MODIFICATIONS AND COBOUNDING MANIFOLDS

ANDREW H. WALLACE

Introduction. The object of this paper is to establish a simple connection between Thom's theory of cobounding manifolds and the theory of modifications. The former theory is given in detail in **(8)** and sketched in **(3)**, while the latter is worked out in **(1)**. In particular in **(1)** it is shown that the only modifications which can transform one differentiable manifold into another are what I call below spherical modifications, which consist in taking out a sphere from the given manifold and replacing it by another. The main result is that manifolds cobound if and only if each is obtainable from the other by a finite sequence of spherical modifications.

The technique consists in approximating the manifolds by pieces of algebraic varieties. Thus if M_1 and M_2 form the boundary of M, the last is taken to be part of an algebraic variety such that M_1 and M_2 are two members of a pencil of hyperplane sections. If this pencil is properly chosen it will cut only finitely many singular sections on M, each of which will correspond to a spherical modification. The converse result is proved by a construction which seeks to bring about the situation just sketched. These results are proved in the first three sections.

The situation described here is essentially the same as arises in the study of critical values of a function on a manifold. Thus if M is embedded in N-space, each modification on the way from M_1 to M_2 corresponds to a critical value of x_N. The main result of § 4 is to show that the embedding can be done in such a way that, as x_N increases from its value on M_1 to its value on M_2, the type numbers of these critical points **(7**, p. 21**)** do not decrease. Whether the theory of critical points could be used more extensively in the present connection is not quite clear. One factor arising here (as for example in § 5) is that M_1 and M_2 are the main objects of interest usually, and the M which they cobound may be altered in some way, whereas the application of critical point theory would require that M should not be changed but should be treated as the underlying space. At any rate so far any application of, say, the Morse inequalities **(7**, p. 85**)** has yielded only trivial results.

Section 5 shows how the same effect may be brought about sometimes by modifications of different types, and the result is applied to give a solution of a problem of Bing **(2)** on the structure of 3-manifolds.

In § 6 it is shown that any differentiable manifold of dimension not less

Received January 19, 1959.

than 3 cobounds a simply connected manifold, while in § 7 a few results are given extending this to higher homology and homotopy groups.

1. Spherical modifications. Throughout this discussion E^n and S^n will denote an n-dimensional cell and an n-dimensional sphere, respectively, subscripts being used where necessary to distinguish between different copies of these sets.

Let M be a differentiable manifold of dimension n, and let S^m be an m-sphere homeomorphically and differentiably embedded in M. It is known that a sufficiently small neighbourhood B of S^m in M can be fibred by $(n-m)$-dimensional cells; B is then the normal bundle of S^m in M. If B can be expressed as the topological product $S^m \times E^{n-m}$, S^m will be said to be directly embedded in M. In this case the frontier of B, or what is the same thing, the frontier of $M-B$ is of the form $S^m \times S^{n-m-1}$. The last product can, however, be identified with the frontier of a product of the type $E^{m+1} \times S^{n-m-1}$. It follows at once that the union of $M-B$ and $E^{m+1} \times S^{n-m-1}$, corresponding points on the frontiers of these sets being identified, can be made into a differentiable manifold M'. The transition from M to M' is a modification **(1)**. Modifications constructed in this particular way from directly embedded spheres will be called spherical modifications. To draw attention to the dimensions involved, the modification from M to M' described above will be called a modification of type $(m, n-m-1)$; it can easily be seen that the inverse operation, going from M' to M, is a spherical modification of type $(n-m-1, m)$. It will also sometimes be convenient to describe the modification from M to M' as a modification which shrinks S^m and introduces S^{n-m-1}.

It is clear from the above description that the manifold M' contains a directly embedded sphere S^{n-m-1} and that $M - S^m$ and $M' - S^{n-m-1}$ are, in a natural way, homeomorphic. This homeomorphism will be said to be induced by the modification.

Still using the above notations, it is not hard to see that the result of a spherical modification does not depend essentially on the way in which B is fibred by cells transversal to S^m. This follows from the fact that every such fibring can be continuously deformed into a canonical fibring by cells made up from geodesic arcs normal to S^m, with respect to some Riemannian metric on M. Similarly, isotopic deformations of S^m will not affect the modifications. On the other hand the mode of expression of B as a product $S^m \times E^{n-m}$, equivalent to the choice of a system of cross-sections of B, may be an essential factor in determining the result of the modification. Thus it is not in general possible to speak of *the* modification shrinking S^m unless reference is also made to the way in which B is written as a product.

2. Cobounding manifolds. A differentiable manifold with boundary is a topological space M with a subspace M_1 such that (1) M_1 is a differentiable manifold; (2) each point of $M - M_1$ has a neighbourhood homeomorphic to

https://doi.org/10.4153/CJM-1960-045-7 Published online by Cambridge University Press

an n-cell (n the same for each point); (3) each point of M_1 has a neighbourhood in M homeomorphic to a solid n-dimensional hemisphere, the base of the hemisphere corresponding to the part of the neighbourhood on M_1; and (4) the transition functions between one neighbourhood and another of the types just described are differentiable. When M and M_1 are related in this way, M_1 will be said to be the boundary of M, and M_1 will be said to be a bounding manifold. In all this there is no need for the manifolds to be connected. Two differentiable manifolds M_1 and M_2 will be said to be cobounding if their union is a bounding manifold.

In the case of orientable manifolds the idea of bounding can be made a bit stronger. If M_1 is orientable it will be said to be an oriented bounding manifold if it is the boundary of an oriented manifold whose orientation induces a preassigned orientation of M_1. The set of all orientable manifolds is now taken as the set of generators of an additive abelian group. Each connected manifold is supposed to be given a preassigned orientation, and the minus sign denotes change of this orientation. The manifolds M_1 and M_2 are now said to be cobounding if $M_1 - M_2$ is an oriented bounding manifold.

From the algebraic point of view, the notion of cobounding introduced at the beginning of this section can be described as cobounding modulo 2.

The first main result to be proved is the following connection between the ideas of cobounding and of spherical modifications.

THEOREM 1. *Let M_1 and M_2 be two given compact differentiable manifolds, the question of orientation being for the moment ignored. Then M_1 and M_2 are cobounding if and only if each can be obtained from the other by a finite sequence of spherical modifications.*

Proof. The "if" part of the theorem will be established if it is shown that M_1 and M_2 cobound whenever one is obtained from the other by a single spherical modification, since the relation of cobounding is transitive. This will be proved now as part (a) of the proof, part (b) being the proof of the converse.

(a) Suppose then that M_1 is obtained from M_2 by a spherical modification of type $(r, n - r - 1)$, n being the dimension of the manifolds. Thus there are spheres S^r and S^{n-r-1} contained respectively in M_1 and M_2, with normal bundles $B_1 = S^r \times E^{n-r}$ and $B_2 = E^{r+1} \times S^{n-r-1}$ in these manifolds such that $M_1 - B_1$ and $M_2 - B_2$ are homeomorphic. Assume now that $M_1 - B_1$ and $M_2 - B_2$ are identified with $(M_1 - B_1) \times \{0\}$ and $(M_1 - B_1) \times \{1\}$, respectively, in the set $(M_1 - B_1) \times I$, where I is the unit interval $0 \leqslant t \leqslant 1$. Form the union $[(M_1 - B_1) \times I] \cup B_1 \cup B_2$, B_1 and B_2 being inserted where they belong in $(M_1 - B_1) \times \{0\}$ and $(M_1 - B_1) \times \{1\}$ according to the identification just made. The subset $B_1 \cup B_2 \cup (\mathrm{Fr} B_1 \times I)$ in the space so constructed is an n-sphere and so can be identified with the boundary of an $(n + 1)$-cell E^{n+1}. Adding E^{n+1} to $[(M_1 - B_1) \times I] \cup B_1 \cup B_2$ with suitable identifications on the boundaries, an $(n + 1)$-dimensional manifold M is obtained, and can easily be adjusted along the boundary of E^{n+1} so as to be differentiable. Moreover, it is clear that the boundary of M is the union of M_1 and M_2. Thus M_1 and M_2 are cobounding manifolds as was to be shown.

(b) The idea of the converse is as follows. Suppose M is a differentiable manifold with boundary, the boundary being the union of M_1 and M_2. It is to be shown that M_2 can be obtained from M_1 by a finite sequence of spherical modifications. To show this, M is first to be approximated by part of a real algebraic variety in N-space in such a way that M_1 and M_2 are parts of the sections by the hyperplanes $x_N = 0$ and $x_N = 1$, respectively. This can be done in such a way that the family of hyperplanes $x_N = c$, for $0 \leqslant c \leqslant 1$, cuts the approximation of M in non-singular sections with just a finite number of exceptions, on each of which there is exactly one singular point at which the tangent cone is a non-degenerate quadric cone. Then it will be shown that the transition from one side of a singular section to the other is locally the same as the transition from negative to positive values of t in a family of quadrics

$$\sum_{i=1}^{n} a_i x_i^2 = t$$

in n-space, and hence it will be verified that each such transition is carried out by means of a spherical modification.

The details of the proof just sketched will now be worked out. In the first place M is to be embedded in a Euclidean N-space E_N, which can be done if N is large enough. Also it is clear that the embedding can be done in such a way that M_1 and M_2 lie in the hyperplanes $x_N = 0$ and $x_N = 1$, respectively, while the rest of M lies entirely between these hyperplanes. The algebraic approximation mentioned above could be made already at this stage, but to ensure that the approximating variety will have no points near M except those which are actually approximating points of M it is convenient to carry out the following additional construction. Take second copies in E_N of M, M_1, M_2, respectively, namely M', M_1', M_2', and suppose that M_1' and M_2' lie in the hyperplanes $x_N = 0$ and $x_N = 1$, respectively, and that the rest of M' lies between these hyperplanes; also assume that $M \cap M' = \emptyset$. M' can be constructed in this way by a translation in E_N for example. In addition M and M' can be adjusted so that they cut the hyerplanes $x_N = 0$ and $x_N = 1$ orthogonally. By adding to $M \cup M'$ sets homeomorphic to $M_1 \times I$ and $M_2 \times I$, lying in the parts of E_N where $x_N \leqslant 0$ and $x_N \geqslant 1$, respectively, a compact differentiable manifold M'' can be constructed. M'' has the property that there is a neighbourhood U of M in E_N such that $U \cap M''$ is homeomorphic to M; in fact it is equal to M with, so to speak, a narrow fringe added along M_1 and M_2. Now it is known **(4; 9)** that there is a real algebraic variety V in E_N with an isolated sheet approximating M'' arbitrarily closely. This approximation is not only in the pointwise sense, but also the tangent linear varieties at corresponding points of M'' and V approximate one another arbitrarily closely **(4; 9)**. In particular it follows that M itself is approximated arbitrarily closely by the part of $V \cap U$ which lies between $x_N = 0$ and $x_N = 1$, while M_1 and M_2 are approximated by the intersections of $V \cap U$ with these hyperplanes.

At this stage it is convenient to make a change of notation, simply replacing M by its approximation. Thus from now on in this proof it will be assumed that M lies on a real algebraic variety V in E_N and that there is a neighbourhood U of M such that M is the part of $V \cap U$ lying between the hyperplanes $x_N = 0$ and $x_N = 1$, while M_1 and M_2 are the intersections of these hyperplanes with M.

Some properties of an algebraic variety in relation to a pencil of hyperplane sections are now to be applied to the present situation. In the first place, if V is an algebraic variety in real projective space and Π is a generic hyperplane pencil only a finite number of members of Π will contain the tangent linear variety at some simple point of V, and each of these will contain the tangent linear variety at exactly one point of V. In addition, each of these finitely many points of contact for members of Π is a generic point of V. This can all be proved as in **(10**, ch. I**)**. The fact that V may not be non-singular makes no essential difference to the technique of the dual variety used there. Now choose homogeneous co-ordinates $(x_1, x_2, \ldots, x_N, x_{N+1})$ in the space containing V such that $x_{N+1} = 1$ and the equations of the members of Π are of the form $x_N =$ constant, and also such that, if V is of dimension m, the projection of V into the linear subspace $x_{m+1} = x_{m+2} = \ldots = x_{N-1} = 0$ is one-one around a generic point. When this is done the equations of V (in affine form) will be

$$(1) \qquad \left.\begin{aligned} F(x_1, x_2, \ldots, x_m, x_N) &= 0 \\ x_i &= R_i(x_1, x_2, \ldots, x_m, x_N) \end{aligned}\right\}$$

where $i = m+1, m+2, \ldots, N-1$, F being a polynomial and the R_i rational functions with coefficients which are real when V is a real variety. Also, making a shift of origin to one of the points at which a member of Π contains the tangent linear space to V, and remembering that such a point is generic on V over the real numbers it turns out that equations (1) can be written in the form

$$(2) \qquad \begin{aligned} x_N &= f(x_1, x_2, \ldots, x_m) \\ x_i &= g_i(x_1, x_2, \ldots, x_m) \end{aligned}$$

where $i = m+1, m+2, \ldots, N-1$, and the functions f and the g_i are real analytic in a sufficiently small neighbourhood of the origin. Also, since the new origin started off as a generic point of V the power series expansion for f around that point is of the form

$$(3) \qquad f = \sum_{i,j=1}^{m} a_{ij} x_i x_j + \ldots$$

where the dots denote terms of order greater than two and the determinant

$|a_{ij}|$ is not zero. The linear terms are of course zero because the tangent linear variety to V at the origin is contained in $x_N = 0$.

Now in what has just been said the pencil Π is generic, that is to say, the coefficients of the linear equations defining the axis of Π are indeterminates over the real numbers. The conditions that the choice of Π and of co-ordinates as above should not give equations for V of the type (2) and (3) at each of the points where a member of Π contains the tangent linear variety is algebraic in these indeterminates. It follows that the coefficients of the equations of the axis of Π can be given real values in such a way that the equations of V can be brought into the form described above. A final point is that, since the pencils which are unfavourable lie in an algebraic family, then whatever co-ordinate system is given in the space containing V, a linear change of co-ordinates with a matrix whose elements are arbitrarily close to those of the identity matrix will yield a co-ordinate system in which the equations of V can be written in the manner just described.

The discussion just carried out is now to be applied to the variety V of dimension $n + 1$ introduced in the earlier part of this proof, namely the real variety containing the manifold with boundary M whose sections with $x_N = 0$ and $x_N = 1$ are the manifolds M_1 and M_2 respectively. Then a small displacement of the given co-ordinate system will give a system with the following properties. There is a neighbourhood U of M such that the intersection of $U \cap V$ with $x_N = c$ is non-singular for all except a finite set of values $c_1, c_2, \ldots, c_k$ of c; for each i, $x_N = c_i$ intersects $U \cap V$ in a section with exactly one singular point, say P_i; if P_i is taken as origin the equations of V can be written in the form (2) and (3) around P_i. Since V was approximately orthogonal to $x_N = 0$ and $x_N = 1$ at points of M_1 and M_2 in terms of the original co-ordinates, and since the displacement of co-ordinates is supposed to be small, it follows that the intersections of $x_N = 0$ and $x_N = 1$ with $U \cap V$ in the new co-ordinates are respectively homeomorphic to M_1 and M_2. Again it is convenient to change the notation and simply to say that these intersections are M_1 and M_2.

To complete the proof of the theorem it will be shown that the transition from the intersection of $U \cap V$ with $x_N = c_i - \epsilon$ to its intersection with $x_N = c_i + \epsilon$, for some small positive ϵ can be made by means of a spherical modification. To do this fix attention on one of the P_i and take it as origin. Then in a neighbourhood of the origin V will have equations of the type (2), with f of the form (3). With this new arrangement of the co-ordinates the section $M(c)$ of M by the hyperplane $x_N = c$, for sufficiently small c, will have equations in a neighbourhood of the origin of the form

$$\sum a_{ij}x_i x_j + \phi = c \tag{4}$$

where ϕ is a power series in the variables $x_1, x_2, \ldots, x_{n+1}$ of order not less than three, along with further equations which express $x_{n+2}, x_{n+3}, \ldots, x_{N-1}$ as analytic functions of $x_1, x_2, \ldots, x_{n+1}$. By a linear change of the variables

$x_1, x_2, \ldots, x_{n+1}$ the quadratic terms in (4) can be diagonalized. Assuming that this has been done, (4) will be of the form

$$\sum_{i=1}^{n+1} a_i x_i^2 + \phi = c. \tag{5}$$

Since ϕ contains only terms of degree greater than two, a theorem of Samuel **(5)** shows that, for sufficiently small values of the variables, an analytic change of co-ordinates from $x_1, x_2, \ldots, x_{n+1}$ to a new set $y_1, y_2, \ldots, y_{n+1}$ can be made by formulae of the type $x_i = y_i + h_i(y)$, where the h_i are power series of order not less than two, in such a way that

$$\sum a_i x_i^2 + \phi = \sum a_i y_i^2.$$

By orthogonal projection from $(x_1, x_2, \ldots, x_N)$-space into $(x_1, x_2, \ldots, x_{n+1})$-space followed by a change to the y-co-ordinates it is then clear that a neighbourhood of the origin on V, that is to say on M, can be mapped analytically and homeomorphically on a neighbourhood of the origin in $(y_1, y_2, \ldots, y_{n+1})$-space, and the parts of the $M(c)$ near the origin in N-space will be mapped into the family of quadrics $Q(c)$, or at least the parts of these quadrics near the origin, in $(y_1, y_2, \ldots, y_{n+1})$-space, where $Q(c)$ has the equation

$$\sum_{i=1}^{n+1} a_i y_i^2 = c. \tag{6}$$

Now it can be explicitly verified that if $r+1$ of the a_i in (6) are positive and the rest negative (none are zero) and if c_0 is positive then the transition from $Q(c_0)$ to $Q(-c_0)$ can be made by a spherical modification of type $(r, n-r+1)$. In addition the homeomorphism induced by this modification can be constructed in a particular way. Namely, if small neighbourhoods, more precisely normal bundles, of the spheres

$$\sum_{i=1}^{r+1} a_i y_i^2 = c_0, \; y_j = 0 \qquad (j \geqslant r+2)$$

on $Q(c_0)$ and

$$\sum_{i=r+2}^{n+1} a_i y_i^2 = -c_0, \; y_j = 0 \qquad (j \leqslant r+1)$$

on $Q(-c_0)$ are removed (here it is assumed that $a_1, a_2, \ldots, a_{r+1}$ are the positive a_i) then the corresponding points on the remaining sets of $Q(c_0)$ and $Q(-c_0)$ are joined to each other by members of the family F of orthogonal trajectories to the family of $Q(c)$.

Returning to the variety V and more specifically to M, it has already been seen that $y_1, y_2, \ldots, y_{n+1}$ can be taken as a set of local analytic co-ordinates on M around the origin. Also the ordinary Euclidean metric in $(y_1, y_2, \ldots, y_{n+1})$-space induces a Riemannian metric on M in a neighbourhood of the origin. By means of a partition of unity a Riemannian metric can be set up on the whole of M so as to agree with this induced metric in a sufficiently small neighbourhood of the origin on M. Then the image on M of the family F of orthogonal trajectories to the $Q(c)$ can be extended to the family F' of orthogonal trajectories to the family of sections $M(c)$ of M, at least in a neighbourhood of $M(0)$. It is thus clear that, for c_0 sufficiently small and positive, if the images on M of the spheres on $Q(c_0)$ and $Q(-c_0)$ mentioned above are removed, then the remaining sets on $M(c_0)$ and $M(-c_0)$ are homeomorphic, corresponding points being joined by members of the family F'. Apart from this the spherical modification carrying $Q(c_0)$ into $Q(-c_0)$, in so far as it affects points near the origin, is carried into a similar modification taking $M(c_0)$ into $M(-c_0)$. And this completes the proof of the theorem.

It is possible to give part (a) of the above theorem a more precise form. Namely, if M_2 is obtained from M_1 by a single spherical modification, then the manifold M can be constructed in such a way that M_1 and M_2 belong to a pencil of hyperplane sections of M containing exactly one singular section. In other words the given modification can be made to arise in the same way as the modifications shown to exist in part (b) of the theorem. To prove this, the cell E^{n+1} which appeared in the course of the proof of part (a) must be constructed in a special way. For values of t such that $-1 \leqslant t \leqslant 1$, let $Q(t)$ be the quadric hypersurface $x_1^2 + x_2^2 + \ldots + x_{r+1}^2 - x_{r+2}^2 - \ldots - x_{n+1}^2 = t$ in $(n+1)$-space. The section of $Q(1)$ by the linear space $x_{r+2} = x_{r+3} = \ldots = x_{n+1} = 0$ is an r-sphere S^r whose normal bundle of some convenient radius in $Q(1)$ is a set B_1' homeomorphic to $S^r \times E^{n-r}$, and so to B_1 (in the notation of part (a) of the above theorem). Construct the family of orthogonal trajectories F to the family $Q(t)$. Then the set of points on curves of F meeting $Q(1)$ at points of B_1' is an $(n+1)$-cell $E'^{(n+1)}$. It is clear that, apart from the curves of F starting at points of S^r, all of which end at the origin, all the members of F starting at points of B_1' reach $Q(-1)$ at points in the normal bundle B_2' of the sphere S^{n-r-1} in which $Q(-1)$ is cut by the linear space $x_1 = x_2 = \ldots = x_{r+1} = 0$, and similarly the other way round. B_2' is homeomorphic to $S^{n-r-1} \times E^{r+1}$, that is to say, to B_2. Now, referring to the proof of part (a) of the above theorem, it will be seen that the frontier of E^{n+1} first appeared as the frontier of $(M_1 - B_1) \times I$ with the sets B_1 and B_2 added in the appropriate way, M_1 and M_2 being identified with $(M_1 - B_1) \times \{0\} \cup B_1$ and $(M_1 - B_1) \times \{1\} \cup B_2$, respectively. The frontiers of E^{n+1} and $E'^{(n+1)}$ are now to be identified. To do this define a mapping f of the frontier of $E'^{(n+1)}$ onto that of E^{n+1} as follows: first f is to be defined as a homeomorphism of B_1' onto B_1 preserving the product structure. Then if (p, t) is the point of parameter t (that is, the point lying on $Q(t)$) on the curve of the family F which passes through p on B_1', $f(p, t)$ will be defined as the point $(f(p), \frac{1}{2} - \frac{1}{2}t)$ in $\mathrm{Fr}B_1 \times I$ (this makes sense as $f(p)$ is already defined). In particular f is now defined as a homeomorphism of $\mathrm{Fr}B_2'$ onto $\mathrm{Fr}B_2$, preserving the product structure, and so it can be extended over the whole of B_2', carrying this set homeomorphically onto B_2. f is now defined on the whole of $\mathrm{Fr}E'^{(n+1)}$, and so can be extended to a homeomorphism of $E'^{(n+1)}$ onto $E^{(n+1)}$.

Using the mapping f just defined, the family $M(t)$ of sets will now be defined. For each t such that $-1 \leqslant t \leqslant 1$ set

$$M(t) = f(Q(t) \cap E'^{(n+1)}) \cup (M_1 - B_1) \times \{s\}$$

where $s = \frac{1}{2} - \frac{1}{2}t$. Then, for each $t \neq 0$, $M(t)$ is a manifold, and $M(0)$ has a single isolated singular point corresponding to the vertex of the cone $Q(0)$. In particular $M(1) = M_1$ and $M(-1) = M_2$.

If the family $M(t)$ is in $(x_1, x_2, \ldots, x_{N-1})$-space, then the manifold M of part (a) of Theorem 1 can be constructed in $(x_1, x_2, \ldots, x_N)$-space by taking it as the set whose intersection with $x_N = t$ is $M(t)$. As the construction has been done here, M and the $M(t)$ may not be differentiable, but they can clearly be arranged to be so by taking suitable precautions when the boundary of E^{n+1} and that of $(M_1 - B_1) \cup B_1 \cup B_2$ are identified, and when the mapping f is extended into the interior of $E'^{(n+1)}$.

A further point to notice is the existence of a family F of curves on M consisting of the image under f of the orthogonal trajectories to the $Q(t)$ lying in $E'^{(n+1)}$ along with all the curves of the form $\{p\} \times I$ for p in $M_1 - B_1$. These curves have the following properties:

(1) Exactly one of them passes through each point of M different from P, the image under f of the origin in $(x_1, x_2, \ldots, x_{n+1})$-space.

(2) The curves starting on S^r in M_1 all end at P; so also do those which start at points of S^{n-r-1} in M_2.

(3) The set of points on the members of F starting on S^r is an $(r + 1)$-cell E^{r+1} in M. Thus E^{r+1} is an $(r + 1)$-cell in M with boundary S^r on M_1. Similarly there is an $(n - r)$-cell E^{n-r} in M with boundary S^{n-r-1} on M_2.

Suppose that, in addition to the modification ϕ carrying M_1 into M_2, a second modification ϕ' is applied to M_2, taking it into M_3, and suppose that a manifold M' having M_2 and M_3 as its boundary and containing a family F' of curves with properties similar to (1), (2), and (3) above has been constructed in the manner just described for M and F. Then M and M' can be joined together along M_2, and if suitable precautions are taken the result will be a differentiable manifold. Also the families F and F' can be combined, each curve of F being joined to the curve of F' starting at its end point on M_2. Now it has been remarked that a displacement of the sphere shrunk in a modification does not affect the result, and so if ϕ' is of type $(s, n - s - 1)$ with $s \leqslant r$ it can always be arranged that the S^s shrunk by ϕ' does not meet the S^{n-r-1} introduced by ϕ. It follows that the curves of F' starting on S^{n-r-1} can be added to E^{n-r} to give a larger $(n - r)$-cell in $M \cup M'$ with its boundary in M_3. A similar remark can be made concerning any sequence of modifications of suitable types.

It should be remarked here that, in the proof of part (b) of the above theorem, there is an extreme case which may occur, corresponding fo the values -1 or n for r. This arises when a section $x_N = c$ of M has a singularity which is an isolated point. Although, strictly speaking this should be allowed as a modification with the appropriate alteration to the statement of Theorem 1, it will turn out (cf. § 4, Theorem 4) that these extreme cases can be avoided by suitably transforming the manifold M.

3. The oriented case. For the present purpose the most convenient way of fixing the orientation of a connected orientable differentiable manifold is by means of sets of local co-ordinates. Namely, having fixed a co-ordinate system in a neighbourhood U, a second system in U will be called positively or negatively oriented according as the Jacobian of the co-ordinate transformation is positive or negative. For a connected orientable manifold there is a covering by co-ordinate neighbourhoods with co-ordinates chosen so that, in the overlap of any two of the neighbourhoods the Jacobian of the corresponding co-ordinate transformation is positive. If the restriction to U of any one of these co-ordinate systems is positively oriented then the whole collection of local co-ordinate systems defines on the manifold the orientation induced by the fixed system in U.

The following lemmas prepare the way for the main result of this section.

LEMMA 3.1. *Let M be a connected orientable differentiable manifold in Euclidean N-space, and let H be a hyperplane such that $H \cap M$ is a connected differentiable manifold. Then $H \cap M$ is orientable.*

Proof. Local co-ordinates can be taken on M in a neighbourhood U of a point of $H \cap M$ in such a way that, if the Euclidean co-ordinates have been arranged so that H has the equation $x_N = 0$, then x_N is one of the local co-ordinates. It is clear then that x_N can be included among the local co-ordinates around every point of $H \cap M$, and so the orientation induced on M by the selected co-ordinate system in U automatically defines an orientation on $H \cap M$, which is therefore orientable.

COROLLARY. *If M is an orientable differentiable manifold with a connected boundary which is also a differentiable manifold, then this boundary is also orientable.*

Proof. For the given manifold can be so arranged that the boundary is a hyperplane section.

In the above lemma it should be noted (and this observation also applies to the corollary) that, if $H \cap M$ is not connected, orientability holds for each of the connected components separately.

LEMMA 3.2. *Let M and M' be connected orientable differentiable manifolds having a common boundary which is a connected differentiable manifold M_1. Then $M \cup M'$ is orientable.*

Proof. Embed M and M' in N-space so that M is in the set $x_N \leqslant 0$ and M' in the set $x_N \geqslant 0$, M_1 thus being the section of $M \cup M'$ by $x_N = 0$. It is then easy to see that local co-ordinates in $M \cup M'$ can be chosen around each point of M_1 so that x_N is always included as one of the co-ordinates, while the rest of $M \cup M'$ can be covered by co-ordinate neighbourhoods in M and M' separately. Since M and M' are orientable and M_1 is connected it follows at once that the co-ordinates can be chosen in each of these neighbourhoods so that an orientation is defined on $M \cup M'$ as required.

LEMMA 3.3. *Let M_1 be a connected orientable differentiable n-manifold, and let M_2 be obtained from M_1 by a spherical modification of type $(r, n - r - 1)$ with r not equal to 0 or $n - 1$. Then M_2 is orientable.*

Proof. Suppose that the modification in question shrinks the sphere S^r with normal bundle B_1 in M_1. Then $M_1 - B_1$ isan oriented manifold with a connected boundary. Also B_2 (the set to be added to $M_1 - B_1$ in the modification) is oriented with the same connected boundary. Then by Lemma 3.2 $M_2 = (M_1 - B_1) \cup B_2$ is orientable.

The condition on r in the last lemma cannot be dropped. For it is possible for a $(0, n - 1)$- or $(n - 1, 0)$-modification to change the orientability or otherwise of a manifold, as, for example, in the case of a (0,1)-modification applied to the surface of a sphere to make it into a Klein surface. Of course there are two ways in which a (0, 1)-modification can be applied to a sphere, the one giving a torus and the other a Klein surface. A similar situation holds in general. For the effect of a $(0, n - 1)$-modification on a manifold M_1 is to remove two disjoint n-cells from M_1 (namely the normal bundle of the S^0 to be shrunk) and to identify the points of the two $(n - 1)$-spheres which are their boundaries. Clearly there are essentially two different ways of making this identification, and if M_1 is orientable one of these ways will give an orientable M_2 and the other a non-orientable one. If the $(0, n - 1)$-modification carries an orientable manifold into another orientable manifold, then the modification itself will be said to be orientable.

The following theorem now gives the necessary complement to Theorem 1 for the case of orientable manifolds.

THEOREM 2. *Let M_1 and M_2 be two orientable differentiable manifolds. Then, with suitable orientations of their connected components, they cobound in the oriented sense if and only if they are related by a finite sequence of spherical modifications of which each modification of type $(0, n - 1)$ or $(n - 1, 0)$ is orientable.*

Proof. If M_1 and M_2 cobound in the oriented sense, then, by definition, their union constitutes the boundary of an orientable manifold M, and the orientations of the various components of M_1 and M_2 are supposed to be those induced by some selected orientation of M. As in Theorem 1, M is to be taken as part of a real algebraic variety in N-space such that M_1 and M_2 are the sections of M by the hyperplanes $x_N = 0$ and $x_N = 1$, while the rest of M lies between these hyperplanes. Also just a finite number of the hyperplanes $x_N = c$ are to cut M in singular sections, each with exactly one singular

https://doi.org/10.4153/CJM-1960-045-7 Published online by Cambridge University Press

point as in Theorem 1. By Lemma 3.1 and the remark following it, each hyperplane $x_N = c$, except those cutting singular sections, cuts M in a differentiable manifold whose components are orientable,with orientations induced by that of M. It follows at once, by considering sections on either side of a singular section corresponding to a $(0, n - 1)$- or $(n - 1, 0)$-modification that each such modification must be orientable (noting that this terminology makes sense whether the modification affects one component only or has the effect of joining two components together, for these components all have well defined orientations). This completes the proof in one direction.

To prove the converse, let M_2 be obtained from M_1 by a sequence of spherical modifications in which each of type $(0, n - 1)$ or $(n - 1, 0)$ is orientable. Here it is assumed that the components of M_1 are given preassigned orientations. Then Lemma 3.3 along with the assumed orientability of the $(0, n - 1)$- and $(n - 1, 0)$-modifications ensures that, as each modification is performed, the result is orientable with a naturally induced orientation on each component. The final result is supposed to be M_2 with suitable orientations on its components. The object now is to show that M, constructed as in Theorem 1, part (a), is orientable, and that it can be oriented in such a way that the correct orientations are induced on the components of M_1 and M_2. Clearly it is sufficient to carry out the proof in the case where M_1 and M_2 are related by one spherical modification.

Consider then the construction of M in the proof of Theorem 1, part (a). If the modification in question is of type $(r, n - r - 1)$ with r not 0 or $n - 1$ it may as well be assumed that M_1 is connected, since such a modification will affect just one component. Then, in the notation of part (a) of Theorem 1, $(M_1 - B_1) \times I$ is orientable and it is not hard to see that its frontier along with B_1 and B_2 will make up an oriented S^n, the orientation induced by that of $(M_1 - B_1) \times I$. It follows at once that when the cell E^{n+1} is added to form M the latter will be orientable and its orientation will induce that of M_1 and M_2. In the case of a $(0, n - 1)$-modification, assumed orientable, this assumption turns out to be exactly what is wanted to ensure that $(\mathrm{Fr}B_1 \times I) \cup B_1 \cup B_2$ will be an oriented n-sphere, M_1 and M_2 having been suitably oriented. Then as before the addition of an $(n + 1)$-cell gives an orientable manifold as required.

4. Rearrangement of modifications. In general there is no guarantee that the members of a sequence of modifications can be commuted among themselves, for the spheres introduced by the earlier modifications may intersect those to be shrunk in the later ones and it may be impossible to disentangle them. There are, however, certain ways in which the order of a sequence of modifications can be changed, and these will be examined in this section.

THEOREM 3. *Let M_1 and M_2 be n-dimensional differential manifolds related by a sequence of spherical modifications of types $(n - p - 1, p)$ for various values of p not less than r. Then the order of these modifications can be changed*

https://doi.org/10.4153/CJM-1960-045-7 Published online by Cambridge University Press

in such a way that all the $(n - r - 1, r)$*-modifications are done last* (M_1 *being counted as the initial state*).

Proof. The assumption on p is vacuous if r is zero, but otherwise the proof in this case is the same. The situation of § 2 will be assumed to hold here, in particular as described in the remarks at the end of the section following the proof of Theorem 1. Namely, M_1 and M_2 will be assumed to form the boundary of a differentiable manifold M in N-space, and in fact to be the sections of M by the hyperplanes $x_N = 0$ and 1, the rest of M lying between these. And among the sections of M by the family $x_N = c$ there are to be finitely many with a singularity, each corresponding to a spherical modification. It will also be assumed for the moment that none of these singular sections has an isolated point corresponding to an n-sphere which shrinks to a point and vanishes as the section $x_N = c$ varies from $c = 0$ to $c = 1$. This restriction will be removed later (cf. Theorem 4). The section of M by $x_N = t$ is to be denoted by $M(t)$, and as in § 2 there is to be a family F of curves in M cutting across the non-singular $M(t)$ transversally.

Starting from M_1 let ϕ be the first modification of type $(n - r - 1, r)$, corresponding to a section $M(c)$ of M with a singularity at the point P. Then, as remarked in § 2, it can be assumed that the spheres shrunk in later modifications do not meet the members of F which meet the r-sphere introduced by ϕ, since all other modifications are of type $(n - p - 1, p)$ with $p \geqslant r$. The points of all the curves of F starting at P and lying in the part of M for which $x_N \geqslant c$ form an $(r + 1)$-cell E^{r+1} with boundary S^r contained in M_2. The idea of this proof is to deform the family $M(t)$ in a neighbourhood of E^{r+1}, so obtaining a new family of submanifolds, some with singularities. M is then to be deformed so that this new family becomes a pencil of hyperplane sections, a finite number being singular. These singular sections will correspond to a sequence of spherical modifications leading from M_1 to M_2, and it will turn out that the modifications are all the same as those in the given sequence, but that ϕ now appears last.

The details of the idea just sketched will now be filled in. There is a neighbourhood U of P on M which is the homeomorphic image, under a mapping f, of a neighbourhood of the origin on the quadric Q in $(n + 2)$-space with the equation

$$z = y_1^2 + y_2^2 + \ldots + y_{r+1}^2 - y_{r+2}^2 - \ldots - y_{n+1}^2.$$

By means of this mapping the section $M(c + t)$ of M is locally identified with the section $Q(t)$ of Q given by $z = t$ (cf. the end of § 2, with the appropriate changes of notation). Also under this homeomorphism f the sphere introduced by the modification ϕ is the image of the sphere on Q given by $y_1^2 + y_2^2 + \ldots + y_{r+1}^2 = z, y_{r+2} = 0, \ldots , y_{n+1} = 0$, for some sufficiently small $z > 0$, and the family F restricted to U is the image of the family of orthogonal trajectories to the $Q(t)$ in a neighbourhood of the origin.

The next step is to construct a neighbourhood in M of the set E^{r+1} in a rather special way. First, in the neighbourhood U, take the smaller neighbourhood $f(U_0)$, image under f of the set in Q defined by the inequalities

$$|z| \leqslant \epsilon, y_{r+2}^2 + y_{r+3}^2 + \ldots + y_{n+1}^2 \leqslant \delta$$

for sufficiently small positive ϵ and δ. It is not hard to see that U_0 is an $(n + 1)$-cell with boundary consisting of the following three sets:

(1) The part of $z = - \epsilon$ on Q such that

$$\sum_{r+2}^{n+1} y_i^2 \leqslant \delta.$$

(2) The set $|z| \leqslant \epsilon$ satisfying

$$\sum_{r+2}^{n+1} y_i^2 = \delta.$$

This is homeomorphic to $S^r \times S^{n-r-1} \times I$.

(3) The part of $z = \epsilon$ on Q with

$$\sum_{r+2}^{n+1} y_i^2 \leqslant \delta.$$

This is homeomorphic to $S^r \times E^{n-r}$.

The image of the set (3) under f is a neighbourhood B_0 of S_0^r in $M(c + \epsilon)$, S_0^r being the sphere introduced by ϕ. If B_0 is small enough all the curves of F meeting it can be continued up to M_2; let B_1 be the set of points on all these curves. Then define B as the union of B_1 and $f(U_0)$. Clearly B is a neighbourhood in M of E^{r+1} and is an $(n + 1)$-cell with boundary consisting of the sets:

(1)′ The image under f of the set (1) above.

(2)′ The union of the image under f of (2) above with the set of points on curves of F meeting $\text{Fr}B_0$ on $M(c + \epsilon)$.

(3)′ $B \cap M$.

Note that the set (2)′, like (2), is homeomorphic to $S^r \times S^{n-r-1} \times I$. In (2) I is identified with the interval $|z| \leqslant \epsilon$, and in (2)′ with the interval $c - \epsilon \leqslant t \leqslant 1$, t being the parameter specifying the sections $M(t)$.

The switching of the order of modifications so that ϕ comes last is carried out by constructing a new mapping g of U_0 in Q into M, this time mapping it onto the whole of B. This mapping will be defined by identifying the sets (1), (2), and (3) on the frontier of U with the sets (1)′, (2)′, and (3)′ on the frontier of B, and then extending into the interiors of these sets.

The mapping $g : \text{Fr}\,U_0 \to \text{Fr}B$ is defined as follows:

(a) The restriction of g to the set (1) is to coincide with f.

(b) g is to map (2) onto (2)′. It has been noted that both sets are homeomorphic, and in a natural way, to $S^r \times S^{n-r-1} \times I$. g will be defined by giving a homeomorphism h of the interval I in (2), namely $-\epsilon \leqslant z \leqslant \epsilon$, and the interval I in (2)′, namely $c - \epsilon \leqslant t \leqslant 1$. h is to be defined in such a way that the interval $-\epsilon \leqslant z \leqslant 0$ is mapped on the interval $c - \epsilon \leqslant t \leqslant 1 - \eta$, where η is chosen so that all the sections $M(t)$ with $t \geqslant 1 - \eta$ are homeomorphic to M_2. Apart from this condition h can be arbitrary.

(c) g as defined in (a) and (b) is to be extended in the obvious way to map the set (3) on the set (3)′.

Finally, since U_0 and B are $(n+1)$-cells, g can be extended into the interior of U_0 to give a homeomorphism of U_0 onto B.

To define a new sequence of modifications relating M_1 and M_2, construct a family $M'(t)$ of subsets of M as follows:

For $t \leqslant c - \epsilon$, $M'(t) = M(t)$.

For $c - \epsilon \leqslant t \leqslant 1 - \eta$, $M'(t)$ is the union of the part of $M(t)$ outside B with $g(Q(h^{-1}(t)) \cap U_0)$.

For $t \geqslant 1 - \eta$, $M'(t) = M(t)$.

The $M'(t)$ as so defined may not be differentiable but can be made so (apart from a finite number each of which will have one singularity) by a suitable adjustment, or by a suitable definition of g in the first place. Define M' to be the set in $(x_1, x_2, \ldots, x_{N+1})$-space such that $M(t)$ is the section by $x_{N+1} = t$ (M of course is supposed to be in $(x_1, x_2, \ldots, x_N)$-space). In particular $M'(0) = M_1$ and $M'(1) = M_2$, and so M_1 and M_2 cobound the new manifold M', which, incidentally, is clearly homeomorphic to M.

Consider now the set of modifications corresponding to the singular members of the family $M'(t)$. For $t < 1 - \eta$, the only singular $M'(t)$s are those corresponding to all the original modifications relating M_1 and M_2 except ϕ. $M'(1 - \eta)$ is a singular section of M' corresponding to a $(n - r - 1, r)$-modification ϕ'. And there are no further modifications.

ϕ' can be thought of as the modification ϕ shifted to the end of the sequence of modifications. To complete the proof of the theorem, each $(n - r - 1, r)$-modification is to be shifted to the end in this way, and this can be done in a finite number of steps as above.

There are a number of remarks and corollaries connected with the theorem just proved. In the first place it must be emphasized that M', as constructed in the course of the proof, is homeomorphic to M; this point is of importance in certain applications where the main object of interest is not the pair of manifolds M_1 and M_2 but the manifold M which they bound. Another point is that ϕ was taken as the first $(n - r - 1, r)$-modification starting from M_1. It is quite clear however that M_1 and M_2 could be replaced by two intermediate sections M_1' and M_2' of M, when the same method of proof would show that any $(n - r - 1, r)$-modification can be moved to any later stage in the sequence of modifications leading from M_1 to M_2.

An essential result which must now be obtained is the possibility of removing the restriction imposed in Theorem 1, that no section of M by a hyperplane $x_N = c$ should have an isolated point.

https://doi.org/10.4153/CJM-1960-045-7 Published online by Cambridge University Press

THEOREM 4. *Let M_1 and M_2 cobound M, these manifolds being arranged as in Theorem 1 in Euclidean N-space, singular sections by hyperplanes $x_N = c$ corresponding to spherical modifications leading from M_1 to M_2. Then the embedding of M can be done in such a way that no section by a hyperplane $x_N = c$ has an isolated point.*

Proof. Proceeding from M_1 to M_2 let $M(c)$ (notation of Theorem 1) be the last section of M with an isolated point P corresponding to a vanishing sphere. That is to say $M(c)$ has the isolated point P and for small ϵ $M(c - \epsilon)$ has a small isolated sphere near P, while $M(c + \epsilon)$ has no points near P. Varying t from c downwards, the n-sphere introduced at P becomes joined to some other component of a section of M by a $(0, n - 1)$-modification (possibly after some modifications have been applied to the sphere itself). Let ϕ be the inverse of this $(0, n - 1)$-modification, corresponding to a singular section $M(c')$ of M, and then, for a sufficiently small ϵ, apply Theorem 3 to the part of M between $M(c' - \epsilon)$ and $M(c - \epsilon)$. The result is that it can be assumed that, in the sequence of modifications leading from M_1 to M_2, the last modification before the vanishing of the n-sphere at P is an $(n - 1, 0)$-modification which isolates that sphere. This modification will still be called ϕ, and the corresponding singular section of M will be $M(c')$.

For t near c' but less than it, $M(t)$ contains an $(n - 1)$-sphere $S^{n-1}(t)$ which is to be shrunk by the modification ϕ. The part of $M(t)$ on one side of $S^{n-1}(t)$ is an n-cell $E^n(t)$. As t tends to c', $E^n(t)$ closes up to form an n-sphere, and $M(t)$, for $c' \leqslant t < c$, contains this detached sphere $S^n(t)$ which shrinks to a point as t tends to c. It is clear that, for a sufficiently small positive ϵ, the union of all the $E^n(t)$ for $c' - \epsilon \leqslant t < c'$ and all the $S^n(t)$ for $c' \leqslant t \leqslant c$ is an $(n + 1)$-cell E^{n+1}, having on its boundary the n-cell E^n formed by the union of all the $S^{n-1}(t)$ for $c' - \epsilon \leqslant t \leqslant c'$ ($S^{n-1}(t)$ reduces to a point for $t = c'$). E^{n+1} is homeomorphic to a solid $(n + 1)$-dimensional hemisphere, E^n corresponding to the solid n-sphere forming the base, and so, corresponding to the fibring of the hemisphere by concentric n-dimensional hemispheres, E^{n+1} can be fibred by a family of n-cells $E_1^n(t)$ such that $S^{n-1}(t)$ is the frontier of $E_1^n(t)$.

Now define the family $M'(t)$ of subsets of M as follows:

$$M'(t) = M(t) \text{ for } t \leqslant c' - \epsilon;$$

$$M'(t) = (M(t) - E^n(T)) \cup E_1^n(t) \text{ for } c' - \epsilon \leqslant t < c';$$

$$M'(t) = M(t) - S^n(t) \text{ for } c' \leqslant t \leqslant c;$$

$$M'(t) = M(t) \text{ for } t > c.$$

Having done this, let M' be the set in $(N + 1)$-space such that $M'(t)$ is its section by the hyperplane $x_{N+1} = t$. It is clear that M' can be adjusted to become a differentiable manifold, and that M_1 and M_2 will form its boundary. The singular sections of M' by members of the pencil $x_{N+1} = c$ correspond to a sequence of spherical modifications leading from M_1 to M_2. These modifications are the same as the original ones (corresponding to the singular sections of M) with the exception that ϕ has now dropped out, and the section corresponding to the isolated point at P is no longer there. By means of a finite number of steps as just described, all singular sections with isolated points can be removed.

In connection with the proof of this theorem it should be noted that the manifold M' is homeomorphic to M.

The results of Theorems 3 and 4 can now be combined to give a stronger form of Theorem 1.

THEOREM 5. *Let the n-dimensional differentiable manifolds M_1 and M_2 form the boundary of the differentiable manifold M. Then M can be embedded in N-space, for sufficiently large N, as part of a real algebraic variety, M lying entirely between the hyperplanes $x_N = 0$ and $x_N = 1$. Only a finite number of sections by hyperplanes $x_N = c$ $(0 \leqslant c \leqslant 1)$ will have singular points, one point on each such section, and none of these singular points will be an isolated point of the section in question. Finally the embedding can be arranged in such a way that, in the sequence of modifications leading from M_1 to M_2, corresponding to the singular sections of M, all the $(r, n - r - 1)$-modifications come before the $(s, n - s - 1)$-modifications for each pair of integers r, s with $r < s$.*

Proof. The first part of the theorem is simply Theorem 1. The absence of isolated points on the singular sections of M can be brought about by Theorem 4, and the ordering of the modifications according to type can be done by repeated application of Theorem 3.

A further point to notice in connection with the last theorem is that the modifications of any one type can be rearranged freely among themselves. For consider the modifications of type $(r, n - r - 1)$ with $2r \leqslant n$ (this inequality imposes no restriction since in the contrary case one can look at the sequence of modifications the other way round, starting from M_2). Repeated application of Theorem 3 will rearrange these modifications in any preassigned way. The question of identifying the modifications as they are permuted is settled by noting that, since $2r \leqslant n$, there is a set of disjoint r-spheres each to be shrunk by one of the modifications, and the modifications can be named according to the sphere shrunk.

Theorem 5 is a generalization of a well-known result concerning orientable 3-manifolds. Let M be an orientable 3-manifold with boundary formed by M_1 and M_2, arranged as in Theorem 5; no generality is lost here since a 3-manifold can be triangulated and then smoothed to give a differentiable manifold. The only modifications leading from M_1 to M_2 as in Theorem 5 will be of types $(0, 1)$ and $(1, 0)$, all those of the former type being done first. If now M is a closed manifold, it can be assumed to be contained between the hyperplanes $x_N = 1 + \epsilon$ and $x_N = -\epsilon$, for small positive ϵ, while the sections of M by the hyperplanes $x_N = 1$ and $x_N = 0$ will be 2-spheres M_2

and M_1, boundaries of 3-cells E_2 and E_1 which lie respectively in the sets $1 \leqslant x_N \leqslant 1 + \epsilon$ and $-\epsilon \leqslant x_N \leqslant 0$. Theorem 5 then implies that M is obtained by applying $(0, 1)$-modifications to the surfaces of E_1 and E_2, filling the surfaces in as one goes to obtain two solids, whose boundaries are then identified. Since M is orientable, all the (0,1)-modifications are of orientable type (Theorem 2), and so the solids obtained are solid spheres with handles. That is to say the manifold M is constructed by taking the union of two solid handled spheres (necessarily of the same genus) and identifying their boundaries (**6**, p. 219).

Clearly Theorem 5 gives a similar way of constructing a non-orientable 3-manifold. In this case, however, at least one of the modifications applied to the surfaces M_1 and M_2 must be of non-orientable type. Thus the two solids which are to be put together to form M must each have at least one handle twisted (in the manner of the Klein surface).

To formulate Theorem 5 as a generalization of this classical result on 3-manifolds, a generalized handled sphere can be defined as an $(n + 1)$-dimensional solid obtained from a solid $(n + 1)$-sphere by applying to its surface $(r, n - r - 1)$-modifications, with $r \leqslant n - r - 1$, filling out the surface at each stage to form an $(n + 1)$-solid. Then Theorem 5 implies that any differentiable $(n + 1)$-manifold can be expressed as the union of two generalized handled spheres with boundaries identified. In particular if M is orientable, all the $(0, n - 1)$-modifications involved will be of orientable type.

5. Complementary modifications. Let M_1 be a differentiable n-manifold and let S^r be a directly embedded r-sphere to be shrunk by a spherical modification ϕ. Suppose also that S^r is the boundary of an $(r + 1)$-cell non-singularly and differentiably embedded in M_1. When $B_1 = S^r \times E^{n-r}$ is removed from M_1, the remaining set will contain an $(r + 1)$-cell E_1^{r+1} with boundary $S^r \times \{p\}$ for some $p \in S^{n-r-1} = \text{Fr}E^{n-r}$. When $B_2 = E_2^{r+1} \times S^{n-r-1}$ is added to make the modification ϕ, E_2^{r+1} joins up with E_1^{r+1} to form a sphere S^{r+1} in M_2. S^{r+1} is not necessarily directly embedded in M_2, but a sufficient condition for direct embedding is that the natural (the precise meaning of this overworked word in this context is explained below) product structure of the normal bundle of E^{r+1} in M_1 should induce the product structure on B_1 associated with the modification ϕ. If this condition is satisfied, a second modification ϕ' can be performed, shrinking S^{r+1} and transforming M_2 into a manifold M_3.

LEMMA 5.1. *Under the conditions just described M_1 and M_3 are homeomorphic.*

Proof. A normal neighbourhood (union of normal geodesic elements) of E^{r+1} in M_1 is an n-cell E_1^n, and it can be assumed that, in the modification ϕ carrying M_1 into M_2, the complement of E_1^n in M_1 is left unchanged. In the proof of this lemma, therefore, nothing is lost if $M_1 - E_1^n$ is replaced by a second n-cell E_2^n so that $E_1^n \cup E_2^n$ is an n-sphere. The modifications are to be carried out on this sphere in such a way that E_2^n is left unchanged, that is to say, so that a neighbourhood of some point is left unchanged.

At this stage the phrase used above, "natural product structure" in a neighbourhood of E^{r+1}, can be explained. The idea is that, when $M_1 - E_1^n$ is replaced by E_2^n to form the sphere $S^n = E_1^n \cup E_2^n$, and then when the neighbourhood B_1 of S^r is removed, the remainder of S^n will be a product $E_1^{r+1} \times S^{n-r-1}$ having the cell E_1^{r+1} as one of its cross-sections. The normal neighbourhood of E^{r+1} will then be the product $E^{r+1} \times U$, where U is a cellular neighbourhood on S^{n-r-1}, with B_1 added on.

The proof of the lemma will now be completed by performing the modifications ϕ and ϕ', related as described above, on the n-sphere S^n, and showing that the final result is again S^n.

S^n can be written as $B_1 \cup (E_1^{r+1} \times S^{n-r-1}) = (S^r \times E^{n-r}) \cup (E_1^{r+1} \times S^{n-r-1})$, where the boundaries of the two products are identified. A point (p, q) is to be selected in the interior of $(E_1^{r+1} \times S^{n-r-1})$, and it is to be checked that at each stage a neighbourhood of (p, q) is left invariant. The modification ϕ replaces B_1 by a product $E_2^{r+1} \times S^{n-r-1}$. Thus, with the boundaries of the products identified, $M_2 = (E_1^{r+1} \times S^{n-r-1}) \cup (E_2^{r+1} \times S^{n-r-1})$. It is clear that a neighbourhood of (p, q) has been left invariant here. Also the identification of the boundaries of the products is such that M_2 is homeomorphic to $S^{r+1} \times S^{n-r-1}$. Now S^{n-r-1} can be written as a union $E_1^{n-r-1} \cup E_2^{n-r-1}$ of two cells, with q in the interior of E_2^{n-r-1}. Thus, the boundaries of the products being identified, $M_2 = (S^{r+1} \times E_1^{n-r-1}) \cup (S^{r+1} \times E_2^{n-r-1})$, and the first product is the normal bundle of S^{r+1}. Thus ϕ' consists in replacing this product by $(E^{r+2} \times S^{n-r-2})$, and the result is S^n; also in the process a neighbourhood of (p, q) is left invariant, and so the proof is completed.

If the situation described in the above lemma holds, the modification ϕ' will be called complementary to ϕ.

One case in which this situation will always hold is where ϕ is a $(0, n-1)$-modification of orientable type. Thus if only the result of a sequence of modifications is of interest (and not the manifold bounded by the initial and final states) every orientable $(0, n-1)$-modification can be replaced by a $(n-2, 1)$-modification.

An important special case of this result is obtained by taking M_1 to be an orientable 3-dimensional manifold. According to Thom's theory of cobounding manifolds, M_1 is the boundary of an orientable 4-dimensional manifold. Hence, by Theorem 2, M_1 can be obtained from a 3-sphere M_2 by a sequence of $(0, 2)$-, $(1, 1)$-, and $(2, 0)$-modifications, those of types $(0, 2)$ and $(2,0)$ all being orientable. By the result just obtained, the modifications of types $(0, 2)$ and $(2, 0)$ can all be replaced by modifications of type $(1, 1)$. Translating into simple geometrical language the meaning of a $(1,1)$-modification, the following theorem is proved, giving an affirmative answer to a problem of Bing **(2)**:

THEOREM 6. *Any orientable 3-manifold can be obtained from a 3-sphere by removing a finite number of disjoint tori and refilling the resulting holes by tori with suitable identification of the boundary surfaces.*

6. Killing the fundamental group. The object of this section is to show that a manifold which is orientable and of dimension n can always be carried into a simply connected manifold by a finite sequence of spherical modifications of type $(1, n-2)$. This having been done, the next section will show how, under certain conditions, this process can be extended to one which will kill all the homotopy, or what in this context is the same thing, the homology groups up to the dimension $n-1$.

The results of this section will be obtained by comparing the fundamental groups of two orientable n-dimensional manifolds M_1 and M_2 which are related by a single spherical modification ϕ of type $(r, n-r-1)$ (necessarily orientable in case $r = 0$ or $n-1$). As in §2, M_1 and M_2 together will constitute the boundary of an $(n+1)$-dimensional manifold M which can be assumed to lie on an $(n+1)$-dimensional real algebraic variety in Euclidean N-space. It is convenient here to arrange the co-ordinates in such a way that M_1 and M_2 are, respectively, the sections of M by the hyperplanes $x_N = -1$ and $x_N = 1$, while $x_N = 0$ is the singular section of M corresponding to the modification leading from M_1 to M_2. The singular point P of this section can be taken as origin. As in § 2 there will be a family F of curves cutting transversally across the sections of M by the hyperplanes $x_N = c$, except at P. The members of the family F passing through P form two cells E^{r+1} and E^{n-r}, the former lying in the set $x_N \leqslant 0$ and having as its boundary the sphere S^r in M_1 shrunk by the modification ϕ, while the latter lies in $x_N \geqslant 0$ and has as its boundary the sphere S^{n-r-1} in M_2 introduced by ϕ.

The most convenient way of comparing the fundamental groups of M_1 and M_2 is to compare them both with that of M_0, the section of M by $x_N = 0$. This will be done by means of the two mappings $f_i : M_i \to M_0$ $(i = 1, 2)$ defined by setting $f_i(p)$ equal to the point on M_0 and on the curve of F through p. These are continuous mappings (**10**, ch. II), and so induce homomorphisms $f_{i*}: \pi_1(M_i) \to \pi_1(M_0)$ $(i = 1, 2)$. Here π_1 denotes the fundamental group, and in the meantime M_1 and M_2 will be assumed to be connected. The following lemma will now be proved.

LEMMA 6.1. (1) *For* $1 < r \leqslant n-1$, f_{1*} *is an isomorphism onto.*

(2) *For* $r = 1$, f_{1*} *is onto and its kernel is generated by the image of* $\pi_1(S')$ *in* $\pi_1(M_1)$ *induced by the inclusion mapping.*

(3) *For* $r = 0$, f_{1*} *is an isomorphism into.*

Proof. Let α be a closed path on M_1 beginning and ending at a base point p on S^r, and suppose that $f_1(\alpha)$ is homotopic to a constant on M_0 with respect to the fixed base point $P = f_1(p)$. It is clear then that α is homotopic to a constant on M with respect to the fixed base point p. That is to say, there is a continuous mapping h of a 2-cell E^2 into M such that the restriction of h to the circumference S^1 of E^2 coincides with α (S^1 is being identified with a line segment with ends joined; this is really a description of free homotopy, but for the present purpose no distinction need be made). Now h can be assumed to be an algebraic mapping. This is done by noting that, under the given h, co-ordinates in the ambient N-space are given as continuous functions of the co-ordinates in a 2-space containing E^2. Approximating these functions by polynomials and then projecting normally into M the required result is obtained **(4)**. At the same time h can be adjusted so that $h(E^2)$, now a piece of algebraic surface in M, bears a simple relation to E^{r+1} and E^{n-r}, which can themselves be assumed to be pieces of algebraic subvarieties of M. Namely, it can be assumed that, if $0 < r < n-1$, $h(E^2)$ meets $E^{r+1} \cup E^{n-r}$ in at most finitely many points, while if $r = 0$ or $n - 1$ the intersection may also include some arcs of algebraic curves. These cases will now be considered in more detail.

First take the case where $0 < r < n-1$. When the adjustments described above have been made it can be assumed that there is at most a finite set $P_1, P_2, \ldots, P_m$ of points in the interior of E^2 such that $h(P_i)$ is on E^{r+1} or E^{n-r}. If the adjustment to h is sufficiently small the new path α will of course be homotopic to the original one. It can also clearly be arranged that exactly one point q of the boundary S^1 of E^2 is mapped on p by h. Now let U be a small preassigned neighbourhood of P in M, and let W be the point-set union of all the curves of the family F which meet U. It is not hard to see that W is a neighbourhood of $E^{r+1} \cup E^{n-r}$ in M. Since h is continuous it follows that there are neighbourhoods $U_1, U_2, \ldots, U_m$ of $P_1, P_2, \ldots, P_m$ in E^2, which can in fact be assumed to be non-overlapping circular discs, such that for each i, $h(U_i) \subset W$. From q draw an arc β_i to some point on the circumference of U_i, for each i, arranging that the β_i do not meet each other except at q. Let β be the closed path on E^2 starting at q and going along β_i, round the circumference of U_i and back along β_i for each i in turn. This can be done so that β is homotopic on $E^2 - \cup U_i$ to the path which makes a single circuit of S^1. It then follows that $\alpha' = h(\beta)$ is a path on M homotopic in $M - E^{r+1} - E^{n-r}$ to α, with respect to the fixed base point p. In fact the deformation of α into α' is carried out in $M - W$, with possibly a small neighbourhood of p added on. But, making use of the family F of curves, it can be seen that $M_1 - (M_1 \cap W)$, along with a small neighbourhood of p, is a deformation retract of this set (cf. **(10)**, p. 17), and from this it follows that α is homotopic in M_1, with respect to the base point p, to the path $g(\alpha')$, where g maps a point t of $M - W$ on the end point, on M_1, of the curve of F through t. $g(\alpha')$ is a product of paths of the type $\gamma_i\alpha_i\gamma_i^{-1}$ where $\gamma_i = gh(\beta_i)$, and the α_i are closed paths in a small neighbourhood of S^r, a neighbourhood which can be assumed to be a product of S^r by a cell. Since $r > 0$, an easy transformation makes the γ_i into closed paths based on p.

Then if $r > 1$, the α_i are all homotopic to a constant on M_1 (in fact in a neighbourhood of S^r), and so in this case it has been shown that the kernel of f_{1*} is the identity. On the other hand, if $r = 1$, the α_i represent elements of the injection image of $\pi_1(S^1)$, as required in the statement of the lemma.

The kernel of f_{1*} must now be shown to be the identity in the cases $r = 0$ and $r = n - 1$. When $r = 0$, $h(E^2)$ can be assumed to meet E^{r+1}, a 1-cell, only at the point p, $h(S^1)$ will not meet E^{n-r}, but $h(E^2)$ may meet E^{n-r} in some curves. In this case, in addition to the points P_i appearing in the above discussion, there may be some algebraic curves in the interior of E^2 carried by h into E^{n-r}. Since h is continuous, it is in this case possible to find a finite number of simple closed loops C_i in E^2, each surrounding one or more of these curves, and each lying within such a small neighbourhood of these curves that $h(C_i) \subset W$ for each i. The P_i not already surrounded by the C_j are to be given neighbourhoods U_i as before, and the U_i and C_j are not to meet each other. The argument as above is then repeated, using the C_i along with the circumferences of the U_j.

The case $r = n - 1$ is a little more complicated. $h(E^2)$ will meet E^{n-r} at most in a finite number of points (and this need only happen if $n = 2$), but it may meet E^{r+1} in both isolated points and in pieces of algebraic curves, some of which may be arcs with end points on α. The inverse images of these arcs will be arcs of algebraic curves with end points on S^1. A preliminary adjustment will be made this time, deforming the mapping h in such a way that all these end points coincide with q. There are now in E^2 isolated points, isolated curves in the interior of E^2, and a set of curves forming a connected set containing q, all mapped into $E^{r+1} \cup E^{n-r}$ by h. The isolated points and curves in the interior of E^2 are to be treated as in the case $r = 0$, and the remaining curve is to be surrounded by a simple closed loop beginning and ending at q and lying in such a small neighbourhood of the curve that it is mapped by h into W. This loop is to be included in the product of paths forming β, and the rest of the argument is the same as before.

To complete the proof of the lemma it must be shown that f_{1*} is onto except in the case $r = 0$; it obviously will fail to be onto in this case. If then $r \neq 0$, let α be a closed path on M_0, and it is convenient this time to take as base point for closed paths a point Q different from P. α is then homotopic in M to a path α_1 not meeting E^{n-r}; this is possible since $r \neq 0$. Let α_2 and α_3 be the projections of α_1 on M_0 and M_1 respectively along the curves of F. The point $f_1^{-1}(Q)$ is well defined and will be taken as base point for closed paths on M_1. Clearly $\alpha_2 = f_1(\alpha_3)$. On the other hand, α_2 is homotopic in M, with respect to the base point Q, to α_1 and hence to α. But, using the curves of F, M_0 is a deformation retract of M (**10**, ch. I, § 4) and so α_2 and α are homotopic in M_0. Hence f_{1*} carries the homotopy class of α_3 in M_1 into that of α in M_0, and this shows f_{1*} to be onto for $r \neq 0$. This completes the proof of the lemma.

The case in which M_1 (or similarly M_2) is not connected is dealt with as follows:

LEMMA 6.2. *Continuing with the notation of the last lemma, let ϕ be a $(0, n-1)$-modification, let M_2 be connected but let M_1 consist of two connected components M_1' and M_1''. Then the fundamental group of M_0 is the free product of the images under f_{1*} of those of M_1' and M_1''.*

Proof. This is a well known result, but is also easy to derive in the manner of the last lemma.

Applying the above lemmas also to f_{2*} and putting the results together, the following theorem is at once obtained.

THEOREM 7. *Let M_1 and M_2 be two n-dimensional orientable differentiable manifolds related by a spherical modification ϕ of type $(r, n-r-1)$. Then*

(1) *if $1 < r < n-2$, $\pi_1(M_1)$ and $\pi_1(M_2)$ are isomorphic under $f_{2*}^{-1}f_{1*}$. (This can only happen if $n > 4$.)*

(2) *If $r = 1$ and $n > 3$, $f_{2*}^{-1}f_{1*}$ is a homomorphism of $\pi_1(M_1)$ onto $\pi_1(M_2)$ with kernel generated by the image of $\pi_1(S^1)$ induced by the inclusion of S^1 in M_1, S^1 being the 1-sphere shrunk by ϕ.*

(3) *If $r = 0$ and $n > 2$, and M_1 is connected, $f_{2*}^{-1}f_{1*}$ is an isomorphism into. If M_1 has two components, $\pi_1(M_2)$ is the free product of their fundamental groups.*

Complementary results to (2) and (3) can of course be obtained by taking $r = n-1$ or $n-2$. The condition $n > 2$ in (3) is no great obstacle, as modifications on a surface are rather a trivial matter. On the other hand the restriction $n > 3$ in (2) shows up one of the essential difficulties of the 3-dimensional case, where a modification which shrinks one circle simply has the effect of introducing another.

Suppose now that M_1 is a compact orientable differentiable manifold of dimension $n > 3$. $\pi_1(M_1)$ is a finitely generated group in this case, and the generators can be assumed to be carried by a finite collection of disjoint 1-spheres differentiably and, of course, directly embedded in M_1. Performing the modifications which shrink these 1-spheres, and using part (2) of the above theorem, we have the following theorem.

THEOREM 8. *An orientable compact differentiable manifold of dimension > 3 can be made simply connected by a finite sequence of $(1, n-2)$-modifications.*

Note that, according to Theorem 6 and the remarks preceding it, the condition $n > 3$ can be dropped. But there is no guarantee in the case of $n = 3$ that the modifications involved correspond to a systematic killing of the generators of the fundamental group.

7. Killing the homology groups. The aim of this section is to give a partial extension of the results of the last section to the homology and homotopy groups of dimension higher than the first. The idea is that a cycle carried by a directly embedded sphere can be annulled by the modification

https://doi.org/10.4153/CJM-1960-045-7 Published online by Cambridge University Press

which shrinks that sphere. But the condition imposed here on the cycle is a rather strong one, and so no sort of complete theory is possible until the situation has been analysed in much greater detail. The ideal result would be to achieve a complete "killing" by adding to the given manifold suitable auxiliary manifolds, namely representatives of the generators of the Thom cobounding groups but, in the meantime, a few of the simpler cases will be treated.

Let ϕ be a spherical modification of type $(r, n-r-1)$ carrying the differentiable manifold M_1 into M_2, shrinking the sphere $S^r \subset M_1$ and introducing $S^{n-r-1} \subset M_2$. Let B_1 and B_2 be the normal bundles of S^r and S^{n-r-1} in M_1 and M_2, both, of course, topological products. Using singular homology with integral coefficients, an application of the homotopy and excision theorems shows that $H_p(M_1, S^r) \cong H_p(M_1 - B_1, \mathrm{Fr}B_1)$, and $H_p(M_2, S^{n-r-1}) \cong H_p(M_2 - B_2, \mathrm{Fr}B_2)$, for all p. On the other hand ϕ induces a homeomorphism between $M_1 - B_1$ and $M_2 - B_2$, and so it follows that $H_p(M_1, S^r) \cong H_p(M_2, S^{n-r-1})$ for all p. The results to be obtained now depend on the examination of the following diagram, in which the horizontal lines are the appropriate homology sequences:

$$\tag{7} \begin{array}{ccccccccc} \to H_p(S^r) & \xrightarrow{i_p} & H_p(M_1) & \xrightarrow{j_p} & H_p(M_1, S^r) & \xrightarrow{\partial_p} & H_{p-1}(S^r) \to \\ & & & & \wr\| & & \\ \to H_p(S^{n-r-1}) & \xrightarrow[i'_p]{} & H_p(M_2) & \xrightarrow[j'_p]{} & H_p(M_2, S^{n-r-1}) & \xrightarrow[\partial'_p]{} & H_{p-1}(S^{n-r-1}) \to. \end{array}$$

The proofs of the following lemmas are immediate.

LEMMA 7.1. *In the above notation if* $2r < n - 1$ (*that is* $r < n-r-1$) *then* $H_p(M_1) \cong H_p(M_2)$ for $p < r$ *and* $H_r(M_2) \cong H_r(M_1)/i_r H_r(S^r)$.

Obviously there is a complementary result for $r > n - r - 1$, amounting simply to looking at ϕ as leading from M_2 to M_1. If in the lemma just proved S^r carries a representative of some generator of $H_r(M_1)$, then the lemma shows that the effect of ϕ is to annul that generator.

LEMMA 7.2. *If* $r + 1 < p < n - r - 1$, $H_p(M_1) \cong H_p(M_2)$, *and except when* n *is even and equal to* $2(r+1)$, $H_{r+1}(M_1)$ *can be identified with a subgroup of* $H_{r+1}(M_2)$ *and the quotient group is isomorphic to the kernel of* i_r.

In particular, this shows that, if the cycle carried by S^r is homologous to zero in M_1 or is a torsion cycle, the effect of the modification (with the exception noted) is to add another generator to the $(r+1)$st homology group. These two lemmas show two of the characteristic ways in which a modification can affect homology. Note that, if M_1 and M_2 are simply connected and $r > 1$ (in addition to the conditions already imposed on it) then the above results, by the Hurewicz isomorphism theorem, can be interpreted in terms of the homotopy groups provided that all the lower dimensional homology groups are already known to be zero.

The cases in which the condition $2r < n - 1$ fails will now be examined. This will have to be done separately in the two cases n odd and n even. First consider an odd value $2m + 1$ of the dimension; the case to be looked at then corresponds to the value m of r.

LEMMA 7.3. *In the situation just described, $H_p(M_2) \cong H_p(M_1)$ for all $p < m$. If the image of i_m is not of finite order in $H_m(M_1)$ then the effect of ϕ on M_1 is to reduce the* mth *Betti number by* 1, *but possibly to introduce a new torsion cycle.*

Proof. The first statement, concerning $p < m$ follows at once from the diagram (7). Next, if the image of i_m is not of finite order in $H_m(M_1)$, the fundamental cycle α of S^m will be homologous in M_1 to $k\alpha_1$, where k is an integer and α_1 belongs to a Betti basis for M_1. Using a dual basis, it follows that there is a cycle β on M_1 such that $\beta \,.\, \alpha = k$. Now β can be chosen as a linear combination of singular simplexes on M_1 each of which either does not meet S^m or has exactly one interior point in common with S^m. If the latter simplexes are removed a relative cycle of $M_1 - B_1$ modulo $\mathrm{Fr}B_1$ is obtained whose boundary is easily seen to be $k\gamma$, where γ is a fundamental cycle of the m-sphere $S_1{}^m$ in M_2 introduced by the modification. Clearly $k\gamma$ is homologous to zero in M_2. Now the diagram (7) gives the isomorphism

$$H_m(M_1)/i_mH_m(S^m) \cong H_m(M_2)/i'_mH_m(S_1^m).$$

Since the images of i_m and $i_m{}'$ have generators represented respectively by α and γ, the result stated follows at once.

There is obviously a complementary result to the above, starting with the assumption that α is a torsion cycle in M_1; this is not an essentially different result, but simply consists in reversing the parts played by M_1 and M_2 in the above.

Consider next an even value $2m$ for n. The inequality $2r < n - 1$ is equivalent to $r < m$, and so again the case requiring special attention is $r = m$.

LEMMA 7.4. *If the image of i_m is not of finite order the effect of the modification is to decrease the* mth *Betti number by* 2.

Proof. If α is the fundamental cycle of S^m there is a cycle β on M_1 such that $\beta \,.\, \alpha \neq 0$. Then, reasoning as in the last lemma, it follows that the modification annuls the homology classes, over rational numbers, of α and β; these classes are certainly different, for, since S^m is directly embedded, $\alpha . \alpha = 0$.

Lemma 7.4 could be formulated more completely by describing the effect of the modification on torsion, but as there is no immediate application it does not seem worth while. In any case the most suitable situation for applying this result would be where the lower dimensional homology groups were all zero, when the m-dimensional torsion group would also automatically vanish.

References

1. A. Aeppli, *Modifikationen von reellen und komplexen Mannigfaltigkeiten*, Comm. Math. Helv., *31* (1957), 219–301.

2. R. H. Bing, *Necessary and sufficient conditions that a 3-manifold be* S^3, Ann. Math., *68* (1958), 17–37.

3. F. Hirzebruch, *Neue topologische Methoden in der algebraischen Geometrie*, Ergebnisse der Mathematik (Berlin, 1956).

4. J. Nash, *Real algebraic manifolds*, Ann. Math., *56* (1952), 405–421.

5. P. Samuel, *Sur l'algébricité de certains points singuliers*, J. math. pures et appl., *35* (1956), 1–6.

6. H. Seifert and W. Threlfall, *Lehrbuch der Topologie* (New York, 1934).

7. ——— *Variationsrechnung im Grossen* (New York, 1951).

8. R. Thom, *Quelques propriétés globales des variétés différentiables*, Comm. Math. Helv., *28* (1954), 17–86.

9. A. H. Wallace, *Algebraic approximation of manifolds*, Proc. Lond. Math. Soc., *7* (1957), 196–210.

10. ——— *Homology theory on algebraic varieties*, (London, 1957).

Indiana University

https://doi.org/10.4153/CJM-1960-045-7 Published online by Cambridge University Press