5

Theory of handle decompositions

A handle decomposition is perhaps the simplest way to build a manifold from elementary pieces. The existence of such decompositions is obtained by analysing the geometry associated to a non-degenerate function on the manifold.

In the first section we prove the existence of handle decompositions for compact manifolds: in the next few sections we will show how to operate on such decompositions. In §5.2 we normalise the decomposition; then, after a section on the homology of handles, we manipulate the decompositions: there are results on adding handles, and on removing or introducing complementary pairs of handles. The technical details use the results treated in Chapter 2.

The definition of a handle decomposition is analogous to that of a CW complex. Also the results we establish run in parallel with operations on finite CW complexes that can be performed in homotopy theory. We will see below that up to a point the theory of handle presentations parallels that of cell decompositions and even to an important extent to that of algebraic operations on chain complexes.

The high point of this development is the h-cobordism theorem, which gives an effective criterion for diffeomorphism of compact manifolds. We prove this result in §5.5. Then we give a number of applications, discuss what is known in low dimensions, and outline what modifications need to be made to the theory when the fundamental group is non-trivial. In some places we anticipate Theorem 6.4.11, but Chapter 6 is independent of this chapter.

In this chapter, all manifolds will be compact unless otherwise stated.

5.1 Existence

Let W be a manifold, and suppose $\partial_- W$ and $\partial_+ W$ disjoint manifolds with union ∂W. Then we call the pair $(W, \partial_- W)$ a *cobordism* and the pair $(W, \partial_+ W)$ the

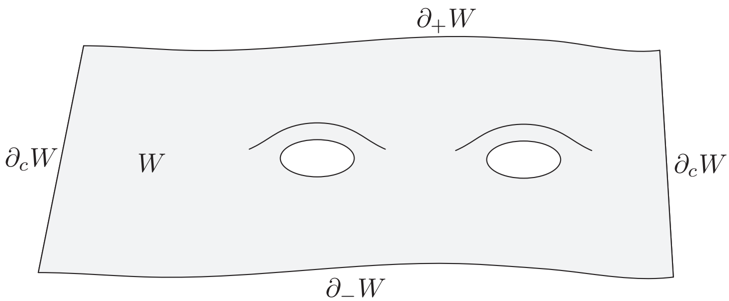

Figure 5.1 A cobordism

dual cobordism; we also call W a cobordism of $\partial_- W$ to $\partial_+ W$, and say that $\partial_- W$ and $\partial_+ W$ are *cobordant*. If W is a manifold with corner, and $\partial_- W$, $\partial_c W$, $\partial_+ W$ are parts of the boundary such that $\partial_- W$ and $\partial_+ W$ are disjoint and

$$\partial \partial_c W = \angle W = \partial(\partial_- W \cup \partial_+ W),$$

we still call W a cobordism of $\partial_- W$ to $\partial_+ W$. We shall usually denote a cobordism by a single letter and often just call it a manifold. A picture of a cobordism is offered in Figure 5.1. For example, we usually regard a product $M \times I$ as a cobordism, with $\partial_-(M \times I) = M \times 0$, $\partial_+(M \times I) = M \times 1$; if M has boundary, write $\partial_c(M \times I) = \partial M \times I$.

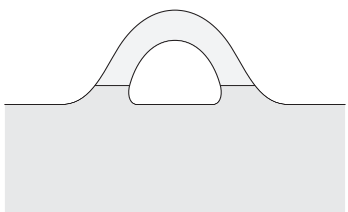

Figure 5.2 A handle

Suppose W^n a cobordism, $f : S^{r-1} \times D^{n-r} \to \partial_+ W$ an embedding. Introduce a corner (as in Lemma 2.6.3) along $f(S^{r-1} \times S^{n-r-1})$. Now glue $D^r \times D^{n-r}$ to W by f. The result is unique up to diffeomorphism, and is denoted $W \cup_f h^r$; it has the same corners as W. We describe it as W *with an r-handle attached by* f. We call f the *attaching map* of the handle, and r the *dimension* of the handle. Figure 5.2 offers a picture of a handle. We define $\partial_+(W \cup h^r) = (\partial_+ W \setminus \operatorname{Im} f) \cup (D^r \times S^{n-r-1})$.

If we have a sequence of attached handles:

$$W^* = W \cup_{f_1} h^{r_1} \cup \cdots \cup_{f_k} h^{r_k},$$

we describe this as a *handle presentation* of W^* on W; if the maps f_i are not specified, as a *handle decomposition*. In particular, if $W = M \times I$, we speak of a handle decomposition of W^* with *base* M (here, M may be empty).

To prove existence, we use non-degenerate functions.

Lemma 5.1.1 *Any cobordism W with $\partial_c W = \emptyset$ admits a non-degenerate function f, with all critical values distinct, attaining an absolute minimum on $\partial_- W$ only, and an absolute maximum on $\partial_+ W$ only. The same holds if $\partial_c W$ is a product $M \times I$.*

Proof Let $\partial_- W \times I$, $\partial_+ W \times I$ be disjoint collar neighbourhoods of $\partial_- W$ and $\partial_+ W$. Define $g : W \to [0, 1]$ by:

$$g(x, t) = \begin{cases} \frac{1}{3}t & \text{for } x \in \partial_- W, \\ 1 - \frac{1}{3}t & \text{for } x \in \partial_+ W, \end{cases} \tag{5.1.2}$$

and extend to a continuous function taking only values between $\frac{1}{3}$ and $\frac{2}{3}$ elsewhere: this is possible since W is normal. By Proposition 1.1.7, we can approximate g by a smooth function h, agreeing with g near ∂W. Now approximate h by a non-degenerate function f with distinct critical values, agreeing with h, and so g, near ∂W. This is possible by Proposition 4.5.10 since g and h have no critical points in a neighbourhood of ∂W.

For the case $\partial_c W = M \times I$ we use the same argument: here we use Proposition 1.5.6 to find the collars, and to ensure that along $\angle W$ they agree with the product structure on $\partial_c W$, and Proposition 1.1.7 allows us to suppose that h and hence f agree on $M \times I$ by the projection on I; the proof now concludes as above. □

The proof shows that we may suppose that for x close to ∂W, f is defined by the formula (5.1.2).

Now we give W a Riemannian structure adapted to the boundary; for convenience we suppose, as in Proposition 2.3.7 that it is a product metric in some neighbourhood of ∂W. Then the differential 1-form df induces at each $P \in W$ an element df_P of $T_P W^\vee$; using the Riemannian structure, this is identified with an element of $T_P W$, a tangent vector. Thus df gives a vector field, which we call ∇f.

In $\mathring{W}$, we can use Theorem 1.4.2 to integrate f and obtain a flow $\varphi_t(P)$, defined for certain values of (t, P). Near a point of $\partial_- W$, we can take coordinates $x_1, \ldots, x_n$ such that W is defined by $x_1 \geq 0$, x_1 is the t-coordinate in the tubular neighbourhood, so that $f(x) = x_1 - 1$ and the Riemannian structure is of the form $ds^2 = dx_1^2 + \sum_{i,j=2}^{n} g_{i,j} dx_i dx_j$. Hence ∇f agrees with $\partial/\partial x_1$ in such a neighbourhood, and orbits are of the form

$$\varphi_t(x_1, \ldots, x_n) = (x_1 + t, x_2, \ldots, x_n) \qquad x_1 \geq 0, x_1 + t \geq 0.$$

Thus for $P \in \partial_- W$, $\varphi_t(P)$ is defined for small positive values of t, and φ is defined on a neighbourhood of $\partial_- W \times 0$ in $\partial_- W \times \mathbb{R}^+$ and gives a collar neighbourhood of $\partial_- W$. Similarly for $\partial_+ W$.

If we regard $\varphi_t(P)$ as a point parametrised by t, it is smooth, and we have a metric, so can speak of speed.

Lemma 5.1.3 *We have (a)* $\frac{d}{dt} f(\varphi_t(P))|_{t=0} = \|df_P\|^2$,
(b) The speed of $\varphi_t(P)$ at $t = 0$ *is* $\|df_P\|$.

Proof (a)
$$\begin{aligned} \frac{d}{dt} f(\varphi_t(P))\bigg|_{t=0} &= \nabla f(f)|_P \quad \text{by definition of} \quad \varphi \\ &= df(\nabla f)|_P \\ &= \langle df_P, df_P \rangle = \|df_P\|^2 \end{aligned}$$

in the Riemannian inner product on $T_P W$, since this defined ∇f.

(b) Take coordinates $(x_1, \ldots, x_m)$ at P, so that P has coordinates $(0, \ldots, 0)$ and at P the Riemannian metric agrees with the standard metric in $\mathbb{R}^n$. Let $df = \sum a_i dx_i$: then $\nabla f = \sum a_i \partial/\partial x_i$ (at P). Thus, at P, $\frac{\partial \varphi_t(P)}{\partial x_i} = a_i$, so the speed of $\varphi_t(P)$ is just $(\sum a_i^2)^{1/2} = \|df_P\|$. □

Now suppose $P \in \mathring{W}$, and that the maximum range of t in which $\varphi_t(P)$ is defined is (a, b).

Lemma 5.1.4 *Suppose W is compact. Then either*
a is finite and as $t \to a$, $\varphi_t(P)$ tends to a point on $\partial_- W$, or
$a = -\infty$ and, for any K, the closure of $\varphi_t^{-1}(-\infty, -K)$ contains a critical point of f.
Similarly for b.

Proof If a is finite, by Lemma 5.1.3 (b), the points $\varphi_t(P)$ form a Cauchy sequence as $t \to a$ (since W is compact, $\|df_P\|$ is bounded); since W is complete, they tend to a limit point Q. If Q was interior to W, it would follow that Q was on the orbit, which could then be extended: thus Q is on ∂W. Since by Lemma 5.1.3 (a), f increases along each orbit, $f(Q) < f(P)$, so Q is on $\partial_- W$.

Now let $a = -\infty$. Then by Lemma 5.1.3 (a),

$$\int_{-\infty}^{0} \|df_{\varphi_t(P)}\|^2 dt$$

converges. So $\|df_{\varphi_t(P)}\|$ has infimum zero as $t \to -\infty$. Outside any open neighbourhood of the set of critical points, $\|df\|$ is non-zero, and attains its lower bound (by compactness), so $\varphi_t(P)$ meets any such neighbourhood. But the set of critical points is compact, and so meets the closure of the orbit. □

We are now ready to analyse the function f of Lemma 5.1.1. For $a \in \mathbb{R}$, write

$$W_a = \{P \in W : f(P) \leq a\}$$
$$M_a = \{P \in W : f(P) = a\}$$

thus for

$a = 0$	$W_a = \partial_- W$	$M_a = \partial_- W$
$a = \varepsilon$	$W_a = \partial_- W \times [0, \varepsilon]$	$M_a = \partial_- W \times \varepsilon$
$a = 1 - \varepsilon$	$W_a = W \setminus (\partial_+ W \times [0, \varepsilon))$	$M_a = \partial_+ W \times \varepsilon$
$a = 1$	$W_a = W$	$M_a = \partial_+ W$

provided that ε is so small that $\partial_\eta W \times [0, \varepsilon]$ ($\eta = +, -$) are contained in the collar neighbourhoods described earlier. Clearly, for $a < b$, $W_a \subset W_b$; we next investigate how W_b is formed from W_a.

Lemma 5.1.5 *Suppose that for $a \leq c \leq b$, c is not a critical value of f. Then $f^{-1}[a, b]$ is diffeomorphic to $M_a \times [a, b]$ and W_b is diffeomorphic to W_a.*

Observe that since a, b are not critical values, it follows from Lemma 4.5.1 that M_a, M_b and $f^{-1}[a, b]$ are submanifolds.

Proof The first assertion follows at once by applying Theorem 1.5.4 to the vector field ∇f.

Thus W_b is obtained from W_a by glueing on $M_a \times I$ along M_a. The result now follows from Lemma 2.7.2. □

This shows that 'as long as a does not pass through a critical value, the diffeomorphism type of W_a remains constant'. We now have to investigate the critical value.

Theorem 5.1.6 *Suppose that for $a \leq f(P) \leq b$ there is just one critical point Q, which is non-degenerate and with $f(Q) = c$ ($a < c < b$). Then W_b is diffeomorphic to W_a with a handle attached.*

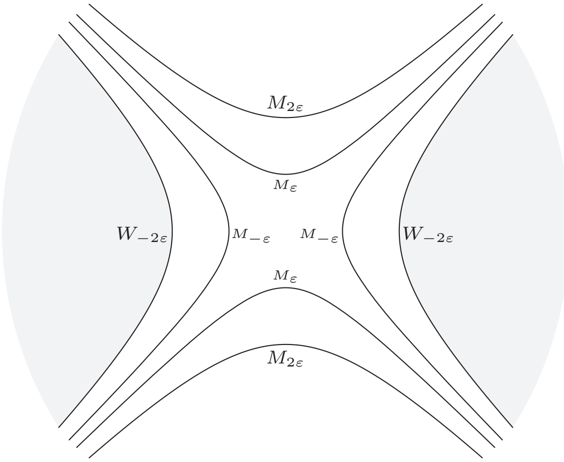

Figure 5.3 Level sets

Proof Our discussion of orbits in Theorem 1.5.4 remains valid except for those orbits with Q as a limit point. We must therefore investigate a neighbourhood of Q. By the Morse Lemma 4.8.2 there exist local coordinates $x_1, \ldots, x_n$ such that in a neighbourhood of Q f is given by

$$f(x) = c - x_1^2 - \cdots - x_\lambda^2 + x_{\lambda+1}^2 + \cdots + x_n^2.$$

The integer λ is called the *index* of f of the critical point 0. Using a partition of unity, we choose a Riemannian structure which agrees with the Euclidean structure in this coordinate system. With respect to this, the gradient vector field ∇f is given by

$$\nabla f = \sum_1^\lambda -x_i \frac{\partial}{\partial x_i} + \sum_{\lambda+1}^n x_i \frac{\partial}{\partial x_i}.$$

For example, if $f(x) = -x_1^2 + x_2^2$, the curves M_a are hyperbolae with asymptotes $x_1^2 = x_2^2$, except for M_0 which is this line-pair, and as a increases up to zero, W_a increases without essential change, but it engulfs the origin when $a = 0$. Figure 5.3 shows the evolution of M_a for $a = -2\epsilon,\ -\epsilon,\ \epsilon$, and 2ϵ.

Choose ε so small so that for $\|x\| \leq 5\varepsilon$, the above formulae are valid. We now modify $W_{-\varepsilon}$ by first introducing a corner, then attaching a handle, to obtain something close to W_ε: the procedure is illustrated in Figure 5.4.

More precisely, write $x = (\xi, \eta)$, where $\xi = (x_1, \ldots, x_\lambda)$, $\eta = (x_{\lambda+1}, \ldots, x_n)$, $f(x) = c - \|\xi\|^2 + \|\eta\|^2$. Define the handle H to be the set $\|\eta\| \leq \varepsilon$, $\|\xi\| \leq \varepsilon$. Let V be a smooth manifold with corner which

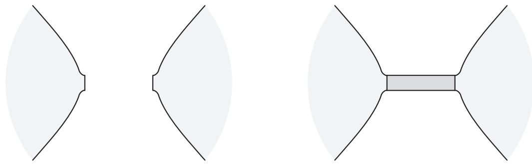

Figure 5.4 Attaching a handle

(i) coincides with $\|\xi\| \leq \varepsilon$, $\|\eta\| \geq \varepsilon$ near $\|\xi\| = \varepsilon$ (this includes the corner $\|\xi\| = \|\eta\| = \varepsilon$);

(ii) coincides with $W_{-\varepsilon}$ when $\|x\| \geq 5\varepsilon$, and contains $W_{-\varepsilon}$;

(iii) has ∂V everywhere transverse to the orbits of ∇f.

Such a V may be constructed using a bump function. Then by Proposition 2.6.4, $M_{-\varepsilon}$ is obtained from V by straightening the corner – or equivalently, by Lemma 2.6.3, V from $M_{-\varepsilon}$ by introducing one. Now H is diffeomorphic to the product $D^\lambda \times D^{n-\lambda}$, and $\partial_- H := H \cap V$ is given by $\|\xi\| = \varepsilon$, $\|\eta\| \leq \varepsilon$, hence is a copy of $S^{\lambda-1} \times D^{n-\lambda}$ in $\partial H \cap \partial V$. Since the union $H \cup V$ is smooth, and H and V are defined by cutting it along $H \cap V$, it follows by Proposition 2.7.3 that $H \cup V$ is obtained by glueing these.

Now $H \cup V$ is a smooth manifold, transverse to the orbits, with no critical points between it and M_b; thus it follows from Theorem 1.5.4 that W_b is diffeomorphic to $H \cup V$. But $H \cup V$ consists of W_a with a λ-handle attached. □

The following are immediate consequences.

Corollary 5.1.7 *(i) If the Hessian of f at c has index λ, we attach a λ-handle.*

(ii) If there are several non-degenerate critical points at level c, we attach several handles.

(iii) W has a handle decomposition based on $\partial_- W$,

for we can apply the above argument in a neighbourhood of each critical point.

With a little care, the arguments may also be applied to non-compact manifolds: we give one sample result.

Lemma 5.1.8 *Any manifold W can be expressed as the union of an infinite sequence of handles attached one at a time.*

Proof By Corollary 2.2.10 there is a proper map $f : W \to \mathbb{R}$; as before, we may suppose f smooth and non-degenerate. Then each set W_a is compact, and we may apply the same arguments as above. □

It is also possible to proceed in the opposite direction.

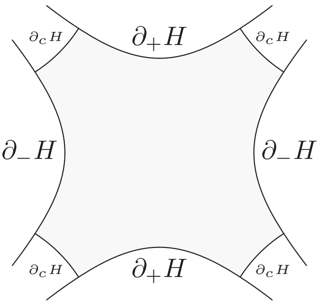

Figure 5.5 Alternative picture of a handle

Theorem 5.1.9 *Given a handle decomposition of W on $\partial_- W$, there is a non-degenerate function f on W (as in Lemma 5.1.1) with just one critical point of index λ for each λ-handle.*

Proof The result is proved by induction on the number of handles: if there are none, $W \cong \partial_- W \times I$, and we take f as the projection on I. Now let V be defined by attaching all but the last handle: by the induction hypothesis, f can be defined on V, constant on $\partial_+ V$. So if we can define f on $(\partial_+ V \times I) \cup h^\lambda$ we can glue back (using collar neighbourhoods of $\partial_+ V$ on which f reduces to a projection) to make f smooth. Hence it suffices to consider the case when W is formed from $\partial_- W \times I$ by attaching one handle.

Now let $g : S^{\lambda-1} \times D^{n-\lambda} \to \partial_- W$ be the attaching map of a λ-handle. Write K for the closure of the complement of the image. Write H for the subset of $\mathbb{R}^\lambda \times \mathbb{R}^{n-\lambda}$ defined by

$$-1 \leq -\|x\|^2 + \|y\|^2 \leq 1, \quad \|x\|^2 \|y\|^2 \leq 2.$$

Then the function defined on H by $F(x, y) = -\|x\|^2 + \|y\|^2$ attains its minimum value -1 on $\partial_- H$, say, and its maximum $+1$ on $\partial_+ H$. Write $\partial_c H$ for the subset given by $\|x\|^2 \|y\|^2 = 2$.

We have a diffeomorphism $G_- : \partial_- H \to S^{\lambda-1} \times D^{n-\lambda}$ given by $G_-(x, y) = \left(\frac{x}{\|x\|}, y\right)$: its inverse is given by $G_-^{-1}(u, v) = \left((1 + \|v\|^2)^{1/2} u, v\right)$. We also have a diffeomorphism $G_c : \partial_c H \to S^{\lambda-1} \times S^{n-\lambda-1} \times [-1, 1]$ given by $G_c(x, y) = \left(\frac{x}{\|x\|}, \frac{y}{\|y\|}, F(x, y)\right)$: its inverse is $G_c^{-1}(u, v, t) = (au, bv)$, where $b^2 - a^2 = t$ and $b^2 + a^2 = \sqrt{t^2 + 8}$. This description goes with the picture of a handle offered in Figure 5.5.

Now attach H to $K \times [-1, 1]$ by the map G_c to form a manifold W'. The function $f : W' \to [-1, 1]$ defined by $f|H = F$, $f|K =$ projection is a smooth function, whose only critical point is the non-degenerate one in H. We have a diffeomorphism h of $\partial_- W'$ to $\partial_- W$ given by the identity on $K \times -1$, and by $g \circ G_-$ on $\partial_- H$.

Finally, we have a diffeomorphism of W' on W. For each is obtained by attaching a λ-handle to the lower boundary: W by hypothesis and W' by Theorem 5.1.6. By construction, the attaching maps of the handles correspond under h, so the identity map of $\partial_- W$ extends to a diffeomorphism. $\square$

5.2 Normalisation

We could proceed immediately to make various deductions about smooth manifolds from the existence of a handle decomposition. First, however, it is convenient to normalise a presentation. We have defined $W \cup_f h^r$ by attaching $D^r \times D^{n-r}$ to W using an embedding $f : S^{r-1} \times D^{n-r} \to M := \partial_+ W$; however it will usually be more convenient to regard the handle H as consisting of a collar $M \times I$ to which $D^r \times D^{n-r}$ is attached. The *attaching sphere* (or *a*-sphere) of H is the sphere $f(S^{r-1} \times 0)$ in $\partial_- H$. The *belt sphere* (or *b*-sphere) is the sphere $0 \times S^{n-r-1}$ in $\partial_+ H$. The *core* is the disc $D^r \times 0$.

It follows at once from Theorem 2.4.2 (diffeotopy extension) that $W \cup_f h^r$ is determined up to diffeomorphism by the diffeotopy class of f, for if g is a diffeomorphism of W, g induces a diffeomorphism of $W \cup_f h^r$ with $W \cup_{gf} h^r$. By Theorem 2.5.5 (tubular neighbourhood), it is even determined by the diffeotopy class of $\bar{f} = f | S^{r-1} \times 0$ together with a homotopy class of normal framing of $f(S^{r-1} \times 0)$ in $\partial_+ W$.

Lemma 5.2.1 *Let $r \le s$. Then $(W \cup_f h^s) \cup_g h^r$ is diffeomorphic to manifolds obtained from W by attaching the handles simultaneously, or in the reverse order.*

Proof Let $n = \dim W$, $Q = \partial_+(W \cup_f h^s)$. Then we have in Q the *a*-sphere S^{r-1} of h^r and the *b*-sphere S^{n-s-1} of h^s. Since

$$(r-1) + (n-s-1) = n - 1 - (s+1-r) < n - 1 = \dim Q,$$

by Theorem 4.5.6, S^{r-1} may be approximated by a sphere not meeting S^{n-s-1}: by Proposition 4.4.4 if the approximation is close enough, we still have an imbedded sphere, diffeomorphic to the old one. By further diffeotopies, we may make S^{r-1} avoid the tubular neighbourhood $D^s \times S^{n-s-1}$ (using the diffeotopy extension theorem, and the fact that the tubular neighbourhood may be shrunk to avoid S^{r-1}) and shrink the tubular neighbourhood $S^{r-1} \times D^{n-r}$ so that, this, too, avoids $D^s \times S^{n-s-1}$. But now the attaching map of the r-handle is disjoint from the s-handle: its image lies in ∂W, and the handles may be added in either order. $\square$

Corollary 5.2.2 *Any W has a handle decomposition on $\partial_- W$ with the handles arranged in increasing order of dimension.*

This follows at once by induction.

From now on we shall generally assume that handles have been arranged in order of increasing dimension: this is in some sense the usual case. Indeed, since we can always reduce to this case, a handle decomposition without this property carries extra information.

We now introduce the notation

$$W_{r+\frac{1}{2}} = (\partial_- W \times I) \cup \text{ all } s - \text{handles for } s \leq r,$$

where we use Lemma 5.2.1 and attach all r-handles simultaneously. Also set $M_{r+\frac{1}{2}} = \partial_+ W_{r+\frac{1}{2}}$.

This is related to our previous notation as follows. It follows from Lemma 5.2.1, in conjunction with the relation between handles and non-degenerate functions on W, that there exist non-degenerate functions f with the property that, for each critical point P of f of index λ, we have $f(P) = \lambda$. Such functions are called *self-indexing*. If we use a self-indexing function f on W, the two definitions of $W_{r+\frac{1}{2}}$ coincide.

In $M_{r+\frac{1}{2}}$ we have the a-spheres S^r of the $(r+1)$-handles and the b-spheres S^{n-r-1} of the r-handles, which have complementary dimensions. By Theorem 4.5.6, the embedding of a sphere S^r may be approximated by a map transverse to S^{n-r-1}, and if the approximation is close enough, we have merely altered the embedding by a diffeotopy. Since the dimensions are complementary, and the map transverse, intersections are isolated points; since S^r is compact, there are only finitely many. We can thus modify the presentation by a diffeotopy so that all these a-spheres are transverse to all these b-spheres. We will say that a presentation with this property is in normal position. We have shown

Lemma 5.2.3 *Any handle presentation of $(W, \partial_- W)$ may be modified by diffeotopies so that the handles are arranged in increasing order of dimension, and any two handles of consecutive dimensions are in normal position.*

In this situation, for each such transverse intersection P of S^r with S^{n-r-1}, it follows from Lemma 4.8.1 that there is a chart for $M_{r+\frac{1}{2}}$ meeting the a-sphere in $D^r \times \{0\}$ and the b-sphere in $\{0\} \times D^{n-r-1}$. Since the tubular neighbourhoods are unique up to diffeotopy, we may suppose that they both meet this chart in $D^r \times D^{n-r-1}$, with the projections being those on the factors.

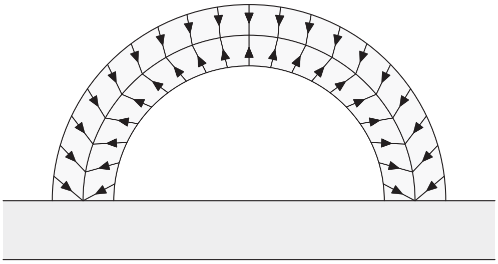

Figure 5.6 Retracting a handle on its core

5.3 Homology of handles and manifolds

For each r−handle attached to W, using a deformation retraction of $D^r \times D^{n-r}$ on $(S^{r-1} \times D^{n-r}) \cup (D^r \times \{0\})$ (which may be obtained from a deformation retraction of $I \times I$ on $(\{1\} \times I) \cup (I \times \{0\})$ by rotating about both axes), we have a deformation retraction of $W \cup_f h^r = W \cup_f (D^r \times D^{n-r})$ on $W \cup_f (D^r \times \{0\})$. Thus, up to homotopy, attaching a handle is the same as attaching a cell (its core). The deformation retraction is pictured in Figure 5.6.

This gives a very close connection between handle decompositions and cell complexes. In particular, we deduce the following from Corollary 5.2.2.

Proposition 5.3.1 *If W is closed, it has the homotopy type of a finite CW complex. In general, $(W, \partial_- W)$ has the homotopy type of a finite CW pair.*

Proof The first statement follows by taking a normalised handle decomposition of W and replacing each handle by an equivalent cell. In fact it is not difficult to show that W is homeomorphic to an appropriate finite CW complex.

For the second statement, note that by the first, we can regard $\partial_- W$ as a finite cell complex, and again apply Corollary 5.2.2. □

Before continuing, it is convenient to recall some basic results about the homology of manifolds: we focus on the simplest case when M^m is closed, connected, and oriented. Then $H_m(M; \mathbb{Z})$ is infinite cyclic (this is a special case of the Poincaré duality Theorem 5.3.5 below). If M is triangulated and we use simplicial homology, a generator is represented by the sum of the m-simplices, where each must be given the orientation induced by that of M. We denote this generator by $[M]$.

A map $f : M \to N$, where M and N are both closed, oriented, and connected has degree d, where d is the integer such that $f_*[M] = d[N]$. The degree may be determined as follows. Suppose f smooth and transverse to a point $Q \in N$. Then $f^{-1}(Q)$ is a finite set of points P_i and (by Lemma 4.5.1) for each i, the tangent map induces an isomorphism $T_{P_i}M \to T_Q N$. Set $\varepsilon_i = \pm 1$ according as this map preserves the given orientations or not. Then $d = \sum_i \varepsilon_i$.

We can see directly that this is independent of choices: a homotopy F of f_0 to f_1 may be made transverse to Q, and the preimage of Q is then a collection of loops (which do not contribute), arcs from $M \times 0$ to $M \times 1$ (whose two end points make the same contribution for f_0 and f_1) and arcs with both ends on $M \times 0$ (or on $M \times 1$) (whose two end points contribute opposite signs ε, so cancel each other).

To see that $d = \sum_i \varepsilon_i$, choose a disc neighbourhood D of Q such that its preimage consists of discs D_i round the P_i each mapped by a diffeomorphism. Inclusion induces isomorphisms $H_m(D, \partial D) \to H_m(N, N \setminus \mathring{D} \to H_m(N)$. In M there are similar isomorphisms, but now $H_m(M, M \setminus f^{-1}(\mathring{D}))$ is the direct sum of the $H_m(M, M \setminus \mathring{D}_i))$, and the result follows by adding up the local contributions.

The same considerations apply to intersection numbers: again we describe only the simplest case when we have compact oriented submanifolds V_1, V_2 of the oriented manifold M, of dimensions v_1 and v_2 with $m = v_1 + v_2$. At a point P where V_1 and V_2 intersect transversely we define a local intersection number $\varepsilon(P)$ to be ± 1 according as a base tor $T_P V_1$ giving the chosen orientation of V_1, followed by a corresponding base for $T_P V_2$ defines the given orientation of M or not. If V_1 and V_2 meet transversely everywhere, $\sum_{P \in V_1 \cap V_2} \epsilon(P)$ gives the intersection number $V_1.V_2$. Again, arguing by making a homotopy transverse, we see that this depends at most on the diffeotopy classes of V_1 and V_2.

Each compact oriented submanifold V^v of M^m defines a homology class $i_*[V] \in H_v(M : \mathbb{Z})$ and hence by duality 5.3.5 a cohomology class in $H^{m-v}(M; \mathbb{Z})$, which we temporarily denote by $\{V\}$. In the situation of the preceding paragraph, the cup product $\{V_1\}\{V_2\}$ is equal to $V_1.V_2$ times the class $\{P\}$ of a point. More generally we can see that if V_1 and V_2 (still closed and oriented, with $v_1 + v_2 \geq m$) intersect transversely along a submanifold W, then $\{V_1\}\{V_2\} = \{W\}$. This principle extends in a natural way (subject to appropriate technical conditions) when we allow boundaries and cease to require orientations.

It follows from the remark preceding Proposition 5.3.1 that, up to homotopy, we may replace handles by cells, and may calculate homology using the chain groups

$$C_r(W, \partial_- W) = \oplus \, \mathbb{Z},$$

where the summands are indexed by the r-handles. We will denote by α_r the number of r-handles, equal to the rank of $C_r(W, \partial_- W)$. We need to calculate the boundary homomorphism

$$\partial : C_{r+1}(W, \partial_- W) \to C_r(W, \partial_- W).$$

This is determined by incidence numbers, one for each r- and each $(r+1)$-handle.

Lemma 5.3.2 *The incidence number of handles h^{r+1} and h^r equals the intersection number in $M_{r+\frac{1}{2}}$ of the a-sphere of h^{r+1} and the b-sphere of h^r.*

Proof We need some care with signs: a choice of orientation of the cell $(D^{r+1} \times 0)$ in the cell complex induces orientations of the bounding a-sphere S^r and of the *normal bundle* of the corresponding b-sphere. If an a-sphere S^r and a b-sphere S^{n-r-1} meet transversely at a point, we take the sign $+$ or $-$ according as the orientation of S^r does or does not agree with that in the normal bundle of S^{n-r-1}. If W (and hence M) is oriented, orienting the normal bundle of a b-sphere is equivalent to orienting the sphere, and we can count multiplicities in the usual way.

We may suppose that S^r meets S^{n-r-1} transversely: then the intersection number agrees with the (local) degree of the projection of S^r on the normal disc D^r. But this degree coincides with the incidence number in the cell complex. □

If F is a field of coefficients (for example, $\mathbb{Q}$ or $\mathbb{Z}_2$), we define the Betti numbers β_i (strictly, $\beta_i(W, \partial_- W; F)$) as the ranks of the F-vector spaces $H_i(W, \partial_- W; F)$. Since these may be calculated from the chain groups

$$C_i(W, \partial_- W; F) := C_i(W, \partial_- W) \otimes F,$$

which have ranks α_i, we have

Lemma 5.3.3 (Morse inequalities) *We have*

$$\sum_0^n (-1)^i \alpha_i = \sum_0^n (-1)^i \beta_i$$

and, for each $0 \leq j \leq n$,

$$\sum_0^j (-1)^{j-i} \alpha_i \geq \sum_0^j (-1)^{j-i} \beta_i.$$

Proof Write r_i for the rank of the boundary map

$$C_i(W, \partial_- W; F) \to C_{i-1}(W, \partial_- W; F).$$

The definition of homology gives $\alpha_i = r_{i+1} + \beta_i + r_i$. Hence

$$\sum_0^j (-1)^{j-i}\alpha_i = r_{j+1} + \sum_0^j (-1)^{j-i}\beta_i. \qquad \square$$

We now discuss duality. Observe that with f, $-f$ is also non-degenerate. Its critical points coincide with those of f, but if f has index λ at 0, it has locally the form

$$f(x) = c - x_1^2 - \cdots - x_\lambda^2 + x_{\lambda+1}^2 + \cdots + x_n^2$$

and $-f$ has index $n - \lambda$. Using the correspondence (Theorems 5.1.6 and 5.1.9) between non-degenerate functions and handle decompositions, we find the following.

Proposition 5.3.4 *Suppose W has a handle decomposition on $\partial_- W$ with α_r r-handles for $0 \le r \le n$. Then it also has one on $\partial_+ W$, with α_r $(n-r)$-handles.*

If we ignore corners, we may identify the handles in the two cases, and observe that in the reversal, a- and b-spheres are interchanged.

Theorem 5.3.5 (Lefschetz Duality Theorem)
Suppose either that W is orientable or we use $\mathbb{Z}_2$ for coefficients: then we have isomorphisms $H_r(W, \partial_- W) \cong H^{n-r}(W, \partial_+ W)$.
In particular, $H_r(W) \cong H^{n-r}(W, \partial W)$ and $H^r(W) \cong H_{n-r}(W, \partial W)$.
If $\partial W = \varnothing$, then (Poincaré Duality) $H_r(W) \cong H^{n-r}(W)$.

Proof By Proposition 5.3.4 we can identify the chain groups of $(W, \partial_- W)$ with the chain or cochain groups of $(W, \partial_+ W)$. By Lemma 5.3.2 the incidence numbers are the same up to sign (only a-spheres and b-spheres are interchanged) and the isomorphism identifies the one boundary with the other coboundary. $\square$

The proof above is reminiscent of the earliest proofs of the result (see, for example, the account in [84]), but of course is only valid for compact smooth manifolds.

As a special case of homology groups, we mention connectivity. We retain the notation of Lemma 5.3.2. The a-sphere S^{-1} of a 0-handle is the empty set; in fact a 0-handle consists precisely of an n-disc, disjoint from $\partial_- W \times I$. The a-sphere S^0 of a 1-handle is a pair of points: these may or may not be in the same component of $W_{1/2}$. If not, the 1-handle connects the two components; but if they are, the corresponding handle does not affect connectivity.

If $\partial_- W$ is non-orientable then so, of course, is W. If, however, $\partial_- W$ is orientable, so is $W_{\frac{1}{2}}$, since adding a disjoint set of discs has no effect. Nor does adding a set of 1-handles which connect different components of $W_{\frac{1}{2}}$ (here we think of 1-handles as being added in turn, not simultaneously). However, the attaching map for a 1-handle is a map of $S^0 \times D^{n-1}$ – i.e. of a pair of discs. If these are mapped into the same component of $W_{\frac{1}{2}}$ with opposite orientations, then the orientation of $W_{\frac{1}{2}}$ can be extended over the handle; but if with the same orientation, $W_{1\frac{1}{2}}$ is non-orientable. Thus if, say, $W_{\frac{1}{2}}$ is connected and orientable, we may speak of orientable and of non-orientable 1-handles. Now r-handles for $r \neq 1$ do not affect orientability; for they introduce no new (potentially orientation-reversing) elements of the fundamental group.

For a 1-handle with both ends in the same component of $W_{\frac{1}{2}}$, we can deform both components of $S^0 \times D^{n-1}$ into a disc in $M_{\frac{1}{2}}$: as for the Disc Theorem, the diffeotopy class is determined by the orientations. Attaching an orientable 1-handle to D^n gives $S^1 \times D^{n-1}$, so we have $W_{1\frac{1}{2}} = W_{\frac{1}{2}} + (S^1 \times D^{n-1})$. In the non-orientable case, we have the sum with a non-orientable bundle over S^1 with fibre D^{n-1}.

5.4 Modifying decompositions

In this section we discuss several modifications that can be made to handle decompositions. We will see that (under suitable hypotheses) any elementary change of the chain complex $C_*(W, \partial_- W)$ can be effected by a change in the handle decomposition. The basic moves are introduction or cancellation of a complementary pair of handles, and addition of handles. We suppose throughout that W is a compact manifold, perhaps with boundary.

The results are simplest for 0-handles. If W has α_i i-handles, then $W_{\frac{1}{2}} \cong (\partial_- W \times I) \cup_{\alpha_0} D^n$. Attaching a 1-handle affects connectivity only if its a-sphere S^0 has the two points in different components of $W_{1/2}$.

Suppose that W is connected: since r-handles for $r \geq 2$ do not affect connectedness, $W_{\frac{1}{2}}$ is connected. Rearrange the 1-handles (Lemma 5.2.1) such that the first few each connect different components of $W_{\frac{1}{2}}$. For each of these, we have two manifolds with boundary, and a disc imbedded in the boundary of each. Attaching $D^{n-1} \times I$ is the same (§2.7) as glueing along the $(n-1)$-discs, i.e. forming the boundary sum. Moreover, by Proposition 2.7.6, for any manifold N^n, $N^n + D^n \cong N^n$. So the 0-handles are just cancelled out, and the remaining components of $\partial_- W \times I$ added together. We observe that each use of $N^n + D^n \cong N^n$ to simplify the decomposition removes just one 0-handle and one 1-handle.

We have shown

Proposition 5.4.1 *A connected manifold W admits a handle presentation of the following kind.*

If $\partial_- W = \varnothing$, there is just one 0-handle D^n.

If $\partial_- W$ has components $M_{(i)}$, $1 \leq i \leq k$, there are no 0-handles, then $(k-1)$ 1-handles connecting the components to give $(M_{(1)} \times I) + \cdots + (M_{(k)} \times I)$, then a further number of 1-handles.

We turn to cancellation of handles in general, and first describe a model.

Lemma 5.4.2 *Let $\varphi : D^{n-r-1} \to D^{n-r}$ be the embedding, by stereographic projection from $(0, \ldots, 0, -1)$ on the boundary of the upper hemisphere. Then $(S^r \times D^{n-r}) \cup_{1\times\varphi} h^{r+1} \cong D^n$.*

Proof If we attach D^{r+1} along the boundary to $S^r \times I$, we clearly have another $(r+1)$-disc. Multiplying by D^{n-r-1} shows that there exists a homeomorphism of the desired type. However to obtain a result up to diffeomorphism requires care with rounding corners systematically.

We first give the proof for $r = 0$, $n = 2$. Let E be the ellipse $\frac{1}{2}x^2 + y^2 = 1$ and H the confocal hyperbola $2x^2 - 2y^2 = 1$. Write Int and Ext for the (closed) interior and exterior regions of E. We shall show that $\operatorname{Int} E \cap \operatorname{Ext} H$ is obtained from $S^0 \times D^2$ by introducing a corner along $S^0 \times D^1$; that $\operatorname{Int} E \cap \operatorname{Ext} H$ is diffeomorphic to $D^1 \times D^1$, and that the attaching map $1 \times \varphi$ becomes the identity. It follows that the required manifold is diffeomorphic to $\operatorname{Int} E$, which is diffeomorphic to D^2 by $(x, y) \mapsto (2^{-1/2}x, y)$. Now E meets H at $(\pm 1, \pm 1/\sqrt{2})$. Consider the component of $\operatorname{Int} E \cap \operatorname{Ext} H$ in $x > 0$; it has the focus $(1, 0)$ as interior point.

Rays through the focus define a vector field everywhere transverse to the boundary, which may therefore be used for straightening the corner. A smooth cross-section is given by $(x-1)^2 + y^2 = 1/4$, which meets the rays through the corner in $(1, \pm 1/2)$. Thus the disc component is obtained from a disc by introducing corners at opposite ends of a diameter, as stated. It may be helpful to imagine these constructions using Figure 5.7.

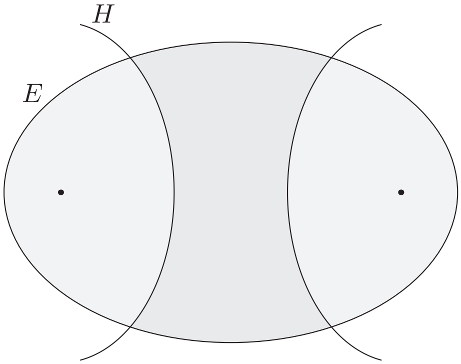

Figure 5.7 A confocal ellipse and hyperbola

In $\operatorname{Int} E \cap \operatorname{Ext} H$ we use confocal coordinates. Each point (x, y) of the plane with $xy \neq 0$ lies on just two of the conics

$$x^2/(\lambda+1) + y^2/\lambda = 1 :$$

one hyperbola, given by $-1 < \lambda_1 < 0$, and one ellipse, given by $0 < \lambda_2$. However, these meet in 4 points. So we write $\mu^2 = a + \lambda_1$, $\nu^2 = \lambda_2$, and obtain

$$x = \mu\sqrt{1+\nu^2} \quad y = \nu\sqrt{1-\mu^2},$$

where the positive square roots are to be taken, and $-1 < \mu < 1$. It is easy to verify that this transformation is smooth, with non-zero Jacobian, injective, and onto the whole plane except for $y = 0$, $x^2 \geq 1$. Hence, in particular, it induces a diffeomorphism of the rectangle $|\mu| \leq 1/\sqrt{2}$, $|\nu| \leq 1$ onto $\operatorname{Int} E \cap \operatorname{Ext} H$, as required.

Now return to the case of general r and n, which is obtained by rotating the figures about x- and y-axes. Write

$$\begin{aligned} x &= (x_1, \ldots, x_{r+1}) & y &= (y_1, \ldots, y_{n-r-1}) \\ \mu &= (\mu_1, \ldots, \mu_{r+1}) & \nu &= (\nu_1, \ldots, \nu_{n-r-1}) \end{aligned}$$

and $\|x\|^2 = \sum_1^{r+1} x_i^2$, etc. Then the transformation given by

$$x_i = \mu_i\sqrt{1+\|\nu\|^2}, \qquad y_i = \nu_i\sqrt{1-\|\mu\|^2}$$

induces a diffeomorphism of the $D^{r+1} \times D^{n-r-1}$ given by $\|\mu\|^2 \leq 1/2$, $\|\nu\|^2 \leq 1$ onto the intersection $\frac{1}{2}\|x\|^2 + \|y\|^2 \leq 1$, $2\|x\|^2 - 2\|y\|^2 \leq 1$.

Likewise in the intersection $\frac{1}{2}\|x\|^2 + \|y\|^2 \leq 1$, $2\|x\|^2 - 2\|y\|^2 \leq 1$, consider the field formed by rays through the r-sphere $y = 0$, $\|x\| = 1$ and perpendicular to it (and not produced beyond their intersection with $x = 0$). This certainly is a vector field (except on the sphere and on $x = 0$), and is transverse to the boundary, so can be used for rounding the corner. Rounding it, we obtain the manifold

$$(\|x\| - 1)^2 + \|y\|^2 \leq 1/4,$$

where the corner is to be introduced along $\|x\| = 1$, $\|y\| = 1/2$ (in fact $S^r \times S^{n-r-2}$).

Consider $S^r \times D^{n-r} \subset \mathbb{R}^{r+1} \times \mathbb{R}^{n-r-1} \times \mathbb{R}^1$ with coordinates (u, w, t), so $\|u\| = 1$, $\|w\|^2 + |t|^2 \leq 1$. We define inverse diffeomorphisms between this and the manifold above by

$$\begin{aligned} &u = x/\|x\| \qquad w = 2y \qquad t = 2(\|x\| - 1) \\ &x = u(1 + t/2) \qquad y = w/2. \end{aligned}$$

Since $\|x\|$ is nowhere zero, both it and its inverse are smooth. The corner $\|x\| = 1$, $\|y\| = 1/2$ becomes the locus $\|w\| = 1, t = 0$.

Finally we must identify the attaching map. The sphere $S^r \times 0$ given by $\|\mu\|^2 = 1/2, \nu = 0$ maps (via $x_i = \mu_i$) to $\|x\|^2 = 1/2, y = 0$, then rounding the corner multiplies x_i by $2^{-1/2}$ and leaves y at 0. Finally we obtain $u = x/\|x\| = \mu/\|\mu\|$ and $v = (w, t) = (0, -1)$; modulo the obvious identifications, we have the identity map. The attaching map is a tubular neighbourhood of this, and a normal direction $\partial/\partial\nu_i$ maps to some positive multiple of $\partial/\partial v_i$; using the tubular neighbourhood theorem, it follows that the attaching map is, up to a diffeotopy, as stated. □

Theorem 5.4.3 (Handle Cancellation Theorem) *Suppose that for $W^n \cup_f h^r \cup_g h^{r+1}$, the a-sphere of h^{r+1} meets the b-sphere of h^r transversely in one point. Then we can suppose $\partial_+ W$ contains a disc D^{n-1} to which both handles are added. Thus we can write $W^n \cong W^n + D^n$, with the handles added to D^n, and so $W^n \cup h^r \cup h^{r+1} \cong W^n + (D^n \cup h^r \cup h^{r+1}) \cong W^n + D^n \cong W^n$.*

Proof It clearly suffices to consider the case $W = M \times I$. By hypothesis, in $M_{r+\frac{1}{2}}$ the a-sphere and b-sphere of the handles meet transversely at a single point P. It follows from Lemma 4.8.1 that there is a chart for $M_{r+\frac{1}{2}}$ at P meeting the a-sphere in $D^r \times \{0\}$ and the b-sphere in $\{0\} \times D^{n-r-1}$. Since the tubular neighbourhoods are unique up to diffeotopy, we may suppose that they both meet this chart in $D^r \times D^{n-r-1}$, with the projections being those on the factors.

The r-handle is attached to M by an embedding $f : S^{r-1} \times D^{n-r} \to M$; W is formed from $M \times I$ by attaching $D^r \times D^{n-r}$ to $M \times \{1\}$ by f and rounding the corner. Here Figure 5.8 represents an r-handle with an $(r + 1)$-handle being sewn on as a patch.

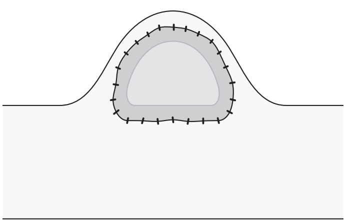

Figure 5.8 Cancelling a handle

The $(r + 1)$ handle is attached by an embedding $g : S^r \times D^{n-r-1} \to M_{r+\frac{1}{2}}$, so there is an embedding of D^r as a hemisphere of S^r, which we may take as S^r_+: thus $S^r_+ \times D^{n-r-1}$ maps onto the outer edge of the r-handle. The closed complementary region $S^r_- \times D^{n-r-1}$ is mapped to the closed complement in M

of the image of f. As before, we may choose the maps to identify D^{n-r-1} with a hemisphere S_+^{n-r-1}.

Thus the subset of M affected by the handles is the union of the embedded images $f(S^{r-1} \times D^{n-r})$ and $g(S_-^r \times D^{n-r-1})$, modulo rounding the corner. The latter image is a disc; since we can isotope D^{n-r} inside a neighbourhood of a point $X \in S_+^{n-r-1}$ and hence $S^{r-1} \times D^{n-r}$ inside a neighbourhood of $S^{r-1} \times X$, there is a disc in M containing both the embedded images. The result now follows from Lemma 5.4.2. □

A pair of handles of consecutive dimensions, with the a-sphere of the second meeting the b-sphere of the first transversely in one point, is called a *complementary pair.*

We can thus paraphrase Theorem 5.4.3 briefly by saying that a complementary pair of handles may always be cancelled. The converse result is now trivial.

Theorem 5.4.4 *At any point of a handle decomposition of a manifold, a complementary pair of handles can be introduced.*

Proof 'At any point' means when we have constructed some manifold W, say. Now $W \cong W + D$ by Proposition 2.7.5 and by Lemma 5.4.2, we can add a complementary pair of handles to D, hence also to W. □

We will see that adding two complementary handles in succession to W has the effect on $V = \partial_+ W$ of performing consecutively spherical modifications of types $(r, n-r)$, leading to V', say, and $(r+1, n-r-1)$: returning to V. 'Reversing' the second of these shows that we can also go from V to V' by a modification of type $(n-r-1, r+1)$. The condition on the first modification necessary for this replacement to be possible was the existence of a complementary handle; arguing as above shows that this is equivalent to requiring the a-sphere to span a disc in V, such that the inward normal vector to the sphere in the disc agrees with the first vector of the chosen normal framing of the a-sphere.

Since $\partial_- W$ need not be simply-connected, an $(r-1)$-sphere in it does not necessarily have a well-defined homotopy class. Here we will ignore this point and focus on homology. This allows us to give a much simpler account, and still obtain full results in the simply-connected case. We reserve comments on the general case until §5.7.

We next discuss 'addition' of handles in a homology sense.

Theorem 5.4.5 (Handle Addition Theorem) *Suppose $\partial_+ W = M$ connected, $2 \leq r \leq m-2$. Let $f, g : \partial D^r \times D^{m-r} \to M$ be disjoint embeddings, determining homology classes x, $y \in H_{r-1}(M; \mathbb{Z})$. Then for $\varepsilon = \pm 1$ there is an* *embedding* $h_\varepsilon : \partial D^r \times D^{m-r} \to M$, *disjoint from* f, *and determining* $y + \varepsilon x \in H_{r-1}(M; \mathbb{Z})$ *such that* $W \cup_f h^r \cup_g h^r \cong W \cup_f h^r \cup_{h_\varepsilon} h^r$.

Moreover, if the classes of the handles in $H_r(W \cup_f h^r \cup_g h^r, W; \mathbb{Z})$ *are* ξ, η *for the first decomposition, those for the second are* ξ, $\eta + \varepsilon\xi$.

Proof We observe that x maps to zero in $H_{r-1}(W \cup_f h^r; \mathbb{Z})$; the idea of the proof is to deform the second handle 'across' the first, by a diffeotopy of the attaching map in $\partial_+(M \cup_f h^r)$; we know that this will not affect the diffeomorphism class of the result.

Since M is connected, there is a path λ joining $f(1 \times 1)$ and $g(1 \times 1)$ (in the non-simply-connected case, it is important to note that this path may be taken in any homotopy class). By the general position arguments of §4.5, we can make the path an embedding, disjoint from the images of $\bar{f}$ and $\bar{g}$; we can choose it to start along the outward normals to $\operatorname{Im} f$ and $\operatorname{Im} g$, and we can deform it off tubular neighbourhoods of $\operatorname{Im} \bar{f}$ and $\operatorname{Im} \bar{g}$, so that it meets $\operatorname{Im} f$ and $\operatorname{Im} g$ only at its ends.

Choose a normal framing $e_1, \dots, e_{m-2}$ for λ so that $e_1, \dots, e_{r-1}$ gives the standard orientation of $g(S^{r-1} \times 1)$ at $g(1 \times 1)$. Since $r \leq m - 2$, we can also change this framing so that $e_1, \dots, e_{r-1}$ agrees with the opposite orientation of the $(r-1)$-sphere. By Proposition 1.5.6 (ii) we can choose a Riemannian metric in which $f(S^{r-1} \times 1)$ and $g(S^{r-1} \times 1)$ are totally geodesic. Then exponentiating normal vectors to λ gives an embedding $\varphi' : I \times D^{r-1} \to M$ with

$$\varphi'(0 \times D^{r-1}) \subset g(S^{r-1} \times 1), \qquad \varphi'(1 \times D^{r-1}) \subset f(S^{r-1} \times 1).$$

Extend λ by a diameter of $D^{r-1} \times 1$ in $\partial_+(M \cup_f h^r)$, and φ' correspondingly to an embedding $\varphi : [0, 2] \times D^{r-1} \to \partial_+(M \cup_f h^r)$.

The properties of the bump function ensure that the formulae

$$\begin{aligned} \bar{g}_t(x) &= x \quad \text{if} \quad x \notin \varphi(0 \times D^{r-1}), \\ \bar{g}_t\varphi(0, y) &= \varphi(2tBp(1 - \|y\|), y) \end{aligned}$$

fit to give a smooth diffeotopy of $\bar{g}$. This 'pulls' the cell $\varphi(0 \times D^{r-1}) \subset g(S^{r-1} \times 1)$ across part of the disc $D^r \times 1$, covering the central point.

This procedure (with $r = 1$) is illustrated in Figure 5.9: here, to add the handle, deform the attaching sphere of the left handle along the dotted path to give a new attaching sphere.

Since $g(S^{r-1} \times 0)$ is diffeotopic to $g(S^{r-1} \times 1)$, we also obtain a diffeotopy of $\bar{g}$, which we can extend to one of g such that the final embedding h is disjoint from $0 \times S^{n-r-1}$. But we can think of the (f-) handle as shrunk to a small neighbourhood of this b-sphere (c.f. proof of 5.2.3), so $h(S^{r-1} \times D^{n-r})$ lies in M again.

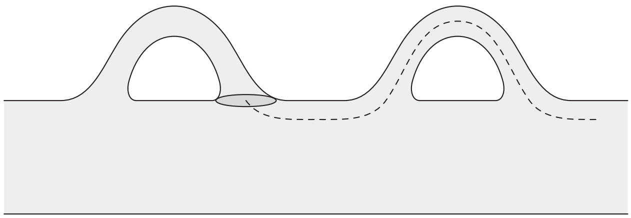

Figure 5.9 Adding one handle to another

Since our diffeotopy has degree 1 on the attached cell, the homology class of h is that of g plus or minus that of f, the sign depending on an orientation chosen earlier. $\square$

5.5 Geometric connectivity and the h-cobordism theorem

In the last section we gave methods of changing handle decompositions under geometric assumptions. We now obtain corresponding results under algebraic hypotheses: this will enable us to operate with handles using only homotopy data. We recall that a CW-pair (Y, X) is called r-connected if any map $f : (B, A) \to (Y, X)$ with $\dim(B) \leq r$ is homotopic relative to A to a map into X; equivalently, if the relative homotopy set $\pi_i(Y, X)$ is trivial for $0 \leq i \leq r$. Moreover this holds if and only if the pair (Y, X) is homotopy equivalent to a pair with Y' obtained from X by attaching cells of dimension $> r$.

We focus first on results showing the existence of handle decompositions without i-handles for $i \leq r$: if W admits such a decomposition, we say that $(W, \partial_- W)$ is *geometrically r-connected*.

We start with a technique of handle replacement. It is interesting to note that this closely resembles a technique of Whitehead, with CW complexes. Although it may seem that it would be more efficient to simply cancel handles, handle replacement bypasses arguments involving fundamental groups, which otherwise would confuse the issue for low-dimensional cases.

Proposition 5.5.1 *Suppose* $n \geq 2r + 3$, $W^n = (M \times I) \cup h^r \cup lh^{r+1}$, *and* $\pi_r(W, M) = 0$. *Then* $W \cong (M \times I) \cup lh^{r+1} \cup h^{r+2}$.

Proof The case $r = 0$ follows from Proposition 5.4.1; otherwise we may suppose M connected.

We identify h^r with $D^r \times D^{m-r}$. Since $n \geq 2r + 2$, we can perform a diffeotopy to ensure that the attaching maps of the h^{r+1} avoid $D^r \times 1$. The disc $D^r \times 1$ determines an element of $\pi_r(W, M)$, which is zero by the hypothesis. Hence this disc is homotopic in W (relative to its boundary) to one in M; i.e.

there is a map $F : D^{r+1} \to W$, which takes the upper hemisphere of S^r onto $D^r \times 1$ and the lower into M.

Since $n \geq 2r + 3$, we may suppose that $\operatorname{Im} F$ is disjoint from the cores of the handles, which have dimensions r and $(r + 1)$. We can therefore also deform F off tubular neighbourhoods of the cores, and thus suppose $\operatorname{Im} F \subset \partial_+ W$.

We may suppose $F \,|\, S^r$ an embedding of S^r in $\partial_+ W$: this embedding is homotopic to zero, hence also diffeotopic (since $n \geq 2r + 3$, a map $S^r \times I \to \partial_+ W \times I$ may be supposed an embedding). So by Theorem 5.4.4, we can use $F(S^r)$ for the a-sphere of the first of a complementary pair of handles h_A^{r+1}, h_B^{r+2}, where h_A^{r+1} is disjoint from the other h^{r+1}. But h_A^{r+1} is also complementary to h^r, so

$$\begin{aligned} W &\cong (M \times I) \cup h^r \cup lh^{r+1} \cup (h_A^{r+1} \cup h_B^{r+2}) && \text{(Theorem 5.4.4)} \\ &\cong (M \times I) \cup (h^r \cup h_A^{r+1}) \cup lh^{r+1} \cup h_B^{r+2} \\ &\cong (M \times I) \cup lh^{r+1} \cup h_B^{r+2} && \text{(Theorem 5.4.3).} \end{aligned}$$

□

It is possible, with some difficulty, to sharpen the proof of Theorem 5.5.1 to cover also the case $n = 2r + 2$, $r \neq 1$: the points to be addressed are the deformation of F off the cores of the h^{r+1} and obtaining a diffeotopy.

Theorem 5.5.2 *If $W = V \cup kh^r \cup lh^{r+1}$, $\pi_r(W, V) = 0$, $\pi_1(\partial_+ V) \cong \pi_1(V)$, $n \geq 2r + 3$, then $W \cong V \cup lh^{r+1} \cup kh^{r+2}$.*

Proof Write $V' := V \cup (k - 1)h^r$, $M' := \partial_+(V')$, so W is the union of V' and $W' := (M' \times I) \cup h^r \cup lh^{r+1}$. Since $\pi_r(W, V)$ and $\pi_{r-1}(V', V)$ vanish, so does $\pi_r(W, V')$. If we show that $\pi_r(W', M') = 0$, we can apply Proposition 5.5.1 to replace the r-handle in W' by an $(r + 2)$-handle, so $W \cong V \cup (k - 1)h^r \cup lh^{r+1} \cup h^{r+2}$. Since the $(r + 2)$-handle does not affect the calculation of π_r, the result will follow by induction.

Now $W = V' \cup W'$ and $M' = V' \cap W'$. Since $n \geq r + 4$, the fundamental groups of $\partial_+ W$, W', W, M' and hence of V' are isomorphic. Thus the universal covers of W', V', M' (which we denote by affixing a tilde) are induced from that of W.

Thus by the Hurewicz isomorphism theorem (see B.3 (i)),

$$\pi_r(W', M') \cong H_r(\tilde{W}', \tilde{M}') \cong H_r(\tilde{W}, \tilde{V}) \cong \pi_r(W, V) = 0,$$

where the middle isomorphism holds by excision. □

We next extend the result by a more direct application of the Handle Cancellation Theorem. To avoid technicalities, we restrict to the simply-connected case: in §5.7 we indicate what is needed to remove this restriction.

Proposition 5.5.3 *Proposition 5.5.1 continues to hold if the hypothesis '$n \geq 2r + 3$' is replaced by '$n \geq r + 4$ and M is simply-connected.'*

Proof Since $H_r(W, M) = 0$, the class y of the r-handle in the chain complex $C_*(W, M)$ is a boundary, so if the $(r+1)$-handles have classes x_i, there are coefficients $c_i \in \mathbb{Z}$ with $\partial(\sum_i c_i x_i) = y$.

Use Theorem 5.4.4 to add a complementary pair of handles h_A^{r+1}, h_B^{r+2} away from the existing handles. Now use Theorem 5.4.5 to add c_i copies of the ith $(r+1)$-handle to h_A^{r+1} for each i. The a-sphere of the resulting handle has intersection number 1 with the b-sphere S_b of h^r.

Hence by Theorem 6.3.2 (i), provided $r \geq 3$ and $n \geq r + 4$, we can perform a diffeotopy to reduce the number of intersections to one. But then h^r and $h_A'^{r+1}$ are complementary, so can be cancelled by Theorem 5.4.3.

The cases $r \leq 1$ follow from Proposition 5.5.1, also the case $r = 2$ except if $n = 6$. But if $r = 2$, we can use (ii) of the theorem, provided we show that the complement of S_b is simply-connected. But we have a diffeomorphism of $\partial_+((M \times I) \cup h^r) \setminus S_b$ with $M \setminus S^1$, where S^1 is the a-sphere of h^2. By hypothesis M is simply-connected, and deleting an embedded circle does not affect this property. □

We can go a little further.

Proposition 5.5.4 *The result also holds if $n = r + 3$, provided $r \geq 3$ and ∂_+W is simply connected.*

Proof The above argument remains valid, except in the use of Theorem 6.3.2 (i). Again, we can use (ii) of the theorem, provided we show that the complement of S_a is simply-connected.

In the dual decomposition, we attach to ∂_+W first h_B^1, then a complementary h_A^2 and other 2-handles, then a 3-handle. Thus the fundamental group remains trivial at each stage after the second, hence the boundaries are simply-connected. But the complement of S_a in $\partial_+((M \times I) \cup h^r)$ is diffeomorphic to the complement of the belt sphere (a circle) in $\partial_+((M \times I) \cup h^r \cup h_A^{r+1})$, so is simply-connected. □

Theorem 5.5.5 *Suppose (W, V) r-connected, ∂_+V, V and W simply-connected, and either $n \geq r + 4$ or $n = r + 3$, $r \geq 3$ and ∂_+W is simply-connected.*

Then W has a handle decomposition on V with no i-handles for $i \leq r$.

Proof By induction on r, we may suppose there are no i-handles for $i < r$. A second induction shows that it will suffice to remove or replace a single r-handle.

Write W' for the union of V and all but one of the r-handles, W'' for the union of the remaining r-handle and all $(r+1)$-handles, and W''' for the rest. The conclusion will follow if we show that Proposition 5.5.3 can be applied to W''.

We thus need to show that $\partial_- W''$ is simply-connected and $H_r(W'', \partial_- W'') = 0$. As V is simply-connected and W' is obtained from V by attaching r-handles with $r \neq 1$, W' is simply-connected. Since W' has no handle of index $n-2$, $\partial_+ W' = \partial_- W''$ also is simply-connected.

From the exact homology sequence

$$0 = H_r(W, V) \to H_r(W, W') \to H_{r-1}(W', V) = 0$$

we see that $H_r(W, W') = 0$; since handles of index $> r+1$ do not change H_r, we have $0 = H_r(W' \cup W'', W') = H_r(W'', \partial_- W'')$. □

The culmination of the theory developed in this chapter is the so-called h-cobordism theorem. Let W be a cobordism. If the inclusions of $\partial_- W$, $\partial_+ W$ in W are homotopy equivalences, W is called an *h-cobordism*. Provided all of $\partial_- W$, $\partial_+ W$, and W are simply-connected, it suffices if the relative homology groups $H_i(W, \partial_- W; \mathbb{Z})$ vanish, since by duality the $H_i(W, \partial_+ W; \mathbb{Z})$ also vanish, so both inclusions are homotopy equivalences. We have

Theorem 5.5.6 (h-cobordism Theorem) *If W^n is a simply-connected h-cobordism with $n \geq 6$, then $W \cong \partial_- W \times I$, so $\partial_+ W \cong \partial_- W$.*

Proof Take a handle decomposition, and choose r with $2 \leq r \leq n-3$. By Theorem 5.5.5, we can inductively replace each i-handle for $i < r$ by an $(i+2)$-handle. Now apply the same argument to the dual handle decomposition to eliminate all j-handles with $j > r+1$. Observe that if $r = 2$ or $r = n-3$ we need to use Theorem 6.3.2 (ii), so cannot allow both equalities together.

We now only have r- and $(r+1)$-handles, so the chain complex $C_*(W, \partial_- W)$ reduces to a single map $\partial : C_{r+1} \to C_r$ which, since we have an h-cobordism, is an isomorphism. Performing handle additions has the effect of row operations on the matrix of ∂. Consider the first column: by the Euclidean algorithm, we can repeatedly subtract smaller from larger entries to reduce until there is a single non-zero entry, which must be ± 1.

This shows that one of the a-spheres and one of the b-spheres have intersection number ± 1. Now if $2 < r < n-3$ we can use Theorem 6.3.2 (i) to reduce the number of intersections to one, and then cancel the corresponding handles using Theorem 5.4.3. As above if one (but not both) equality holds, we can instead use (ii). Repeating the argument, we can remove all the handles. □

5.6 Applications of h-cobordism

We can formulate the h-cobordism theorem a little more generally as follows.

Theorem 5.6.1 *If $V^n \subset W^n$ is a homotopy equivalence with $\partial_- V = \partial_- W$, V, $\partial_+ V$ and $\partial_+ W$ are all simply-connected and $n \geq 6$, then $V^n \cong W^n$.*

Proof We may write W as the union of two cobordisms V and V' with a common boundary $\partial_+ V$; since this and W are simply-connected, so is V'. It now follows as each $H_i(V', \partial_- V') \cong H_i(W, V) = 0$ that V' is an h-cobordism, hence by Theorem 5.5.6 is diffeomorphic to $\partial_+ V \times I$. By Lemma 2.7.2, W is diffeomorphic to V.

The argument applies even allowing $\partial_- V$ to have a boundary X. Here we need first to adjust corners so that $\partial_c V \cong X \times I$ and also $\partial_c V' \cong X \times I$. □

We have as simple application,

Theorem 5.6.2 (Disc Bundle Theorem) [139] *Suppose M^{n-c} a submanifold of W^n, $\partial M = \varnothing$, $c \geq 3$, $n \geq 6$, $M \subset W$ a homotopy equivalence, and M, ∂W simply-connected. Then W has the structure of a disc bundle with M as zero cross-section.*

Proof Take V as a tubular neighbourhood of M. Since $c \geq 3$, ∂V is simply connected. The result thus follows from the preceding theorem. □

Taking M to be a point gives

Corollary 5.6.3 *If W^n is contractible, $n \geq 6$, $\pi_1(\partial W) = 0$, then $W^n \cong D^n$.*

We call a closed manifold a *homotopy sphere* if it is homotopy equivalent to a sphere.

Corollary 5.6.4 *If Σ^n is a homotopy sphere, $n \geq 6$, then Σ^n may be obtained by glueing two discs together along the boundary. Thus Σ^n is homeomorphic to S^n.*

Proof Let W^n be the closure of the complement of a disc D^n in Σ^n. Then W is homotopic to $\Sigma^n \setminus \{point\}$, so is simply-connected, and its reduced homology groups vanish, so W is contractible. By Corollary 5.6.3, $W^n \cong D^n$.

Since D^n is homeomorphic to the cone over S^{n-1}, any homeomorphism of S^{n-1} extends, by taking the cone, to a homeomorphism of D^n. Since Σ^n is homeomorphic to the union of two copies of D^n glued by a homeomorphism of S^{n-1}, it follows that we have a homeomorphism on S^n. □

The Generalised Poincaré Conjecture states that any homotopy sphere Σ^n is homeomorphic to the sphere S^n: the original conjecture referred to the case $n = 3$. We have just proved this if $n \geq 6$. The cases $n \leq 5$ are discussed in the next section §5.7. We will return to the question of diffeomorphism in the final section §8.8.

Proposition 5.6.5 *(i) Suppose M, M' compact, simply-connected and without boundary, $f : M \to M'$ a homotopy equivalence and $2c \geq m$. Then $M' \times D^c$ is a disc bundle over M.*

(ii) Suppose in addition that $c \geq m + 1$ and $f^*(\mathbb{T}(M') + 1) \cong \mathbb{T}(M) + 1$. Then $M \times D^c \cong M' \times D^c$.

Proof If $c < 3$ then $m \leq 1$, M and M' are homotopy equivalent to a circle or a point, and the result is trivial. Now let $c \geq 3$. Then by Theorem 6.4.11, we can approximate f by an embedding of M in $M' \times D^c$. The result now follows from Theorem 5.6.2.

(ii) In this case, the normal bundle of $g(M)$ in $M' \times D^c$ is stably trivial and of fibre dimension $\geq m + 1$, hence (by §B.3(xi)) is trivial. □

Proposition 5.6.6 *Let Σ^{n-c} be a homotopy sphere embedded in S^n ($n \geq 6$, $c \geq 3$), N a tubular neighbourhood of Σ, V the closure of its complement. Then V is diffeomorphic to $S^{c-1} \times D^{n-c+1}$.*

Proof Let N' be a tubular neighbourhood of Σ with N in its interior, D^c a fibre, S^{c-1} its boundary. Since S^{c-1} bounds the contractible D^c, its normal bundle is trivial. We assert that the inclusion of S^{c-1} in V is a homotopy equivalence; indeed, both are simply-connected (V since S^n is, and $S^n \setminus \Sigma^{n-c}$ since $c \geq 3$) and the complement of $V \cup D^c$ is the interior of $N \setminus D^c$, a cell bundle over a cell and so contractible. By duality, $V \cup D^c$ is contractible, and $0 = H_r(V \cup D^c, D^c) = H_r(V, V \cap D^c)$. But $V \cap D^c$ is an annulus with S^{c-1} as deformation retract, hence $H_r(V, S^{c-1}) = 0$.

If $c \neq n - 1$, $\partial V = \partial N$ is simply-connected, and $n - c + 1 \geq 3$, so the result follows by applying Theorem 5.6.2 to $S^{c-1} \subset V$. If $c = n - 1$, Σ is a circle, and unknots, so the result is trivial. □

We can adapt some of the above arguments to give a relative result.

Theorem 5.6.7 *(i) Suppose W^n a simply-connected h-cobordism, $n \geq 6$, V^{n-c} a submanifold, $c \geq 3$, such that $V^{n-c} \cong \partial_- V \times I$. Then $(W, V) \cong (\partial_- W, \partial_- V) \times I$.*

(ii) Two h-cobordant pairs of homotopy spheres $(\Sigma_i^{n+c}, \Sigma_i^n)(i = 0, 1)$ with $n \geq 5$, $c \geq 3$ are diffeomorphic.

Proof As in Lemma 5.1.1, we can find a non-degenerate function on W whose restriction to V has no critical points; the proof of Lemma 5.1.1 is only changed by using the given product structure to define g near V. We can now carry out all the handle decomposition and cancellation arguments in $W \setminus V$.

More precisely, write N for a tubular neighbourhood of V in W, $\mathring{N}$ for its interior, $X = W \setminus \mathring{N}$ and $Y = N \cap X = \partial_c N = \partial_c X$.

Since $c \geq 3$ is the codimension of V in W (and of $\partial_- V$ in $\partial_- W$, $\partial_+ V$ in $\partial_+ W$), removing V does not alter the fundamental groups.

So it is enough to check that $\partial_- X \subset X$ is a homotopy equivalence, and so enough to show that $H_*(X, \partial_- X) = 0$. Since $\partial_- V$ is a deformation retract of V, and N is a disc bundle, $\partial_- N$ is a deformation retract of N, also of $\partial_- N \cup Y$. Hence $0 = H_*(N, \partial_- N \cup Y) \cong H_*(W, X \cup \partial_- W)$ by excision. But $H_*(W, \partial_- W)$ is trivial, so using the homology exact sequence of the triple $\partial_- W \subset X \cup \partial_- W \subset W$, we deduce that $H_*(X \cup \partial_- W, \partial_- W)$ is trivial. It follows by excision that $H_*(X, \partial_- X) = 0$. The result follows.

(ii) By the h-cobordism theorem, the h-cobordism of the Σ_i^n is a product, so the result follows from (i). □

A different relative form of these results can also be obtained, giving a topological unknotting theorem for pairs of spheres.

Proposition 5.6.8 *(i) Let $M^m \subset W^{m+c}$ be a proper embedding of contractible manifolds with $c \geq 3$, $m + c \geq 6$. Assume that either $M^m \cong D^m$ or $m \geq 6$. Then the pair (W^{m+c}, M^m) is diffeomorphic to (D^{m+c}, D^m).*

(ii) Let $T^m \subset \Sigma^{m+c}$ be an embedding of homotopy spheres with $c \geq 3$, $m + c \geq 6$: assume either that $T^m \cong S^m$ or that $m \geq 6$. Then the pair (Σ^{m+c}, T^m) is homeomorphic *to (S^{m+c}, S^m).*

Proof (i) Take a tubular neighbourhood V of M in W: then V is contractible, so we can apply Theorem 5.6.1 (where we set $\partial_- V = \partial_- W = V \cap \partial W$) to the inclusion $V \subset W$ to infer that W is obtained from V by adding a collar.

(ii) Choose an embedding $(D^{m+c}, D^m) \to (\Sigma^{m+c}, T^m)$ (it is essentially unique by Lemma 2.5.11), and delete the interior to give a pair as in (i): by that result, we have another copy of (D^{m+c}, D^m). These copies are attached by a diffeomorphism of the boundary (S^{m+c-1}, S^{m-1}). But as in Corollary 5.6.4, any such diffeomorphism extends, taking the cone, to a *homeomorphism* of (D^{m+c}, D^m). □

We now proceed to obtain minimal handle decompositions in general.

Theorem 5.6.9 *Suppose W^n $(n \geq 6)$ such that $\partial_- W$, $\partial_+ W$ and W are simply-connected. Let $H_i(W, \partial_- W) \cong F + T$, where F is a free abelian group of rank*

β_i and T is a finite group with $\tau_{i+\frac{1}{2}}$ generators. Then W has a handle decomposition on $\partial_- W$ with $\tau_{i-\frac{1}{2}} + \beta_i + \tau_{i+\frac{1}{2}}$ i-handles for each i.

Proof By Corollary 5.4.1, there is a handle decomposition with no 0- or 1-handles. Similarly, we can dispense with $(n-1)$- and n-handles. This gives a chain-complex of free abelian groups whose homology is that of $H_*(W, \partial_- W)$. By making changes of basis of the chain groups, we can put this chain-complex into normal form, i.e. a direct sum of elementary subcomplexes, each with rank 1 or 2, and differential either

$$0 \to \mathbb{Z} \to 0 \qquad \text{or} \qquad 0 \to \mathbb{Z} \xrightarrow{\theta} \mathbb{Z} \to 0.$$

Now the required changes of base can be induced by a sequence of elementary automorphisms of the chain groups, and by Theorem 5.4.5, each of these can be induced by a change in handle decomposition. It remains only to remove the elementary subcomplexes with $\theta = 1$. But it follows as above from Theorem 5.4.3 that such pairs of handles may be cancelled. □

This allows us in favourable cases to obtain classifications up to diffeomorphism. It follows at once from Theorem 5.6.9 that

Lemma 5.6.10 *Suppose M^m, with $m \geq 6$, such that M and ∂M are simply-connected, $H_r(M)$ is free abelian of rank k, and $\tilde{H}_i(M) = 0$ for $i \neq r$. Then M admits a handle decomposition with one 0-handle, k r-handles, and no others.*

Such a manifold is called a *handlebody* . By Lemma 5.2.1, it can be obtained from D^m by simultaneous attachment of all k r-handles, so is determined by an embedding

$$F : \bigcup_{i=1}^{k} (S^{r-1} \times D^{m-r})_i \to S^{m-1}.$$

We can take this in two stages: first study the restriction $\overline{F}$ of F to the union of the spheres $S^{r-1} \times \{0\}$, and then thicken the spheres up to their tubular neighbourhoods.

The classification in the case $m > 2r$ is straightforward.

Theorem 5.6.11 *A handlebody M with $m > 2r$ is a boundary sum of k $(m-r)$–disc bundles over S^r. M is determined up to diffeomorphism by the values of k, r, m, and the subgroup of $\pi_{r-1}(SO)$ generated by the classes of the bundles.*

Proof Since $m > 2r$, it follows by general position that any two embeddings $\overline{F}$ are diffeotopic. In particular, the components of the image are contained in disjoint $(m-1)-$discs in S^{m-1}. It follows that the handlebody is a boundary sum. Each summand is obtained by attaching a single handle, so (for example, by Theorem 5.6.2) is a disc bundle over S^r.

Such disc bundles are classified by $\pi_{r-1}(SO_{m-r})$. Since $m > 2r$, this is isomorphic to the stable homotopy group $\pi_{r-1}(SO)$. Hence the bundle is determined by the restriction to the central sphere of the (stable) tangent bundle of M, which in turn is determined by the classifying map $M \to B(SO)$. Since M is homotopy equivalent to a bouquet of r-spheres, this comes to the same as a collection of maps $S^r \to B(SO)$. If we change the handle decomposition using the Handle Addition Theorem 5.4.5, the elements of $\pi_r(B(SO))$ add correspondingly.

We now recall the result of Bott [21] (see §B.3(xii)) that the group $\pi_{r-1}(SO)$ is cyclic, infinite if $r \equiv 0$ $(mod\ 4)$, of order 2 if $r \equiv 1, 2$ $(mod\ 8)$, and zero otherwise. Thus we can change the basis of $H_r(M)$ to ensure that all but the first basis element map to zero, and the image of the first generates the subgroup of $\pi_{r-1}(B(SO))$. □

In the case $m = 2r$, there are two extra points: the embedding $\overline{F}$ is no longer unique up to diffeotopy, and the group $\pi_{r-1}(SO_{m-r}) = \pi_{r-1}(SO_r)$ lies in exact sequences (see B.3.2):

$$\mathbb{Z} = \pi_r(S^r) \xrightarrow{\partial} \pi_{r-1}(SO_r) \xrightarrow{i_*} \pi_{r-1}(SO_{r+1}) = \pi_{r-1}(SO).$$

$$\xrightarrow{\partial} \pi_{r-1}(SO_{r-1}) \xrightarrow{i_*} \pi_{r-1}(SO_r) \xrightarrow{\pi_*} \pi_{r-1}(S^{r-1}) = \mathbb{Z}.$$

If r is even, the first map in the first sequence is injective, and both points are accommodated by taking into account the intersection pairing on $H_r(M)$. We have

Theorem 5.6.12 *Let M^{2n} be a manifold with M and ∂M simply-connected, $n \geq 3$, with $\tilde{H}_r(M)$ vanishing for $r \neq n$ and free abelian for $r = n$. The diffeomorphism type of M is determined by the following invariants:*

a free abelian group $H := H_n(M; \mathbb{Z})$,

a $(-1)^n$-symmetric bilinear map $H \times H \to \mathbb{Z}$ given by intersection numbers,

a map $\alpha : H \to \pi_{n-1}(SO_n)$.

These satisfy

(i) $x.x = \pi(\alpha(x))$ for $x \in H$, and

(ii) $\alpha(x+y) = \alpha(x) + \alpha(y) + xy(\partial\iota_n)$ for $x, y \in H$.

Proof We first define α. It follows from Theorem 6.4.11 that each $x \in H_n(M) \cong \pi_n(M)$ is represented by an embedding $f_x : S^n \to M$, and that for

$n \geq 4$ such an embedding is unique up to diffeotopy. We may thus define $\alpha(x)$ as the characteristic class of the normal bundle of $f_x(S^n)$. In the case $n = 3$, the group $\pi_{n-1}(SO_n) = \pi_2(SO_3)$ is trivial, so α is unique.

To see (i), note that $\pi : \pi_{n-1}(SO_n) \to \mathbb{Z}$ coincides with the natural map to $\pi_{n-1}(S^{n-1})$. Now $x.x$ is the intersection number of $f_x(S^n)$ with a nearby perturbation. Since f_x is an embedding, this is the primary obstruction to finding a cross-section of the bundle with fibre S^{n-1} associated to the normal bundle, hence with the image of $\alpha(x)$ under π. As to (ii), we may join the embedded spheres f_x and f_y by a tube to obtain an immersed sphere representing $x + y$. This has normal bundle given by $\alpha(x) + \alpha(y)$ and self-intersection $x.y$. Now as in §6.3 performing a homotopy to remove a single self-intersection will add $\partial\iota_n$ to the normal bundle.

We must now show that these invariants determine M up to diffeomorphism. Choose a handle presentation as above: it will suffice to show that F is determined up to diffeotopy. First consider $\overline{F}$, and note that classifying embeddings into S^{2n-1} is equivalent to classifying embeddings into $\mathbb{R}^{2n-1}$. It follows from Theorem 6.4.11 that an embedding $S^{n-1} \to \mathbb{R}^{2n-1}$ is unique up to diffeotopy.

According to Corollary 6.4.10, if $2m > 3(v + 1)$, diffeotopy classes of smooth embeddings $f : V^v \to \mathbb{R}^m$ correspond bijectively to equivariant homotopy classes of equivariant maps $V \times V \setminus \Delta(V) \to S^{m-1}$, where an embedding f determines the equivariant map f_δ defined by $f_\delta(x, y) = (f(x) - f(y))/\|f(x) - f(y)\|$. Taking $V = \bigcup_{i=1}^k S_i^{n-1}$ and $m = 2n - 1$, we see that the dimension condition is $n > 2$; the result for $k = 1$ shows that we can ignore the components $(S_i^{n-1} \times S_i^{n-1})$; and if $i \neq j$, an equivariant homotopy of a map of $(S_i^{n-1} \times S_j^{n-1}) \cup (S_j^{n-1} \times S_i^{n-1})$ is equivalent to a homotopy of $(S_i^{n-1} \times S_j^{n-1})$.

Since homotopy classes of maps $S^{n-1} \times S^{n-1} \to S^{2n-2}$ are determined by their degree, an integer, for each pair $i \neq j$ we have an integer $c_{i,j}$, which can be interpreted as the linking number of S_i^{n-1} and S_j^{n-1} in S^{2n-1}, so is equal to the intersection number of the corresponding n-spheres in M, and is $(-1)^n$-symmetric.

For each component, the choice of extension of the map $\overline{f}$ on S^{n-1} to an embedding f of $S^{n-1} \times D^n$ is equivalent to choosing an element of $\pi_{n-1}(SO_n)$, and making an appropriate normalisation, this element coincides with the characteristic class $\alpha(x_i)$ of the normal bundle of the corresponding sphere S^n in M. Hence indeed the invariants determine F up to diffeotopy, hence M up to diffeomorphism. □

It follows by a short calculation that if (and only if) the intersection form is nonsingular, the boundary of M is homotopy equivalent – and hence homeomorphic – to S^{2n-1}. This was the case of prime interest in [159], where I also considered the question of when ∂M is diffeomorphic to S^{2n-1}.

There is a corresponding classification for handlebodies in the metastable range. The proof is essentially the same, but the arguments for (i) and (ii) are somewhat more delicate, and we omit the details.

Theorem 5.6.13 *Let M^m be a handlebody with handles of dimension $s \geq 2$, $m \geq 6$, $2m \geq 3s+3$. Then the diffeomorphism type of M is determined by invariants $H := H_s(M; \mathbb{Z})$, a $(-1)^s$-symmetric bilinear map $\lambda : H \times H \to \pi_s(S^{m-s})$, and a map $\alpha : H \to \pi_{s-1}(SO_{m-s})$, satisfying*

(i) $\lambda(x, x) = S\pi_*\alpha(x)$ for $x \in H$, and

(ii) $\alpha(x+y) = \alpha(x) + \alpha(y) + \partial_*\lambda(x, y)$ for $x, y \in H$.

Here ∂_* is the boundary map in the homotopy exact sequence (B.3.2) of the fibre bundle $SO_{m-s} \to SO_{m-s+1} \to S^{m-s}$; $\pi : SO_{m-s} \to S^{m-s-1}$ is the projection, and $S : \pi_{s-1}(S^{m-s-1}) \to \pi_s(S^{m-s})$ the suspension map.

5.7 Complements

In this section we first summarise (without proofs) what is known in dimensions $n \leq 5$. Then we indicate what changes need to be made if we drop the simply-connected hypothesis.

The cases $n \leq 1$ are trivial. For $n = 2$ it follows from Proposition 5.4.1 that a connected closed 2-manifold M has a handle decomposition with just one 0- and one 2-handle. Write α_1 for the number of 1-handles, so that $\chi(M) = 2 - \alpha_1$. In particular, if M is a homotopy sphere, $\alpha_1 = 0$. Since any diffeomorphism of S^1 is diffeotopic to the identity (or a reflection) and hence extends to one of D^2, it follows that $M \cong S^2$.

Otherwise we analyse M by induction on α_1. If $\alpha_1 \geq 1$, choose a 1-handle $D^1 \times D^1$, join the ends $\partial D^1 \times \{0\}$ of the arc $D^1 \times \{0\}$ (the a-sphere) by a smooth arc in D^2 to form an embedded circle C, and cut M along this circle to give N'. There are three possibilities which are illustrated in Figure 5.10.

(a) C is 1-sided, so $\partial N'$ is a circle, the double cover of C. Adding a disc along this boundary gives a closed surface N with $\chi(N) = 1 + \chi(M)$, so $\alpha(N) = \alpha(M) - 1$. Moreover the procedure to recover M from N shows that we have a connected sum $M = N \# P^2(\mathbb{R})$.

(b) C is 2-sided and separates M into two pieces, N_1' and N_2', each with boundary C. Adding a disc to the boundary yields closed surfaces N_1, N_2 with $M \cong N_1 \# N_2$. Since $\chi(M) = \chi(N_1) + \chi(N_2) - 2$, we have $\alpha(M) = \alpha(N_1) + \alpha(N_2)$. It follows from our construction of C that neither N_i can be S^2, so each $\alpha(N_i) < \alpha(M)$.

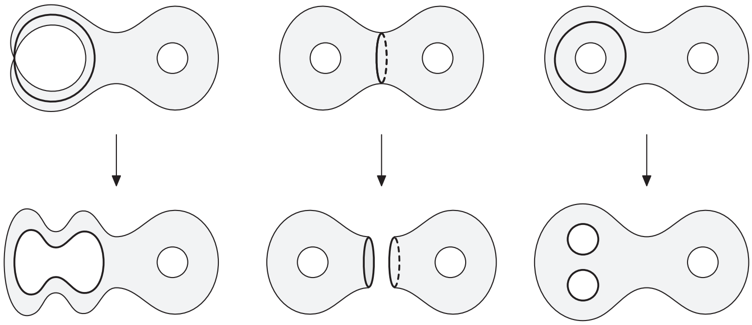

Figure 5.10 The effect of cutting a surface

(c) C is 2-sided, but does not separate M. Filling each component of $\partial N'$ by a disc gives a closed surface N; now we can choose a disc in N containing each of these in its interior, and so express N as connected sum of N^*, say, with a manifold obtained from S^2 by removing two discs, and identifying the two boundaries together. This yields a torus $S^1 \times S^1$ if orientable or a Klein bottle K^2 if not. Calculating as above gives $\chi(M) = \chi(N^*) + 2$.

It follows by induction that M is the connected sum of a collection (possibly empty) of copies of $P^2(\mathbb{R})$, $S^1 \times S^1$ and K^2. The classification is completed by the easy proof that $K^2 \cong P^2(\mathbb{R})\#P^2(\mathbb{R})$ and $P^2(\mathbb{R})\#K^2 \cong P^2(\mathbb{R})\#(S^1 \times S^1)$. The conclusion can also be formulated by saying that Theorem 5.6.12 applies in this case.

For $n = 3$, a decomposition with just one 0- and one 3-handle is essentially equivalent to a Heegard decomposition, i.e. expressing M as the union of two handlebodies, which by itself does not tell us much. However the theory of (compact) 3-manifolds is highly developed, and the principal structural result is Thurston's Geometrisation Principle, which was established by Grigori Perelman [120] in 2003, and which includes the original Poincaré Conjecture. An account of the proof in book form was given by Morgan and Tian [107]. Thurston's own work [154] gives a more leisurely and very geometric account giving some insight into how he was led to the Principle.

The case $n = 4$ is the one where our methods yield the least. To avoid repetition below, let us write $\mathcal{C}_4$ for the class of closed, simply-connected 4-manifolds. The obvious invariant of any $X \in \mathcal{C}_4$ is the symmetric bilinear form λ given by intersection numbers on $H_2(X; \mathbb{Z})$; this has rank $\beta_2(X)$ and signature $\sigma(\lambda) = \sigma(X)$; it follows from duality that λ is nonsingular. The type of λ is even if each $\lambda(x, x)$ is even (equivalently if $w_2(X) = 0$, thus iff X has a spinor structure) and odd otherwise.

It is known from the general theory of quadratic forms that two indefinite nonsingular forms of the same rank, signature, and type are isomorphic (this fails badly for definite forms), so the matrix of an indefinite odd form can be diagonalised. For even forms, $\sigma(\lambda)$ is divisible by 8 (see Proposition 7.3.3) and an example with $\sigma = 8$ is given by the E_8 matrix (7.3.4). A result going back to Rokhlin, and corresponding to the 4-dimensional case of Proposition 8.8.6, tells us that for a closed spinor 4-manifold, σ is divisible by 16. A well-known example with $\sigma = 16$ (and with $\beta_2 = 22$) is a so-called K3 surface, for example, the one given in $P^3(\mathbb{C})$ by $z_0^4 + z_1^4 + z_2^4 + z_3^4 = 0$.

The author proved in [162] that if $X_1, X_2 \in \mathcal{C}_4$ have isomorphic intersection forms they are h-cobordant, and I deduced that they become diffeomorphic after taking connected sums with a number of copies of $S^2 \times S^2$. Up to 1980 it still seemed plausible that this implied diffeomorphism, and that any nonsingular symmetric bilinear form could occur. Indeed the *topological* triviality of an h-cobordism of manifolds $X_1, X_2 \in \mathcal{C}_4$ (and hence the $n = 4$ case of the Generalised Poincaré Conjecture) was proved by Michael Freedman [53] in 1982, and he also proved that indeed any nonsingular symmetric bilinear form is the intersection form of some $X \in \mathcal{C}_4$ (but not in general smooth): see also [54].

The picture changed dramatically with the work of Donaldson. His first paper [42] proved that if $X \in \mathcal{C}_4$ is smooth and its intersection form λ is positive definite, then λ can be diagonalised (and so agrees with the intersection form of a connected sum of copies of $P^2(\mathbb{C})$); in particular, unless $\beta_2 = 0$, λ cannot be even. Next in [43] Donaldson proved non-existence of diffeomorphisms for certain pairs $X_1, X_2 \in \mathcal{C}_4$ of smooth manifolds with isomorphic intersection forms, and hence h-cobordant: thus the h-cobordism theorem fails for such manifolds. Donaldson's techniques are well outside the scope of this book (and beyond the competence of this author), but here is a brief indication of what is involved.

Let X be a closed oriented 4-manifold; write $\beta_2^+(X)$ for the dimension of a maximal subspace of $H_2(X; \mathbb{R})$ which is positive definite for the intersection form. The details require $\beta_2^+(X)$ to be odd, and extra complications arise if $\beta_2^+(X) = 1$. Principal SU_2-bundles P over X are classified by $k = \langle c_2(P), [X]\rangle$. Choose a Riemannian metric g on X; and consider the space of connections A on P. The so-called Yang–Mills equations require that the self-dual part of the curvature tensor of A vanishes. The quotient of the set of solutions of these equations by the group of bundle automorphisms of P ('gauge equivalence') is the moduli space $M_k(g)$. It is shown (with some effort) that for a generic metric g this moduli space is a smooth manifold of dimension $2d_k = 8k - 3(1 + \beta_2^+(X))$, An orientation of this maximal subspace induces one of

$M_k(g)$, and the homology class of $M_k(g)$ in the space of gauge equivalence classes of connections determines a symmetric multilinear function $q(d_k)$ of degree d_k on $H^2(X;\mathbb{R})$ which is independent of the metric. These functions give new diffeomorphism invariants for X.

The paper [81] by Kronheimer and Mrowka assembled these invariants into a generating function $q = \sum_k q(d_k)/(d_k)!$ which is regarded as a power series over $H_2(X;\mathbb{R})$; their first main theorem states that if X is 1-connected and of simple type, there exists a finite list $K_1, \dots, K_p \in H^2(X;\mathbb{Z})$ and non-zero $a_1, \dots, a_p \in \mathbb{Q}$ such that

$$q = \exp\left(\tfrac{1}{2}\lambda \textstyle\sum_{s=1}^{p} a_s e^{K_s}\right),$$

where λ denotes the intersection form. The 'simple type' condition was somewhat ad hoc, but at least allowed large families of examples. The classes K_s are called *basic classes*. They all satisfy $K.K = 2\chi(X) + 3\sigma(\lambda)$. If X is a minimal complex algebraic surface of general type, then the only basic classes are $\pm K$, where K is the canonical class: thus K (up to sign) is a diffeomorphism invariant, and the basic classes in general can be regarded as a diffeomorphism invariant version of the canonical class.

A formula relating the invariants of X to those of the blow-up $X\#P^2(\mathbb{C})$ was obtained in general by Fintushel and Stern [51]. It involves elliptic functions which, when X is of simple type, specialise to the trigonometric functions in the above formula.

Shortly afterwards, a new theory was introduced by Witten [181], based on the so-called Seiberg–Witten equations. Here there is an additional element of structure. We start with a $Spin^c$-structure on X: this induces a pair of vector bundles $W^{\pm}$, a complex line bundle L over X, and isomorphisms $\Lambda^2 W^+ \cong \Lambda^2 W^- \cong L$. The Seiberg–Witten equations define a subset of the space of pairs (A, ψ) with A a unitary connection on L and ψ a section of W^+. Again we form the space M of equivalence classes of solutions under gauge equivalence into a moduli space and need to show that for a generic metric on X, M is smooth of the expected dimension $2s(L)$, where $s(L) = \frac{1}{8}(c_1(L)^2 - (2\chi(X) + 3\sigma(\lambda)))$ (there are in fact possible isolated singularities corresponding to 'reducible solutions'), compact, oriented, and deforms well under change of metric; in fact it seems these points are somewhat easier to deal with here than in the preceding case. There is a canonical class $h \in H^2(M)$ and we obtain an invariant $n_L = \langle h^s, [M]\rangle$. There are only finitely many line bundles L with $n_L \neq 0$. This description assumes $\beta_2^+ > 1$; otherwise the invariant depends on a choice of a chamber in the cohomology of X and there is a wall-crossing formula for moving to a neighbouring chamber.

This method led to a flurry of papers which are surveyed in [45]. The review of this paper mentions conjectures of Witten that the list of basic classes coincides with the set of first Chern classes of $Spin^c$-structures on X having invariant $n_L \neq 0$, that the corresponding coefficients a_s agree (up to a normalising factor) with the Seiberg–Witten invariants, and also hints at a formula for q valid without restriction. Although nearly twenty years have now elapsed, and there is by now a large literature in this area, these questions still seem to be open.

These developments led to refinements of Donaldson's original theorems restricting the intersection form, with the simple-connectivity hypothesis weakened. The best result known in 2015 seems to be that of Furuta [55]: that if the intersection form λ of a spinor 4-manifold X is not definite, then $\beta_2(X) \geq \frac{5}{4}|\sigma(X)| + 2$ (if λ *is* definite, a theorem of Donaldson implies $\beta_2(X) = \sigma(X) = 0$). (The conjecture that $\beta_2(X) \geq \frac{11}{8}|\sigma(X)|$ remains open.)

The second major result in [81], again for X simply-connected and of simple type, asserts that if Σ is a connected surface of genus g, smoothly embedded in X, and with $\Sigma.\Sigma > 0$, then $2g - 2 \geq \Sigma.\Sigma + max_s(K_s.\Sigma)$. This gives a clear indication that there is no simple substitute for the Whitney trick of §6.3 for obtaining embeddings of surfaces in 4-folds. In particular it establishes (as was conjectured by Thom) that no surface smoothly embedded in $P^2(\mathbb{C})$ has lower genus than a smooth projective curve of the same degree.

In contrast to all these results, NO effective general technique is known (in 2015) for proving that two given closed smooth 4-manifolds are diffeomorphic.

For $n = 5$, in the presence of simple connectivity, we can cancel 1- and 4-handles, but the Whitney trick does not apply to allow us to cancel 2- and 3-handles. However any closed oriented 5-manifold with $w_2 = 0$ is the boundary of a 6-manifold, and (see Chapter 7) we can simplify the 6-manifold by surgery. In particular, one can show that any homotopy sphere Σ^5 bounds a contractible W^6, and hence is diffeomorphic to S^5. Similar arguments lead to a complete classification of closed simply-connected 5-manifolds up to diffeomorphism: see §7.9.

We next give brief indications of the changes needed to be made in the main results of this chapter to accommodate the fundamental group.

First, where we use the Whitney trick to remove intersections of spheres of complementary dimensions, it does not suffice to measure intersections in M by a single number: we must take account of the paths joining intersection points. Each intersection is then associated to a sign ± 1 and an element of $\pi_1(M)$, and we add to obtain an element of the integer group ring $\Lambda := \mathbb{Z}[\pi_1(M)]$.

In the discussion of the homology of handles, we must now consider chains in the universal cover of W, giving chains with coefficients in $\mathbb{Z}[\pi_1(W)]$.

We can partly compensate for this by improving the handle addition Theorem 5.4.5 to incorporate a path following an arbitrary element of $\pi_1(M)$.

Now Proposition 5.5.2 goes through without the hypothesis of simple connectivity, and Theorem 5.5.5 with the additional requirement $\pi_1(\partial_+ V) \cong \pi_1(V)$, and if $n = r + 3$ also $\pi_1(\partial_+ W) \cong \pi_1(W)$.

The Euclidean algorithm used in Theorem 5.5.6 fails. Here we have the matrix of ∂ over Λ and moves of the following kinds can be realised geometrically:

(a) Add some multiple of a row to another row (use handle addition, Theorem 5.4.5).

(b) Multiply some row by an element of π, or by -1 (change the path from * to an a-sphere, or the orientation of a cell).

(c) Take the direct sum of the matrix with (1) (insert a complementary pair of handles, Theorem 5.4.4).

The operations (a) and (b) generate a normal subgroup $E_N(\Lambda)$ of the general linear group $GL_N(\Lambda)$; we stabilise using (c) to obtain $E_\infty(\Lambda) \lhd GL_\infty(\Lambda)$, and the quotient defines the Whitehead group $Wh(\pi_1(M))$. There is an obstruction in this group to completing the proof of Theorem 5.5.6. An h-cobordism is called an *s-cobordism* (and the map $\partial_- W \to W$ a simple homotopy equivalence) if this obstruction vanishes.

It is known that $Wh(\pi)$ vanishes if π is free or free abelian, or an elementary 2-group or if $\pi \cong \mathbb{Z}_3, \mathbb{Z}_4$, and many other calculations are known: a survey of results for π finite is given by Oliver [116].

The results in §5.6 remain valid if the simple connectivity hypotheses are replaced as follows:

Theorem 5.6.1 (i) $\partial_+ V \subset V$ and $\partial_+ W \subset W$ induce isomorphisms of π_1; (ii) $M \subset W$ a simple homotopy equivalence, and $\partial W \subset W$ induces an isomorphism of π_1.

For Theorem 5.6.7 it suffices to require that W is an s-cobordism.

For Theorem 5.6.9 there is no direct analogue: the same argument shows that any chain complex chain homotopy equivalent to $C_*(W, \partial_- W)$ can be realised by a handle presentation, subject to compatibility with presentations of $\pi_1(W)$. To formulate this precisely comes to saying that we can imitate construction of a CW-complex of the desired (simple) homotopy type by a handle presentation. The most satisfactory results in this direction are the following, due to Mazur [90], which can be regarded as generalisations of Theorem 5.6.1 (ii).

Let M^m be a compact manifold, K^k a finite complex. We call an embedding $f : K \subset M$ *tame* if M is covered by coordinate neighbourhoods $\varphi_\alpha : U_\alpha \to \mathbb{R}^m$ such that each $\varphi_\alpha | f^{-1}(U_\alpha) : f^{-1}(U_\alpha) \to \mathbb{R}^m$ is linear on each simplex.

A submanifold U^m of $\mathring{M}^m$ is a *simple neighbourhood* of $f(K)$ if $K \subset \mathring{U}$, the inclusion $K \subset U$ is a simple homotopy equivalence, and $\pi_1(\partial U) \cong \pi_1(U \setminus K)$.

Theorem 5.7.1 (Simple Neighbourhood Theorem) *(i) A smooth regular neighbourhood is a simple neighbourhood.*

(ii) A smooth regular neighbourhood has a finite handle decomposition with one h^i corresponding to each simplex σ^i of K.

(iii) Let $m \geq 6$, codim $K \geq 3$. Then if U_1, U_2 are simple neighbourhoods of K, there is a diffeotopy of M, constant near K and away from $U_1 \cup U_2$, which moves U_1 to U_2.

Theorem 5.7.2 (Non-stable Neighbourhood Theorem) *Suppose W^n has a handle decomposition with no i-handles for $i > n - 2$. Assume $\pi_1(W) \cong \pi_1(\partial_+ W)$, $n \geq 6$. Let X be a CW complex with no i-cells for $i > n - 2$ and $f : X \to W$ a simple homotopy equivalence. Then W has a handle decomposition with cells corresponding to those of X.*

There is also a relative version.

5.8 Notes on Chapter 5

§5.1 Although this decomposition has its roots in the nineteenth century, and a version was used by Poincaré, the modern version is essentially due to Morse [108]; however the accurate formulation first appeared in work of Smale [138] and Wallace [171].

§5.3 The Poincaré duality theorem has its origins in work of Poincaré, though in his time homology groups had yet to be invented, so the result obtained was an equality of Betti numbers $\beta_r = \beta_{n-r}$. The Morse inequalities 5.3.3 are due to Morse, who in [109] applied the existence theorem to obtain results on the homology. See [98, I] for a similar account. The extension of duality to manifolds with boundary is due to Lefschetz.

§5.2, §5.4, §5.5 Apparently h-cobordism was first defined by Thom.

This development in these sections is due to Smale [138], [139]: the first paper proved the Generalised Poincaré Conjecture, the second went on to the h-cobordism theorem. Smale had been working on dynamical systems, and was seeking to simplify them.

The preprint version had an error (which was soon corrected but annoying) in the treatment of the fundamental group; in the above account we have bypassed the difficulty by using the handle replacement technique (Proposition 5.5.1).

Another account of the proof of the h-cobordism theorem (in terms of functions rather than handles) is given in the little book [101] by Milnor.

§5.6 We have included examples to illustrate that the h-cobordism theorem is an effective tool for obtaining classification results up to diffeomorphism. These are taken from the author's papers [159] and [160].

§5.7 In the lecture notes from which this book originated, I was at pains to obtain results in maximum generality, and in particular, to remove all restrictions on the fundamental group. Here I have tried to supply enough to give the interested reader a taste of what is involved.