Notes on the classification of surfaces

Matt Day

Fall 2018

These notes are tailored for my fall 2018 MATH 5703 class.

1 Basic examples

Definition 1.1. Let n be a positive integer. An *n-manifold* is a topological space X such that:

- For every $x \in X$, there is an open neighborhood U of x in X that is homeomorphic to an open ball in $\mathbb{R}^n$.
- X is Hausdorff and second countable.

If X is an n-manifold for some n, then X is a manifold.

The term *topological manifold* is a synonym for manifold, and is sometimes used to distinguish topological manifolds from spaces with even more structure (e.g. smooth manifolds, Riemannian manifolds, complex manifolds, etc.). Remarks:

- The second condition is a technical condition to exclude certain bad examples; the content of the definition is in the first condition. The first condition may be summarized by saying "X is locally homeomorphic to $\mathbb{R}^n$."
- Here is one of the bad examples. Let X be $\mathbb{R} \cup \{0'\}$. The topology on X is the same as that of $\mathbb{R}$, except that 0 and $0'$ are in exactly the same open sets. (Specifically, X has a basis consisting open intervals not containing 0, and sets of the form $(a, b) \cup \{0'\}$ where (a, b) is an open interval containing 0.) This X is the *line with a doubled point*. Notice that X is locally homeomorphic to $\mathbb{R}$, but is not Hausdorff. We do not want to consider X to be a 1-manifold.
- Another bad example is the *long line*, which is described in many standard references including Munkres. This is a space that is locally homeomorphic to $\mathbb{R}$, but is not second countable. The long line is a space where sequences do not always detect limit points, and we do not want to consider it a 1-manifold.

- Two exercises: finite products of manifolds are manifolds, and countable disjoint unions of n-manifolds are n-manifolds. Why is an uncountable disjoint union of manifolds not a manifold?

- An open subspace of an n-manifold is an n-manifold.

- Glueings of manifolds are usually not manifolds; for example, a wedge sum of two manifolds is never a manifold.

- Fact: every connected 1-manifold is homeomorphic to the circle S^1 or to the line $\mathbb{R}^1$.

From here on, we focus on surfaces.

Definition 1.2. A *surface* is a 2-manifold.

Examples:

- The plane $\mathbb{R}^2$ is a surface. It is connected and noncompact. It is contractible. The open disk is homeomorphic to $\mathbb{R}^2$.

- The open annulus $A = \{(x, y) \in \mathbb{R}^2 | 1 < x^2 + y^2 < 2\}$ is a surface. It is connected and noncompact. It has the same homotopy type as S^1, so $\pi_1(A) = \mathbb{Z}$. It is homeomorphic to $\mathbb{R}^1 \times S^1$.

  Another model for A: let $S = [0, 1] \times (0, 1)$; then A is homeomorphic to $S/\sim$, where $\sim$ identifies $(0, t)$ with $(1, t)$ for all $t \in (0, 1)$.

- The open Möbius strip M is a surface. This is easiest to describe as a quotient: let $S = [0, 1] \times (0, 1)$; then M is homeomorphic to $S/\sim$, where $\sim$ identifies $(0, t)$ with $(1, 1-t)$ for all $t \in (0, 1)$. Notice that M is connected and noncompact, and M is homotopy equivalent to the annulus A.

  However, M and A are not homeomorphic. For contradiction, suppose $f\colon M \to A$ is a homeomorphism. Suppose $g\colon S^1 \to M$ is a simple closed curve with g_* an isomorphism of fundamental groups. Then $f \circ g$ is a simple closed curve in A, and f restricts to a homeomorphism from $M - f(S^1)$ to $A - f \circ g(S^1)$. However, $M - f(S^1)$ is connected, but $A - f \circ g(S^1)$ is not, which is a contradiction.

- The sphere $S^2 = \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 = 1\}$ is a surface. Exercise: show this from the definition. S^2 is compact and connected, and $\pi_1(S^2) = 1$. However, S^2 is not contractible, as we will see next semester.

- The torus $T^2 = S^1 \times S^1$ is a surface, since it is a product of 1-manifolds. It is compact and connected, and $\pi_1(T^2) = \mathbb{Z}^2$ (since π_1 respects products).

  There is a standard cell complex structure on T^2. In this, the 0-skeleton is 1 point $*$, there are two edges in the 1-skeleton, and there is one cell in the two-skeleton. If we direct the edges and label them a and b, then the 2-cell is glued in so that its boundary reads out $aba^{-1}b^{-1}$. Figure 1 is a proof-by-picture that this cell complex is homeomorphic to T^2. Then Van

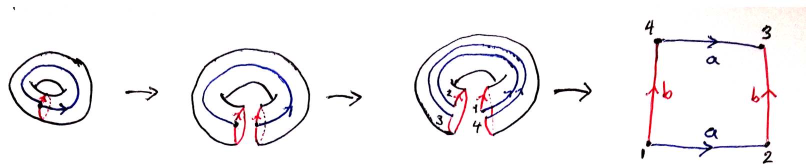

Figure 1: Cutting up a torus to get a square, or, in reverse, glueing up a square to get a torus.

Kampen's theorem gives another proof that $\pi_1(T^2) = \mathbb{Z}^2$, since this cell complex clearly has fundamental group $\langle a, b | aba^{-1}b^{-1} = 1 \rangle$.

What does it look like if we delete a disk from T^2? Let N be a disk that is a neighborhood of $*$, and let S_a and S_b be strips that are neighborhoods of a and b respectively. Then $T^2 - N \cup S_a \cup S_b$ is a disk. We can visualize $N \cup S_a \cup S_b$ as a disk in the plane with strips attached. A careful inspection of the structure shows that the strips are attached untwisted, so that they cross over each other. This is illustrated in Figure 2.

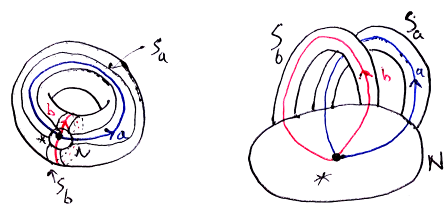

Figure 2: A torus with a disk deleted, expressed as a disk with strips attached.

- The *real projective plane* $\mathbb{R}P^2$ is an important example of a surface. $\mathbb{R}P^2$ is $S^2/\sim$, where $x \sim -x$ for all $x \in S^2$ ($\mathbb{R}P^2$ is S^2 with antipodal points identified). It follows that $\mathbb{R}P^2$ is also $D^2/\sim$, where opposite points on ∂D^2 are identified. This description gives $\mathbb{R}P^2$ a CW-complex structure, as illustrated in Figure 3. This structure has one vertex $*$, one edge a, and one 2-cell, glued in so that its boundary reads off $a \cdot a$. From this structure, it follows that $\pi_1(\mathbb{R}P^2) = \langle a | a^2 = 1 \rangle = \mathbb{Z}/2\mathbb{Z}$.

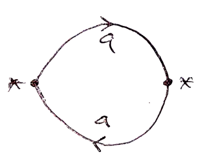

Figure 3: Glue up this bigon to get $\mathbb{R}P^2$.

As we did we the torus, we delete a disk from $\mathbb{R}P^2$ and express the resulting space as a disk with strips glued on. Let N be a disk that is a neighborhood of $*$ above. Let S be a strip that is a neighborhood of a. Then $N \cup S$ is $\mathbb{R}P^2 - D^2$, where D^2 is a disk in the interior of the 2-cell. By inspecting the cell structure, it is clear that S is glued to N with a twist. It follows that $N \cup S$ is a Möbius strip. See Figure 4. In other words, $\mathbb{R}P^2$ is homeomorphic to a Möbius strip with boundary glued to a disk.

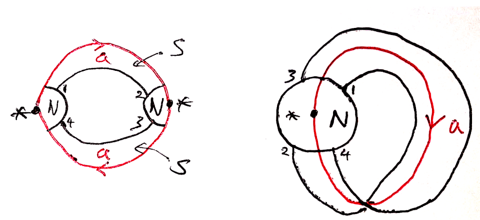

Figure 4: $\mathbb{R}P^2$ with a disk deleted can be rearranged to be a Möbius strip.

2 Connected sums

Definition 2.1. Suppose X and Y are surfaces. The *connected sum* of X and Y, denoted $X\#Y$, is the following space: pick open disks $A \subset X$ and $B \subset Y$ with $\mathrm{cl}(A)$ and $\mathrm{cl}(B)$ both closed disks. Then $X\#Y$ is $X - A$ glued to $Y - B$ by a homeomorphism $f\colon \partial A \to \partial B$.

Proposition 2.2. *For any surfaces X and Y and any choices of A, B and f, $X\#Y$ is a surface.*

Proof. From the construction, it is straightforward to show that $X\#Y$ is Hausdorff and second countable. Suppose $x \in X\#Y$. If x is not on ∂A (the glueing locus) then x has an open neighborhood entirely contained in X or Y, and this neighborhood contains a smaller neighborhood of x that is homeomorphism to an open disk, since X and Y are surfaces. If x is on ∂A, then a neighborhood of x is a glueing of a neighborhood of x in $X - A$ with a neighborhood of x in $Y - B$. Since the closures of A and B are closed disks, these neighborhoods can be arranged so that we are glueing two halves of an open disk together, to form an open disk as a neighborhood of x in $X\#Y$. $\square$

Proposition 2.3. *Suppose X and Y are connected, and at least one of X or Y is T^2 or $\mathbb{R}P^2$. Then the homeomorphism type of $X\#Y$ does not depend on the choices of A, B, or f.*

In fact more is true; all we really need for $X\#Y$ to be well defined is that both X and Y must be connected. However, the only proof that I can see for this uses the classification of surfaces.

Proof. We only sketch the proof.

Claim: if A and B are two different open disks in X, whose closures are closed disks, then $X - A$ is homeomorphic to $X - B$.

We pick points $x \in A$ and $y \in B$. We pick a simple path p from x to y, which must exist since X is connected. We find an open neighborhood U of p that is homeomorphic to a disk. If $A \not\subset U$, then we find a homeomorphism $X \to X$, supported on a neighborhood of A, that sends A to a disk contained in U. Therefore we can assume that $A, B \subset U$, and further, that $A \cap B = \varnothing$. It is then enough to construct a homeomorphism $f\colon X \to X$ that sends A to B. We do this by making f equal to the identity outside of U, and using ideas from analysis to build the restriction of f to U, working with it as if it were a subset of $\mathbb{R}^2$. This shows the first claim.

Now fix $A \subset X$ and $B \subset Y$ as in the definition. Claim: if $f, g\colon \partial A \to \partial B$ are homeomorphisms and f and g are homotopic, then $(X - A) \cup_f (Y - B)$ is homeomorphic to $(X - A) \cup_g (Y - B)$. ∂A has a neighborhood in $(X - A) \cup_f (Y - B)$ that is an open annulus U. We build a homeomorphism $h\colon (X - A) \cup_f (Y - B) \to (X - A) \cup_f (Y - B)$ that is the identity outside of U, and we use the homotopy between f and g to define h on U. We leave the details as an exercise.

Claim: If X is T^2 or $\mathbb{R}P^2$, then there is a homeomorphism $X - A \to X - A$ that reverses the orientation of ∂A. This can be done explicitely, so we leave this claim as an exercise.

Now we put these claims together. We fix standard choices A_0, B_0, and f_0; let Z denote $X\#Y$ using these choices, and let W denote $X\#Y$ using some other choices A, B, and f. Using the first claim, we have homeomorphisms $X - A$ to $X - A_0$ and $Y - B$ to $Y - B_0$, so we can assume $A = A_0$ and $B = B_0$. Since f and f_0 are homeomorphisms between copies of S^1, it must be that they are homotopic, or that $f \circ f_0^{-1}$ reverses orientation. In the second case, we use the third claim to compose f with a self-homeomorphism of X or Y that reverses the orientation of the boundary of the disk. This puts us in the first case, and then the second claim shows that Z is homeomorphic to W. □

So we have shown that connected sums are well defined when both spaces are connected, and at least one of the spaces is T^2 or $\mathbb{R}P^2$.

Proposition 2.4. *The following holds for connected sums:*

- *For any connected surface X, we have that $X\#S^2$ is homeomorphic to X.*
- *For any connected surfaces X and Y, we have that $X\#Y$ is homeomorphic to $Y\#X$.*
- *For any connected surfaces X, Y, and Z, we have that $(X\#Y)\#Z$ is homeomorphic to $X\#(Y\#Z)$.*

In particular, taking connected sums of homeomorphism types of connected surfaces defines a commutative monoid with S^2 as the identity.

The proof is left as an exercise.

How do we compute with connected sums of surfaces?

Suppose we have a surface X expressed as a quotient of an n-gon P by glueing together edges by homeomorphisms. We want to delete a disk A from X, which we can choose wherever we want. We choose A so that it touches exactly one corner of P. This then gives us an $(n+1)$-gon P' that has $X - A$ as a quotient, with the extra edge becoming ∂A.

We do the same thing with a second surface Y, expressed as a quotient of an m-gon Q. We get an $(m+1)$-gon Q' that has $Y - B$ as a quotient. Then we glue P' and Q' together, along the edges that become ∂A and ∂B, to get a polygon R that has $X \# Y$ as a quotient. This is illustrated for tori in Figure 5.

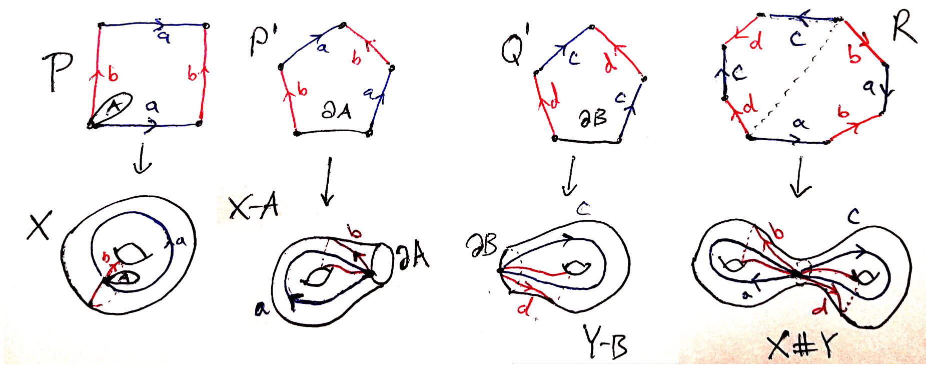

Figure 5: A torus with a disk deleted is a quotient of a pentagon. The connected sum of two tori is a quotient of an octogon, gotten by glueing two pentagons together.

We can also compute connected sums with surfaces represented as disks with strips attached. Suppose C is a disk in X and $X - C$ is represented as a disk with strips. We can choose A anywhere, so we choose it so that it intersects C in a smaller disk, and crosses $X - C$ away from the strips. We suppose that D is likewise a disk in Y, and $Y - D$ is represented as a disk with strips. Again, we choose B to intersect D in a disk. Then when we glue $X - A$ to $Y - B$, we can choose to do it so that we glue $X - A \cup C$ to $Y - B \cup D$, and so that we glue $C - A \cap C$ to $D - B \cap D$. The result is this: to represent $X \# Y$ as the union of a disk and a disk with strips, simply glue $X - C$ to $Y - D$ by attaching two intervals in their boundaries. This is illustrated for tori in Figure 6.

3 The classification

The following theorem has been known since the 1860's.

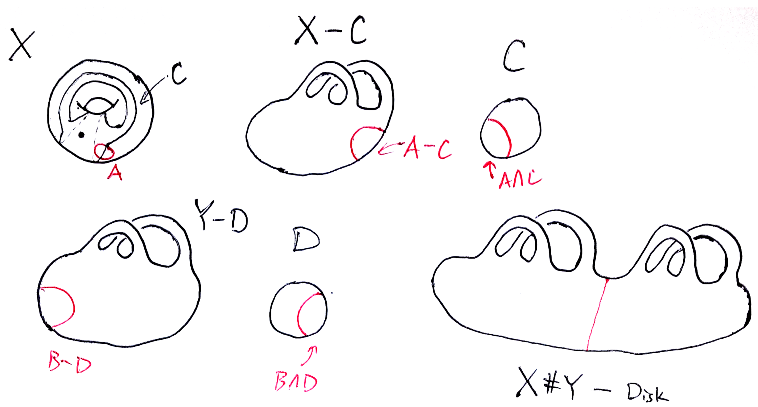

Figure 6: A connected sum of tori, minus a disk, expressed as a disk with strips attached.

Theorem 3.1 (The classification of surfaces)**.** *Let X be a compact, connected surface. Then X is homeomorphic to exactly one of the following:*

- S^2,
- $\#_{i=1}^{g} T^2$, *for some* $g \geq 1$,
- $\#_{i=1}^{g} \mathbb{R}P^2$, *for some* $g \geq 1$.

The connected sum of g copies of T^2 is called the *orientable surface of genus* g. We will sometimes use the notation M_g for it. The connected sum of g copies of $\mathbb{R}P^2$ is called the *non-orientable surface of genus* g, and we will sometimes refer to it as N_g.

It is important to note that we will not prove this theorem starting from the definition of surface, but rather, we will quote the following result as a black box. This result is usually stated in terms of "triangulations", which are CW-complex structures on surfaces satisfying some extra conditions.

Theorem 3.2 (Rado, 1925)**.** *Every compact surface is homeomorphic to a CW-complex.*

First we want to show that the surfaces on the list are all distinct. To do this, we compute their fundamental groups. These computations are interesting in their own right.

Lemma 3.3. *For any* $g \geq 1$,

$$\pi_1(M_g) = \langle a_1, b_1, \dots, a_g, b_g \,|\, a_1 b_1 a_1^{-1} b_1^{-1} \cdots a_g b_g a_g^{-1} b_g^{-1} = 1 \rangle, \text{ and}$$

$$\pi_1(N_g) = \langle a_1, a_2, \dots, a_g \,|\, a_1^2 a_2^2 \cdots a_g^2 = 1 \rangle.$$

Proof. First we show the statement for orientable surfaces. We use the procedure for taking connected sums of surfaces expressed as quotients of polygons. Inductively, we assume that the genus-g surface M_g is a quotient of a $4g$-gon P, with the edge glued according to the labeling $a_1, b_1, \bar{a}_1, \bar{b}_1, \ldots, a_g, b_g, \bar{a}_g, \bar{b}_g$. This appears in Figure 7. Notice that this is true if $g = 1$, and this is our base case. We add an extra edge to P to form P', as in the discussion of connected sums; we add this extra edge between the edge labeled $\bar{b}_g$ and the edge labeled a_1. Let Q be a square with edges glued by $a_{g+1}, b_{g+1}, \bar{a}_{g+1}, \bar{b}_{g+1}$, so that T^2 is the corresponding quotient of Q. Let Q' be the pentagon gotten by adding an extra edge to Q between the edge labeled $\bar{b}_{g+1}$ and the edge labeled a_{g+1}. Then $M_{g+1} = M_g \# T^2$ is the quotient of the labeled polygon that we get by gluing P' to Q' along the extra edges. Since this is the same form, we have inductively proven that M_g is the quotient of the labeled polygon P.

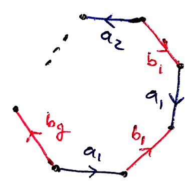

Figure 7: The orientable surface of genus g is this quotient of this $4g$-gon.

By inspecting P, we notice that all vertices of P become a single vertex in M_g, and therefore M_g gets a CW-complex structure with 1 vertex $*$, and $2g$ edges $a_1, \ldots, a_g, b_1, \ldots, b_g$, and a single 2-cell glued in according to the edge labels of the polygon. The given presentation immediately follows by Van Kampen's theorem.

The argument for nonorientable surfaces is similar. An inductive argument, with $\mathbb{R}P^2$ as the base case, shows that the genus-g non-orientable surface N_g is the quotient of the labeled $2g$-gon P with edge labels $a_1, a_1, \ldots, a_g, a_g$. This is shown in Figure 8. All vertices become the same vertex in N_g, which then has a CW-complex structure with one vertex, $2g$ edges, and one 2-cell. The 2-cell is glued in according to the boundary of P. The lemma immediately follows by Van Kampen's theorem. $\square$

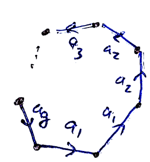

Figure 8: The non-orientable surface of genus g is this quotient of this $2g$-gon.

Remark: Just because these presentations look different, it does not immediately follow that the groups are not isomorphic. For example, it turns out that the groups

$$\langle a, b, c | a^2 b^2 c^2 = 1 \rangle \quad \text{and}$$
$$\langle a, b, c | a^2 b c b^{-1} c^{-1} = 1 \rangle$$

are isomorphic. (These are presentations of the fundamental groups of two different cell complexes that are homeomorphic to the same surface, as we will see below.)

Proposition 3.4. *No two surfaces in the list from the classification are homeomorphic.*

Proof. In fact, no two surfaces in that list are homotopy equivalent, and no two of them even have isomorphic fundamental groups. To see this, we take the abelianizations of the fundamental groups. We know $\pi_1(S^2) = 1$, so its abelianization is the trivial group.

To get a presentation for the abelianization of G from a presentation of G, we simply add relations stating that all pairs of generators commute. If we add these commutators to the presentation for $\pi_1(M_g)$, then the one relation becomes a consequence of the commutators. Essentially, it is an abelian group with $2g$ generators and no relations. Then $\mathrm{Ab}(\pi_1(M_g)) = \mathbb{Z}^{2g}$.

On the other hand, the relation from the presentation for $\pi_1(N_g)$ persists after abelianizing. Let A be the free abelian group spanned by $[a_1], \ldots, [a_g]$, and let K be the subgroup of A spanned by $2[a_1] + \cdots + 2[a_g]$. Then $\mathrm{Ab}(\pi_1(M_g)) = A/K$. We use the following basis for A: $[a_1] + \cdots + [a_g], [a_2], \ldots, [a_g]$. In this basis, it is clear that $A/K \cong (\mathbb{Z}/2\mathbb{Z}) \times \mathbb{Z}^{g-1}$. So this is our description of $\mathrm{Ab}(\pi_1(M_g))$.

Since no two of these abelianizations are isomorphic, it follows that no two surfaces from the list are homeomorphic. □

One immediate question that should arise after reading the classification statement is "what about $T^2 \# \mathbb{R}P^2$?" Since it is a compact, connected surface, it must be homeomorphic to one on the list. However, it does not appear in that form in the list.

Lemma 3.5. *We have $T^2 \# \mathbb{R}P^2 \cong \mathbb{R}P^2 \# \mathbb{R}P^2 \# \mathbb{R}P^2$.*

Proof. We represent the complement of a disk in $T^2 \# \mathbb{R}P^2$ as a disk with strips attached. Call this X. Then we manipulate X by sliding ends of strips along the boundary of X. Since a slide move like this does not change the homeomorphism type of X, we get back a surface homeomorphic to $T^2 \# \mathbb{R}P^2$ if we re-attach a disk to X. However, a short sequence of slide moves turns X into a disk with strips attached that we can recognize as the complement of a disk in $\mathbb{R}P^2 \# \mathbb{R}P^2 \# \mathbb{R}P^2$. In the first and last step, we recognize the complement of a disk in a connected sum using the procedure given shortly after the definition of connected sum. The proof of this lemma is given by pictures in Figure 9. □

Finally we can give the proof of the classification theorem. I first saw a proof like this one in an undergraduate class taught by Michael Starbird.

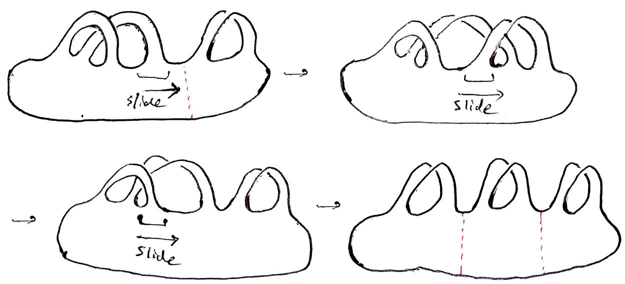

Figure 9: Rearranging $T^2\#\mathbb{R}P^2$ to recognize it as $\mathbb{R}P^2\#\mathbb{R}P^2\#\mathbb{R}P^2$.

Proof of classification. Let M be a compact, connected surface. By Rado's theorem, there is a CW-complex structure X on M. Since M is compact, this X has finitely many cells. Since M is connected, the one-skeleton X^1 is connected.

Let T be a maximal tree for X^1, and let $N \subset M$ be a closed neighborhood that has a deformation retraction to T. Since T is a tree, N is a closed disk. For each edge e in $X^1 - T$, let S_e be a closed neighborhood of e that has a deformation retraction to e, and such that $S_e - N$ consists of a single strip (a disk). Let Y be the union of N with all of the S_e for all e in $X^1 - T$. Then Y is a disk with strips. Further, $M - Y$ consists only of interiors of 2-cells from X, and therefore $M - Y$ is a union of disks.

Now we show that M is homeomorphic to some surface in the list. We show this by induction on the number of strips in Y (or equivalently, by the number of edges in X^1 outside of T). We break into three cases, depending on what Y looks like.

Case 1: some strip in Y is glued in with a half-twist. We visualize Y as N with a twisted strip S_1 glued in, along with some collections of ends of strips glued in on either side of S_1. Call these collections A and B. We slide A to the right until it exits the arch of S_1. This is illustrated in Figure 10. Now we can recognize M as a connected sum of $\mathbb{R}P^2$ and a surface M'. However, a complement of disks in M' can be built with fewer strips, so the inductive hypothesis applies, and M' is a surface from our list. Since $M = \mathbb{R}P^2\#M'$, this is in our list if M' is S^2 or a nonorientable surface of genus g. If M' is an orientable surface of genus g, then we need to apply Lemma 3.5 repeatedly to recognize M as a non-orientable surface of genus $2g + 1$.

Case 2: no strips in Y are twisted, but some pair of strips cross each other. We visualize Y as N with two untwisted strips S_1 and S_2 glued in, along with some collections of ends of strips glued in in between and outside the ends of S_1 and S_2. Call these collections A, B, C, and D. In several steps, we slide these collections out of the region under the arches of S_1 and S_2. This is illustrated in Figure 11. Now we can recognize M as a connected sum of T^2 and a surface M'. Again, a complement of disks in M' can be built using fewer strips than M, so the inductive hypothesis applies and M' is a surface from the list. A

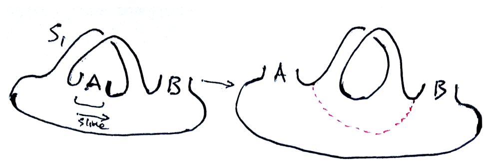

Figure 10: Processing a surface with a twisted strip to find a Möbius strip.

priori, M' might be a non-orientable surface (although it turns out that this is never the case if there are no half-twists on the strips). If M' is S^2 or an orientable surface of genus g, then $M \cong T^2 \# M'$ is a surface from our list. If M' is a non-orientable surface of genus g, then we apply Lemma 3.5 once, and we recognize M as a non-orientable surface of genus $g + 2$.

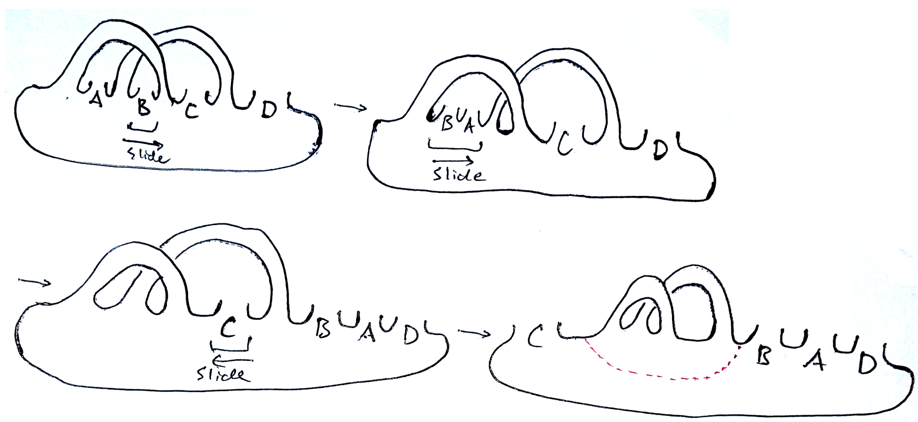

Figure 11: Processing a surface with two untwisted, overlapping strips to find a torus with boundary.

Figure 12: Left: an untwisted, empty arch. Right: since no strips cross, we can recognize a connected sum.

Case 3: no strips in Y are twisted, and no pair of strips in Y cross each other. Pick some strip S_1. If one side of S_1 has no strips at all, then since S_1 is untwisted, that side of S_1 consists of a disk in M. By gluing this disk in, we express M using fewer strips, and the inductive hypothesis applies to M. If both sides of S_1 have strips, then we can recognize M as a connected sum. One of the summands consists of S_1 together with the strips on one side of it, and the other summand consists of the strips on the other side of S_1. So we recognize M as a connected sum of two surfaces that can be expressed using fewer strips. Then M is a connected sum of two surfaces on the list. A priori, this could be a mix of orientable and nonorientable surfaces from the list, and therefore we would have to apply Lemma 3.5 to recognize M. (It turns out that M would have to be S^2 to end up in case 3, so in practice, we would never have to apply that Lemma in this case.) Both parts of this case are illustrated in Figure 12.

$\square$