'$ is a solid torus T. This solid torus helps us to construct an isotopy of F to the surface $F_1 = (F \setminus \partial T) \cup A$ which is simpler than F (see Fig. 4.15).

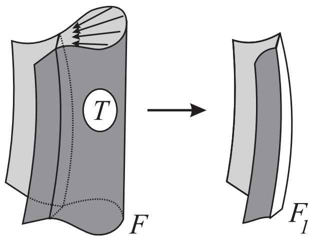

Fig. 4.15. Simplifying isotopy

The case when l is a proper arc is similar. Consider a disc $\Delta \subset (M, \Gamma)$ composed of $\tilde{D}$ and a connected component of $A \setminus (A \cap \tilde{D})$. Clearly, Δ is a clean boundary compressing disc for F. Since F is boundary incompressible and (M, Γ) is irreducible and boundary irreducible, there is a clean ball $B \subset M$ bounded by the union of Δ, a disc in F, and a disc in ∂M. A similar ball is placed on the other side of Δ. The union $T \sim \tilde{D} \times I$ of these two balls helps us to construct an admissible isotopy of F to the surface $F_1 = (F \setminus \partial T) \cup A$ which is simpler than F. □

Lemma 4.1.34. *Let an incompressible boundary incompressible normal surface F in a manifold (M, Γ) be presented in the form $F = G_1 + G_2$ such that there are no clean disc patches. Assume that $G_1 \cup G_2$ admits a compressing or boundary compressing disc D such that D has no bad angles and at least one good angle. Then for $i = 1, j = 2$ or for $i = 2, j = 1$ there exists a disc $D_0 \subset G_i$ such that ∂D_0 consists of an arc in G_j, an arc in ∂D, and maybe an arc in ∂M while Int D_0 has no common points with $G_j \cup D \cup \partial M$.*

Proof. Since there are no bad angles, D determines a compressing or boundary compressing disc $\tilde{D}$ for F. The curve $\alpha = \partial \tilde{D} \cap F$ must cut off a disc $\tilde{D}'$ from F. Recall that trace curves decompose F into patches. Consider the induced decomposition of $\tilde{D}'$. The boundary of every region of the induced decomposition consists of arcs contained in trace curves and of arcs in $\partial \tilde{D}'$. The "no clean disc patch" assumption assures us that $\tilde{D}'$ contains no closed trace curves and proper trace arcs having both endpoints on ∂M. Therefore there exists a region bounded by a trace arc, an arc in α, and maybe an arc in ∂M. Any such region (in Fig. 4.16 they are marked by stars) determines a disc D_0 satisfying the conclusion of the lemma. □

Let a proper normal surface F in a triangulated 3-manifold M be presented in the form $F = G_1 + G_2$ and let Δ be a compressing or boundary compressing disc for G_1. We will assume that Δ is in general position with respect to G_2. Then $\Delta \cap G_2$ is a proper 1-dimensional submanifold of Δ. Define *the weight* of Δ to be the number $c(\Delta) + c_\partial(\Delta)$, where $c(\Delta)$ is the number of components of $\Delta \cap G_2$ and $c_\partial(\Delta)$ is the number of points in $\partial \Delta \cap G_2$ which are not on ∂M. Our goal is to decrease the weight.

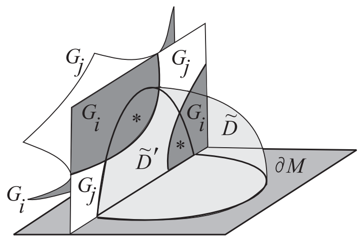

Fig. 4.16. Trace curves decompose F as well as $\tilde{D}'$

Lemma 4.1.35. *Let $F = G_1 + G_2$ be a proper normal surface in a triangulated irreducible boundary irreducible 3-manifold (M, Γ) such that all patches of F are incompressible and boundary incompressible. Let Δ be an essential compressing or boundary compressing disc for G_1. Assume that there is a disc $D_0 \subset M$ such that at least one of the following holds:*

(1) *$D_0 \subset \Delta$ and either $D_0 \cap G_2 = \partial D_0$ or $D_0 \cap G_2$ is an arc in ∂D_0 while the complementary arc of ∂D_0 is in ∂M.*
(2) *$D_0 \subset G_1$, $\partial D_0 = D_0 \cap (G_2 \cup \Delta \cup \partial M)$, and ∂D_0 consists of an arc in G_2, an arc in $\partial \Delta$, and maybe an arc in ∂M.*
(3) *$D_0 \subset G_2$ and ∂D_0 consists of an arc in G_1, an arc in $\partial \Delta$, and maybe an arc in ∂M, while Int D_0 is disjoint from $G_1 \cup \Delta \cup \partial M$.*

Then there is an essential compressing or boundary compressing disc Δ' for G_1 having a strictly smaller weight.

Proof. Case 1. It follows from the assumptions that $\partial D_0 \cap G_2$ is contained in a patch $P \subset G_2$ of $G_1 \cup G_2$. Since P is incompressible and boundary incompressible, the curve $\partial D_0 \cap G_2$ cuts off a clean disc D from P. Recall that M is irreducible and boundary irreducible. It follows that $D \cup D_0$ cuts off a clean ball from M. Using the ball, construct an isotopy of D_0 to the other side of P and get a new disc Δ' having a smaller weight.

Case 2. We use D_0 for constructing an isotopy of Δ that shifts the arc $l = D_0 \cap \Delta$ to the other side of G_2. Clearly, the isotopy decreases the weight. See Fig. 4.17a,b for the cases $\partial l \cap \partial M = \emptyset$ and $\partial l \cap \partial M \neq \emptyset$.

Case 3. Compressing Δ along D_0, we get two simpler compressing or boundary compressing discs for G_1 with strictly smaller weights. It is clear that if Δ is essential, then so is at least one of the new discs. See Fig. 4.18. □

Theorem 4.1.36. *Let a minimal connected normal surface F in an irreducible boundary irreducible 3-manifold (M, Γ) be presented in the form $F = G_1 + G_2$. If F is incompressible and boundary incompressible, then so are G_1, G_2. Moreover, neither G_1 nor G_2 is a sphere, a projective plane, a clean disc or a disc whose boundary crosses Γ exactly once.*

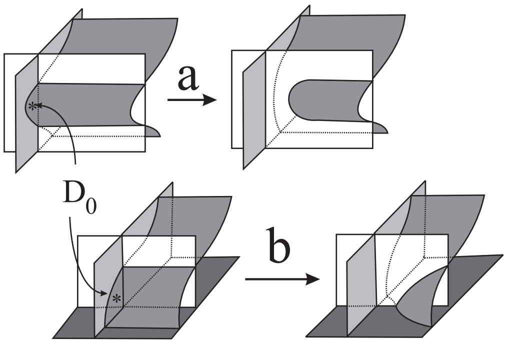

Fig. 4.17. Simplifying Δ by shifting $\partial\Delta$ along G_1

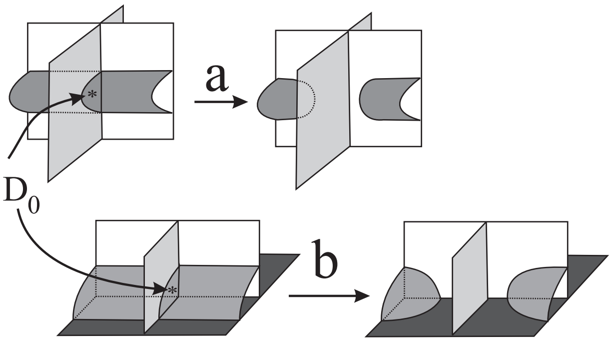

Fig. 4.18. Compressing Δ

Proof. It is sufficient to prove the theorem with the additional assumption that $F = G_1 + G_2$ is in reduced form. Then Lemma 4.1.8 tells us that $G_1 \cup G_2$ has no clean disc patches and all the patches of $G_1 \cup G_2$ are incompressible and boundary incompressible. Since any decomposition of S^2, RP^2, or D^2 with no more than one point in $\partial D^2 \cap \Gamma$ into patches contains a clean disc patch, G_1, G_2 are different from these surfaces. By the symmetry, we only need to prove that one surface, say, G_1, is incompressible and boundary incompressible.

On the contrary, suppose that G_1 admits an essential compressing or boundary compressing disc Δ. The idea is to decrease the weight $c(\Delta)+c_\partial(\Delta)$.

Assume that $\Delta \cap G_2$ contains a circle or an arc with the endpoints on ∂M. Using an innermost circle or an outermost arc argument, we find a disc $D_0 \subset \Delta$ satisfying condition (1) of Lemma 4.1.35. Applying the lemma, we find a simpler compressing or boundary compressing disc. So for the remainder of the argument we may assume that $\Delta \cap G_2$ contains no circles and no arcs with endpoints in ∂M. It follows that the arcs $\Delta \cap G_2$ cut Δ into regions which are homeomorphic to the standard disc and thus can be considered as curvilinear polygons. Every region is a compressing or boundary compressing disc for $G_1 \cup G_2$.

Let us count the number of good and bad angles in these polygons. Any point in $\partial\Delta \cap G_2$ is a common vertex of two angles belonging to different polygons. One of the angles is good, the other is bad. We relate the total number of polygons in Δ and the total number of their bad angles. Denote by m the number of arcs in $\Delta \cap G_2$ that are disjoint from ∂M, and by n the number of arcs having one endpoint in ∂M. Then there are $m+n+1$ polygons and $2m+n$ bad marks. Since $2(m+n+1) > 2m+n$, there is a polygon having ≤ 1 bad angles.

Note that there are no polygons having precisely one bad angle; otherwise F would be not minimal by Lemma 4.1.33. Therefore there is a polygon D that has no bad angles. By Lemma 4.1.35, there exists a disc $D_0 \subset M$ satisfying either assumption (2) or assumption (3) of Lemma 4.1.35. Applying the lemma, we get a simpler essential compressing or boundary compressing disc again.

Continuing this simplification procedure for as long as possible, we get an essential compressing or boundary compressing disc Δ for G_1 that has zero weight and thus does not intersect G_2. This contradicts the assumption that all patches of $G_1 \cap G_2$ are incompressible and boundary incompressible. □

Corollary 4.1.37. *Let a minimal connected normal surface F in an irreducible boundary irreducible 3-manifold (M, Γ) be presented as a sum $F = \sum_{i=1}^{n} G_i$ of $n > 0$ nonempty normal surfaces. If F is incompressible and boundary incompressible, then so are all G_i. Moreover, no G_1 is a sphere, a projective plane, a clean disc or a disc whose boundary crosses Γ exactly once.*

Proof. Rewrite F in the form $F = G_1 + G'$, where $G' = \sum_{i=2}^{n} G_i$. By Theorem 4.1.36, G_1 is incompressible, boundary incompressible, and different from S^2, D^2, and RP^2. The same trick works for all other surfaces G_i. □

Remark 4.1.38. If the surface F in the statements of Theorem 4.1.36 and Corollary 4.1.37 is closed, then the assumption that M is boundary irreducible is superfluous. Indeed, we have used it in the proofs of the theorem, corollary, and preceding lemmas only when $\partial F \neq \emptyset$ (if $\partial F = \emptyset$, then all events are going on strictly inside M).

Proof of Theorem 4.1.30 (On Recognizing Sufficiently Large Manifolds). Let us triangulate a given manifold M and construct the finite set $\{G_i, 1 \leq i \leq n\}$ of all closed fundamental surfaces. Suppose that M contains a two-sided closed incompressible surface $F \neq S^2, RP^2$. We may assume that F is minimal. Present F as a sum $F = \sum_{i=1}^{n} k_i G_i, k_i > 0$, of some fundamental surfaces. It follows from Theorem 4.1.36 that all surfaces $G_i, 1 \leq i \leq n$, are incompressible and different from S^2 and RP^2. However, this is not good enough to develop the desired algorithm. It may be not true that at least one of G_i is two-sided.

To overcome this obstacle, consider the presentation $2F = \sum_{i=1}^{n} k_i(2G_i)$. Since $2F$ is also minimal and incompressible, all surfaces $2G_i$ are incompressible by Corollary 4.1.37. Certainly, they are two-sided.

The desired algorithm can be described as follows. We apply the algorithm of Corollary 4.1.15 to the check whether the double $2G$ of each fundamental surface $G \neq S^2, RP^2$ is incompressible. If we get a positive answer at least once (i.e., if we find a fundamental surface $G \neq S^2, RP^2$ with an incompressible double $2G$), then M is clearly sufficiently large. If all the doubles turn to be compressible, then M is not sufficiently large. □

Remark 4.1.39. The algorithm constructed above can be sharpened by reducing the set of all fundamental surfaces to a much smaller subset of so-called *vertex surfaces*. Every vertex surface G_j corresponds to an admissible vertex solutions $\bar{V}_j$ to the matching system E for the triangulation of M (see the proof of Theorem 3.2.8 for a description of vertex surfaces). Any integer solution $\bar{x}$ to E can be written in the form $\bar{x} = \sum_{j=1}^{N} \alpha_j \bar{V}_j$, where α_j are non-negative rational numbers. Obviously, if $\bar{x}$ is admissible, then so are all $\bar{V}_j$ with $\alpha_j > 0$. It follows that for every normal surface F an appropriate multiple $2kF$ can be presented in the form $2kF = \sum_{j=1}^{m} k_j(2G_j)$, where all k_j are integer and G_j are vertex surfaces. If F is minimal and incompressible, then Corollary 4.1.37 tells us that all these vertex surfaces are incompressible. Therefore M is sufficiently large if and only if the double of at least one vertex surface is incompressible.

4.2 Cutting 3-Manifolds along Surfaces

In this section we investigate what happens to the complexity of a 3-manifold when we cut the manifold along an incompressible surface.

4.2.1 Normal Surfaces and Spines

Let M be a 3-manifold with nonempty boundary. There is a close relationship between handle decompositions and spines of M. Indeed, let ξ be a handle decomposition of M into balls, beams, and plates, without handles of index 3. Collapsing the balls, beams, and plates of ξ onto their core points, arcs, and discs, we get a spine P of M. By construction, P is equipped with a natural cell decomposition into the core cells of the handles. Conversely, let P be a *cellular spine* of M, i.e., a spine equipped with a cell decomposition. Replace each vertex of P by a ball, each edge by a beam, and each 2-cell by a plate (we used a similar construction in Sect. 1.1.4). We get a handle decomposition ξ of a regular neighborhood $N(P)$ of P in M. Since $\partial M \neq \emptyset$, $N(P)$ can be identified with the whole manifold M, so ξ can be considered as a handle decomposition of M.

Consider a normal surface $F \subset M$ and denote by M_F the 3-manifold obtained by cutting M along F. Let us investigate the behavior of ξ and P under the cut.

Since F is normal, it decomposes the handles of ξ into handles of the same index. The new handles form a handle decomposition ξ_F of M_F. Denote by P_F the corresponding cellular spine of M_F. We can think of each handle of ξ_F as being contained in the corresponding handle of ξ. This inclusion relation induces a cellular map $\varphi: P_F \to P$ (a map is *cellular* if it takes cells to cells). For any vertex v of P_F, the map φ, being cellular, induces a map $\varphi_v: \mathrm{lk}(v, P_F) \to \mathrm{lk}(w, P)$ between links, where $w = \varphi(v)$.

Another way for describing φ_v consists in the following. Collapsing each island in ∂B_v to its core point and each bridge to its core edge, we get a graph isomorphic to $\mathrm{lk}(v, P_F)$. The same is true for the ball B of ξ containing w: the island-beam configuration in ∂B has the shape of $\mathrm{lk}(w, P)$. Then φ_v is induced by the inclusion relation between islands and bridges of ξ_F and ξ.

Lemma 4.2.1. *Suppose that a handle decomposition ξ of a 3-manifold M with nonempty boundary corresponds to an almost simple cellular spine P of M. Let F be a normal surface in M, and let ξ_F, P_F, and φ be, respectively, the induced handle decomposition of M_F, cellular spine of M_F, and cellular map $P_F \to P$. Then P_F is almost simple, and for any vertex v of P_F and the corresponding vertex $w = \varphi(v)$ of P the induced map φ_v: $lk(v, P_F) \to lk(w, P)$ is an embedding.*

Proof. Let v, $w = \varphi(v)$ be vertices of P_F, P, and $B_v \subset B_w$ the balls of ξ_F, ξ containing them. We observe the following:

(a) *Any bridge of B_v is contained in a bridge of B_w, and any bridge of B_w contains no more than one bridge of B_v.* The first statement is evident. Let us prove the second (it is true for all handle decompositions, not only for those arising from almost simple spines).

Suppose, on the contrary, that a bridge b of B_w contains two bridges b', b'' of B_v. We can assume that b', b'' are neighbors, that is, the strip $S \subset b$ between them contains no other bridges of B_v. Note that the region $U = \partial B_v \cap \partial B_w$, just as every connected region in $\partial B_w \approx S^2$ bounded by disjoint circles, has exactly one common circle with each connected component of $\mathrm{Cl}(\partial B_w \setminus U)$. It follows that the lateral sides of S lie in the same circle $C \subset \partial U$. Since C is the boundary of an elementary disc in $F \cap B_w$ and since the boundary of any elementary disc crosses each bridge no more than once, we get a contradiction.

(b) *Any island of B_v is contained in an island of B_w, which contains no other islands of B_v.* Since ξ corresponds to an almost simple spine and thus all islands of ξ have valence ≤ 3, the boundary curve of any elementary disc passes through any island no more than once (otherwise condition 7 of Definition 3.4.1 of a normal surface would be violated). Thus the same proof as in (a) does work.

Since φ_v is induced by the inclusion relation between islands and bridges of ξ_F and ξ, it follows from (a), (b) that φ_v is an embedding. Since it is true for all vertices of P_F and P is almost simple, P_F is also almost simple. □

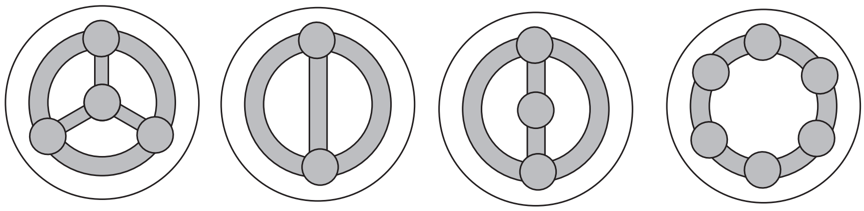

Fig. 4.19. Island-bridge-lake configurations in the boundaries of balls

Suppose P is a simple spine of a 3-manifold M with nonempty boundary. It is convenient to choose a cell decomposition σ of P such that any true vertex of P is incident to only four 1-cells, at any vertex of σ inside a triple line only two or three 1-cells meet together, and inside 2-components there are no vertices of σ incident to only one 1-cell. Denote by ξ the handle decomposition of M induced by σ.

et v be a vertex of σ. Denote by B_v the corresponding ball. The boundary of B_v is decomposed into islands, bridges, and lakes. If v is a true vertex of P, then ∂B_v contains 4 islands, 6 bridges, and 4 lakes such that any two islands are joined by one bridge. If v is a triple point, then ∂B_v contains two or three islands, and if v is a nonsingular point, then the islands and bridges compose an annulus, see Fig. 4.19.

Let us investigate the types of elementary discs in B_v.

Definition 4.2.2. *An elementary disc $D \subset B_v$* has type (k, m) *if its boundary curve ℓ intersects k bridges and m lakes (recall that by definition of an elementary disc, ℓ passes through each bridge and each lake no more than once).*

Lemma 4.2.3. *Let a ball B_v of ξ corresponds to a vertex v of a simple spine P. Then any elementary disc in B_v has one of the following types:*

1. *Types $(4,0), (3,0), (2,1), (1,2), (0,4), (0,3), (0,2)$, if v is a true vertex.*
2. *Types $(2,0), (1,1), (0,2)$, if v lies on a triple line and ∂B_v contains two islands.*
3. *Types $(3,0), (2,1), (2,0), (1,1), (0,2), (0,3)$, if v lies on a triple line and ∂B_v contains three islands.*
4. *Types $(0,2)$ and $(k,0)$, if v is a nonsingular vertex and ∂B_vcontains k islands.*

With a few exceptions, each type determines the corresponding elementary disc in a unique way up to homeomorphisms of B_v taking islands to islands, bridges to bridges, and lakes to lakes. The exceptions are:

1. *If v is a true vertex, then B_v contains two elementary discs of the type $(0,3)$.*
2. *If v is a triple vertex and B_v contains three islands, then there are two elementary discs of the type $(0,2)$.*

3. *If v is a nonsingular vertex of valence k, then there is one elementary disc of the type $(k,0)$ and $[k/2]$ discs of the type $(0,2)$, where $[k/2]$ is the integer part of $k/2$.*

Proof. Let ℓ be the boundary curve of a type (k,m) elementary disc. Then $n = k+m$ is the number of arcs in the intersection of ℓ with the union of all islands. Since each island has valence ≤ 3, ℓ visits each islands no more than once (otherwise condition 7 of Definition 3.4.1 would be violated). Moreover, at least two islands must be visited.

CASE 1. If v is a true vertex, then B_v contains four islands. Therefore, $2 \le k+m \le 4$. It remains to enumerate all possible pairs (k,m) with $2 \le k+m \le 4$ and verify that only the pairs listed in the statement of Lemma 4.2.3 are realizable by elementary discs, and that each of them admits a unique realization except the pair (0,3) that admits two, see Fig. 4.20.

CASE 2. If v is a triple vertex such that B_v contains two islands, then $k+m = 2$. It is easy to verify that only the pairs listed in item 2 are realizable and that the realizations are unique.

CASE 3 is similar. The only difference is that there are two elementary discs of the type $(0,2)$, see Fig. 4.21.

In the last CASE 4 of a nonsingular vertex v the proof is evident. For the annular island-bridge configuration with $k = 6$ islands presented on Fig. 4.21 to the right we show all three discs of type $(0,2)$. □

Our next goal is to show that in many cases cutting a 3-manifold M along an incompressible surface makes M simpler.

Lemma 4.2.4. *Suppose that a handle decomposition ξ of a 3-manifold M with nonempty boundary corresponds to a simple cellular spine P of M. Let F be a connected normal surface in M, and let ξ_F, P_F, and φ be, respectively, the induced handle decomposition of M_F, cellular spine of M_F, and cellular map $P_F \to P$. Then the following holds:*

1. *φ embeds the set of true vertices of P_F into the one of P. This embedding (denote it by φ_0) is bijective if and only if all the elementary discs of F in the balls around true vertices have type $(3,0)$.*
2. *If φ_0 is bijective, then φ embeds the union of triple circles of P_F into the one of P. This embedding (denote it by φ_1) is bijective if and only if all elementary discs of F in the balls around triple vertices have type $(3,0)$ or $(2,0)$.*
3. *If φ_0 and φ_1 are bijective and $\partial F \neq \emptyset$, then F is either an annulus or a Möbius band intersecting only balls of ξ_P that correspond to nonsingular vertices. Moreover, F can intersect these balls only along elementary discs of type $(0,2)$ and $F \cup P$ is a middle circle of F.*

Proof. Let us prove the first conclusion of the lemma. Let v_1 and $v = \varphi(v_1)$ be true vertices and $B_1 \subset B$ the corresponding balls. Then $\mathrm{lk}(v_1, P_F)$ and

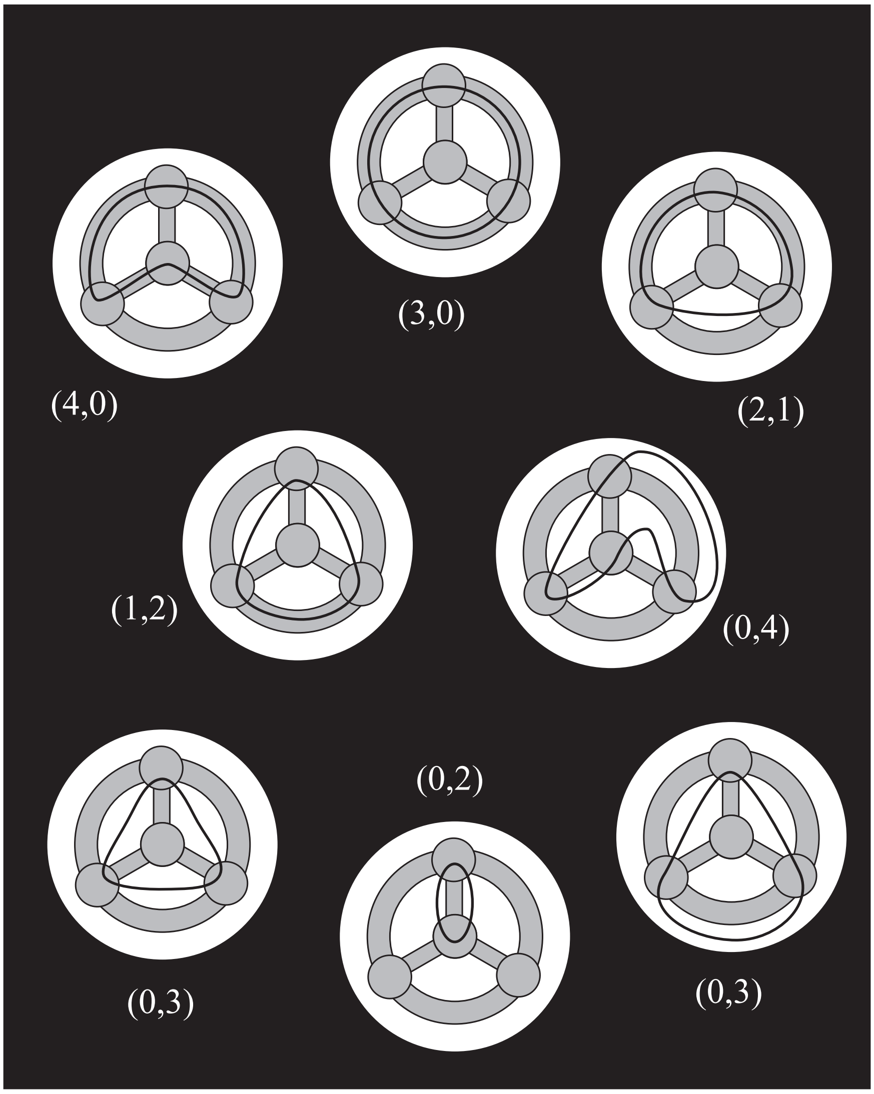

Fig. 4.20. Eight types of elementary discs in a ball neighborhood of a true vertex

$\mathrm{lk}(v, P)$ are homeomorphic. We claim that every elementary disc $D \subset F \cap B$ has type $(3, 0)$. Obviously, the first conclusion follows from the claim.

To prove the claim, suppose that D has type (k, m) with $m > 0$. Then cutting B along D destroys at least one three-valent vertex of $\mathrm{lk}(v, P)$, namely, the one corresponding to the island where ∂D crosses the coast of a lake. Since $\mathrm{lk}(v_1, P_F)$ has four three-valent vertices and $\mathrm{lk}(v_1, P_F)$ and $\mathrm{lk}(v, P)$ are homeomorphic, it is impossible. Cutting along type $(4, 0)$ disc preserves

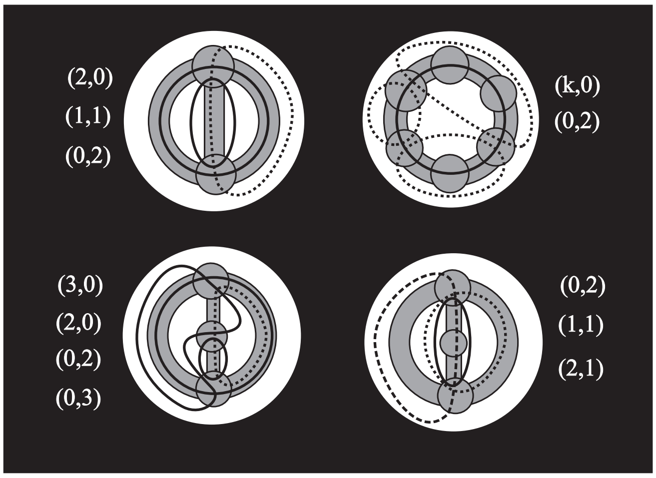

Fig. 4.21. The boundaries of elementary discs in ball neighborhoods of other vertices

three-valent vertices, but decomposes $\text{lk}(v, P)$ into two graphs, each containing two three-valent vertices. Therefore this case is also impossible.

This observation can be easily verified by applying Lemma 4.2.3 and considering Fig. 4.22, which shows why type $(3, 0)$ discs preserve the true vertex and how discs of all other types destroy it.

The proof of the second conclusion of the lemma is similar. Denote by E the union of all true vertices of P with all the triple edges that join them. Let E_F be the similar union for P_F. Suppose that φ_0 is bijective. Then φ induces a homeomorphism between E_F and E. Let us investigate the behavior of φ on the set of triple circles. It follows from Lemma 4.2.3 that the image $\varphi(C)$ of a triple circle $C \subset P_F$ is a triple circle in P if and only if all the elementary discs in the balls around vertices of $\varphi(C)$ have types $(2, 0)$ or $(3, 0)$. Indeed, cutting along such discs preserves the triple circle while cutting along discs of type (k, m) with $m > 0$ destroys it.

Let us prove the last conclusion of the lemma. Suppose that φ_0 and φ_1 are bijective. It means that φ takes the union of singular points of P_F onto the one of P homeomorphically. Then ∂F cannot intersect the balls around true and triple vertices, since otherwise at least one of them would not survive the cut. It follows that ∂F is contained in the union of balls and beams of ξ that correspond to the vertices and edges of P inside a 2-component of P. Each such ball B can be presented as $D^2 \times I$ such that $D^2 \times \partial I$ are the lakes of B. Suppose that B has at least one common point with ∂F. Then any elementary disc $D \subset F \cap B$ has type $(2, 2)$. Therefore D is a quadrilateral having two opposite sides in the islands of B, the other two in the lakes. Similarly, if ∂F passes along a beam $D^2 \times I$, then each strip in $F \cap (D^2 \times I)$ has two opposite sides in the islands, the other two in ∂M.

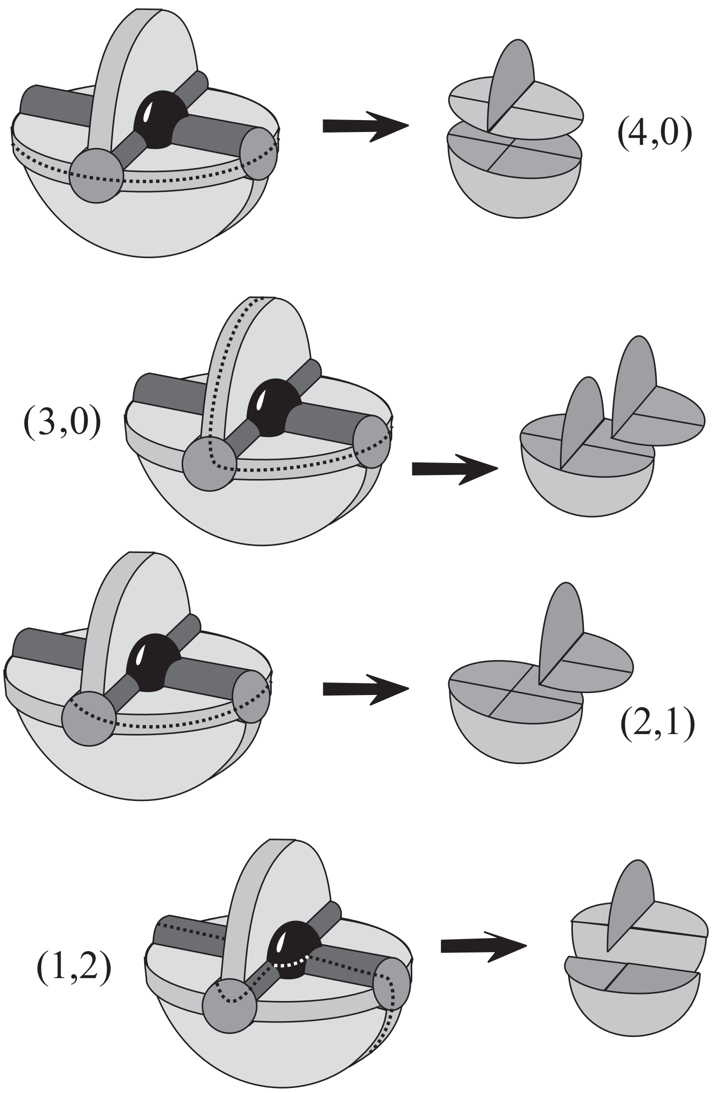

Fig. 4.22. Cutting along discs of all the types except (3,0) destroys the true vertex (see continuation)

Consider the union F_0 of all such quadrilaterals and strips. Since each quadrilateral intersects exactly two strips along its two island sides, F_0 is the disjoint union of annuli and Möbius bands. On the other hand, $\partial F_0 = \partial F$. Taking into account that F is connected, we conclude that $F = F_0$ and F_0 is either an annulus or a Möbius band intersecting P along its middle circle. $\square$

Recall that if P is an almost simple polyhedron, then $c(P)$ denotes the complexity of P, i.e., the number of its true vertices.

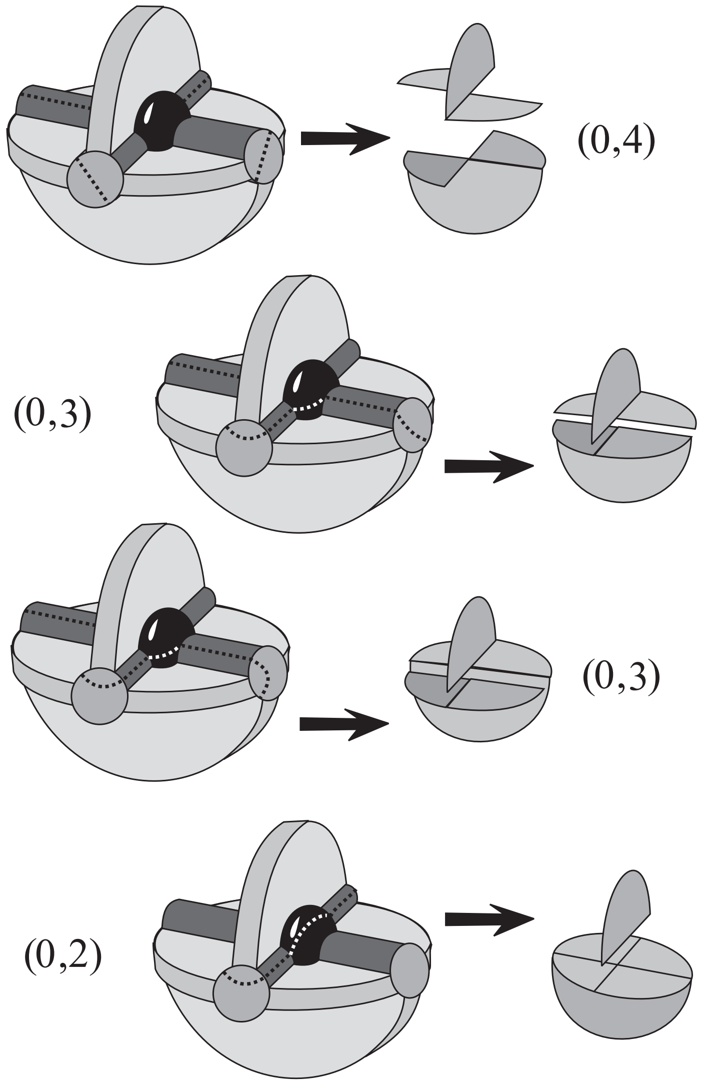

Fig. 4.22. Cutting along discs of the all types except (3,0) destroys the true vertex (continued from the previous page)

Corollary 4.2.5. *Suppose that a handle decomposition ξ of a 3-manifold M with nonempty boundary corresponds to an almost simple cellular spine P of M. Let F be a connected normal surface in M, and let ξ_F, P_F, and φ be, respectively, the induced handle decomposition of M_F, cellular spine of M_F, and cellular map $P_F \to P$. Then $c(P_F) \leq c(P)$ and $c(P_F) = c(P)$ if and only if all elementary discs of F in the balls around true vertices have type* $(3, 0)$.

Proof. Follows from item 1 of Lemma 4.2.4. □

Corollary 4.2.5 tells us that by cutting along any normal surface there appear no new true vertices. However, we need a more complete information. To any almost simple polyhedron we associate a triple (c, c_1, c_2) of non-negative integers, which, in lexicographical ordering, will measure the "extended complexity" of the polyhedron.

Definition 4.2.6. *Let P be an almost simple polyhedron having $c(P)$ true vertices, $c_1(P)$ triple circles, and $c_2(P)$ 2-components. Then the triple $\bar{c}(P) = (c(P), c_1(P), c_2(P))$ is called the* extended complexity *of P.*

Definition 4.2.7. *The* extended complexity $\bar{c}(M) = (c(M), c_1(M), c_2(M))$ *of a compact 3-manifold M is defined as $\bar{c}(M) = min_P \bar{c}(P)$, where the minimum is taken over all almost simple spines of M.*

Thus the extended complexity of M is the triple $(c(M), c_1(M), c_2(M))$, where $c(M)$ is the usual complexity of M as defined in Chap. 2, $c_1(M)$ is the minimum number of triple circles over all almost simple spines of M with $c(M)$ vertices, and $c_2(M)$ is the minimum number of 2-components over all almost simple spines of M having $c(M)$ vertices and $c_1(M)$ triple circles. For example, S^3 has the extended complexity $(0, 0, 0)$ and lens space $L_{3,1}$ has the extended complexity $(0, 1, 1)$. The extended complexity of all I-bundles over closed surfaces is $(0, 0, 1)$.

Corollary 4.2.8. *Suppose that a handle decomposition ξ of a 3-manifold M with nonempty boundary corresponds to an almost simple cellular spine $P = H \cup G$ of M, where H is a simple polyhedron (the 2-dimensional part of P) and G is a graph (the 1-dimensional part of P). Let F be a connected normal surface in M, and let ξ_F, P_F, and φ be, respectively, the induced handle decomposition of M_F, cellular spine of M_F, and cellular map $P_F \to P$. Suppose that $\partial F \neq \emptyset$. Then P_F can be collapsed onto a spine P'_F of M such that $\bar{c}(P'_F) \leq \bar{c}(P)$. Moreover, if F is not a disc, then $\bar{c}(P'_F) < \bar{c}(P)$.*

Proof. Step 1. Assume that P is simple, i.e., that $P = H$. If $c(P_F) < c(P)$ or if $c(P_F) = c(P)$ and $c_1(P_F) < c_1(P)$, we are done. Otherwise we are in the situation of item 3 of Lemma 4.2.4 and hence can conclude that F is an annulus or a Möbius strip such that $F \cap P$ is its middle circle. This means that P_F is obtained from P by cutting along the circle contained in a 2-component α of P. Collapsing P_F, we eliminate α and get a spine $P'_F \subset P_F$ of M_F such that either $c(P'_F) < c(P)$ (if the boundary circles of α pass through at least one true vertex of P), or $c_1(P'_F) < c_1(P)$ (if they are triple circles), or $c_2(P'_F) = c_2(P) - 1$ (if P is a closed surface). In all three cases we get $\bar{c}(P'_F) < \bar{c}(P)$.

Step 2. Assume that P is not simple, i.e., $G \neq \emptyset$. Since F is normal, it does not intersect the handles of ξ that correspond to edges of G. It follows that either:

(a) F lies in a ball of ξ around a vertex of G which is not in H.
(b) F is contained in the union $N(Q)$ of handles that correspond to the cells of H.

Evidently, in Case (a) F is a disc of the required type. Consider Case (b). Recall that H is a simple polyhedron. If F is normal in $N(H)$, then we apply Step 1 and get the desired inequality $\bar{c}(P'_F) < \bar{c}(P)$. Therefore, we may assume that F, being normal in M, is not normal in $N(H)$. The only reason for that phenomenon is violation of condition 5 of Definition 3.4.1. It follows that ∂F is contained in a lake and is nontrivial there. We may conclude that in this case F is a disc in a ball of ξ. □

Remark 4.2.9. Condition $\partial F \neq \emptyset$ in Corollary 4.2.8 is essential. Indeed, consider a normal surface $F \subset M$ which is normally parallel to ∂M. Then P_F consists of a copy of itself and a copy of ∂M. Therefore, $\bar{c}(P'_F) > \bar{c}(P)$.

Lemma 4.2.10. *If D is a proper disc in a 3-manifold M, then $c(M_D) = c(M)$ and $\bar{c}(M_D) = \bar{c}(M)$. In other words, the complexity and extended complexity of a 3-manifold are preserved under removing as well as under attaching a handle of index 1.*

Proof. It is sufficient to prove that $\bar{c}(M_D) = \bar{c}(M)$. First, we note that if P_D is a minimal almost simple spine of M_D, then an almost simple spine of M having the same extended complexity can be obtained from P_D by adding an appropriate arc. It follows that $\bar{c}(M_D) \geq \bar{c}(M)$. Iterating this argument, we can conclude that $\bar{c}(Q) \geq \bar{c}(M_D) \geq \bar{c}(M)$, where Q is the core of M (see Definition 4.1.21).

To prove the inverse inequality $\bar{c}(Q) \leq \bar{c}(M)$ we choose a minimal almost simple spine P of M. Denote by ξ_P the corresponding handle decomposition. If ∂M admits an essential boundary compressing disc, then, normalizing it, we get an essential normal compressing disc D'. Then $\bar{c}(P'_{D'}) \leq \bar{c}(P)$ by Corollary 4.2.8. It follows that $\bar{c}(M_{D'}) \leq \bar{c}(M)$.

Let us perform now boundary compressions along essential normal discs as long as possible. As it is explained in the proof of Proposition 4.1.25, after a finite number of compressions we end up with a core Q' of M. Since Q and Q' are isotopic by Proposition 4.1.25 and since each compression does not increase $\bar{c}$, we can conclude that $\bar{c}(Q) = \bar{c}(Q') \leq \bar{c}(M)$. Combining the inequalities $\bar{c}(Q) \geq \bar{c}(M)$ and $\bar{c}(Q) \leq \bar{c}(M_D) \leq \bar{c}(M)$, we get $\bar{c}(M_D) = \bar{c}(M)$. Therefore any compression along any proper disc preserves both $c(M)$ and $\bar{c}(M)$. □

Corollary 4.2.11. *Let S be a 2-sphere in a 3-manifold M. Then $c(M_S) = c(M)$.*

Proof. Cutting along S can be realized by removing a ball and cutting along a disc whose boundary lies in the boundary of the ball. Both preserve $c(M)$. Or vice-versa: Gluing two boundary spheres together means attaching an index 1 handle followed by attaching a 3-ball. Both preserve $c(M)$. □

Remark 4.2.12. Cutting along a 2-sphere can increase the extended complexity of a 3-manifold. For example, if M is a 3-ball, then $\bar{c}(M) = (0,0,0) < (0,0,1) = \bar{c}(M_S)$

Let ξ be a handle decomposition of a 3-manifold M and $F \subset M$ a proper incompressible surface. Then F can be normalized by the normalization procedure described in Theorem 3.4.7, which consists of tube and tunnel compressions, and eliminating trivial spheres and discs. Let us modify the procedure as follows. Since F is incompressible, each tube compression results in appearance of a 2-sphere. If this sphere is inessential, then we accomplish the compression by throwing it away. Other normalization moves remain the same. Of course, the modified normalization procedure transforms F into actually the same normal surface F' as the unmodified one. The only difference is that we get a fewer number of trivial spherical components. Note also that if F is a closed surface different from a sphere, then a connected component of F' is homeomorphic to F.

Lemma 4.2.13. *Suppose a handle decomposition ξ of a 3-manifold M corresponds to an almost simple cellular spine P of M. Let $F \subset M$ be a connected incompressible surface and let a surface $F' \subset M$ be obtained from F by the modified normalization procedure described above. Then $c(M_F) = c(M_{F'})$ and, if F is not a 2-sphere, $\bar{c}(M_F) < \bar{c}(M_{F'})$.*

Proof. Let us analyze the behavior of $\bar{c}(M_F)$ under the modified normalization moves.

1. Let us show that the modified tube compressions preserve $\bar{c}(M_F)$. Indeed, denote by D the compressing disc of a tube and by D' the disc bounded by ∂D in F. Then the surface F' resulting from the tube compression can be presented as $F' = (F \setminus \text{Int } D') \cup D$. Let $W = (M_F)_D = (M_{F'})_{D'}$ be the manifold obtained from M by cutting along $F \cup D = F' \cup D'$. Then W is obtained from M_F by cutting along D, and simultaneously it is obtained from M'_F by cutting along D', see Fig. 4.23. It follows from Lemma 4.2.10 that $\bar{c}(M_F) = \bar{c}(W) = \bar{c}(M_{F'})$.
2. Let us show that tunnel compressions also preserve $\bar{c}(M_F)$. Denote by D a boundary compressing disc for F, which can be considered as a partition wall inside a tunnel. Let $W = (M_F)_D$ be the 3-manifold obtained from M_F by cutting it along D. Consider the surface F' obtained from F by compressing along D. Then $\bar{c}(M_F) = \bar{c}(W)$ by Lemma 4.2.10, and $\bar{c}(W) = \bar{c}(M_{F'})$, since W and $M_{F'}$ are homeomorphic. It follows that $\bar{c}(M_F) = \bar{c}(M_{F'})$, see Fig. 4.24.
3. Crossing out an inessential disc component of F preserves $\bar{c}(M_F)$ by Lemma 4.2.10.
4. We may conclude that the first three modified normalization moves preserve the extended complexity and hence the complexity of the manifold. Consider the last normalization move (removing an inessential 2-sphere).

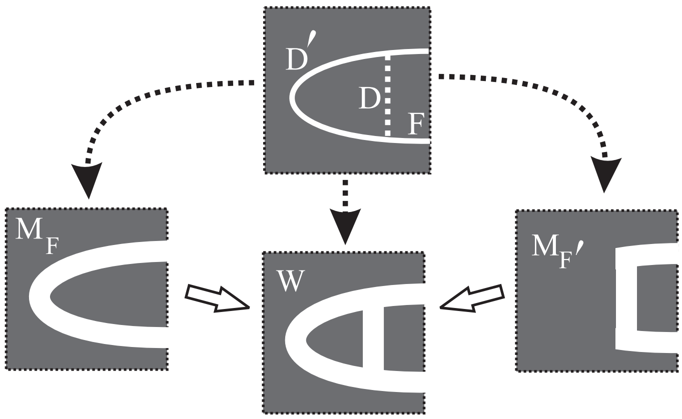

Fig. 4.23. W is obtained from M_F and $M_{F'}$ by cutting along discs

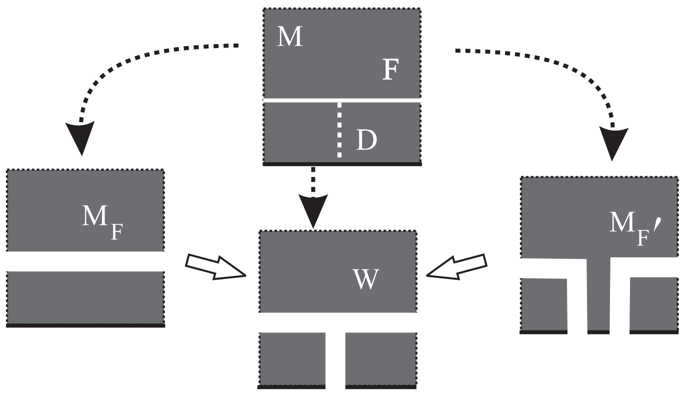

Fig. 4.24. W is homeomorphic to $M_{F'}$ and can be obtained from M_F and by cutting D

This move preserves $c(M_F)$ by Corollary 4.2.11, but can decrease $\bar{c}(M_F)$. Nevertheless, if F is not a sphere, then, thanks to the modification of the first normalization move, spherical components do not appear at all. So we do not need the last move at all. □

Theorem 4.2.14. *Let F be a connected incompressible surface in a 3-manifold M. Then $c(M_F) \le c(M)$. Moreover, if M is closed and contains no projective planes, and F is not a 2-sphere, then $c(M_F) < c(M)$.*

Proof. Consider a handle decomposition of M corresponding to its minimal almost simple spine P and a connected normal surface F' obtained from F by the modified normalization procedure. By Lemma 4.2.13, $c(M_F) = c(M_{F'})$. Therefore, the first conclusion of the theorem follows from Corollary 4.2.5.

Let us prove the second. Let F'_0 be a connected component of F' which is homeomorphic to F. Consider a decomposition $M = \#_{i=1}^n M_i$ of M into the connected sum of prime summands. We can assume that P is obtained from minimal almost simple spines P_i of M_i by joining them by arcs. Then, since F'_0 is normal and connected, there is $j, 1 \leq j \leq n$, such that F'_0 is contained in the union $(M_j)_0$ of handles that correspond to the cells of P_j. It cannot happen that M_j is $S^1 \times S^2$, $S^1 \tilde{\times} S^2$ or $L_{3,1}$, since these manifolds contain no closed incompressible surfaces except S^2. We cannot also have $M_j = RP^3$, since by assumption M contains no projective planes. Since all other closed manifolds have special minimal spines, we can assume that P_j is special. Therefore, the only normal surface which is contained in $(M_j)_0$ and consists only of elementary discs of the type $(3, 0)$, is a 2-sphere. It follows that F'_0 contains at least one elementary disc whose type is not $(3, 0)$. By Corollary 4.2.5, $c(M_{F'_0}) < c(M)$. □

Theorem 4.2.15. *Let F be a connected proper incompressible surface in a 3-manifold M such that $\partial F \neq \emptyset$. Then $\bar{c}(M_F) \leq \bar{c}(M)$. If, in addition, F is a boundary incompressible surface not homeomorphic to a disc, then $\bar{c}(M_F) < \bar{c}(M)$.*

Proof. Consider a handle decomposition of M corresponding to its minimal almost simple spine P and a connected normal surface F' obtained from F by the modified normalization procedure. By Lemma 4.2.13, $\bar{c}(M_F) = \bar{c}(M_{F'})$. Therefore, the first conclusion of the theorem follows from Corollary 4.2.5. To obtain the second conclusion of the theorem, we note that the above normalization procedure preserves the property of a surface to be boundary incompressible and contain a connected component which has nonempty boundary and is not a disc. Therefore, F' contains such a component F'_0. Applying Corollary 4.2.8, we get $\bar{c}(M_{F'_0}) < \bar{c}(M)$. Since $\bar{c}(M_F) = \bar{c}(M_{F'}) =\leq \bar{c}(M_{F'_0})$, we are done. □

4.2.2 Triangulations vs. Handle Decompositions

As the reader might have observed, the triangulation and handle decomposition versions of the normal surface theory are in a sense parallel. To make the observation precise, we note that the triangulation version works without any changes for closed manifolds equipped with singular triangulations. Recall that by Corollary 1.1.27 one-vertex singular triangulations of a closed manifold M correspond bijectively to special spines of M.

Let T be a one-vertex triangulation of a closed 3-manifold M and P the corresponding dual special spine of M. Recall that P has a natural cell decomposition into true vertices, edges, and 2-components. This decomposition induces a handle decomposition ξ_P of M such that ξ_P has only one handle of index 3, and balls, beams, and plates of ξ_P correspond naturally to the true vertices, edges, and 2-components of P. Since P is dual

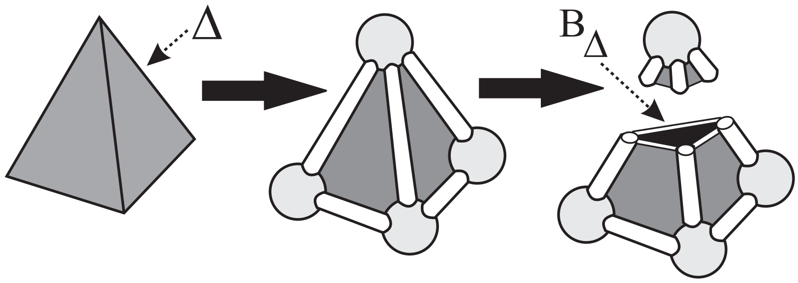

Fig. 4.25. Elementary discs in Δ correspond bijectively to the ones in the ball B_Δ shown *black*

to T, ξ_P is dual to the handle decomposition ξ_T of M obtained by thickening vertices, edges, and triangles of T to balls, beams, and plates, respectively, see Chap. 1. The duality of the handle decompositions means that every index i handle of ξ_T is considered as an index $(3-i)$ handle of ξ_P, $0 \le i \le 3$.

Let T be a one-vertex singular triangulation of a closed 3-manifold M and ξ_P the corresponding handle decomposition of M.

Theorem 4.2.16. *The matching system for* T (*see Sect. 3.3.4*) *coincides with the one for* ξ_P (*see Sect. 3.4*).

Proof. We will think of ξ_P as being obtained from ξ_T by appropriate renumbering of indices of handles. Let Δ be a tetrahedron of T. Denote by B_Δ the ball of ξ_P which is placed inside Δ. See Fig. 4.25, where B_Δ is shown as a black core of Δ. It is evident that any elementary disc for Δ is normally isotopic to a disc that crosses B_Δ along an elementary disc for B_Δ. This gives us a one-to-one correspondence between the variables of the matching systems for T and ξ_P. It means that the matching systems have actually the same variables.

Similarly, each equation of the matching system for T appears by considering an arc l in a triangle having the endpoints in different sides of the triangle. Each such arc l determines a strip in the corresponding beam of ξ_P joining distinct islands, see Fig. 4.26. The strip is responsible for an equation of the matching system for ξ_P. It is easy to see that these equations of the matching systems for ξ_T and ξ_P are the same. This means that the systems are identical. □

Corollary 4.2.17. *Let* T *be a one-vertex singular triangulation of a closed 3-manifold* M *and* ξ_P *the corresponding handle decomposition. Then any surface in* M *normal with respect to* T *is normally isotopic to a surface normal with respect to* ξ_P. *This correspondence determines a bijection between the sets of the normal isotopy classes of normal surfaces in* T *and* ξ_P, *and respects the summation.*

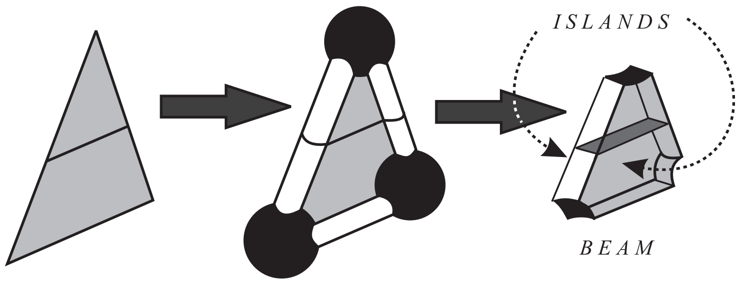

Fig. 4.26. Each arc in a triangle of T determines a strip in the corresponding beam of ξ_P

Proof. Follows directly from Theorem 4.2.16. □

Later on we will use Corollary 4.2.17 to switch from triangulations to handle decompositions and back whenever we find it advantageous. Note that for manifolds with boundary $\neq \emptyset, S^2$ the matching systems are quite different.