Invent. math. 90, 219–242 (1987)

Inventiones mathematicae

The classification of maps of surfaces

David Gabai*,[1] and William H. Kazez**,[2]

[1] Mathematical Sciences Research Institute, Berkeley, CA 94720, USA
[2] Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104, USA

§ 1. Introduction

In this paper we describe the topology of maps of positive degree between closed orientable surfaces. Two maps $f, g\colon M \to N$ are said to be *equivalent* if there exist homeomorphisms $h\colon M \to M$ and $k\colon N \to N$ such that $kf = gh$ (or $kf \simeq gh$ in the homotopy category). If k is homotopic to id_N we say f and g are *strongly equivalent*.

Surface maps of special interest are branched coverings, i.e., $f\colon M \to N$ is a *branched covering* if there exists a finite set of points $B \subset N$ such that $f|M - f^{-1}(B)$ is a covering map. An arbitrary branched covering may be approximated by a *generic branched covering*, i.e., one in which each point of N has d or $d-1$ preimages where $d = \text{degree } f$.

One of the first people to study branched coverings was Riemann, who proved in his thesis [R] in 1851 that Riemann surfaces occur as conformal branched coverings of S^2. In 1871 and 1873 Lüroth [L] and Clebsch [C] showed that generic branched coverings of S^2 are classified up to (strong) equivalence by their degree. In 1891 Hurwitz [Hu] reduced the classification problem for general range N to the algebraic-combinatorial study of representations of $\pi_1(N-B)$ into $S(d)$, the symmetric group on d letters.

A generic branched covering $\phi\colon M \to N$ may be factored uniquely as $p \circ \tilde{\phi}$ where $\tilde{\phi}\colon M \to \tilde{N}$ is a *primitive* (i.e., surjective on π_1) generic branched covering and $p\colon \tilde{N} \to N$ is the covering corresponding to $\phi_\# \pi_1(M) \subset \pi_1(N)$. Primitive generic branched coverings were shown to be classified up to equivalence by Hamilton [Ha] in 1966 for arbitrary N provided that $b \geqq 2d$, where b is the number of branch points and d is the degree. This was improved by Berstein and Edmonds in 1979 and 1984 ([BE1], [BE2]) to $b > d/2$, or if $N = S^1 \times S^1$, then no restriction on b. More importantly, Berstein and Edmonds stressed that primitive generic branched coverings should be classified up to equivalence by their degree, and they conjectured a suggestive normal form.

* Partially supported by NSF grant DMS 84-03645
** Partially supported by NSF grant DMS 84-04535
Current address of both authors: California Institute of Technology, Pasadena, CA 91125, USA

We show that primitive generic branched coverings are actually classified up to strong equivalence by their degree, and consequently we prove the following theorem.

Theorem 9.2. *Two generic branched coverings ϕ, ψ: $M \to N$ of closed orientable surfaces are strongly equivalent if and only if degree ϕ = degree ψ and $\phi_{\#} \pi_1(M) = \psi_{\#} \pi_1(M)$.*

By using results of Nielson [N], Kneser [K1, K2] and Edmonds [E] we obtain as a corollary the homotopy classification of surface maps.

Corollary 9.4. *If f, g: $M \to N$ are maps of closed orientable surfaces of positive degree then f and g are strongly equivalent in the pointed homotopy category if and only if degree f = degree g and $f_{\#} \pi_1(M) = g_{\#} \pi_1(M)$.*

Since surfaces are $K(\pi, 1)$'s Corollary 9.3 also gives a classification of homomorphisms of surface groups.

Corollary 9.5. *If f, g: $G \to H$ are homomorphisms of surface groups of equal topological degree greater than zero such that $f(G) = g(G)$ in H then there exists an isomorphism h: $G \to G$ such that $f = gh$.*

The proof of the theorem starts with an idea introduced by the first author in [G] of factoring a map ϕ: $M \to N$ as a branched immersion s: $M \to N \times I$ followed by projection π: $N \times I \to N$. In this way the branched covering is "identified" with the space $s(M) \subset N \times I$. §3 contains a calculus of double curves which will be used to put s, and hence ϕ, into normal form.

§4 contains a proof of a weak structure theorem which states roughly that any generic branched cover is a connected sum of covering spaces. A useful technical device for exploiting primitivity of a generic branched cover is the virtual graph which is defined in §5 and used in §6 via "shuffling". §7 gives some very useful sufficient conditions for uniqueness (such as the existence of a nonseparating trivially covered curve in N). §8 gives a quick proof of uniqueness in the case of 4 or more branch points and §9 contains the general case.

Acknowledgements. We would like to thank S. Beckmann, A. Casson, P. Edelman, A. Edmonds, W. Jaco, and P. Scott for helpful conversations and Donna Gabai for help with computer computations. We would also like to thank the referee for many constructive criticisms.

§2. Definitions and examples

M and N will be used to denote orientable surfaces. All branched coverings ϕ: $M \to N$ will be assumed to be generic (or in the terminology of [BE1], [BE2], *simple*). Notice that if $b \in N$ is a *branch point* of ϕ then there is a unique *singular point* $x \in \phi^{-1}(b)$, and near x, ϕ is 2 to 1. $b(\phi)$ is used to denote the branch set of ϕ, d is the degree of ϕ, $|X|$ is the number of elements or path components of X, and $\mathcal{N}(X)$ is a closed regular neighborhood of X. All constructions may be performed in the category of pointed spaces and maps. Discussion of the care needed for dealing with base points is contained in

Remarks 2.6, 3.8, and 6.7. Branched coverings of surfaces with boundary are required to be proper maps, i.e. $\phi^{-1}(\partial N)=\partial M$.

If α is a simple closed curve in N and $\phi^{-1}(\alpha)=\alpha_1\cup\ldots\cup\alpha_k$, we say α is *evenly covered* if degree $\phi|\alpha_i=$degree $\phi|\alpha_j$ for all i and j. If this degree is 1 then α is *trivially covered*. A *dual curve* of α is a simple closed curve $\beta\subset N$ such that α and β have 1 transverse point of intersection.

Example 2.1. Let T: $S^1\times I\to S^1\times I$ be the involution given by $T(\theta,t)=(-\theta,1-t)$ where $S^1=\mathbb{R}/(2\pi\mathbb{Z})$ and $I=[0,1]$. Then $S^1\times I/\{(\theta,t)=T(\theta,t)\}$ is homeomorphic to D^2 and the quotient map ϕ: $S^1\times I\to D^2$ is a degree 2 branched covering with two branch points.

Alternatively, $\phi=\pi\circ s$ where s: $S^1\times I\to D^2\times I$ is a "branched immersion" and π: $D^2\times I\to D^2$ is projection. See Fig. 2.1. In Fig. 2.1 x and y are the branch points of ϕ, and the edge e is the image of the double arc of the immersion s.

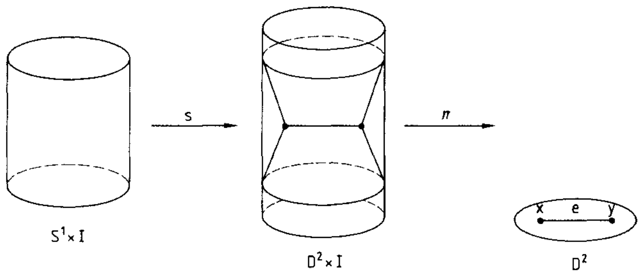

Fig. 2.1

Example 2.2. Let ϕ_1: $X_1\to N$ and ϕ_2: $X_2\to N$ be unbranched covering maps. Let $E\subset X_1$ and $F\subset X_2$ be disks such that $\phi_1(E)=\phi_2(F)=D$ a disk, and let $E_1,\ldots,E_k\subset E$ and $F_1,\ldots,F_k\subset F$ be pairwise disjoint disks such that $\phi_1(E_i)=\phi_2(F_i)$ for $i=1,\ldots,k$. Define

$$X_1\#_k X_2\equiv\left(X_1-\bigcup_{i=1}^{k}\mathring{E}_i\right)\cup\left(\bigcup_{i=1}^{k}S_i^1\times I\right)\cup\left(X_2-\bigcup_{i=1}^{k}\mathring{F}_i\right)$$

where $\partial(S_i^1\times I)=\partial E_i\cup\partial F_i$. Now let $\phi_1\#_k\phi_2$: $X_1\#_k X_2\to N$ by defining

$$\phi_1\#_k\phi_2\left|\left(X_1-\bigcup_{i+1}^{k}\mathring{E}_i\right)=\phi_1,\qquad \phi_1\#_k\phi_2\right|\left(X_2-\bigcup_{i=1}^{k}\mathring{F}_i\right)=\phi_2,$$

and $\phi_1\#_k\phi_2|S_i^1\times I=$the branched covering of Example 2.1.

Example 2.3. In the obvious way for covering maps ϕ_i: $X_i\to N$ define $\phi_1\#_k\phi_2\#\phi_3\#\ldots\#\phi_n$: $X_1\#_k X_2\# X_3\ldots\# X_n\to N$ where $\#\equiv\#_1$.

In §4 we shall show that up to strong equivalence all branched coverings of closed surfaces occur in Example 2.3. See Fig. 4.2.

s: $M \to N \times I$ is a *branched immersion representing* ϕ if the following three conditions hold:

1) $\phi = \pi \circ s$ where π: $N \times I \to N$ is projection, and s is an immersion in general position with respect to itself on M-(singular set of ϕ).

2) Each branch point x is contained in a disk $D \subset N$ such that $\phi^{-1}(D)$ consists of $d-2$ disks which are embedded horizontally and disjointly in $N \times I$ and 1 disk which contains the singular point above x which is mapped into $N \times I$ by s with just one double arc. See Fig. 2.2.

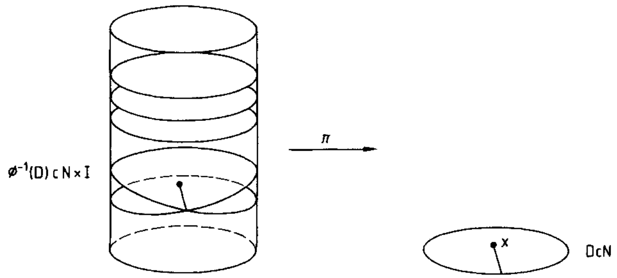

Fig. 2.2

3) The boundary components of M are efficiently immersed in $\partial N \times I$. See Definition 2.5.

It is straightforward to construct a branched immersion s representing any given ϕ. s may be thought of as having almost horizontal image. In § 3, s will be varied in such a way that its image remains almost horizontal and hence $\pi \circ s$ will remain a branched covering. This technique was first used in [G] for studying surface maps.

The branched immersion s has various double curves and triple points in $N \times I$. Call s *regular* if the projection of its double curves have transverse intersections in N. A branched immersion s representing ϕ may be perturbed without changing $\pi \circ s$ so that it is regular. We shall refer to regular branched immersions as immersions.

Let $\Gamma_s \subset N$ be the projection of the double curves of the immersion s representing ϕ. The vertices of Γ_s are those points of Γ_s which do not have neighborhoods homeomorphic to $\mathbb{R}^1$. An edge of Γ_s is the closure of a component of Γ_s-(vertices of Γ_s). Γ_s will be referred to as a graph, though some of its edges may be homeomorphic to S^1.

For each point $x \in N - \Gamma_s$, label the d points of $\phi^{-1}(x)$ 1 through d so that the labeling preserves the order of $\pi \circ s(\phi^{-1}(x)) \subset I = [0, 1]$. Points of M labeled i are said to be at height i. Label each edge e of Γ_s with the transposition $\sigma_i = (i\ i+1)$, where i and $i+1$ are the heights of the points near the double curve which projects to e. (Notice that regularity of s implies that there is a unique double curve projecting to e.)

By regularity of s, the vertices v of Γ_s are one of the following three types:

1) v has six edges leaving it. These are the projection of triple points of s. The edges must be labeled consecutively $\sigma_i, \sigma_{i+1}, \sigma_i, \sigma_{i+1}, \sigma_i, \sigma_{i+1}$ for some i.

2) v has four edges leaving it. These are the intersection of the projections of two double curves occuring at different heights above v. The edges must be labeled consecutively $\sigma_i, \sigma_j, \sigma_i, \sigma_j$ for some i, j with $|i-j|>1$.

3) v has just one edge leaving it. These are either the branch points of ϕ or points in $\Gamma_s \cap \partial N$.

Remark 2.4. This construction may be performed on any branched covering, generic or not. Conversely, given any graph $\Gamma \subset N$ labeled as above, one may construct an immersion s and a branched covering $\phi = \pi \circ s$ such that $\Gamma_s = \Gamma$.

ϕ is determined by its associated unbranched covering, and this is determined by the representation $\rho\colon \pi_1(N-b(\phi)), x) \to S(d)$, where ρ is the map to the symmetric group given by the action of the fundamental group on the labeled fibre $\phi^{-1}(x)$ and $x \in N - \Gamma_s$. ϕ is primitive if and only if ρ is surjective. (See [BE1].) Corollary 5.2 gives another characterization of primitivity.

If $[\alpha] \in \pi_1(N-b(\phi), x)$ and $\alpha \subset N$-(vertices of Γ_s), then $\rho[\alpha]$ equals the product of the σ_i corresponding to the edges of Γ_s that α crosses in the order that α crosses them. In general if $\alpha \subset N$-(vertices of Γ_s) is a path from x to y where $x, y \in N - \Gamma_s$, then denote by $\rho(\alpha)$ the permutation of heights given by following the lifts of α from $\phi^{-1}(x)$ to $\phi^{-1}(y)$. Notice that if α is a simple closed curve and $\phi^{-1}(\alpha)$ is connected then $\rho(\alpha)$ is a *d-cycle*.

Definition 2.5. ∂M is *efficiently immersed* in $\partial N \times I$ by s if Γ_s is transverse to ∂N and if $|\Gamma_s \cap \partial N| \leqq |\Gamma_s' \cap \partial N|$ for all immersions s' representing ϕ. If δ is a component of ∂N and $\rho(\delta) = \rho_1 \dots \rho_k$ is a product of disjoint cycles, then $|\Gamma_s \cap \delta| = \sum_{i=1}^{k} (\text{length } \rho_i - 1)$.

Remark 2.6. All our work takes place in the pointed category. More precisely let $* \in \mathring{M}$ and $* \in \mathring{N}$ be basepoints. All our branched coverings $\phi\colon M \to N$ have the property that $\phi(*) = *$ and $* \notin b(\phi)$. Equivalence and strong equivalence in the pointed category refer to basepoint preserving homeomorphisms and isotopies. In the construction of $s\colon M \to N \times I$, we require additionally that $s(*) = (*, 0)$ and $* \notin \Gamma_s$.

§ 3. Calculus of double curves

In this section we describe some elementary modifications that may be made to the labeled graph $\Gamma_s \subset N$. A modification of Γ_s will consist of first changing ϕ within its strong equivalence class and then rechoosing the immersion s. The goal is to simplify Γ_s, for instance, by decreasing the number of edges or vertices.

In Lemmas 3.1–3.4 all changes to Γ_s are made in a disk $D \subset N$ whose intersection with Γ_s is shown before and after the changes in Figs. 3.1–3.4. No changes are made to s outside $M - \phi^{-1}(D)$ or to Γ_s outside D. Lemmas 3.2, 3.3, and 3.5 were the key steps in the proof of the Simple Loop Conjecture [G].

Lemma 3.1. *Surgery may be performed along an arc α where α cuts Γ_s in successive points labeled σ_i.* See Fig. 3.1.

Proof. Just change the height of s in a neighborhood of α. □

Lemma 3.2. *A branch point and its edge labeled σ_i may be pushed through edges of Γ_s labeled σ_j where $|i-j|>1$.* See Fig. 3.2.

Proof. Since $|i-j|>1$ the component of $\phi^{-1}(D)$ which creates σ_i is disjoint in $N \times I$ from the components creating σ_j. The desired change in Γ_s can be made by changing the immersion on the immersed disk at heights i and $i+1$. □

Lemma 3.3. *A branch point and its edge labeled σ_i may be pushed through a triple point with edges labeled alternately σ_i, σ_{i+1}.* See Fig. 3.3.

Proof. The triple point is formed by the intersection of two components of $\phi^{-1}(D)$. s may be redefined on the component which contains the singular point. □

Lemma 3.4. *Γ_s may be modified by the* pincer move *shown in Fig. 3.4.*

Proof. See Fig. 3.4. □

ϕ: $M \to N$ is *boundary reducible* if there exists a component δ of ∂M and a strongly equivalent branched covering ψ: $M \to N$ such that $|\psi^{-1}(\delta')| > |\phi^{-1}(\delta)|$ where δ' is the other component of $\partial N(\delta)$. Notice that if e is an edge of Γ_s

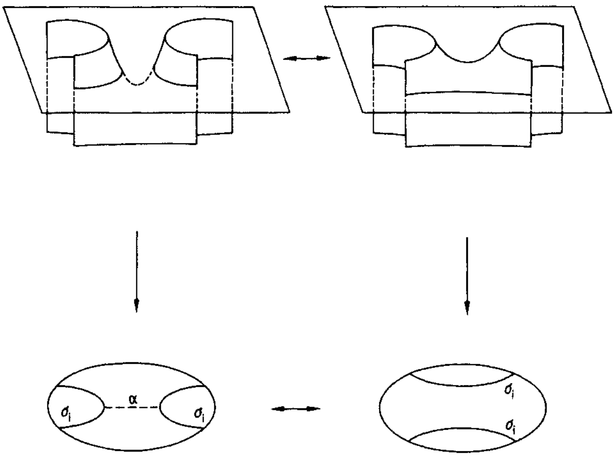

Fig. 3.1

Fig. 3.2

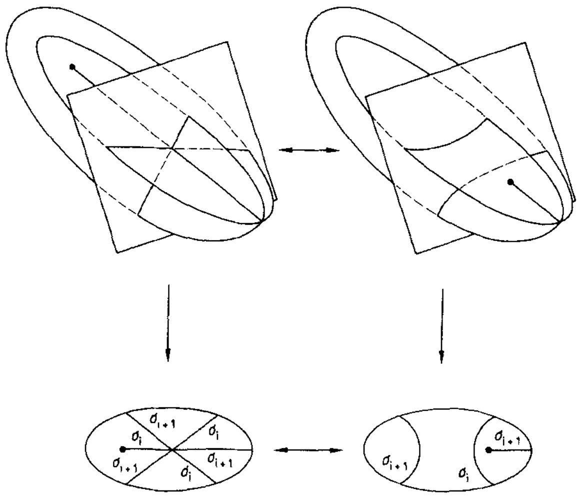

Fig. 3.3

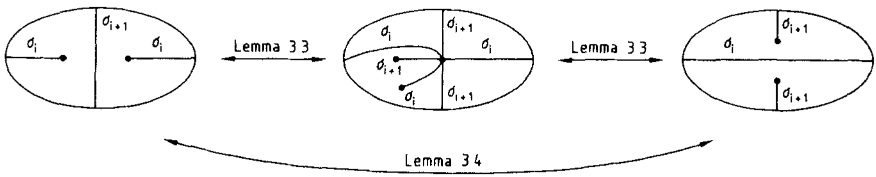

Fig. 3.4

connecting a branch point to ∂N then since ∂N is efficiently immersed by s, ϕ is boundary reducible.

The next lemma shows that for closed surfaces all branching of ϕ can be described locally as a two fold branched covering from a cylinder to a disk. See Fig. 2.1.

Lemma 3.5. *Either ϕ is boundary reducible, or Γ_s may be modified so that all branch points occur in pairs which are connected by single edges of Γ_s.*

Proof. Let x be a branch point. x occurs on a unique edge of Γ_s, and the other vertex of the edge is either a boundary point, a branch point, a double point, or a triple point. The first two cases, are the desired conclusion. In the third and fourth cases either Lemma 3.2 or 3.3 applies, and the argument is completed by induction on the total number of edges and vertices of Γ_s. □

If $D \subset N$ intersects Γ_s in an edge e labeled σ_i connecting two branch points, then $\phi^{-1}(D)$ consists of $d-2$ disks of M and a *cylinder* which when immersed by s connects levels i and $i+1$. By imagining the cylinder as being long and thin as in Fig. 2.1, it makes sense to talk about the *top* and *bottom* of the

Fig. 3.5

cylinder. Given a path $\alpha \subset N$-(vertices of Γ_s) which starts at a boundary point of D and misses $\mathring{D}$, we may *slide the cylinder along* α in such a way that the new cylinder connects heights $\rho(\alpha)(i)$ and $\rho(\alpha)(i+1)$. Furthermore, though sliding a cylinder along α may add concentric circles about the final position of e to Γ_s, the rest of Γ_s is unchanged. This is proved by induction on the number of edges of Γ_s that α crosses and the following lemma.

Lemma 3.6. *An edge surrounded by concentric circles may be slid across an edge labeled σ_i of Γ_s. A concentric circle labeled σ_i will be added to Γ_s but otherwise Γ_s will remain unchanged.* See Fig. 3.5.

Proof. By Lemma 3.1 surgery may be performed on the arc β of Fig. 3.5. □

This proof does not necessarily produce an immersion s with the fewest concentric circles in Γ_s. If the cylinder connects heights $i<j$, then s may be chosen so that the disks at various heights are embedded horizontally and so that the double curve of the cylinder occurs between the disks at height $j-2$ and $j+1$. The concentric circles are then labeled $\sigma_i, \sigma_{i+1}, \ldots, \sigma_{j-2}$ and the edge is labeled σ_{j-1}.

Lemma 3.7. *ϕ: $M \to N$ is boundary reducible if and only if there is a strongly equivalent ψ: $M \to N$ with immersion s' such that there is an edge e of Γ_s' connecting a branch point to ∂N.*

Proof. Details are given only for the only if direction. Suppose $|\psi^{-1}(\delta')| > |\phi^{-1}(\delta)|$ where ψ, δ, and δ' are as in the definition of boundary reducible. Choose an immersion s' so that $|\Gamma_s' \cap \delta'|$ is as small as possible. The proof is by induction on the number n of branch points of ψ in A, where A is annulus bounded by δ and δ'. Notice that $|\psi^{-1}(\delta')| > |\phi^{-1}(\delta)|$ implies $n>0$.

Pick a branch point x in A and use Lemma 3.5 to attempt to form a short double arc. There are three possibilities. First, pushing on x may lead across δ'. This decreases n and increases $|\psi^{-1}(\delta')|$. Second, a short double arc may be formed in A. By Lemma 3.6 this may be slid across δ', thereby decreasing n without effecting $|\psi^{-1}(\delta')|$. Lastly, pushing on x may lead to δ. This is the desired conclusion. □

Remark 3.8. The modifications of ϕ described in this section do not change the strong equivalence type of ϕ in the pointed category, provided only that the arcs along which surgery is performed or cylinders are slid are chosen to miss $*$.

An isotopy g: $N \to N$ corresponding to any of the above moves corresponds to a branch point slide. For example, the isotopy corresponding to the move of Fig. 3.3 is an isotopy which pushes the old branch point to the new one along the obvious arc of Γ_s.

Remark 3.9. If ϕ: $M \to N$ is a generic branched covering and $\phi = p \circ \tilde{\phi}$ where $\tilde{\phi}$: $M \to \tilde{N}$ is a branched covering and p: $\tilde{N} \to N$ is a covering map, then in a natural way s: $M \to N \times I$ induces a map $\tilde{s}$: $M \to \tilde{N} \times I$ such that $\pi \tilde{s} = \tilde{\phi}$. If $\tilde{\phi}$ is modified to $\tilde{\phi}'$ by double curve calculus moves, then by viewing these moves as occurring in $N \times I$ (after projection by p) we see that ϕ is strongly equivalent to $p \circ \tilde{\phi}'$.

§ 4. The weak structure theorem

Lemma 4.1. *A generic branched covering ϕ: $M \to N$ with two branch points is either boundary reducible or is strongly equivalent to a generic branched covering which has a separating cylinder.*

Proof. First use Lemma 3.5 form a cylinder C. If $M - C$ is connected let m be the maximum height of any point in $M - \mathring{C}$ and let $\tilde{\alpha}$ be a path in $M - \mathring{C}$ from the base of C to a point at height m. When C is slid along $\alpha = \phi \circ \tilde{\alpha}$ the vertical orientation of C is changed.

It follows that the vertical orientation of C must have changed at some point where α crosses an edge e of Γ_s. Furthermore, just prior to crossing e, C must connect heights h and $h+1$ where e is labeled σ_h. Now use Lemma 3.1 to surger a, the double arc corresponding to C, and e, then use Lemma 3.5 to reform the short double arc connecting the branch points. The net effect is to reduce the number of edges of Γ_s.

This procedure may be repeated until the new cylinder C has its boundary components contained in different components or until Γ_s has no edges other than a. In the latter case $M - \mathring{C}$ must consist of two components each mapped homeomorphically by $\phi | M - \mathring{C}$, hence C is separating. □

Lemma 4.2 (The weak structure theorem). *If ϕ: $M \to N$ is a generic branched covering of compact surfaces then either ϕ is boundary reducible, or there exist covering maps ϕ_i: $X_i \to N$ such that ϕ is strongly equivalent to*

$$\phi_1 \#_k \phi_2 \# \phi_3 \# \ldots \# \phi_n\colon\ X_1 \#_k X_2 \# X_3 \# \ldots \# X_n \to N.$$

Proof. Apply Lemma 3.5 to ϕ to conclude either ϕ is boundary reducible or we have created $e_1, \ldots, e_t$ short edges of Γ_s connecting branch points. Let $D_i = \mathcal{N}(e_i)$, let C_i be the cylinder component of $\phi^{-1}(D_i)$, let $X = M - \phi^{-1}(\mathring{D}_1 \cup \ldots \cup \mathring{D}_t)$ and let $Y = N - \mathring{D}_1 \cup \ldots \cup \mathring{D}_t$. Let G be the graph with vertices equal to the components $X_1, \ldots, X_u$ of X and edges $C_1, \ldots, C_t$. By Lemma 4.1 it may be assumed that each edge connects distinct vertices.

Let A be the longest embedded arc connecting vertices of G. Relabel the vertices and edges of G so that $A = X_1 \cup C_1 \cup X_2 \cup \ldots \cup X_v$, and so that the bottom and top of C_i are contained in X_i and X_{i+1} respectively.

If $v = t+1$, A contains all of the edges and hence all of the vertices of G, and the result is immediate. If $v < t+1$ there are two cases to consider.

Case 1. There is a vertex of G not in A.

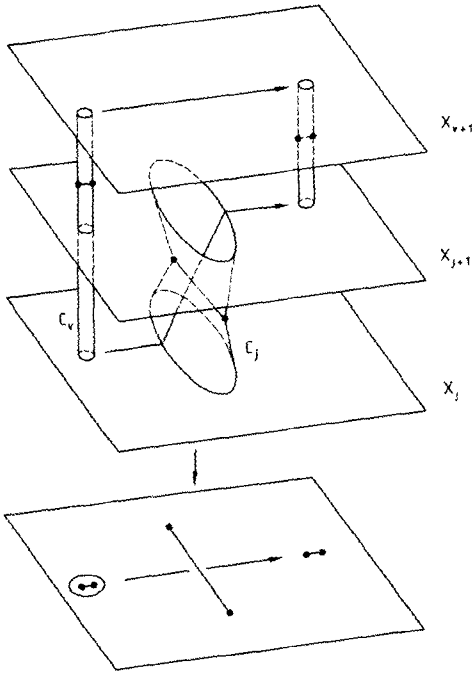

Fig. 4.1

Choose such a vertex X_{v+1} connected to A by an edge C_v and say the base of C_v is contained in the vertex X_j of A. If $j=v$ then $A \cup C_v \cup X_{v+1}$ is a longer embedded path in G. If $j<v$ slide the base of C_v to the base of C_j, then slide C_v up C_j into X_{j+1} so that C_v now connects X_{j+1} and X_{v+1}. See Fig. 4.1. This can be continued until C_v connects X_v and X_{v+1}.

Case 2. A contains all vertices of G but not all edges.

Let C_v be an edge not in A, and let X_i and X_j $(i<j)$ be the vertices containing the bottom and top of C_v. Use the technique of Case 1 to slide the base of C_v into X_1 (and notice that the top stays in X_j). Then slide the top of C_v into X_2 until it is next to the top of C_1. The bottom of C_v will have stayed in X_1, and it may or may not be at the same height as the bottom of C_1. If it is at a different height, the top of C_v may be slid down C_1, and X_1 will contain both ends of C_v. Lemma 4.1 may be applied to find a longer path than A.

If the bottom of C_v is the same height as the bottom of C_1, then C_v has slid into the position required by the Weak Structure Theorem. Repeating Case 2 with the remaining edges finishes the proof. □

Lemma 4.2 gives a quick proof of the following corollary which was first proved by Luroth [L] and Clebsch [C] in 1871 and 1873.

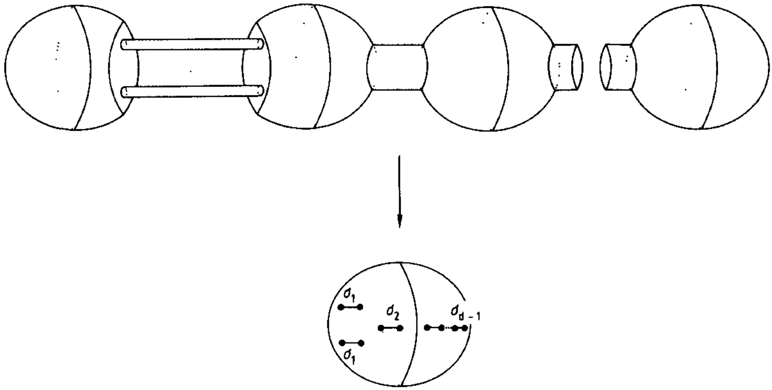

Fig. 4.2

Corollary 4.3. *If ϕ, ψ: $M \to S^2$ are generic branched coverings and degree ϕ $=$degree ψ, then ϕ and ψ are strongly equivalent.*

Proof. By Lemma 4.2 ϕ is strongly equivalent to

$$\phi_1 \#_k \phi_2 \# \dots \# \phi_n\colon\ X_1 \#_k X_2 \# \dots \# X_n \to N.$$

It follows that $X_i = S^2$ for all i, hence $n = \text{degree}\,\phi$ and $k = (\text{genus}\ M) - 1$. Since X_i, n, and k depend only on the degree of ϕ and M, it follows that ϕ and ψ are strongly equivalent. □

§ 5. The virtual graph

Let ϕ: $M \to N$ be a generic branched covering and suppose that $e \subset N$ is a short double arc connecting two branch points and C is the cylinder above $\mathcal{N}(e)$. We shall use the following device to keep track of all positions to which C may be slid.

The *virtual graph* $\Gamma(\phi, C)_x$ is defined for each point $x \in N - b(\phi)$. The vertices of $\Gamma(\phi, C)_x$ are the points in $\phi^{-1}(x)$. Two vertices $a, b \in \phi^{-1}(x)$ are connected by an edge if it is possible to slide C in N so that it connects a and b. The virtual graph depends on the choice of C, but $\Gamma(\phi, C)_x$ may be slid isomorphically to $\Gamma(\phi, C)_y$ along any path in $N - b(\phi)$ so the dependence on x will usually be omitted.

If N_1 is a submanifold of N containing e, then the *restricted virtual graph* $\Gamma(\phi_1, C)_z$ is defined over each $z \in N_1 - b(\phi_1)$ where ϕ_1 is the restriction of ϕ to $\phi^{-1}(N_1)$.

The *induced cover* $\tilde{N}(\phi, C)$ of N, is defined to be $M/\Gamma(\phi, C)$. More precisely, if $x \in N - b(\phi)$ then points of $\phi^{-1}(x)$ in the same component of $\Gamma(\phi, C)_x$ are identified. If $b \in b(\phi)$ and x is a nearby point in N, then points in $\phi^{-1}(b)$ are identified if the nearby points in $\phi^{-1}(x)$ are identified. This is well defined since

ϕ is generic, that is, the two points of $\phi^{-1}(x)$ near the singular point must lie in the same component of $\Gamma(\phi, C)_x$.

It is not hard to see that the canonically defined maps $\tilde{\phi}$: $M \to \tilde{N}(\phi, C)$ and p: $\tilde{N}(\phi, C) \to N$ are respectively branched and unbranched covering maps.

Proposition 5.1. *ϕ may be factored as $p \circ \tilde{\phi}$ where $\tilde{\phi}$: $M \to \tilde{N}(\phi, C)$ is a primitive generic branched covering and p: $\tilde{N}(\phi, C) \to N$ is an unbranched covering map.*

Proof. To show that $\tilde{\phi}$ is primitive it is enough to show that $\tilde{\phi}$ cannot be factored as a branched covering ψ followed by a covering map q. To see this, notice that ψ must also identify the boundary components of C and hence must also identify points of M to which C may be slid. It follows that ψ identifies points in path components of $\Gamma(\phi, C)$. □

Corollary 5.2. *ϕ: $M \to N$ is primitive if and only if $\Gamma(\phi, C)$ is connected.* □

The following definition will be used to measure the extent to which a branched covering ϕ: $M \to N$ can be put into normal form. Let α be a simple closed curve contained in $N - b(\phi)$, and let $\Delta(\alpha) = \gcd\{\text{degree } (\phi|\tilde{\alpha}) \mid \tilde{\alpha}$ is a component of $\phi^{-1}(\alpha)\}$.

Lemma 5.3. *If ϕ: $M \to N$ is a primitive branched covering and $\alpha \subset N - b(\phi)$ is a simple closed curve then either ϕ is boundary reducible or there exists a simple closed curve $\beta \subset N - b(\phi)$ such that*

1) *α and β are isotopic in N*

2) *$\Delta(\beta) \leqq \Delta(\alpha)$*

3) *β is* evenly covered, i.e., *each component of $\phi^{-1}(\beta)$ is mapped by degree $\Delta(\beta)$.*

Proof. The proof will use the techniques of § 3. Start by choosing an immersion s in such a way that when $\rho(\alpha)$ is written as a product of disjoint cycles $\rho_1 \ldots \rho_k$, that $\rho_i = (x_i\, x_i + 1\, x_i + 2 \ldots x_{i+1} - 1)$ and the edges of Γ_s intersecting α are labeled $\sigma_{x_i}, \sigma_{x_i+1}, \sigma_{x_i+2}, \ldots, \sigma_{x_{i+1}-2}$ for $i = 1, \ldots, k$. Among all immersions representing ϕ near α, this has the fewest number of double curves intersecting α transversely. Next, form a short double arc. This may cause some of the double curves intersecting α to disappear and therefore split various ρ_i into products of shorter cycles, but it cannot increase $\Delta(\alpha)$.

If not all components of $\phi^{-1}(\alpha)$ are mapped by equal degree, then by connectivity of the virtual graph, the cylinder corresponding to the short double arc may be assumed to connect two components of $\phi^{-1}(\alpha)$ which are mapped by different degrees. Lemma 5.4 shows how to find an isotopic curve β such that $\Delta(\beta) = \Delta(\alpha)$, but $|\phi^{-1}(\beta)| = |\phi^{-1}(\alpha)| + 1$. Continuing this leads to a curve with the desired properties. □

Lemma 5.4. *Let $\phi = \phi_1 \# \phi_2$: $S^1 \times I \# S^1 \times I \to S^1 \times I$ where degree $\phi_1 = a <$ degree $\phi_2 = b$. There exists a simple closed curve $\beta \subset S^1 \times I - b(\phi)$ such that β is isotopic to $S^1 \times 0$ and $\phi^{-1}(\beta)$ has exactly three components mapped by degrees a, $b - a$, and a respectively.*

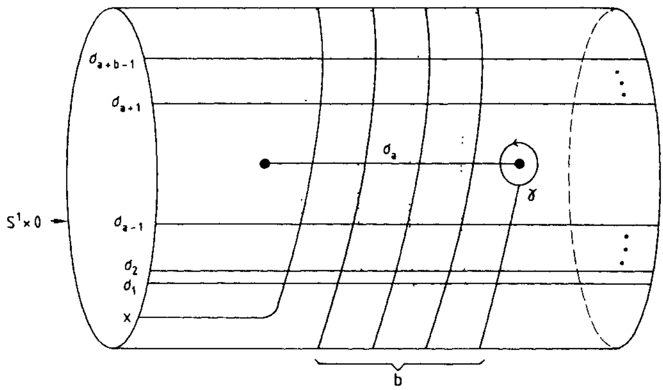

Fig. 5.1

Proof. Figure 5.1 shows the double curves of ϕ and a closed curve γ starting at a point x. Notice that $\rho(\gamma)=(b\,a+b)$ and $\rho(S^1\times 0)=(12\ldots a)(a+1\;a+2\ldots a+b)$, hence if β is chosen to be a simple closed curve near $\gamma*(S^1\times 0)$ then $\rho(\beta)=(12\ldots a)(a+1\ldots b)(b+1\ldots a+b)$. $\phi^{-1}(\beta)$ now has the desired properties. □

§ 6. Shuffling

This section describes a more global method of varying a branched covering in its strong equivalence class than the local moves described in § 3. Though the terminology of § 3 is used, no effort is made to keep track of the changes in Γ_s.

A good example of shuffling is the following. Let C be the separating cylinder in $\phi_1 \# \phi_2$: $M_1 \# M_2 \to N$, let $\alpha \subset N$ be a trivially covered nonseparating simple closed curve, and let $\tilde{\alpha}_i \subset M_i$ be the components of $\phi_i^{-1}(\alpha)$ bear C. Let M be the connected surface obtained by cutting M_i along $\tilde{\alpha}_i$ and then glueing $\tilde{\alpha}_1^{\pm}$ to $\tilde{\alpha}_2^{\mp}$. The effect of shuffling is to replace $\phi_1 \# \phi_2$ by the canonical map ϕ' from M to N.

More precisely and more generally, let α be a simple closed curve in $M-b(\phi)$ with base point a parametrized by $\alpha(t)$: $[0,1]\to M$ such that $\alpha(0)=\alpha(1)=a$. If $b\in\phi^{-1}(\phi(a))$ then there exists a unique lift $\beta(t)$: $[0,1]\to M$ of $\phi\circ\alpha$ such that $\beta(0)=b$ and $\phi(\beta(t))=\phi(\alpha(t))$. $\beta[0,1]$ is not in general either closed or embedded.

Now assume that $\beta[0,1]$ is a simple closed curve β. Let $\mathcal{N}(\alpha)=\alpha\times I$ and $\mathcal{N}(\beta)=\beta\times I$ be regular neighborhoods where α (resp. β) is identified with $\alpha\times 1/2$ (resp. $\beta\times 1/2$) in such a way that $\pi(\alpha(t),t')=\pi(\beta(t),t')$. Suppose also that $|(\mathcal{N}(\alpha)\cup\mathcal{N}(\beta))\cap\phi^{-1}(b(\phi))|=0$.

Shuffling Lemma 6.1. *If an edge of the virtual graph connects $a\times 0$ to $b\times 0$, then ϕ is strongly equivalent to ϕ': $M\to N$ whose immersion s' has image equal to*

$$s(M-(\alpha\times[1/4,3/4]\cup\beta\times[1/4,3/4]))\cup A_1\cup A_2$$

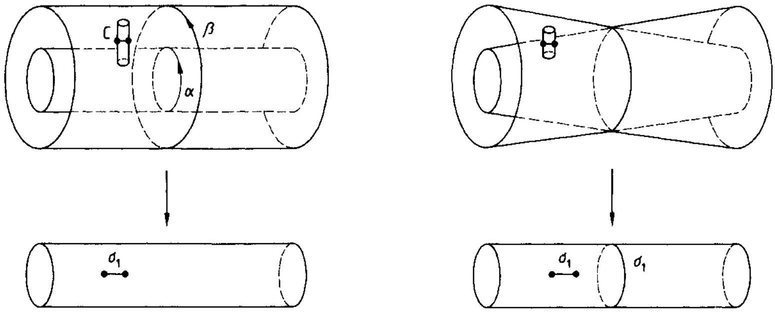

Fig. 6.1

where $A_1 = S^1 \times [1/4, 3/4]$ *satisfies*

$$S^1 \times 1/4 = s(\alpha \times 1/4),$$
$$S^1 \times 3/4 = s(\beta \times 3/4)$$

and

$$\pi(r, t) = \pi(s(\alpha(t))) = \pi(s(\beta(t)))$$

and $A_2 = S^1 \times [1/4, 3/4]$ *satisfies*

$$S^1 \times 1/4 = s(\beta \times 1/4),$$
$$S^1 \times 3/4 = s(\alpha \times 3/4)$$

and

$$\pi(r, t) = \pi(s(\alpha(t))) = \pi(s(\beta(t))).$$

Proof. Figure 6.1 shows how to construct ϕ' in the case that α and β map homeomorphically to a curve in the range. The general case may be sketched as follows.

Let C be the immersed cylinder connecting $a \times 0$ and $b \times 0$. Next modify ϕ through time in such a way that at time t, C is elongated into an immersed cylinder $C(t)$ whose back portion connects $a \times 0$ and $b \times 0$ and whose front portion connects $\alpha(t) \times 0$ and $\beta(t) \times 0$. At the double curve level, the short double arc corresponding to C is stretched along the curve $\phi(\alpha(t))$ while the leading branch point is pushed through double curves as necessary. Since α and β are closed curves, as t approaches 1 the leading portion of $C(t)$ approaches the unmoved back portion, that is, the top and bottom boundary components of $C(t)$ approach the portion of $C(t)$ that was never moved away from $a \times 0$ and $b \times 0$. Therefore a surgery (Lemma 3.1) may be performed. This restores the original cylinder C and completes the *shuffle*. □

We state as a lemma the most common situation for shuffling.

Lemma 6.2. *If* $\gamma \subset N - b(\phi)$ *is evenly covered by* ϕ: $M \to N$, *then we may shuffle along two components of* $\phi^{-1}(\gamma)$ *that are connected by an edge of the virtual graph.* □

Lemma 6.3. *If ϕ: $M \to N$ is a generic branched covering and genus $N \geqq 1$, then either ϕ is boundary reducible or ϕ is strongly equivalent to a branched covering ϕ' for which there exists a cylinder C with $M - C$ connected.*

Proof. Let $\phi = p\tilde{\phi}$ where $\tilde{\phi}$: $M \to \tilde{N}$ is primitive and p: $\tilde{N} \to N$ is a covering map. By Lemma 3.5 $\tilde{\phi}$ and hence ϕ is boundary reducible, or $\tilde{\phi}$ is strongly equivalent to a branch covering $\tilde{\phi}'$ containing a cylinder C. Let $\alpha \subset \tilde{N}$ be a nonseparating curve. By Lemma 5.3 we may assume α is evenly covered. If $M - C$ is not connected then C connects different components of $\tilde{\phi}^{-1}(\alpha)$ and after shuffling along α, $M - C$ will be connected. By Remark 3.9 ϕ is strongly equivalent to $p \circ \tilde{\phi}'$ and the result follows. □

Lemma 6.4. *If ϕ: $M \to N$ is a primitive branched covering, and $\alpha \subset N - b(\phi)$ is a trivially covered nonseparating simple closed curve, and $\beta \subset N - b(\phi)$ is a dual curve to α, then ϕ may be modified within its strong equivalence class so that α is trivially covered and $\phi^{-1}(\beta)$ is connected,* i.e., *$\Delta(\beta) = d = $ degree ϕ or equivalently $\rho(\beta)$ is a d-cycle.*

Proof. The proof is by induction on $|\phi^{-1}(\beta)|$. If $|\phi^{-1}(\beta)| > 1$, then by connectivity of the virtual graph we can choose a cylinder that connects two components of $\phi^{-1}(\beta)$. These two components correspond to two disjoint cycles when $\rho(\beta)$ is written as a product of disjoint cycles. Shuffling over α with this cylinder combines these two cycles without effecting the others, hence $|\phi^{-1}(\beta)|$ is decreased. □

In general, shuffling over an evenly covered curve α changes the virtual graph. An important exception occurs when α is parallel to a boundary component of N. This is exploited in the following technical lemma.

Lemma 6.5. *Let ϕ: $M \to N$ be a branched covering, let α be an evenly covered nonseparating simple closed curve in $N - b(\phi)$, and let β be a dual curve to α. Let ϕ_1: $M_1 \to N_1 = N - \mathcal{N}(\alpha)$ be the restriction of ϕ to $M - \phi^{-1}(\mathcal{N}(\alpha))$, and assume ϕ_1 is primitive. Then ϕ may be modified so that either $\Delta(\alpha)$ is decreased or so that $\Delta(\beta) = 1$.*

Proof. By Lemma 4.1 applied to $\phi | \mathcal{N}(\alpha)$ if an edge of $\Gamma(\phi, C)$ connects a component of $\phi^{-1}(\alpha)$ to itself then $\Delta(\alpha)$ can be decreased.

Say $\partial \mathcal{N}(\alpha) = \alpha^+ \cup \alpha^-$. By a slight abuse of notation identify N with $N_1/\sim$ where $\sim$ identifies corresponding points of α^+ and α^-. Similarly identify $M_1/\sim$ with M in such a way that the map induced by ϕ_1 is ϕ.

Let $\beta_1 = \beta \cap N_1$ and choose a lift $\tilde{\beta}_1$ of β_1. Let $x \in \phi^{-1}(\alpha^-)$ and $y \in \phi^{-1}(\alpha^+)$ be the endpoints of $\tilde{\beta}_1$. Choose $z \in \phi^{-1}(\alpha^-)$ so that $z = y$ in $M_1/\sim$. If $x = z$ then $\phi^{-1}(\beta)$ has a component mapped homeomorphically to β and hence $\Delta(\beta) = 1$.

Otherwise by primitivity of ϕ_1, x and z lie in the same component of $\Gamma(\phi_1, C)_w$ (the restricted virtual graph over $w = \alpha \cap \beta$), and are hence connected by a sequence of edges $e_1, \ldots, e_k$. If ϕ_1 is first modified by this sequence of shuffles over a curve parallel to α^+, it will follow that $x = z$.

This is schematically pictured in Fig. 6.2. □

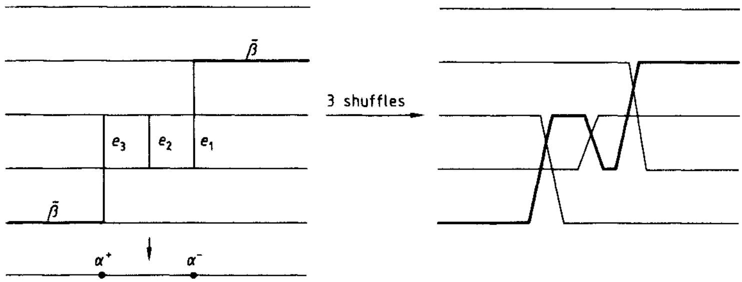

Fig. 6.2

Remark 6.6. The proof of 6.5 shows that the assumption that ϕ_1 is primitive may be replaced by the assumption that x and z lie in the same component of $\Gamma(\phi_1, C)_w$.

Remark 6.7. The modifications to ϕ necessary to make a curve α evenly covered and to then perform shuffles over α, have no effect on the pointed strong equivalence type of ϕ as long as α misses $*$. In the following sections, α and various dual curves should be chosen to miss $*$.

§ 7. Sufficient conditions for uniqueness

For $N \neq S^2$ uniqueness is proved by showing that a branched covering ϕ: $M \to N$ may be put into *normal form*, that is, Γ_s consists of precisely $d-1$ parallel nonseparating simple closed curves labeled $\sigma_1, \ldots, \sigma_{d-1}$ where d $=$degree ϕ, and $\frac{|b(\phi)|}{2}$ short double arcs labeled σ_1. In particular if ϕ is in normal form, the boundary components of N are trivially covered. See Fig. 7.1.

Lemma 7.1. *If ϕ: $M \to N$ is a primitive generic branched covering, and if there exists a nonseparating simple closed curve $\alpha \subset N - b(\phi)$ such that $\Delta(\alpha) = 1$, then either ϕ is boundary reducible or ϕ is strongly equivalent to a map in normal form.*

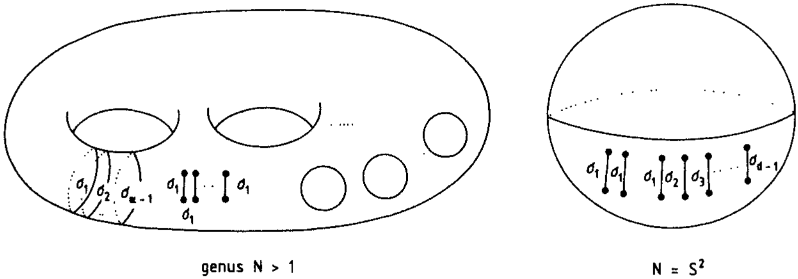

Fig. 7.1

Proof. By Lemma 5.3 we may assume α is trivially covered. α has a dual curve β, and by Lemma 6.3 we may assume $\rho(\beta)$ is a d=degree ϕ cycle. Thus $\phi|\mathcal{N}$ (where $\mathcal{N}=\mathcal{N}(\alpha\cup\beta)$) is just the standard cyclic covering of a punctured torus. It is therefore possible to choose an immersion s such that $\Gamma_s\cap\mathcal{N}$ consists of $d-1$ circles parallel to α that are labeled $\sigma_1, \sigma_2, \ldots, \sigma_{d-1}$ consecutively. Next slide a cylinder near $\phi^{-1}(\partial\mathcal{N})$ in such a way that is connects heights 1 and $1+\ell$, and hence has length ℓ.

Claim. *We may assume that $\ell|d$.*

Proof of Claim. Let $x=\gcd(\ell,d)$. It is not hard to show that there is an integer u such that

$$ux\equiv\ell \bmod(d)$$

$$\gcd(u,d)=1.$$

Choose r,s such that $ru+sd=1$ and consider the pair of dual curves $\bar{\alpha}, \bar{\beta}$ such that $\bar{\alpha}$ is an (r,d) curve and $\bar{\beta}$ is a $(-s,u)$ curve. Then $\rho(\bar{\alpha})=\rho(\beta)^d=(1)$ and $\rho(\bar{\beta})=\rho(\beta)^u$ is a d-cycle. Furthermore $\rho(\bar{\beta})^x(1)=\rho(\beta)^{ux}(1)=\rho(\beta)^{\ell}(1)=\ell+1$, hence with respect to $\bar{\alpha}$ and $\bar{\beta}$ the cylinder has length x and $x|d$.

Claim. *Either $\ell=1$ or some double curves of $\Gamma_s\cap(N-\mathcal{N})$ can be eliminated.*

Proof of Claim. The proof is by induction on the set of lexicographically ordered pairs $(|\text{edges of } \Gamma_s\cap(N-\mathcal{N})|, \ell)$. Notice that if $|\Gamma_s\cap(N-\mathcal{N})|=0$, then primitivity implies $\ell=1$.

Consider the virtual graph restricted to $\mathcal{N}$, i.e., $\Gamma(\phi|\mathcal{N}, C)$. Since $\rho(\beta)=(1\,2\ldots d)$ and $\ell|d$, it follows that $\Gamma(\phi|\mathcal{N}, C)$ has ℓ components and contains the set of edges $S=\{(i,j)\,|\,|i-j|=\ell\}$.

Let $\gamma\subset N-\mathring{\mathcal{N}}$ be a closed curve starting at a point in $\partial\mathcal{N}$. If the entire set of edges S (each oriented from bottom to top) is slid around γ the following cases can occur:

Case 1. The orientation of some edge is reversed. As in Lemma 4.1, a double curve of $\Gamma_s\cap(N-\mathcal{N})$ may be eliminated.

Case 2. The length of some edge is decreased. This new edge has length less than ℓ and the first claim allows us to assume its length divides d and is less than ℓ without changing $\Gamma_s\cap(N-\mathcal{N})$.

Case 3. The length of some edge is increased. If Case 1 does not hold, Case 3 implies Case 2.

Case 4. The length of every edge is preserved. Suppose that this is the case for all γ. Then S is preserved by sliding over all of $N-\mathcal{N}$. It follows that $\Gamma(\phi|\mathcal{N}, C)$ is preserved by sliding over $\mathcal{N}\cup(N-\mathcal{N})=N$. $\Gamma(\phi|\mathcal{N}, C)$ has $\ell>1$ components and $\Gamma(\phi, C)$ is generated by sliding C over N, therefore $\Gamma(\phi, C)$ is not connected. This contradicts primitivity of ϕ.

Completion of the proof. The cylinder has length 1 and hence projects to a short double arc e labeled σ_i (as opposed to a short double arc surrounded by concentric circles) near $\partial\mathcal{N}(\alpha\cup\beta)$. We now eliminate all double curves in $\Gamma_s\cap(N-\mathcal{N})$ other than the other short double arcs as follows.

Suppose that $c\subset N-\mathcal{N}$ is labeled σ_j and is the closest double curve to e. Consider the effect of sliding e once around β. The cylinder corresponding to e connects heights i and $i+1$, $\rho(\beta)=(1,2\ldots d)$, and hence sliding e around β creates a cylinder connecting heights $i+1$ and $i+2$. At the double curve level, the label σ_i on e is changed to σ_{i+1}. Slide e around β until it is labeled σ_j. Perform a surgery to kill c, reform the short double arc e, and continue until the only remaining double curves are short double arcs. These may be relabeled σ_1, as in Fig. 7.1, by sliding around β.

We can now choose a homeomorphism of N taking the double curves of ϕ to the double curves in the normal form with the same labeling. This implies equivalence up to homeomorphism of domain and range. Equivalence up to isotopy of N follows from the next two lemmas. □

Lemma 7.2. *If α' intersects α transversely in one point, then ϕ may be modified so that the double curves $c_1,\ldots,c_{d-1}$ labeled $\sigma_1,\ldots,\sigma_{d-1}$ parallel to α are made parallel to α'.* (Notation as in 7.1.)

Proof. Let e be a short double arc labeled σ_1 of Γ_s, and use it to do a surgery with c_1. Stretch this double arc from one side of c_2, parallel to α', to the other side of c_2. Use Lemma 3.4 to do a pincer move thereby creating the first double curve c_1' parallel to α'. See Fig. 7.2. The pincer move also creates a new double arc labeled σ_2 out of a portion of c_2. This double arc in turn may be stretched from one side of c_3, parallel to α', to the other side of c_3. Continue the process of using pincer moves and moving double arcs until all c_i are converted to c_i'. □

Lemma 7.3. *Given two nonseparating simple closed curves $\alpha,\bar{\alpha}$ there exists a sequence of nonseparating simple closed curves $\alpha_1,\ldots,\alpha_n$ such that $\alpha_1=\alpha$, $\alpha_n=\bar{\alpha}$, and $|\alpha_i\cap\alpha_{i+1}|=1$.*

Proof. We first show how to construct a sequence $\alpha_1,\ldots,\alpha_n$ of nonseparating simple closed curves where $|\alpha_i\cap\alpha_{i+1}|=0$ or 1 by induction on $|\alpha\cap\bar{\alpha}|$.

Choose an arc $\beta\subset\bar{\alpha}$ with endpoints p,q such that $\alpha\cap\beta=\{p,q\}$. Let δ,γ be arcs contained in α such that $\alpha=\delta\cup\gamma$ and $\delta\cap\gamma=\{p,q\}$.

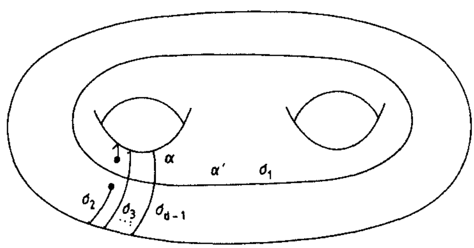

Fig. 7.2

Case 1. α and $\bar{\alpha}$ intersect algebraically with the same sign at p and q. In this case $\delta \cup \beta$ may be isotoped so that $|\alpha \cap (\delta \cup \beta)| = 1$ and $|(\delta \cup \beta) \cap \bar{\alpha}| < |\alpha \cap \bar{\alpha}|$.

Case 2. α and $\bar{\alpha}$ intersect algebraically with different signs at p and q. Notice that homologically $(\delta \cup \beta) + (\gamma \cup -\beta) = \alpha \neq 0$, hence without loss of generality we may assume $\delta \cup \beta$ is non zero in homology and hence is nonseparating. Now $\delta \cup \beta$ may be isotoped so that $|\alpha \cap (\delta \cup \beta)| = 0$ and $|(\delta \cup \beta) \cap \bar{\alpha}| < |\alpha \cap \bar{\alpha}|$.

To finish the proof of the lemma it is enough to show that if $|\alpha \cap \bar{\alpha}| = 0$, then there exists a curve k such that $|\alpha \cap k| = |\bar{\alpha} \cap k| = 1$. It is easy to construct k by considering separately the cases when $N - (\alpha \cup \bar{\alpha})$ has one or two components. □

Since the moves used in Lemma 7.2 only use isotopies of N, it follows that the labeled double curves of ϕ may be put in normal form by a homeomorphism of M and an isotopy of N.

Corollary 7.4. *If ϕ, M, N, and α are as in* 7.1, *but $\Delta(\alpha) = d = \text{degree } \phi$, then either ϕ is boundary reducible or ϕ may be put into normal form.*

Proof. Let β be a dual curve and suppose $\rho(\beta)(1) = i$. Choose k such that $\rho(\beta\alpha^k)(1) = 1$, so that the nonseparating simple closed curve $\beta\alpha^k$ has $\Delta(\beta\alpha^k) = 1$. □

Remarks 7.5. The idea of proving uniqueness by finding a curve α such that $\rho(\alpha)$ is a d-cycle, is due to Berstein and Edmonds [BE1], [BE2] who show that if $|b(\phi)| > 2$ and $\rho(\alpha)$ is a d-cycle, then ϕ may be put into normal form.

In practice it is much easier to find curves α such that $\Delta(\alpha) = 1$. For instance, if degree $\phi = p$ a prime, then for any curve α, $\Delta(\alpha) | p$ hence $\Delta(\alpha) = 1$ or $\Delta(\alpha) = p$. This proves uniqueness for maps of prime degree with N as in 7.1 and with no condition on $b(\phi)$.

§8. Uniqueness for $|b(\phi)| > 2$

Theorem 8.1. *If ϕ: $M \to N$ is a primitive generic branched covering of orientable surfaces with genus $N > 0$, and if $|b(\phi)| > 2$, then either ϕ is boundary reducible or ϕ is strongly equivalent to a generic branched covering in normal form.*

Proof. The goal is to find a nonseparating curve α and a generic branched covering ϕ' strongly equivalent to ϕ such that, with respect to ϕ', $\Delta(\alpha) = 1$. The result will then follow from Lemma 7.1. By Lemma 4.2 we may assume $\phi = \phi_1 \#_k \phi_2 \# \ldots \# \phi_n$: $X_1 \#_k X_2 \# \ldots \# X_n \to N$.

Case 1. $k > 1$.

Let C denote the k^{th} cylinder between X_1 and X_2. Let α be any nonseparating simple closed curve in $N - b(\phi)$ that starts and ends near the edge e corresponding to C but does not cross e. Let $\phi' = \phi_1 \#_{k-1} \phi_2 \# \ldots \# \phi_n$: $X_1 \#_{k-1} X_2 \# \ldots X_n \to N$. It is easily seen directly that ϕ' is surjective on π_1 i.e., that ϕ' is primitive. Therefore the representation ρ' corresponding to ϕ' is surjective onto $S(d)$.

Say $\rho(\alpha)(1)=i$. Since ρ' is surjective, it is possible to slide C until it returns to its original position above N but now connects heights 1 and i. Now α may be isotoped slightly so that it crosses the double curves corresponding to C. Then $\rho(\alpha)(1)=1$ and therefore $\Delta(\alpha)=1$. □

Case 2. $k=1$.

The proof is by induction on degree ϕ_3. If degree $\phi_3>1$, apply Lemma 6.3 to $\phi_1 \# \phi_2$: $M_1 \# M_2 \to N$ to produce a strongly equivalent branched covering ϕ'_1: $M'_1=M_1 \# M_2 \to N$ with a nonseparating cylinder C_1. Let C_2 be the separating cylinder in $M'_1 \# M_3$ and slide the top of C_1 through M'_1 up C_2, and into M_3. Then slide the base of C_1 to the base of C_2. There are two possibilities.

First, C_1 and C_2 are adjacent and connect the same heights. This case was handled in Case 1.

Otherwise, C_1 and C_2 are adjacent but have only the bottom height in common. In this situation the bottom of C_1 may be slid up C_2 thereby placing all of C_1 in M_3. Lemma 4.1 can then be applied to split M_3 into $M'_2 \# M'_3$ thereby lowering the degree of ϕ_3. □

The next theorem was first proved by Berstein and Edmonds in [BE1].

Theorem 8.2. *Let ϕ: $M \to T$ be a primitive generic branched covering where T is the torus. Then ϕ is strongly equivalent to a generic branched covering in normal form.*

Proof. By Theorem 8.1, we may assume $|b(\phi)|=2$. Form a cylinder $C \subset M$ and choose a nonseparating simple closed curve $\alpha \subset T$ which misses the double arc of ϕ. Modify ϕ so that α is evenly covered, and shuffle over α if necessary so that $M-C$ is connected.

$\phi|M-C$ is just the restriction of a covering map of the torus, hence there exist dual curves $\alpha, \beta \subset T$ such that $\Delta(\alpha)=a$, $\Delta(\beta)=b$, $ab=$degree ϕ, and $a|b$. Consider

$$H_1(M) \xrightarrow{\phi_*} H_1(T)=\mathbb{Z}\oplus\mathbb{Z} \xrightarrow{\nu} \mathbb{Z}_a\oplus\mathbb{Z}_b$$

where $\nu[\alpha]=(1,0)$ and $\nu[\beta]=(0,1)$. ϕ_* is surjective since ϕ is primitive. $H_1(M)$ has a basis consisting of a component of $\phi^{-1}(\alpha)$, a component of $\phi^{-1}(\beta)$, a nontrivial separating curve in C, and a curve from the bottom of C up through C and then through $M-\mathring{C}$ to the bottom of C. $\nu \circ \phi_*$ is zero on the first three basis elements, hence $\mathbb{Z}_a\oplus\mathbb{Z}_b$ is cyclic. This implies $\Delta(\alpha)=1$ which implies uniqueness by Lemma 7.1. □

§ 9. Classification theorems

Theorem 9.1. *If ϕ: $M \to N$ is a primitive generic branched covering of orientable surfaces such that $|b(\phi)|=2$ and genus $N \geqq 2$, then either ϕ is boundary reducible or ϕ is strongly equivalent to a generic branched covering in normal form.*

Proof. Let $\alpha \subset N - b(\phi)$ be a nonseparating simple closed curve. By 5.3 we may assume α is evenly covered and that s is chosen so that $|\Gamma_s \cap \alpha|$ is as small as possible. Let $N_1 = N - \mathring{\mathcal{N}}(\alpha)$ and say $\partial\mathcal{N}(\alpha) = \alpha^+ \cup \alpha^-$. By creating an evenly covered nonseparating simple closed curve in N_1 and shuffling, we may assume $M_1 = \phi^{-1}(N_1)$ is connected. Creating this evenly covered nonseparating simple closed curve may lower $\Delta(\alpha)$, but the proof is by induction on ordered triples $(\deg \phi, \Delta(\alpha), d - m(\gamma))$ where α is a nonseparating evenly covered curve with $\phi^{-1}(N_1)$ connected, and $m(\gamma)$ is defined as follows. Let $\gamma \subset N_1$ be a nonseparating simple closed curve and let $\phi_1^{-1}(\gamma) = \gamma_1 \cup \ldots \cup \gamma_k$ where $\phi_1 = \phi | M_1$. Say degree $\phi_1 | \gamma_i$ is maximized for $i = 1$, and call this maximum degree $m(\gamma)$.

We now show that either $\Delta(\alpha)$ may be reduced, or while keeping $\Delta(\alpha)$, M_1, and N_1 fixed, $m(\gamma)$ may be increased. This will complete the proof of the theorem, for by Lemma 7.1 and Corollary 7.4, $\Delta(\alpha) = 1$ or $m(\gamma) = d$ imply uniqueness.

By connectivity of the virtual graph of ϕ, there is a cylinder C connecting γ_1 with another component of $\phi_1^{-1}(\gamma)$, say γ_2. By Lemma 6.5 we may assume $\Gamma(\phi_1, C)$ is not connected, i.e., ϕ_1 is not primitive. Therefore ϕ_1 may be factored as

$$\begin{tikzcd} M_1 \arrow[d, "\phi_1"'] \arrow[dr, "\tilde{\phi}_1"] & \\ N_1 & \tilde{N}_1 \arrow[l, "p"] \end{tikzcd}$$

where $\tilde{\phi}_1$ is a primitive branched cover of degree $d_1 < d$ and p is a covering map of degree $n > 1$. Say $\tilde{\phi}_1(\gamma_1) = \tilde{\gamma}$. Since C connects γ_1 and γ_2, it follows that $\tilde{\phi}_1(\gamma_2) = \tilde{\gamma}$, and since γ is a nonseparating simple closed curve in N_1, $\tilde{\gamma}$ is a nonseparating simple closed curve in $\tilde{N}_1$.

We now assume genus $N \geqq 3$.

Since $\chi(\tilde{N}_1) = n\,\chi(N_1)$ it follows that

$$\begin{aligned} \text{genus } \tilde{N}_1 &= 1 + n(\text{genus } N_1 - 1) + \tfrac{1}{2}(n|\partial N_1| - |\partial \tilde{N}_1|) \\ &\geqq 1 + n(\text{genus } N_1 - 1) \geqq 1 + 2(2-1) = 3, \end{aligned} \tag{$*$}$$

and therefore by induction on the degree of ϕ, we may assume that the theorem is true for $\tilde{\phi}_1$.

Case 1. $\tilde{\phi}_1$ may be put into normal form.

In this case since $\tilde{\gamma}$ is a nonseparating simple closed curve, $\tilde{\phi}_1$ may be modified as in Lemmas 7.1 through 7.3 so that $\tilde{\phi}_1^{-1}(\tilde{\gamma})$ is connected. This increases $m(\gamma)$. It is important to see that as $\tilde{\phi}_1$ is varied in its strong equivalence class, that the strong equivalence type of $p \circ \tilde{\phi}_1$ is unchanged. The proof is basically a diagram chase that we defer to the proof of Corollary 9.2.

Case 2. $\tilde{\phi}_1$ is boundary reducible.

We first show that $p\circ\tilde{\phi}_1$ is boundary reducible. Let $\tilde{\delta}_1$ be a boundary component of $\tilde{N}_1$ and let $p(\tilde{\delta}_1)=\delta_1\subset\partial N_1$. Let $\tilde{\delta}'_1$ (resp. δ'_1) be the other component of $\partial\mathcal{N}(\tilde{\delta}_1)$ (resp. $\partial\mathcal{N}(\delta_1)$). We may assume $p(\tilde{\delta}'_1)=\delta'_1$ and that there exists a strongly equivalent $\tilde{\psi}_1$ with $|\tilde{\psi}_1^{-1}(\tilde{\delta}'_1)|>|\tilde{\phi}_1^{-1}(\tilde{\delta}_1)|$. If the regular neighborhoods are suitably chosen then $|\psi_1^{-1}(\tilde{\delta}'_i)|=|\tilde{\phi}_1^{-1}(\tilde{\delta}_i)|$ for other components of $p^{-1}(\delta'_1)$ and $p^{-1}(\delta_1)$. It follows that $|(p\tilde{\psi}_1)^{-1}(\delta'_1)|>|(p\tilde{\phi}_1)^{-1}(\delta_1)|$.

Reattach $\mathcal{N}(\alpha)$ to N_1 to get N. If δ_1 is a boundary component of N then it follows that ϕ is boundary reducible. Otherwise δ_1 is parallel to α. Lemma 3.7 applied to ϕ_1 and δ_1 produces a branch point connected to δ_1 by a double curve. Pushing the branch point across α increases $|\phi^{-1}(\alpha)|$ and hence decreases $\Delta(\alpha)$.

This completes the proof for genus $N\geqq 3$.

Now consider the case genus $N=2$. The argument just given works unless (*) fails to guarantee that genus $\tilde{N}_1\geqq 2$. This occurs only when $\frac{1}{2}(n|\partial N_1|-|\partial\tilde{N}_1|)=0$. In this case the boundary components of N_1 are trivially covered, hence p: $\tilde{N}_1\to N_1$ is an abelian covering.

Choose $x\in\phi_1^{-1}(\alpha^+)$ and $z\in\phi_1^{-1}(\alpha^-)$ that correspond when $\phi_1^{-1}(\mathcal{N}(\alpha))$ is reattached to M_1. Let $\tilde{\beta}$ be an embedded arc from $\tilde{\phi}_1(x)$ to $\tilde{\phi}_1(z)$ in $\tilde{N}_1$. Let $p(\tilde{\beta})=\beta\subset N_1$.

If β is not embedded, let w be the closest transverse self intersection to α^+. Figure 9.1 shows how to choose a curve β' with one less self intersection than β. Since α^+ is trivially covered by p, $\tilde{\beta}'(0)=\tilde{\phi}_1(x)$ and $\tilde{\beta}'(1)=\tilde{\phi}_1(z)$. Continuing this leads to an embedded arc β such that $\tilde{\beta}(0)=\tilde{\phi}_1(x)$, $\tilde{\beta}(1)=\tilde{\phi}_1(z)$.

Now consider the lift $\bar{\beta}\subset M_1$ of $\tilde{\beta}$ or β such that $\bar{\beta}(0)=x$. Since $\tilde{\phi}_1$ is primitive and $\tilde{\phi}_1\bar{\beta}(1)=\tilde{\beta}(1)=\tilde{\phi}_1(z)$, it follows that z and $\bar{\beta}(1)$ lie in the same (unique) component of $\Gamma(\tilde{\phi}_1, C)_{\tilde{\phi}_1(z)}$.

Therefore z and $\bar{\beta}(1)$ lie in the same component of the restricted virtual graph $\Gamma(\phi_1, c)_{\phi_1(z)}$. β naturally extends to a closed curve, also denoted β in N. The argument of Lemma 6.5 and Remark 6.6 shows how to force $\Delta(\beta)=1$ by shuffling.

This completes the proof of the theorem. □

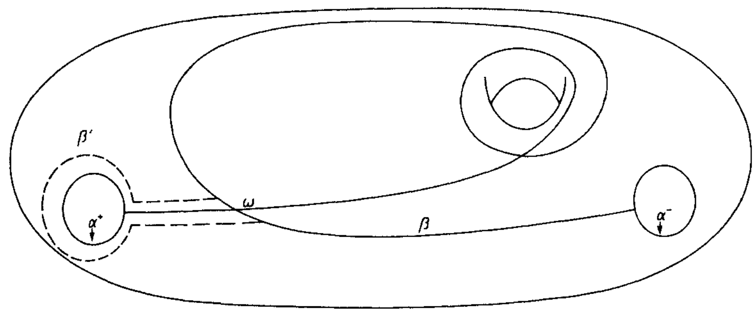

Fig. 9.1

Theorem 9.2. *Two generic branched coverings $\phi, \psi\colon M \to N$ of closed orientable surfaces are strongly equivalent if and only if degree ϕ = degree ψ and $\phi_{\#}\pi_1(M) = \psi_{\#}\pi_1(M)$.*

Proof. We first show that it is enough to consider the case when ϕ and ψ are primitive. If ϕ and ψ are not primitive let $p\colon \tilde{N} \to N$ be the covering corresponding to $\phi_{\#}\pi_1(M) = \psi_{\#}\pi_1(M) \subset \pi_1(N)$. Then $\phi = p \circ \tilde{\phi}$ and $\psi = p \circ \tilde{\psi}$ where $\tilde{\phi}, \tilde{\psi}\colon M \to \tilde{N}$ are primitive branched coverings.

$$\begin{tikzcd} M \arrow[r, "f"] \arrow[d, "\tilde{\phi}"'] & M \arrow[d, "\tilde{\psi}"] \\ \tilde{N} \arrow[r, "g"] \arrow[d, "p"'] & \tilde{N} \arrow[d, "p"] \\ N \arrow[r, "\mathrm{id}_N"] & N \end{tikzcd}$$

Since $\tilde{\phi}$ and $\tilde{\psi}$ are strongly equivalent ($\tilde{\phi} \cong \tilde{\psi}$) there exist homeomorphisms f, g such that $\tilde{\psi} f = g \tilde{\phi}$ and $g \simeq \mathrm{id}_{\tilde{N}}$. We now claim that $\psi = p\tilde{\psi} \cong pg\tilde{\phi} \cong pg_t\tilde{\phi} \cong p\tilde{\phi} = \phi$ where g_t is an isotopy from g to $\mathrm{id}_{\tilde{N}}$ which may be chosen so that $pg_t\tilde{\phi}$ is a generic branched covering for all t. It is not hard to see that sufficiently close generic branched coverings are strongly equivalent, hence the strong equivalence type of $pg_t\tilde{\phi}$ is independent of t. The claim follows by setting $t = 0$ or 1.

The above, together with 4.3, 8.1, and 8.2, and Theorem 9.1 complete the proof. □

Before giving applications to other categories of surface maps, we summarize results of Nielson [N] (Column 1), Edmonds [E] and Kneser [K1, K2] (Column 2) in Diagram 9.2. The entry in each box is a map which necessarily exists in a given homotopy class of maps from M to N. (A pinch is a map which maps a subsurface of M with connected boundary to a point.)

Diagram 9.2

$\simeq$	injective on π_1	primitive
degree = 1	homeomorphism	pinch
degree > 1	covering map	generic branched covering

Historical Note 9.3. Kneser proved in [K2] (see p. 354) that if $f\colon M \to N$ is a map of closed surfaces then either

1) degree $f = 0$ and f is homotopic to a map g such that $g(M) \subset 1$ - skeleton of N, or

2) degree $f \neq 0$ and f is homotopic to a map g such that $g(P)$ = point for some possibly empty connected subsurface P of M and furthermore if $M' = (M - \mathring{P})/\sim$ (where $a \sim b$ if a, b lie in the same component of ∂P), then $g|M'$ is a branched covering.

It follows that if degree $f = 1$ then f is homotopic to a pinch map. We would like to thank David Epstein for bringing this work to our attention.

Corollary 9.4. (The homotopy classification of surface maps.) *If f, g: $M \to N$ are maps of closed orientable surfaces of positive degree, then f and g are strongly equivalent in the pointed homotopy category if and only if degree f = degree g and $f_\# \pi_1(M) = g_\# \pi_1(M)$.*

Proof. There exist a covering map p: $\tilde{N} \to N$ and primitive maps $\tilde{f}$: $M \to \tilde{N}$ and $\tilde{g}$: $M \to \tilde{N}$ such that $f = p\tilde{f}$ and $g = p\tilde{g}$.

Case 1. degree $\tilde{f}$ = degree $\tilde{g} = 1$. Then we may assume that $\tilde{f}$ and $\tilde{g}$ are pinches. It is easy to find homeomorphism h: $M \to M$ such that $\tilde{f} = \tilde{g}h$ and hence $f = p\tilde{f} = p\tilde{g}h = gh$.

Case 2. degree $\tilde{f}$ = degree $\tilde{g} > 1$. Then we may assume that $\tilde{f}$ and $\tilde{g}$ are branched coverings and the corollary follows immediately from Theorem 9.2. □

Since surfaces are $K(\pi, 1)$'s the last corollary gives a classification of homeomorphisms of surface groups.

Corollary 9.5. *If f, g: $G \to H$ are homomorphisms of surface groups of equal topological degree greater than zero such that $f(G) = g(G)$ in H, then there exists an isomorphism h: $G \to G$ such that $f = gh$.* □

References

[BE1] Berstein, I., Edmonds, A.: On the construction of branched coverings of low dimensional manifolds. Trans. Am. Math. Soc. **247**, 87–124 (1979)

[BE2] Berstein, I., Edmonds, A.: On the classification of generic branched coverings of surfaces. Ill. J. Math. **28**, 64–82 (1984)

[C] Clebsch, A.: Zur Theorie der Riemannschen Flächen. Math. Ann. **6**, 1–15 (1873)

[E] Edmonds, A.: Deformations of maps to branched coverings in dimension two. Ann. Math. **110**, 113–125 (1979)

[G] Gabai, D.: The simple loop conjecture. J. Differ. Geom. **21**, 143–149 (1985)

[Ha] Hamilton, R.: Variation of structure on Riemann surfaces. Princeton University Thesis, 1966

[Hu] Hurwitz, A.: Über Riemannsche Flächen mit gegebenen Verzweigungspunkten. Math. Ann. **39**, 1–60 (1891)

[K1] Kneser, H.: Glättung von Flächenabbildungen. Math. Ann. **100**, 609–617 (1928)

[K2] Kneser, H.: Die kleinste Bedeckungszahl innerhalb einer Klasse von Flächenabbildungen. Math. Ann. **103**, 347–358 (1930)

[L] Lüroth, J.: Note über Verzweigungsschnitte und Querschnitte in einer Riemannschen Fläche. Math. Ann. **4**, 181–184 (1871)

[N] Nielsen, J.: Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen. Acta Math. **50**, 189–358 (1927)

[R] Riemann, G.F.B.: Grundlagen für eine allgemeine Theorie der Funktionen einer veränderlichen komplexen Größe. Gesammelte Mathematische Werke, pp. 3–43. Dover, New York: Dover 1953

Oblatum 3-I-1986 & 14-I-1987