A Guide to the Classification Theorem for Compact Surfaces

Jean Gallier
Department of Computer
and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
jean@cis.upenn.edu

Dianna Xu
Department of Computer Science

Bryn Mawr College
Bryn Mawr, PA 19010
dxu@cs.brynmawr.edu

September 19, 2012

Preface

The topic of this book is the classification theorem for compact surfaces. We present the technical tools needed for proving rigorously the classification theorem, give a detailed proof using these tools, and also discuss the history of the theorem and its various "proofs."

We find the classification theorem for compact surfaces quite fascinating because its statement fits very well our intuitive notion of a surface (given that one recognizes that there are non-orientable surfaces as well as orientable surfaces) but a rigorous proof requires a significant amount of work and machinery. Indeed, it took about sixty years until a rigorous proof was finally given by Brahana [6] in 1921. Early versions of the classification theorem were stated by by Möbius [35] in 1861 and by Jordan [24] in 1866. Present day readers will be amused by the "proofs" given by Möbius and Jordan who did not have the required technical tools at their disposal and did not even have the definition of a (topological) surface. More definite versions and "proofs" were given later by von Dyck [50] in 1888 and by Dehn and Heegaard [9] in 1907. One of our goals is to present a history of the proof as complete as possible. A detailed history seems lacking in the literature and should be of interest to anyone interested in topology.

It is our opinion that the classification theorem for compact surfaces provides a natural and wonderful incentive for learning some of the basic tools of algebraic topology, in particular homology groups, a somewhat arduous task without relevant motivations. The reward for such an effort is a thorough understanding of the proof of the classification theorem. Our experience is that self-disciplined and curious students are willing to make such an effort and find it rewarding. It is our hope that our readers will share such feelings.

The classification theorem for compact surfaces is covered in most algebraic topology books. The theorem either appears at the beginning, in which case the presentation is usually rather informal because the machinery needed to give a formal proof has not been introduced yet (as in Massey [33]) or it is given as an application of the machinery, as in Seifert and Threlfall [45], Ahlfors and Sario [1], Munkres [39], and Lee [30] (the proofs in Seifert and Threlfall [45] and Ahlfors and Sario [1] are also very formal). Munkres [39] and Lee [30] give rigorous and essentially complete proofs (except for the fact that surfaces can be triangulated). Munkres's proof appears in Chapter 12 and depends on material on the fundamental group from Chapters 9 and 11. Lee's proof starts in Chapter 6 and ends in Chapter 10, which depends on Chapter 7 on the fundamental group. These proofs are very nice but we feel that the reader will have a hard time jumping in without having read a significant portion of these books. We make further comparisons between Munkres and Lee's approach with ours in Chapter 6.

We thought that it would be useful for a wider audience to present a proof of the classification theorem for compact surfaces more leisurely than that of Ahlfors and Sario [1] (or Seifert and Threlfall [45] or Munkres [39] or Lee [30]) but more formal and more complete than other sources such as Massey [33], Amstrong [3], Kinsey [26], Henle [21], Bloch [5], Fulton [18] and Thurston [49]. Such a proof should be accessible to readers who have a certain amount of "mathematical maturity." This definitely includes first-year graduate students but also strongly motivated upper-level undergraduates. Our hope is that after reading our guide, the reader will be well prepared to read and compare other proofs of the theorem on the classification of surfaces, especially in Seifert and Threlfall [45], Ahlfors and Sario [1], Massey [33], Munkres [39], and Lee [30]. It is also our hope that our introductory chapter on homology (Chapter 5) will inspire the reader to undertake a deeper study of homology and cohomology, two fascinating and powerful theories.

We begin with an informal presentation of the theorem, very much as in Massey's excellent book [33]. Then, we develop the technical tools to give a rigorous proof: the definition of a surface in Chapter 2, simplicial complexes and triangulations in Chapter 3, the fundamental group and orientability in Chapter 4, and homology groups in Chapter 5. The proof of the classification theorem for compact surfaces is given in Chapter 6, the main chapter of this book.

In order not to interrupt the main thread of the book (the classification theorem), we felt that it was best to put some of the material in some appendices. For instance, a review of basic topological preliminaries (metric spaces, normed spaces, topological spaces, continuous functions, limits, connected sets and compact sets) is given in Appendix C. The history of the theorem and its "proofs" are discussed quite extensively in Appendix D. Finally, a proof that every surface can be triangulated is given in Appendix E. Various notes are collected in Appendix F.

Acknowledgements

We would like to thank Eugenio Calabi, Chris Croke, Ron Donagi, Herman Gluck, David Harbater, Alexander Kirillov, Steve Shatz, Wolfgang Ziller, and Jeff Byrne, for their encouragement, advice, inspiration and for what they taught us.

Philadelphia, September 2012 *Jean Gallier & Dianna Xu*

Contents

1 The Classification Theorem: Informal Presentation **7**
1.1 Introduction 7
1.2 Informal Presentation of the Theorem 10

2 Surfaces **29**
2.1 The Quotient Topology 29
2.2 Surfaces: A Formal Definition 31

3 Simplices, Complexes, and Triangulations **35**
3.1 Simplices and Complexes 35
3.2 Triangulations 41

4 The Fundamental Group, Orientability **47**
4.1 The Fundamental Group 47
4.2 The Winding Number of a Closed Plane Curve 52
4.3 The Fundamental Group of the Punctured Plane 55
4.4 The Degree of a Map in the Plane 56
4.5 Orientability of a Surface 58
4.6 Surfaces With Boundary 58

5 Homology Groups **63**
5.1 Finitely Generated Abelian Groups 63
5.2 Simplicial and Singular Homology 67
5.3 Homology Groups of the Finite Polyhedra 82

6 The Classification Theorem for Compact Surfaces **87**
6.1 Cell Complexes 87
6.2 Normal Form for Cell Complexes 92
6.3 Proof of the Classification Theorem 104
6.4 Connected Sums and The Classification Theorem 105
6.5 Other Combinatorial proofs 106
6.6 Application of the Main Theorem 108

A Viewing the Real Projective Plane in $\mathbb{R}^3$ **113**

B Proof of Proposition 5.1 **121**

C Topological Preliminaries **125**

C.1 Metric Spaces and Normed Vector Spaces 125
C.2 Topological Spaces, Continuous Functions, Limits 129
C.3 Connected Sets 138
C.4 Compact Sets 144

D History of the Classification Theorem **159**

E Every Surface Can be Triangulated **167**

F Notes **175**

Symbol Index **185**

Index **187**

Chapter 1

The Classification Theorem: Informal Presentation

1.1 Introduction

Few things are as rewarding as finally stumbling upon the view of a breathtaking landscape at the turn of a path after a long hike. Similar experiences occur in mathematics, music, art, etc. When we first read about the classification of the compact surfaces, we sensed that if we prepared ourself for a long hike, we could probably enjoy the same kind of exhilarating feeling.

The Problem

Define a suitable notion of *equivalence* of surfaces so that *a complete list of representatives, one in each equivalence class of surfaces, is produced*, each representative having a simple explicit description called a *normal form*. By a suitable notion of equivalence, we mean that two surfaces S_1 and S_2 are equivalent iff there is a "nice" bijection between them.

The *classification theorem* for compact surfaces says that, despite the fact that surfaces appear in many diverse forms, surfaces can be classified, which means that every compact surface is equivalent to exactly one representative surface, also called a surface in *normal form*. Furthermore, there exist various kinds of normal forms that are very concrete, for example, polyhedra obtained by gluing the sides of certain kinds of regular planar polygons. For this type of normal form, there is also a finite set of transformations with the property that every surface can be transformed into a normal form in a finite number of steps.

Of course, in order to make the above statements rigorous, one needs to define precisely

1. what is a surface
2. what is a suitable notion of equivalence of surfaces
3. what are normal forms of surfaces.

Figure 1.1: Tibor Radó, 1895-1965.

This is what we aim to do in this book!

For the time being, let us just say that a surface is a topological space with the property that around every point, there is an open subset that is homeomorphic to an open disc in the plane (the interior of a circle).[1] We say that a surface is *locally Euclidean*. Informally, two surfaces X_1 and X_2 are equivalent if each one can be continuously deformed into the other. More precisely, this means that there is a *continuous bijection*, $f\colon X_1 \to X_2$, such that f^{-1} is also continuous (we say that f is a *homeomorphism*). So, by "nice" bijection we mean a homeomorphism, and two surfaces are considered to be equivalent if there is a homeomorphism between them.

The Solution

Every proof of the classification theorem for compact surfaces comprises two steps:

(1) *A topological step*. This step consists in showing that every compact surface *can be triangulated*.

(2) *A combinatorial step*. This step consists in showing that every triangulated surface can be converted to a normal form in a finite number of steps, using some (finite) set of transformations.

To clarify step 1, we have to explain what is a *triangulated surface*. Intuitively, a surface can be triangulated if it is homeomorphic to a space obtained by pasting triangles together along edges. A technical way to achieve this is to define the combinatorial notion of a 2-dimensional complex, a formalization of a polyhedron with triangular faces. We will explain thoroughly the notion of triangulation in Chapter 3 (especially Section 3.2).

The fact that every surface can be triangulated was first proved by Radó in 1925. This proof is also presented in Ahlfors and Sario [1] (see Chapter I, Section §8).

The proof is fairly complicated and the intuition behind it is unclear. Other simpler and shorter proofs have been found and we will present in Appendix E a proof due to Carsten Thomassen [47] which we consider to be the most easily accessible (if not the shortest).

[1] More rigorously, we also need to require a surface to be Hausdorff and second-countable; see Definition 2.3.

There are a number of ways of implementing the combinatorial step. Once one realizes that a triangulated surface can be cut open and laid flat on the plane, it is fairly intuitive that such a flattened surface can be brought to normal form, but the details are a bit tedious. We will give a complete proof in Chapter 6 and a preview of this process in Section 1.2.

It should also be said that distinct normal forms of surfaces can be distinguished by simple invariants:

(a) Their *orientability* (orientable or non-orientable)

(b) Their *Euler–Poincaré characteristic*, an integer that encodes the number of "holes" in the surface.

Actually, it is not easy to define precisely the notion of orientability of a surface and to prove rigorously that the Euler–Poincaré characteristic is a topological invariant, which means that it is preserved under homeomorphisms.

Intuitively, the notion of orientability can be explained as follows. Let A and B be two bugs on a surface assumed to be transparent. Pick any point p, assume that A stays at p and that B travels along any closed curve on the surface starting from p dragging along a coin. A memorizes the coin's face at the begining of the path followed by B. When B comes back to p after traveling along the closed curve, two possibilites may occur:

1. A sees the same face of the coin that he memorized at the beginning of the trip.

2. A sees the other face of the coin.

If case 1 occurs for all closed curves on the surface, we say that it is *orientable*. This will be the case for a sphere or a torus. However, if case 2 occurs, then we say that the surface is *nonorientable*. This phenomenon can be observed for the surface known as the *Möbius strip*, see Figure 1.2

Orientability will be discussed rigorously in Section 4.5 and the Euler–Poincaré characteristic and its invariance in Chapter 5 (see especially Theorem 5.8).

In the words of Milnor himself, the classification theorem for compact surfaces is a formidable result. This result was first proved rigorously by Brahana [6] in 1921 but it had been stated in various forms as early as 1861 by Möbius [35], by Jordan [24] in 1866, by von Dyck [50] in 1888 and by Dehn and Heegaard [9] in 1907, so it was the culmination of the work of many (see Appendix D).

Indeed, a rigorous proof requires, among other things, a precise definition of a surface and of orientability, a precise notion of triangulation, and a precise way of determining whether two surfaces are homeomorphic or not. This requires some notions of algebraic topology such as, fundamental groups, homology groups, and the Euler–Poincaré characteristic. Most steps of the proof are rather involved and it is easy to lose track.

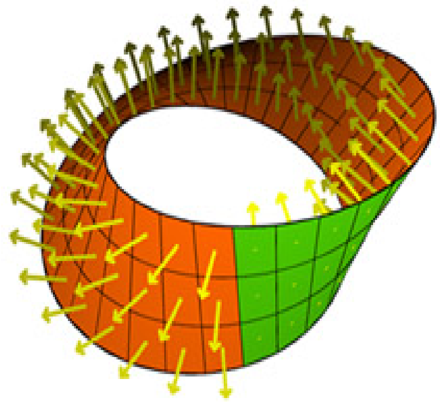

Figure 1.2: A Möbius strip in $\mathbb{R}^3$ (Image courtesy of Prof. Konrad Polthier of FU Berlin).

One aspect of the proof that we find particularly fascinating is the use of certain kinds of graphs (called cell complexes) and of some kinds of rewrite rules on these graphs, to show that every triangulated surface is equivalent to some cell complex *in normal form*. This presents a challenge to researchers interested in rewriting, as the objects are unusual (neither terms nor graphs), and rewriting is really modulo cyclic permutations (in the case of boundaries). We hope that this book will inspire some of the researchers in the field of rewriting to investigate these mysterious rewriting systems.

Our goal is to help the reader reach the top of the mountain (the classification theorem for compact surfaces, with or without boundaries (also called borders)), and help him not to get lost or discouraged too early. This is not an easy task!

We provide quite a bit of topological background material and the basic facts of algebraic topology needed for understanding how the proof goes, with more than an impressionistic feeling.

We also review abelian groups and present a proof of the structure theorem for finitely generated abelian groups due to Pierre Samuel. Readers with a good mathematical background should proceed directly to Section 2.2, or even to Section 3.1.

We hope that this book will be helpful to readers interested in geometry, and who still believe in the rewards of serious hiking!

1.2 Informal Presentation of the Theorem

Until Riemann's work in the early 1850's, surfaces were always dealt with from a local point of view (as parametric surfaces) and topological issues were never considered. In fact, the view that a surface is a topological space locally homeomorphic to the Euclidean plane was

Figure 1.3: James W Alexander, 1888- 1971 (left), Hassler Whitney, 1907-1989 (middle) and Herman K H Weyl, 1885-1955 (right).

Figure 1.4: Bernhard Riemann, 1826-1866 (left), August Ferdinand Möbius, 1790-1868 (middle left), Johann Benedict Listing, 1808-1882 (middle right) and Camille Jordan, 1838-1922 (right).

only clearly articulated in the early 1930's by Alexander and Whitney (although Weyl also adopted this view in his seminal work on Riemann surfaces as early as 1913).

After Riemann, various people, such as Listing, Möbius and Jordan, began to investigate topological properties of surfaces, in particular, *topological invariants*. Among these invariants, they considered various notions of connectivity, such as the maximum number of (non self-intersecting) closed pairwise disjoint curves that can be drawn on a surface without disconnecting it and, the Euler–Poincaré characteristic. These mathematicians took the view that a (compact) surface is made of some elastic strechable material and they took for granted the fact that every surface can be triangulated. Two surfaces S_1 and S_2 were considered *equivalent* if S_1 could be mapped onto S_2 by a continuous mapping "without tearing and duplication" and S_2 could be similarly be mapped onto S_1. This notion of equivalence is a precursor of the notion of a *homeomorphism* (not formulated precisely until the 1900's) that is, an invertible map, $f\colon S_1 \to S_2$, such that both f and its inverse, f^{-1}, are continuous.

Möbius and Jordan seem to be the first to realize that the main problem about the topology of (compact) surfaces is to find invariants (preferably numerical) to decide the equivalence of surfaces, that is, to decide whether two surfaces are homeomorphic or not.

The crucial fact that makes the classification of compact surfaces possible is that every (connected) compact, triangulated surface can be opened up and laid flat onto the plane (as

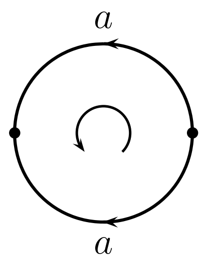

Figure 1.5: A cell representing a sphere (boundary aa^{-1}).

one connected piece) by making a finite number of cuts along well chosen simple closed curves on the surface.

Then, we may assume that the flattened surface consists of convex polygonal pieces, called *cells*, whose edges (possibly curved) are tagged with labels associated with the curves used to cut the surface open. Every labeled edge occurs twice, possibly shared by two cells.

Consequently, every compact surface can be obtained from a set of convex polygons (possibly with curved edges) in the plane, called cells, by gluing together pairs of unmatched edges.

These sets of cells representing surfaces are called *cell complexes*. In fact, it is even possible to choose the curves so that they all pass through a single common point and so, every compact surface is obtained from a single polygon with an even number of edges and whose vertices all correspond to a single point on the surface.

For example, a sphere can be opened up by making a cut along half of a great circle and then by pulling apart the two sides (the same way we open a Chinese lantern) and smoothly flattening the surface until it becomes a flat disk. Symbolically, we can represent the sphere as a round cell with two boundary curves labeled and oriented identically, to indicate that these two boundaries should be identified, see Figure 1.5.

We can also represent the boundary of this cell as a string, in this case, aa^{-1}, by following the boundary counter-clockwise and putting an inverse sign on the label of an edge iff this edge is traversed in the opposite direction.

To open up a torus, we make two cuts: one using any half-plane containing the axis of revolution of the torus, the other one using a plane normal to the axis of revolution and tangential to the torus (see Figure 1.6).

By deformation, we get a square with opposite edges labeled and oriented identically, see Figure 1.7. The boundary of this square can be described by a string obtained by traversing it counter-clockwise: we get $aba^{-1}b^{-1}$, where the last two edges have an inverse sign indicating that they are traversed backwards.

300 VI. TOPOLOGY

in Fig. 281. Once again, we obtain a model of a closed surface; but this time it is easy to reconstruct from the model the surface it represents. To begin with, we bend the rectangle into the form of a circular cylinder (see Figs. 282 and 283) and fasten the sides 1 and 2 together so that identified pairs of points on these sides are actually brought into coincidence. Meanwhile, the sides 3 and 4 have become circles, and by bending the cylinder (see Fig. 284), we can bring them together as prescribed by the identification. Finally, we arrive at the surface of a torus, and the boundary of our rectangle has become a canonical section on the torus, with each of the curves corresponding to two sides of the rectangle (see Figs. 285 and 275b). Conversely, we can begin with a torus and obtain a figure that is topologically equivalent to a rectangle with its sides properly identified in pairs, by slitting the torus along the curves of a canonical section. This procedure can be generalized to all pretzels. For a pretzel of connectivity $2p+1$, the canonical system consists of $2p$ curves, and cutting along these curves results in a $4p$-sided polygon with pairs of sides identified according to a definite rule. Figs. 286 and 287 illustrate the construction for the cases $h=5$ and $h=7$ (i.e. $p=2$ and $p=3$), respectively.

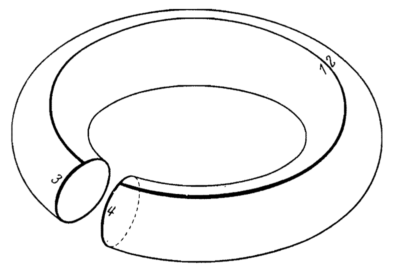

FIG. 284

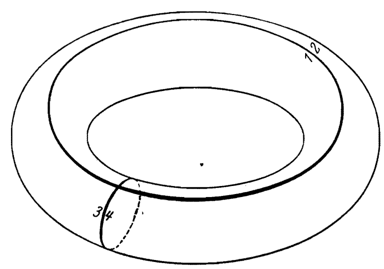

FIG. 285

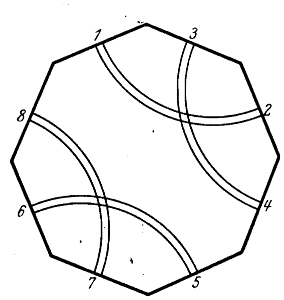

FIG. 286a

The mapping of pretzels into $4p$-sided polygons plays an important part both in the theory of continuous maps (cf. p. 322) and

Figure 1.6: Cutting open a torus, from Hilbert and Cohn-Vossen, page 300.

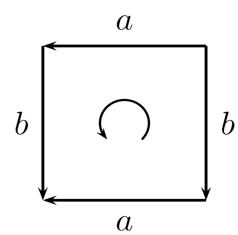

Figure 1.7: A cell representing a torus (boundary $aba^{-1}b^{-1}$).

A surface (orientable) with two holes can be opened up using four cuts. Observe that such a surface can be thought of as the result of gluing two tori together: take two tori, cut out a small round hole in each torus and glue them together along the boundaries of these small holes. Then, we make two cuts to split the two tori (using a plane containing the "axis" of the surface) and then two more cuts to open up the surface. This process is very nicely depicted in Hilbert and Cohn-Vossen [22] (pages 300-301) and in Fréchet and Fan [17] (pages 38-39), see Figure 1.8.

The result is that a surface with two holes can be represented by an octogon with four pairs of matching edges, as shown in Figure 1.9.

A surface (orientable) with three holes can be opened up using 6 cuts and is represented by a 12-gon with edges pairwise identified as shown in Cohn-Vossen [22] (pages 300-301), see Figure 1.8.

In general, an orientable surface with g holes (a surface of *genus* g) can be opened up using $2g$ cuts and can be represented by a regular $4g$-gon with edges pairwise identified, where the boundary of this $4g$-gon is of the form

$$a_1b_1a_1^{-1}b_1^{-1}a_2b_2a_2^{-1}b_2^{-1}\cdots a_gb_ga_g^{-1}b_g^{-1},$$

called type (I). The sphere is represented by a single cell with boundary

$$aa^{-1}, \text{ or } \epsilon \text{ (the empty string);}$$

this cell is also considered of type (I).

The normal form of type (I) has the following useful geometric interpretation: A torus can be obtained by gluing a "tube" (a bent cylinder) onto a sphere by cutting out two small disks on the surface of the sphere and then gluing the boundaries of the tube with the boundaries of the two holes. Therefore, we can think of a surface of type (I) as the result of attaching g handles onto a sphere. The cell complex, $aba^{-1}b^{-1}$, is called a *handle*.

In addition to being orientable or nonorientable, surfaces may have *boundaries*. For example, the first surface obtained by slicing a torus shown in Figure 1.6 (FIG. 284) is a bent cyclinder that has two boundary circles. Similarly, the top three surfaces shown in

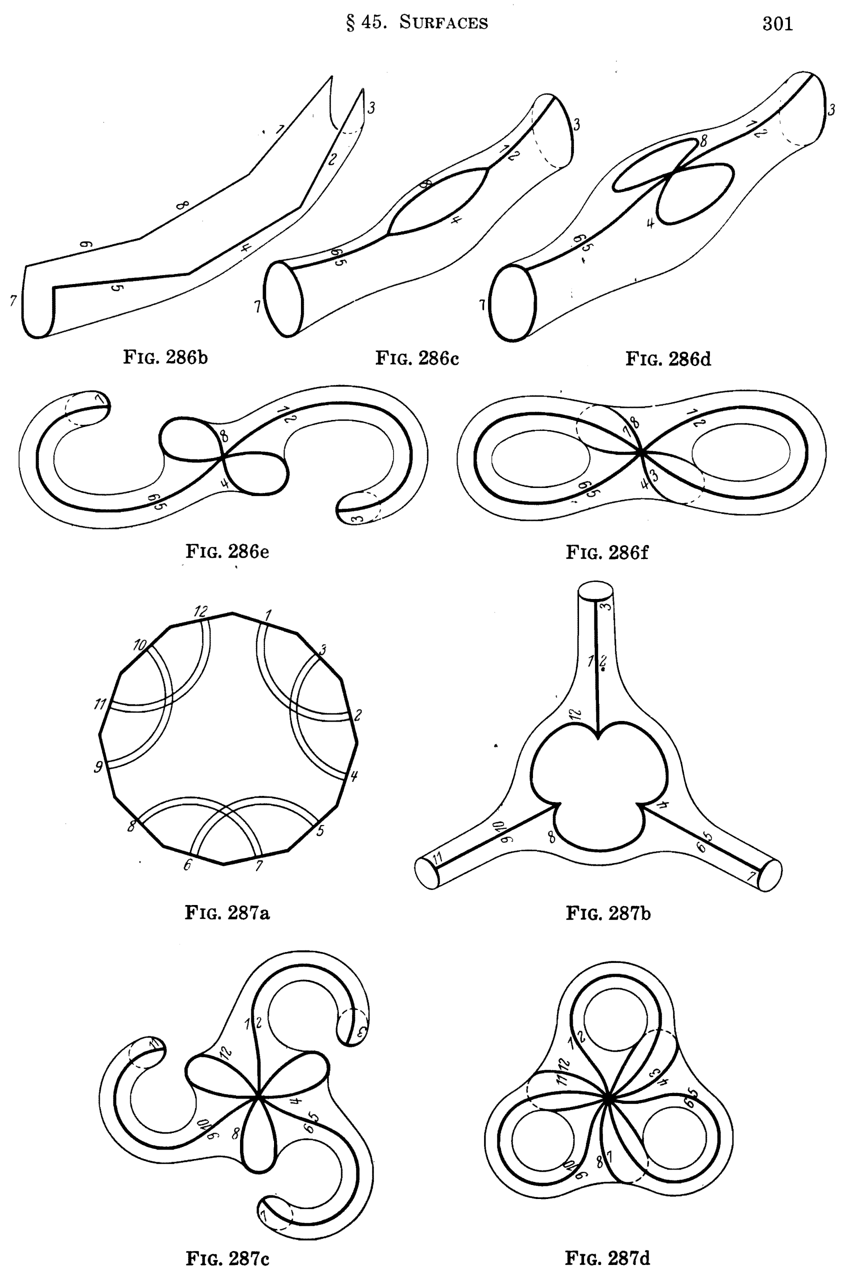

Figure 1.8: Constructing a surface with two holes and a surface with three holes by gluing the edges of a polygon, from Hilbert and Cohn-Vossen, page 301.

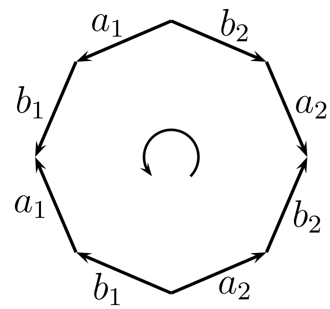

Figure 1.9: A cell representing a surface with two holes (boundary $a_1b_1a_1^{-1}b_1^{-1}a_2b_2a_2^{-1}b_2^{-1}$).

Figure 1.8 (FIG. 286b–d) are surfaces with boundaries. On the other hand, the sphere and the torus have no boundary.

As we said earlier, every surface (with or without boundaries) can be triangulated, a fact proved by Radó in 1925. Then, the crucial step in proving the classification theorem for compact surfaces is to show that every triangulated surface can be converted to an equivalent one in *normal form*, namely, represented by a $4g$-gon in the orientable case or by a $2g$-gon in the nonorientable case, using some simple transformations invoving cuts and gluing. This can indeed be done, and next we sketch the conversion to normal form for surfaces without boundaries, following a minor variation of the method presented in Seifert and Threlfall [45].

Since our surfaces are already triangulated, we may assume that they are given by a finite set of planar polygons with curved edges. Thus, we have a finite set, F, of faces, each face, $A \in F$, being assigned a boundary, $B(A)$, which can be viewed as a string of oriented edges from some finite set, E, of edges. In order to deal with oriented edges, we introduce the set, E^{-1}, of "inverse" edges and we assume that we have a function, $B\colon F \to (E \cup E^{-1})^*$, assigning a string or oriented edges, $B(A) = a_1a_2 \cdots a_n$, to each face, $A \in F$, with $n \geq 2$.[2] Actually, we also introduce the set, F^{-1}, of inversely oriented faces A^{-1}, with the convention that $B(A^{-1}) = a_n^{-1} \cdots a_2^{-1}a_1^{-1}$ if $B(A) = a_1a_2 \cdots a_n$. We also do not distinguish between boundaries obtained by cyclic permutations. We call A and A^{-1} *oriented faces*. Every finite set, K, of faces representing a surface satisfies two conditions:

(1) Every oriented edge, $a \in E \cup E^{-1}$, occurs twice as an element of a boundary. In particular, this means that if a occurs twice in some boundary, then it does not occur in any other boundary.

(2) K is connected. This means that K is not the union of two disjoint systems satisfying condition (1).

A finite (nonempty) set of faces with an assignment of boundaries satisfying conditions (1) and (2) is called a *cell complex*. We already saw examples of cell complexes at the

[2]In Section 6.1, we will allow $n \geq 0$.

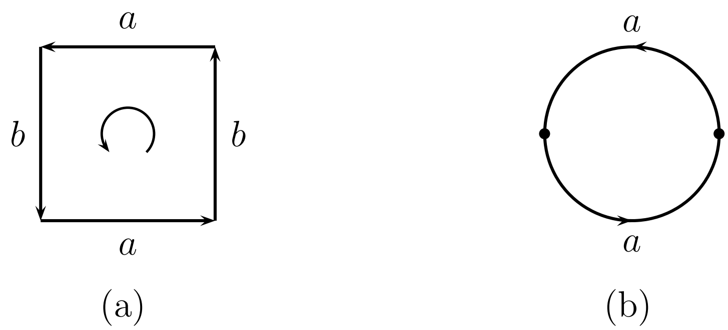

Figure 1.10: (a) A projective plane (boundary $abab$). (b) A projective plane (boundary aa).

beginning of this section. For example, a torus is represented by a single face with boundary $aba^{-1}b^{-1}$. A more precise definition of a cell complex will be given in Definition 6.1.

Every oriented edge has a source vertex a target vertex, but distinct edges may share source or target vertices. Now this may come as a surprise, but the definition of a cell complex allows other surfaces besides the familiar ones, namely *nonorientable* surfaces. For example, if we consider a single cell with boundary $abab$, as shown in Figure 1.10 (a), we have to construct a surface by gluing the two edges labeled a together, but this requires first "twisting" the square piece of material by an angle π, and similarly for the two edges labeled b.

One will quickly realize that there is no way to realize such a surface without self-intersection in $\mathbb{R}^3$ and this can indeed be proved rigorously although this is nontrivial; see Note F.1. The above surface is the *real projective plane*, $\mathbb{RP}^2$.

As a topological space, the real projective plane is the set of all lines through the origin in $\mathbb{R}^3$. A more concrete representation of this space is obtained by considering the upper hemisphere,

$$S^2_+ = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1,\ z \geq 0\}.$$

Now, every line through the origin not contained in the plane $z = 0$ intersects the upper hemisphere, S^2_+, in a single point, whereas every line through the origin contained in the plane $z = 0$ intersects the equatorial circle in two antipodal points. It follows that the projective plane, $\mathbb{RP}^2$, can be viewed as the upper hemisphere, S^2_+, with antipodal on its boundary identified. This is not easy to visualize! Furthermore, the orthogonal projection along the z-axis yields a bijection between S^2_+ and the closed disk,

$$\overline{D} = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\},$$

so the projective plane, $\mathbb{RP}^2$, can be viewed as the closed disk, $\overline{D}$, with antipodal on its boundary identified. This explains why the cell in Figure 1.10 (a) yields the projective plane by identification of edges and so does the circular cell with boundary aa shown in Figure

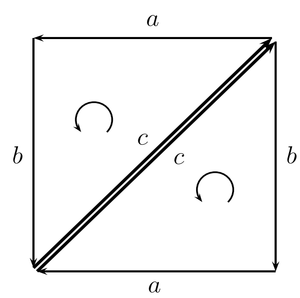

Figure 1.11: An orientable cell complex with $B(A_1) = abc$ and $B(A_2) = bac$.

1.10 (b). A way to realize the projective plane as a surface in $\mathbb{R}^3$ with self-intersection is shown in Note F.2. Other methods for realizing $\mathbb{RP}^2$ are given in Appendix A.

Let us go back to the notion of orientability. This is a subtle notion and coming up with a precise definition is harder than one might expect. The crucial idea is that if a surface is represented by a cell complex, then this surface is orientable if there is a way to assign a direction of traversal (clockwise or counterclockwise) to the boundary of every face, so that when we fold and paste the cell complex by gluing together every edge a with its inverse a^{-1}, no tearing or creasing takes place. The result of the folding and pasting process should be a surface in $\mathbb{R}^3$. In particular, the gluing process does not involve any twist and does not cause any self-intersection.

Another way to understand the notion of orientability is that if we start from some face A_0 and follow a closed path $A_0, A_1, \ldots, A_n$ on the surface by moving from each face A_i to the next face A_{i+1} if A_i and A_{i+1} share a common edge, then when we come back to $A_0 = A_n$, the orientation of A_0 has not changed. Here is a rigorous way to capture the notion of orientability.

Given a cell complex, K, an *orientation of* K is a set of faces $\{A^\epsilon \mid A \in F\}$, where each face A^ϵ is obtained by choosing one of the two oriented faces A, A^{-1} for every face $A \in F$, that is, $A^\epsilon = A$ or $A^\epsilon = A^{-1}$. An orientation is *coherent* if every edge a in $E \cup E^{-1}$ occurs at most once in the boundaries of the faces in the set $\{A^\epsilon \mid A \in F\}$. A cell complex, K, is *orientable* if is has some coherent orientation.

For example, the complex with boundary $aba^{-1}b^{-1}$ representing the torus is orientable, but the complex with boundary aa representing the projective plane is not orientable. The complex with two faces A_1 and A_2 where A_1 has boundary abc and A_2 has boundary bac is orientable, since it has the coherent orientation $\{A_1, A_2^{-1}\}$; see Figure 1.11.

It is clear that every surface represented by a normal form of type (I) is orientable. It turns out that every nonorientable surface (with $g \geq 1$ "holes") can be represented by a $2g$-gon where the boundary of this $2g$-gon is of the form

$$a_1 a_1 a_2 a_2 \cdots a_g a_g,$$

Figure 1.12: Felix C Klein, 1849-1925.

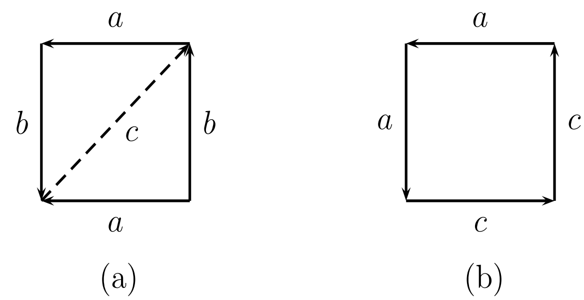

Figure 1.13: (a) A Klein bottle (boundary $aba^{-1}b$). (b) A Klein bottle (boundary $aacc$).

called type (II). All these facts will be proved in Chapter 6, Section 6.3.

The normal form of type (II) also has a useful geometric interpretation: Instead of gluing g handles onto a sphere, glue g projective planes, *i.e.* cross-caps, onto a sphere. The cell complex with boundary, aa, is called a *cross-cap*.

Another famous nonorientable surface known as the *Klein bottle* is obtained by gluing matching edges of the cell showed in Figure 1.13 (a). This surface was first described by Klein [28] (1882). As the projective plane, using the results of Note F.1, it can be shown that the Klein bottle cannot be embedded in $\mathbb{R}^3$.

If we cut the cell showed in Figure 1.13 (a) along the edge labeled c and then glue the resulting two cells (with boundaries abc and $bc^{-1}a^{-1}$) along the edge labeled b, we get the cell complex with boundary $aacc$ showed in Figure 1.13 (b). Therefore, the Klein bottle is the result of gluing together two projective planes by cutting out small disks in these projective planes and then gluing them along the boundaries of these disks. However, it order to obtain a representation of a Klein bottle in $\mathbb{R}^3$ as a surface with a self-intersection is is better to use the edge identification specified by the cell complex of Figure 1.13 (a). First, glue the edges labeled a together, obtaining a tube (a cylinder), then twist and bend this tube to let it penetrate itself in order to glue the edges labeled b together, see Figure 1.14. Other pictures of a Klein bottle are shown in Figure 1.15.

308 VI. TOPOLOGY

but for the neighborhood of a vertex of the heptahedron (Fig. 288) this is not possible. Accordingly, the heptahedron has six singular points. This raises the question of whether there is any one-sided closed surface at all that has no singular points.

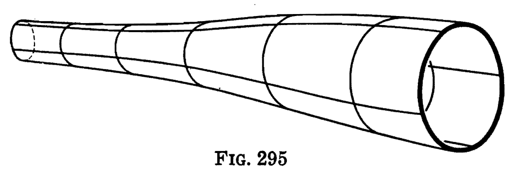

FIG. 295

Such a surface was first constructed by Felix Klein. We begin with an open tube (see Fig. 295). We earlier obtained the torus from such a tube by bending the tube until the ends met and then cementing the boundary circles together. This time we shall put the ends together in a different way. Taking a tube with one end a little thinner than the other, we bend the thin end over and push it through the wall of the tube into the position shown in Fig. 296, where the two circles at the ends of the tube have concentric positions. We now expand the smaller circle and contract the larger one a little until they meet, and then join them together. This does not create any singular points. This construction gives us Klein's surface, also known as the Klein bottle, illustrated in Fig. 297. It is clear that the surface is one-sided and intersects itself along a closed curve where the narrow end was pushed through the wall of the tube.

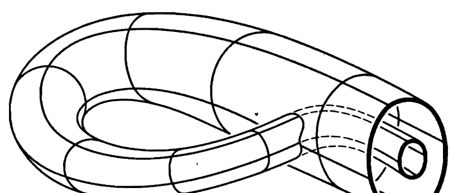

FIG. 296

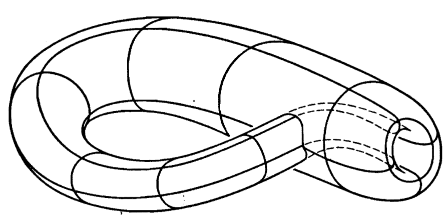

FIG. 297

Our first example of a closed one-sided surface, the heptahedron, differed from the two-sided closed surfaces we have studied thus far also in that it had an even connectivity number. Hence we might expect that the connectivity of the Klein bottle would like-

Figure 1.14: Construction of a Klein bottle, from Hilbert and Cohn-Vossen, page 308.

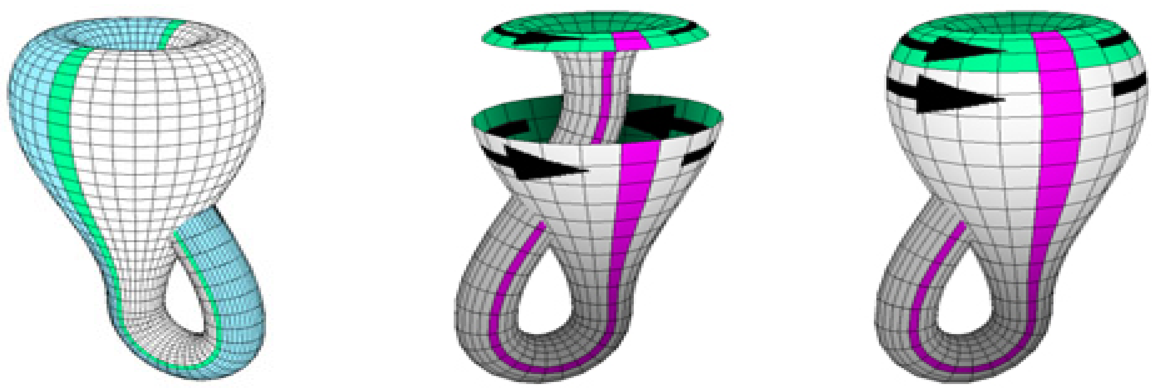

Figure 1.15: Klein bottles in $\mathbb{R}^3$ (Images courtesy of Prof. Konrad Polthier of FU Berlin).

In summary, there are two kinds *normal forms* of cell complexes: These cell complexes $K = (F, E, B)$ in normal form have a single face A ($F = \{A\}$), and either

(I) $E = \{a_1, \ldots, a_p, b_1, \ldots, b_p\}$ and

$$B(A) = a_1 b_1 a_1^{-1} b_1^{-1} \cdots a_p b_p a_p^{-1} b_p^{-1},$$

where $p \geq 0$, or

(II) $E = \{a_1, \ldots, a_p\}$ and

$$B(A) = a_1 a_1 \cdots a_p a_p,$$

where $p \geq 1$.

Observe that canonical complexes of type (I) are orientable, whereas canonical complexes of type (II) are not. When $p = 0$, the canonical complex of type (I) corresponds to a sphere, and we let $B(A) = \epsilon$ (the empty string). The above surfaces have no boundary; the general case of surfaces with boundaries is covered in Chapter 6. Then, the combinatorial form the classification theorem for (compact) surfaces can be stated as follows:

Theorem 1.1. *Every cell complex K can be converted to a cell complex in normal form by using a sequence of steps involving a transformation (P2) and its inverse: splitting a cell complex, and gluing two cell complexes together.*

Actually, to be more precise, we should also have an edge-splitting and an edge-merging operation but, following Massey [34], if we define the elimination of pairs aa^{-1} in a special manner, only one operation is needed, namely:

Transformation P2: Given a cell complex, K, we obtain the cell complex, K', by *elementary subdivision of K* (or *cut*) if the following operation, (P2), is applied: Some face A in K with boundary $a_1 \ldots a_p a_{p+1} \ldots a_n$ is replaced by two faces A' and A'' of K', with

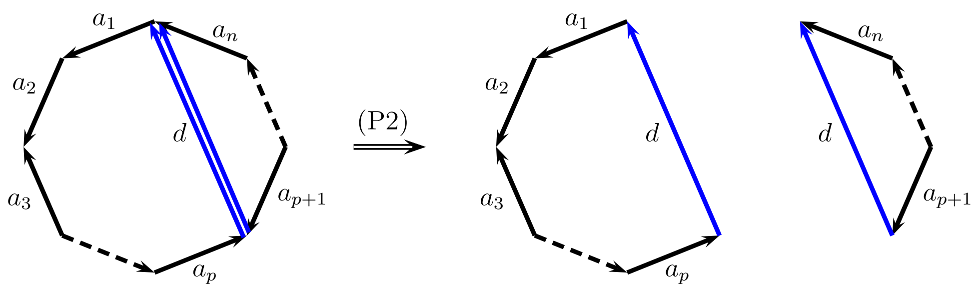

Figure 1.16: Rule (P2).

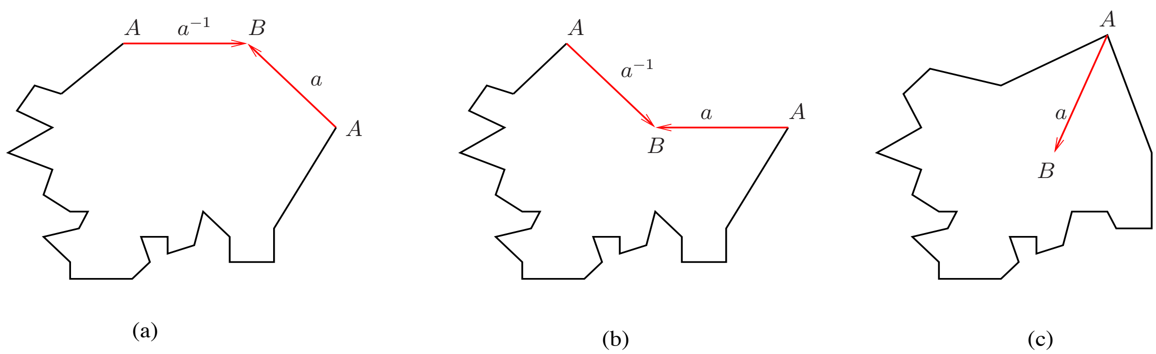

Figure 1.17: Elimination of aa^{-1}.

boundaries $a_1 \ldots a_p d$ and $d^{-1}a_{p+1} \ldots a_n$, where d is an edge in K' not in K. Of course, the corresponding replacement is applied to A^{-1}.

Rule (P2) is illustrated in Figure 1.16.

Sketch of proof for Theorem 1.1. The procedure for converting a cell complex to normal form consists of several steps.

Step 1. Elimination of strings aa^{-1} in boundaries, see Figure 1.17.

Step 2. Vertex Reduction.

The purpose of this step is to obtain a cell complex with a single vertex. We first perform step 1 repeatedly until all occurrences of the form aa^{-1} have been eliminated. If the remaining sequence has no edges left, then it must be of type (I).

Otherwise, consider an inner vertex $\alpha = (b_1, \ldots, b_m)$. If α is not the only inner vertex, then there is another inner vertex β. We assume without loss of generality that b_1 is the edge that connects β to α. Also, we must have $m \geq 2$, since otherwise there would be a string $b_1 b_1^{-1}$ in some boundary. Thus, locate the string $b_1 b_2^{-1}$ in some boundary. Suppose it is of the form $b_1 b_2^{-1} X_1$, and using (P2), we can split it into $b_1 b_2^{-1} c$ and $c^{-1} X_1$ (see Figure

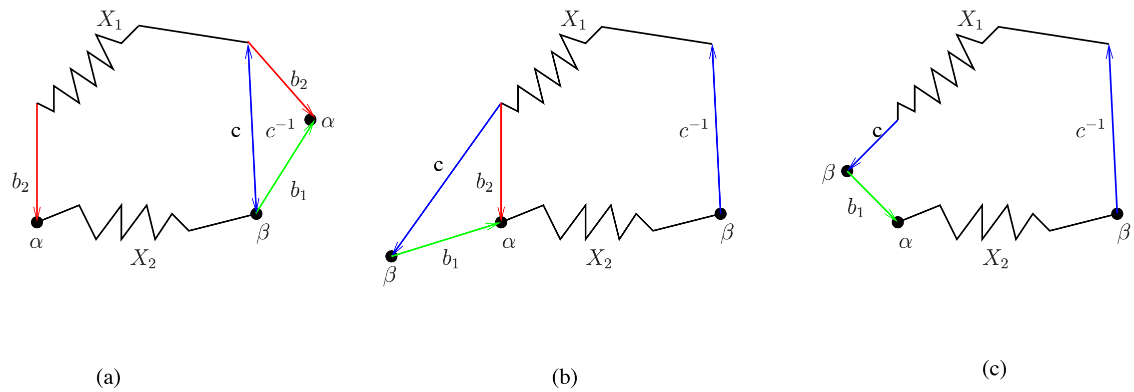

Figure 1.18: Reduction to a single inner vertex.

1.18 (a)). Now locate b_2 in the boundary, suppose it is of the form b_2X_2. Since b_2 differs from b_1, b_1^{-1}, c, c^{-1}, we can eliminate b_2 by applying $(\text{P2})^{-1}$. This is equivalent to cutting the triangle $cb_1b_2^{-1}$ off along edge c, and pasting it back with b_2 identified with b_2^{-1} (see Figure 1.18 (b)).

This has the effect of shrinking α. Indeed, as one can see from Figure 1.18 (c), there is one less vertex labeled α, and one more labeled β.

This procedure can be repeated until $\alpha = (b_1)$, at which stage b_1 is eliminated using step 1. Thus, it is possible to eliminate all inner vertices except one. Thus, from now on, we will assume that there is a single inner vertex.

Step 3. Reduction to a single face and introduction of cross-caps.

We may still have several faces. We claim that for every face A, if there is some face B such that $B \neq A$, $B \neq A^{-1}$, and there is some edge a both in the boundary of A and in the boundary of B, due to the fact that all faces share the same inner vertex, and thus all faces share at least one edge. Thus, if there are at least two faces, from the above claim and using $(\text{P2})^{-1}$, we can reduce the number of faces down to one. It it easy to check that no new vertices are introduced.

Next, if some boundary contains two occurrences of the same edge a, i.e., it is of the form $aXaY$, where X, Y denote strings of edges, with $X, Y \neq \epsilon$, we show how to make the two occurrences of a adjacent. This is the attempt to group the cross-caps together, resulting in a sequence that denotes a cell complex of type (II).

The above procedure is essentially the same as the one we performed in our vertex reduction step. The only differece is that we are now interested in the edge sequence in the boundary, not the vertices. The rule shows that by introducing a new edge b and its inverse, we can cut the cell complex in two along the new edge, and then paste the two parts back by identifying the the two occurences of the same edge a, resulting in a new boundary with a cross-cap, as shown in Figure 1.19 (c). By repeating step 3, we convert boundaries of the form $aXaY$ to boundaries with cross-caps.

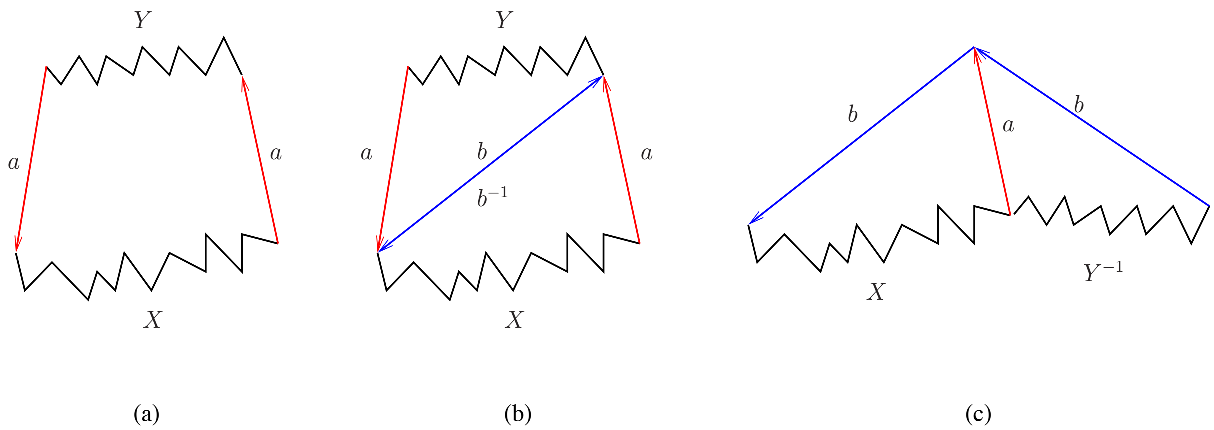

Figure 1.19: Grouping the cross-caps.

Step 4. Introduction of handles.

The purpose of this step is to convert boundaries of the form $aUbVa^{-1}Xb^{-1}Y$ to boundaries $cdc^{-1}d^{-1}YXVU$ containing handles. This is the attempt to group the handles together, resulting in a sequence that denotes a cell complex of type (I). See Figure 1.20.

Each time the rewrite rule is applied to the boundary sequence, we introduce a new edge and its inverse to the polygon, and then cut and paste the same way as we have described so far. Iteration of this step preserves cross-caps and handles.

Step 5. Transformation of handles into cross-caps.

At this point, one of the last obstacles to the canonical form is that we may still have a mixture of handles and cross-caps. If a boundary contains a handle and a cross-cap, the trick is to convert a handle into two cross-caps. This can be done in a number of ways. Massey [34] shows how to do this using the fact that the connected sum of a torus and a Möbius strip is equivalent to the connected sum of a Klein bottle and a Möbius strip. We prefer to explain how to convert a handle into two cross-caps using four applications of the cut and paste method using rule (P2) and its inverse, as presented in Seifert and Threlfall [45] (Section 38).

The first phase is to split a cell as shown in Figure 1.21 (a) into two cells using a cut along a new edge labeled d and then two glue the resulting new faces along the two edges labeled c, obtaining the cell showed in Figure 1.21 (b). The second phase is to split the cell in Figure 1.21 (b) using a cut along a new edge labeled a_1 and then glue the resulting new faces along the two edges labeled b, obtaining the cell showed in Figure 1.21 (c). The third phase is to split the cell in Figure 1.22 (c) using a cut along a new edge labeled a_2 and then glue the resulting new faces along the two edges labeled a, obtaining the cell showed in Figure 1.22 (d). Finally, we split the cell in Figure 1.22 (d) using a cut along a new edge labeled a_3 and then glue the resulting new faces along the two edges labeled d, obtaining the cell showed in Figure 1.22 (e).

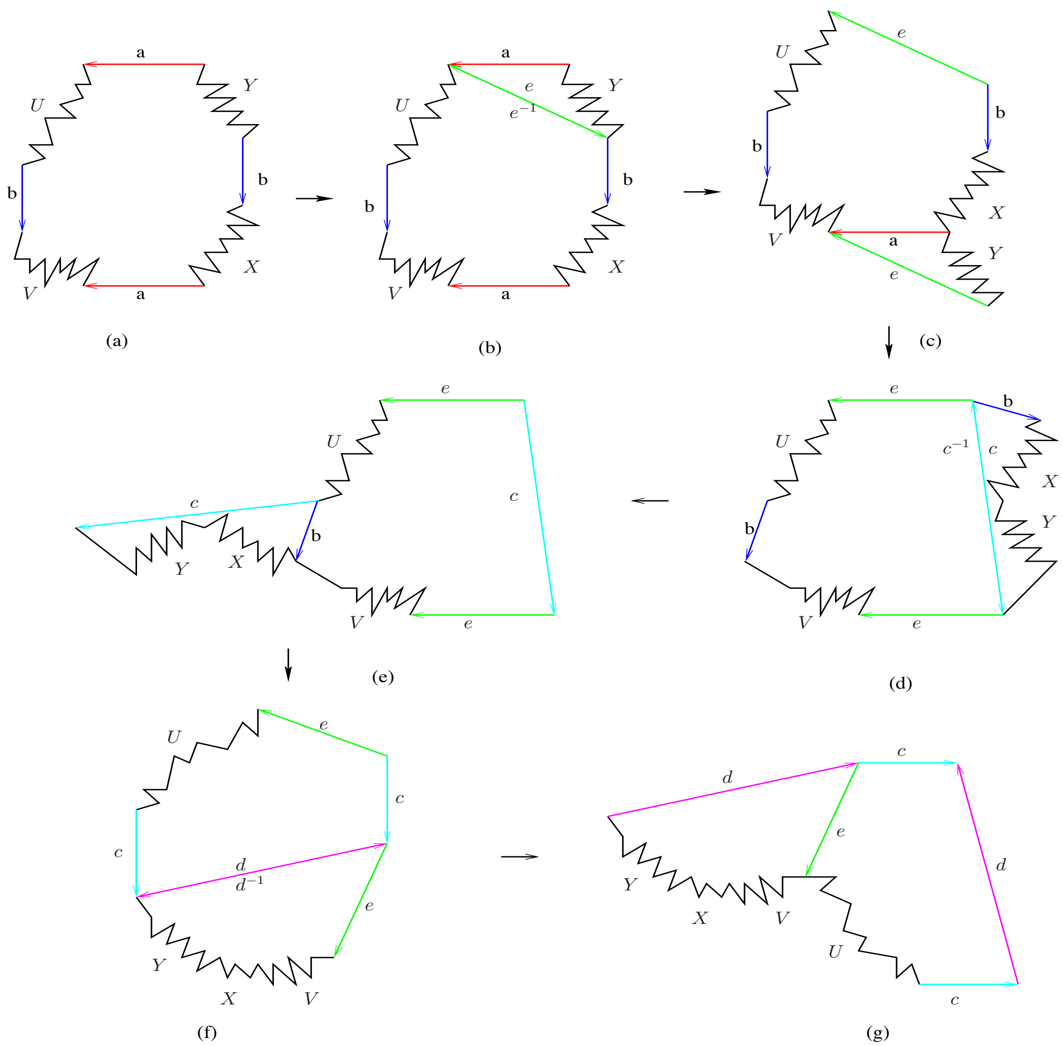

Figure 1.20: Grouping the handles.

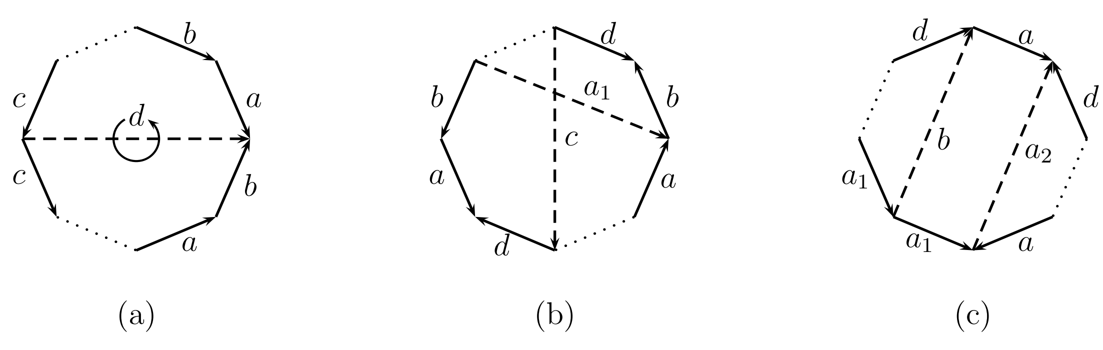

Figure 1.21: Step 5, phases 1 and 2.

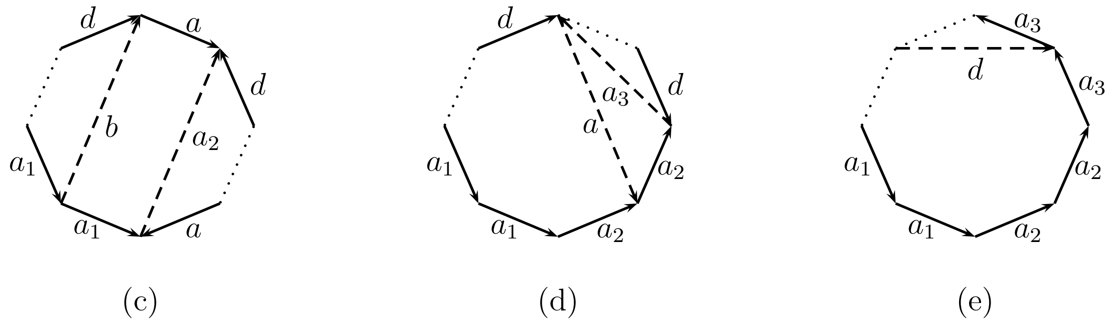

Figure 1.22: Step 5, phases 3 and 4.

Note that in the cell showed in Figure 1.22 (e), the handle $aba^{-1}b^{-1}$ and the cross-cap cc have been replaced by the three consecutive cross-caps, $a_1a_1a_2a_2a_3a_3$.

Using the above procedure, every compact surface represented as a cell complex can be reduced to normal form, which proves Theorem 1.1. □

The next step is to show that distinct normal forms correspond to inequivalent surfaces, that is, surfaces that are not homeomorphic.

First, it can be shown that the orientability of a surface is preserved by the transformations for reducing to normal form. Second, if two surfaces are homeomorphic, then they have the same nature of orientability. The difficulty in this step is to define properly what it means for a surface to be orientable; this is done in Section 4.5 using the degree of a map in the plane.

Third, we can assign a numerical invariant to every surface, its *Euler–Poincaré characteristic*. For a triangulated surface K, if n_0 is the number of vertices, n_1 is the number of edges, and n_2 is the number of triangles, then the Euler–Poincaré characteristic of K is defined by

$$\chi(K) = n_0 - n_1 + n_2.$$

Then, we can show that homeomorphic surfaces have the same Euler–Poincaré characteristic and that distinct normal forms with the same type of orientability have different Euler–Poincaré characteristics. It follows that any two distinct normal forms correspond to inequivalent surfaces. We obtain the following version of the classification theorem for compact surfaces:

Theorem 1.2. *Two compact surfaces are homeomorphic iff they agree in character of orientability and Euler–Poincaré characteristic.*

Actually, Theorem 1.2 is a special case of a more general theorem applying to surfaces with boundaries as well (Theorem 6.4). All this will be proved rigorously in Chapter 6. Proving

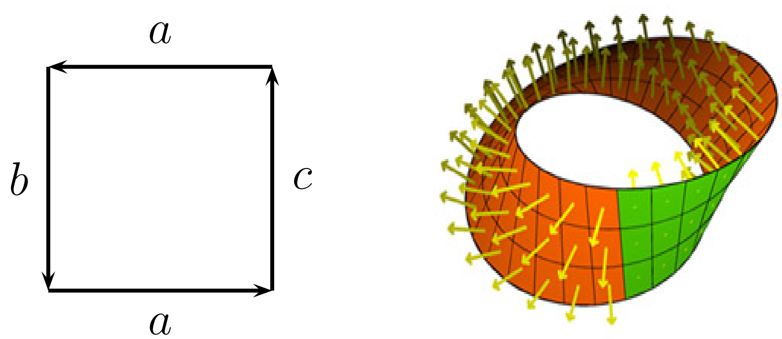

Figure 1.23: Left: A Möbius strip (boundary *abac*). Right: A Möbius strip in $\mathbb{R}^3$ (Image courtesy of Prof. Konrad Polthier of FU Berlin).

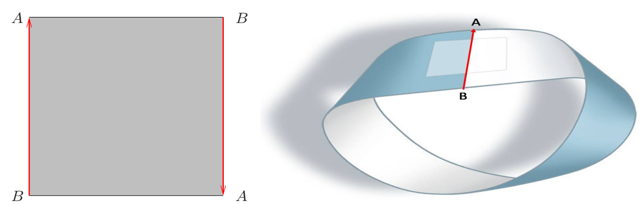

Figure 1.24: Construction of a Möbius strip.

rigorously that the Euler–Poincaré characteristic is a topological invariant of surfaces will require a fair amount of work. In fact, we will have to define homology groups. In any case, we hope that the informal description of the reduction to normal form given in this section has raised our reader's curiosity enough to entice him to read the more technical development that follows.

To close this introductory chapter, let us go back briefly to surfaces with boundaries. Then, there is a well-known nonorientable surface realizable in $\mathbb{R}^3$, the *Möbius strip*. This surface was discovered independently by Listing [32] (1862) and Möbius [37] (1865).

The Möbius strip is obtained from the cell complex in Figure 1.23 by gluing the two edges labeled a together. Observe that this requires a twist by π in order to glue the two edges labeled a properly.

The resulting surface shown in Figure 1.23 and in Figure 1.24 has a single boundary since the two edges b and c become glued together, unlike the situation where we do not make a twist when gluing the two edges labeled a, in which case we get a torus with two distinct boundaries, b and c.

It turns out that if we cut out a small hole into a projective plane we get a Möbius strip. This fact is nicely explained in Fréchet and Fan [17] (page 42) or Hilbert and Cohn-Vossen [22] (pages 315-316). It follows that we get a realization of a Möbius band with a

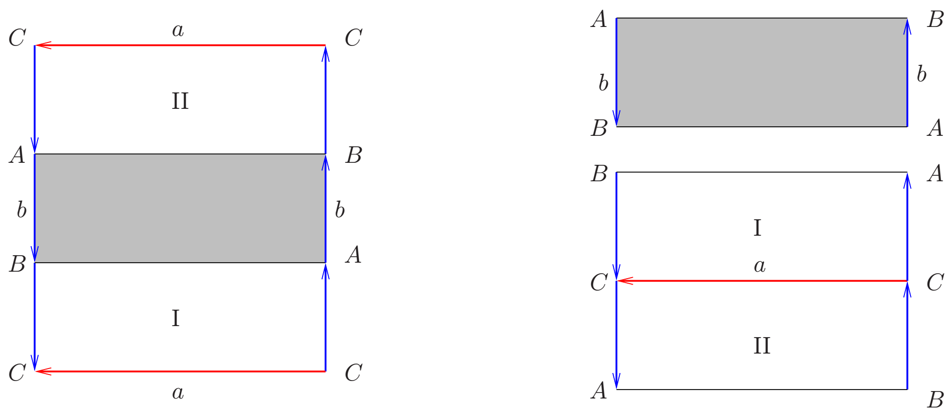

Figure 1.25: Construction of a Klein bottle from two Möbius strips.

flat boundary if we remove a small disk from a cross-cap. For this reason, this version of the Möbius strip is often called a cross-cap. Furthermore, the Klein bottle is obtained by gluing two Möbius strips along their boundaries (See Figure 1.25). This is shown in Massey [34] using the cut and paste method, see Chapter 1, Lemma 7.1.