The Kirby Torus Trick for Surfaces

Allen Hatcher

In the late 1960s the breakthrough that led to the Kirby–Siebenmann classification of PL structures on high-dimensional topological manifolds was the Kirby torus trick, appearing first in the paper [K]. It still seems quite amazing that a construction this simple could so easily convert difficult topological problems into much more manageable ones in the smooth or PL category.

A few years later it was shown in a paper by A.J.S. Hamilton [H] that the torus trick could also be applied in dimension 3 to give new proofs of the earlier theorems of Moise on the existence and uniqueness of PL structures on 3-manifolds. It is no surprise then that the method can be scaled down even further to prove the corresponding facts for surfaces. This is what we do here, and in fact we take the small extra step to produce smoothings instead of triangulations. This approach to smoothing in dimension 2 is undoubtedly well known in some circles, though we do not know of a written source for it, so our purpose is just to provide such a source.

Throughout the paper we take the term "surface" to mean a 2-dimensional manifold, possibly with boundary. We assume manifolds are Hausdorff and are covered by a countable number of coordinate charts.

Theorem A. *Every topological surface has a smooth structure.*

Theorem B. *Every homeomorphism between smooth surfaces is isotopic to a diffeomorphism.*

The latter result implies that smooth structures are unique not just up to diffeomorphism, but up to isotopy. (Just apply the theorem to the identity map between two different smooth structures on the same surface.)

The proof of Theorem A will show that the smooth structure can be chosen to agree with a given smooth structure on a specified open set. Similarly, Theorem B can be refined to the statement that if a homeomorphism is already a diffeomorphism near the boundary of a surface, the isotopy can be chosen to be fixed in a neighborhood of the boundary.

The special feature of the proofs of Theorems A and B using the torus trick is that almost no point set topology is needed. This is replaced instead by a few basic facts about smooth surfaces whose proofs use only smooth techniques. In particular the topological Schönflies theorem, which is an ingredient for other proofs of Theorems A and B, is not needed here. The first proof that surfaces can be triangulated (and hence smoothed) is generally attributed to Radó in 1925 [R]. That proof used the Schönflies theorem without explicitly saying so, but the later exposition of Radó's proof in [AS] makes the dependence clear. The fact that homeomorphisms of surfaces are isotopic to PL homeomorphisms and hence to diffeomorphisms was proved by Epstein in [E], although at least for closed orientable surfaces this can be derived from results of Baer in 1928 [B] on mapping class groups.

The two theorems will be easy to deduce from a result about smoothing handles in surfaces, which we state now. There are three cases, for 0-handles, 1-handles, and 2-handles, and the proofs are somewhat different in each case, with the torus trick being needed only for 0-handles. The three cases could be combined into a single statement, but here are the separate statements.

Handle Smoothing Theorem. *Let S be a smooth surface. Then:*
(0) *An embedding $\mathbb{R}^2 \to S$ can be isotoped to be a smooth embedding in a neighborhood of the origin, staying fixed outside a larger neighborhood of the origin.*
(1) *An embedding $D^1 \times \mathbb{R} \to S$ which is a smooth embedding near $\partial D^1 \times \mathbb{R}$ can be isotoped to be a smooth embedding in a neighborhood of $D^1 \times 0$, staying fixed outside a larger neighborhood of $D^1 \times 0$ and near $\partial D^1 \times \mathbb{R}$.*
(2) *An embedding $D^2 \to S$ which is a smooth embedding in a neighborhood of ∂D^2 can be isotoped to be a smooth embedding on all of D^2, staying fixed in a smaller neighborhood of ∂D^2.*

It should be remembered that a smooth embedding is more than just a topological embedding which is a smooth map, since the differential must also be nonsingular at each point.

Assuming the Handle Smoothing Theorem, let us now deduce Theorems A and B. The proofs will use a couple rather standard results about smooth surfaces, and these will be stated just as facts, with proofs given (or sketched) at the end of the paper for those who would like to see them. The same procedure will be followed later when we prove the Handle Smoothing Theorem.

Proof of Theorem A. For a surface S without boundary, choose a system of coordinate charts $h_i : \mathbb{R}^2 \to S$, $i = 1, 2, \cdots$. We build a smooth structure on $U_n = \cup_{i \le n} h_i(\mathbb{R}^2)$ by induction on n, starting with the smooth structure induced by h_1 on U_1. Consider the induction step of extending the smooth structure from U_{n-1} to U_n. Let $W = h_n^{-1}(U_{n-1})$. This is an open set in $\mathbb{R}^2$.

Fact 1. An open set $W \subset \mathbb{R}^2$ can be triangulated so that the size of the simplices approaches 0 at the frontier of W.

Using the Handle Smoothing Theorem we can then isotope h_n to be smooth on W by first smoothing it near the vertices, then near the 1-skeleton of the triangulation, then on all of W. Since the simplices become small near the frontier of W, we can extend this isotopy on W to an isotopy defined on all of $\mathbb{R}^2$ by taking the constant isotopy on the complement of W. After this has been done, we can use the new h_n to extend the smooth structure from U_{n-1} to U_n.

In case S has a nonempty boundary, we will use the fact that ∂S has a collar neighborhood in S. This is a general fact about topological manifolds, with an elementary proof due to Connelly [C] (included as Proposition 3.42 in the author's algebraic topology textbook) using a partition of unity to piece together local collars. A collar neighborhood of ∂S has a smooth structure since 1-manifolds are smoothable. Then the inductive procedure above extends this to a smooth structure on all of S. $\square$

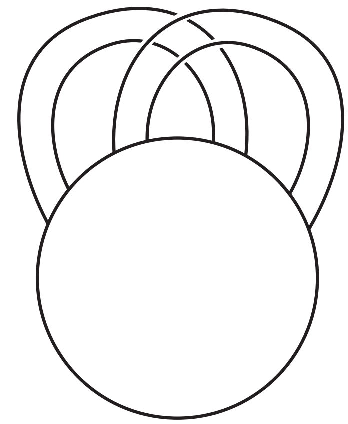
Proof of Theorem B. Let $f: S \to S$ be a homeomorphism of the smooth surface S. Assume first that S has empty boundary.

Fact 2. S has a smooth triangulation.

Using this, we can first apply 0-handle smoothing to isotope f to be smooth near all vertices, then apply 1-handle smoothing to isotope the new f to be smooth near the 1-skeleton of the triangulation, and finally apply 2-handle smoothing to make the resulting f smooth on all of S.

When ∂S is nonempty, we can choose a smooth collar on ∂S. Using this, f can be isotoped to be independent of the collar parameter in a smaller collar. The restriction of f to ∂S is isotopic to a diffeomorphism, and this isotopy can be extended to an isotopy of f, damped down to the constant isotopy as one moves into the collar away from ∂S. This gives a reduction to the case that f is already smooth on a collar neighborhood of ∂S. In this case choose a smooth triangulation of S that restricts to a triangulation of a subcollar neighborhood of ∂S, and then apply the smoothing procedure from the preceding paragraph to smooth f near the simplices not in this subcollar. $\square$

Smoothing 0-handles.



This is the hardest case, where the torus trick is used. We view the torus T as the orbit space $\mathbb{R}^2/\mathbb{Z}^2$, with a basepoint 0 that is the image of $0 \in \mathbb{R}^2$. Deleting some other point $* \in T$ yields a punctured torus T'. We can immerse T' in $\mathbb{R}^2$, with $0 \in T'$ mapping to $0 \in \mathbb{R}^2$, by viewing T' as the interior of the surface obtained from a disk by attaching two 1-handles, then letting the immersion be an embedding of the disk, with the two 1-handles individually embedded so that their images cross in $\mathbb{R}^2$, as shown in the figure.

Let $h:\mathbb{R}^2 \to S$ be a topological embedding into the smooth surface S. Via h, the smooth structure on S induces a smooth structure $\mathcal{S}$ on $\mathbb{R}^2$. This is likely to be quite different from the usual smooth structure if h is far from being a smooth embedding. For example, smooth curves in S that are in the image of h pull back via h^{-1} to curves in $\mathbb{R}^2$ that can be quite wild when looked at with the naked eye but are in fact smooth in the structure $\mathcal{S}$.

The immersion $T' \to \mathbb{R}^2$ pulls back the smooth structure $\mathcal{S}$ on $\mathbb{R}^2$ to a smooth structure on T' that we denote $T'_{\mathcal{S}}$.

Fact 3. There is a compact set in $T'_{\mathcal{S}}$ whose complement is diffeomorphic to $S^1 \times \mathbb{R}$.

This allows us to extend the smooth structure $T'_{\mathcal{S}}$ to a smooth structure $T_{\mathcal{S}}$ on T.

Fact 4. Every smooth structure on a torus $S^1 \times S^1$ is diffeomorphic to the standard smooth structure.

Thus there is a diffeomorphism $g:T_{\mathcal{S}} \to T$. Note that g, viewed just as a homeomorphism from T to T, is likely to be as complicated as the original h locally.

We can normalize g so that it takes 0 to 0 by composing with rotations in the S^1 factors of the target if necessary. We can then further normalize so that g induces the identity homomorphism of $\pi_1(T,0) = \mathbb{Z}^2$ by composing with a diffeomorphism in the target given by a suitable element of $GL_2(\mathbb{Z})$ acting on $\mathbb{R}^2/\mathbb{Z}^2$. After this normalization, g lifts to the universal covers as a diffeomorphism $\widetilde{g}:\mathbb{R}^2_{\mathcal{S}} \to \mathbb{R}^2$ fixing $\mathbb{Z}^2$, where the subscript $\mathcal{S}$ denotes the smooth structure lifted from $T_{\mathcal{S}}$.

The key point of these constructions is that $\widetilde{g}$, as a homeomorphism $\mathbb{R}^2 \to \mathbb{R}^2$ is bounded, meaning that the distance $|\widetilde{g}(x) - x|$ is less than a fixed constant for all $x \in \mathbb{R}^2$. This is certainly true as x varies over the square I^2 by compactness, and then it holds over all of $\mathbb{R}^2$ by periodicity and the fact that $\widetilde{g}$ fixes $\mathbb{Z}^2$.

If we identify $\mathbb{R}^2$ with the interior of D^2 by a radial reparametrization, then $\widetilde{g}$ becomes a homeomorphism of the interior of D^2 that extends via the identity on ∂D^2 to a homeomorphism $G: D^2 \to D^2$, as a result of the boundedness condition. This can be seen by considering polar coordinates (r, θ), since as a disk in $\mathbb{R}^2$ of fixed radius moves out to infinity, the variation in the θ-coordinates of points in the disk approaches zero, and after the radial reparametrization the variation in the r-coordinates also approaches zero. We can choose the identification of $\mathbb{R}^2$ with the interior of D^2 to be the identity near 0, and then $G = \widetilde{g} = g$ near 0. Finally, we can extend $G: D^2 \to D^2$ to a homeomorphism $G: \mathbb{R}^2 \to \mathbb{R}^2$ that is the identity outside D^2.

By the Alexander trick G is isotopic to the identity. This isotopy can be obtained by replacing the unit disk D^2 in the preceding paragraph by disks of progressively smaller radius centered at 0, limiting to radius zero. Reversing the time parameter of this isotopy, we obtain an isotopy G_t with G_0 the identity and $G_1 = G$.

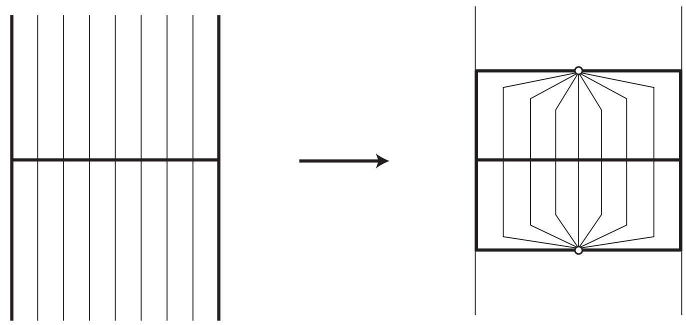
We check now that the desired isotopy of h is given by $h_t = hG_t^{-1}$. We have $h_0 = h$ since G_0 is the identity. Also h_t is stationary outside D^2 since G_t is the identity there. Since G_t fixes the origin for all t, we have $h_t(0) = h(0)$ for all t. Finally, to check that h_1 is smooth near 0 with respect to the standard smooth structure on $\mathbb{R}^2$, observe that $G_1^{-1} = G^{-1}$ is a diffeomorphism from the standard smooth structure to the smooth structure $\mathcal{S}$ near 0, and h carries this smooth structure to the smooth structure that was given on S. $\square$

Smoothing 1-handles.

We are given an embedding $h: D^1 \times \mathbb{R} \to S$ which is smooth near $\partial D^1 \times \mathbb{R}$. Via h the smooth structure on S pulls back to a smooth structure $\mathcal{S}$ on $D^1 \times \mathbb{R}$ which agrees with the standard structure near $\partial D^1 \times \mathbb{R}$.

Fact 5. Every smooth structure on $D^1 \times \mathbb{R}$ that is standard near $\partial D^1 \times \mathbb{R}$ is diffeomorphic to the standard structure via a diffeomorphism that is the identity near $\partial D^1 \times \mathbb{R}$.

Thus there is a diffeomorphism $g: (D^1 \times \mathbb{R})_{\mathcal{S}} \to D^1 \times \mathbb{R}$ which is the identity near $\partial D^1 \times \mathbb{R}$. Choose an embedding $e: D^1 \times \mathbb{R} \to D^1 \times \mathbb{R}$ which is the identity near $D^1 \times 0$ and which has image $(D^1 \times D^1) - (0 \times \partial D^1)$, as indicated in the figure below.



Via e, g defines a homeomorphism of the image of e, and this homeomorphism can be extended via the identity outside this image to yield a homeomorphism $G:\mathbb{R}^2 \to \mathbb{R}^2$ that agrees with g near $D^1 \times 0$ and is the identity outside $D^1 \times D^1$ and near $\partial D^1 \times \mathbb{R}$. If we perform the Alexander isotopy on G, using the square $D^1 \times D^1$ this time instead of the disk D^2, we obtain, after reversing the time parameter, an isotopy $G_t:\mathbb{R}^2 \to \mathbb{R}^2$ from the identity to G. This is stationary near $\partial D^1 \times \mathbb{R}$ and outside $D^1 \times D^1$. We only need the restriction of G_t to $D^1 \times \mathbb{R}$, which we still denote $G_t: D^1 \times \mathbb{R} \to D^1 \times \mathbb{R}$. The desired isotopy of h is then $h_t = hG_t^{-1}$. $\square$

Smoothing 2-handles.

In this case we start with an embedding $h: D^2 \to S$ which is smooth near ∂D^2. The smooth structure on S pulls back to a smooth structure $D^2_{\mathbb{S}}$ which is standard near ∂D^2.

Fact 6. Every smooth structure on D^2 that is standard near ∂D^2 is diffeomorphic to the standard structure via a diffeomorphism that is the identity near ∂D^2.

So there is a diffeomorphism $g: D^2_{\mathbb{S}} \to D^2$ which is the identity near ∂D^2. The Alexander trick yields an isotopy $g_t: D^2 \to D^2$ from the identity to g, and then an isotopy from h to a smooth embedding is given by $h_t = hg_t$, fixing a neighborhood of ∂D^2. $\square$

Proofs of the Facts.

Proof of Fact 1. The horizontal and vertical lines in $\mathbb{R}^2$ through points in the integer lattice $\mathbb{Z}^2$ divide $\mathbb{R}^2$ into closed unit squares. To start, choose all such squares that are contained in the open set W. For the squares that are not contained in W, divide each of these into four squares with sides of length $1/2$ and take all the smaller squares that lie in W. Now repeat this process indefinitely, at each stage subdividing the squares of the preceding stage not contained in W into four subsquares and choosing the ones contained in W. The union of all the squares chosen by this iterative process is W since the distance from any point in W to $\mathbb{R}^2 - W$ is positive. The chosen squares give a cellulation of W, and this can be subdivided to a triangulation by adding a new vertex at the center of each square of the cellulation and joining the vertices in the boundary of this square to its center.

Proof of Fact 2. First we construct a smooth cellulation with 2-cells that are polygons, not necessarily triangles. Choose a Morse function $S \to \mathbb{R}$ that is proper (inverse images of compact sets are compact) and has all its critical points on distinct levels. Cutting S along noncritical levels separating all critical levels then gives rise to a decomposition of S into pieces that are diffeomorphic to disks, annuli, pairs of pants, or twisted pairs of pants (an annulus with a 1-handle attached to one of its boundary circles so as to produce a nonorientable surface). A twisted pair of pants can be further cut along a circle in its interior to produce an ordinary pair of pants and a Möbius band. Thus S is decomposed into disks, annuli, pairs of pants, and Möbius bands, glued together along their boundary circles. From this a cellulation is easily obtained by inserting one vertex in each circle, then in each annulus connecting the two vertices in its boundary by an arc cutting the annulus into a square, in each pair of pants connecting the three vertices in its boundary by two arcs cutting the pair of pants into a polygon (a heptagon in fact), and in each Möbius band inserting an edge to cut it into a triangle.

Having the smooth cellulation, we can subdivide it to a triangulation by coning off to a new vertex in the interior of each polygon.

Proof of Fact 3. As in the proof of Fact 2, $T'_{\mathbb{S}}$ can be cut along a collection of disjoint smooth circles C_i to produce pieces P_j that are disks, annuli, pairs of pants, and Möbius bands, but the last of these cannot occur since $T'_{\mathbb{S}}$ is orientable. (Orientability can be defined purely topologically, either in terms of local homology groups or in terms of local fundamental groups, which are fundamental groups of neighborhoods of a point with the point deleted.) The pattern in which the pieces P_j are assembled to form $T'_{\mathbb{S}}$ is described by a graph G having one vertex for each P_j and one edge for each circle C_i. There is a quotient map $q : T'_{\mathbb{S}} \to G$ projecting a product neighborhood of each C_i onto its interval factor, which becomes the corresponding edge of G, and collapsing the complementary components of these annuli to points, the vertices of G. The induced homomorphism $q_* : \pi_1 T'_{\mathbb{S}} \to \pi_1 G$ is surjective and in fact split since q has a section up to homotopy. Thus the free group $\pi_1 G$ is a quotient of $\pi_1 T'_{\mathbb{S}}$, so it is finitely generated. This implies that there is a finite subgraph $G_0 \subset G$ such that the closure of $G - G_0$ consists of a finite number of trees. Only one of these trees can be noncompact since $T'_{\mathbb{S}}$ has only one end. The one-endedness also implies that this noncompact tree consists of an infinite subtree homeomorphic to $[0, \infty)$ with finite subtrees attached to it. We can eliminate these finite subtrees by deleting the edges leading to vertices corresponding to disk pieces P_j, discarding the circles C_i corresponding to these edges. (These simplifications in G can in fact be realized by simplifications in the Morse function having the C_i's as level curves, by canceling a saddle and a local extremum in the case of deleting a C_i where a disk attaches to a pair of pants, and changing the level of a local extremum in the case that C_i splits a disk from an annulus.)

Thus the end of $T'_{\mathbb{S}}$ consists of an infinite sequence of annuli glued together to form an infinite cylinder. The glueings are smooth so the cylinder is diffeomorphic to a standard cylinder.

Proof of Fact 4. We could just quote the classification of smooth closed surfaces, or we can give an argument similar to the one for Fact 3. The associated graph G is now a finite graph with $\pi_1 G$ a quotient of the abelian group $\pi_1 T_{\mathbb{S}}$, so $\pi_1 G = \mathbb{Z}$. We can reduce G to a circle by eliminating vertices corresponding to disk pieces as before, and then we see that $T_{\mathbb{S}}$ consists of annuli glued together in a cyclic pattern, so $T_{\mathbb{S}}$ is diffeomorphic to a torus or a Klein bottle, but the latter is ruled out since $\pi_1 T_{\mathbb{S}}$ is abelian.

Proof of Fact 5. The projection $D^1 \times \mathbb{R} \to \mathbb{R}$ can be perturbed to a Morse function $f : (D^1 \times \mathbb{R})_{\mathbb{S}} \to \mathbb{R}$ staying fixed near $\partial D^1 \times \mathbb{R}$ where it is already smooth without critical points. The function f is proper, and we can assume all its critical points lie in distinct levels. Noncritical levels then consist of a single arc and finitely many circles. Cutting along these curves produces pieces that are disks, annuli, or pairs of pants as in the earlier cases, and there are now also two new kinds of pieces: rectangles, and rectangles with an open disk removed. The associated graph G is a tree since $\pi_1(D^1 \times \mathbb{R}) = 0$, and as in the proof of Fact 3 we can modify f staying fixed near $\partial D^1 \times \mathbb{R}$ until G is homeomorphic to $\mathbb{R}$. Thus f is a proper Morse function on $(D^1 \times \mathbb{R})_{\mathbb{S}}$ without critical points, so f is the second coordinate of a diffeomorphism $g : (D^1 \times \mathbb{R})_{\mathbb{S}} \to D^1 \times \mathbb{R}$ which is the identity near $\partial(D^1 \times \mathbb{R})_{\mathbb{S}}$. The first coordinate of g is obtained by flowing along the gradient vector field of f.

Proof of Fact 6. Here we extend the radial coordinate in D^2 near ∂D^2 to a Morse function $f : D^2_{\mathbb{S}} \to (0,1]$ with $f^{-1}(1) = \partial D^2_{\mathbb{S}}$ and all critical points in the interior of $D^2_{\mathbb{S}}$. The associated graph is a tree, and it can be simplified as before until it is a point, so f has a single critical point, of index 0. Then f together with the flow lines of the gradient field can be used to construct a diffeomorphism $g : D^2_{\mathbb{S}} \to D^2$ which is the identity near $\partial D^2_{\mathbb{S}}$.

References

[AS] L. V. Ahlfors and L. Sario, *Riemann Surfaces*, 1960.

[B] R. Baer, Isotopie von Kurven auf orientierbaren, geschlossenen Flächen und ihr Zusammenhang mit der topologischen Deformation der Flächen, *J. Reine u. Angew. Math.* 159 (1928), 101–116.

[C] R. Connelly, A new proof of Brown's collaring theorem, *Proc. A.M.S.* 27 (1971), 180–182.

[E] D. B. A. Epstein, Curves on 2-manifolds and isotopies, *Acta Math.* 115 (1966), 83–107.

[H] A. J. S. Hamilton, The triangulation of 3-manifolds, *Quart. J. Math. Oxford* 27 (1976), 63–70.

[K] R. C. Kirby, Stable homeomorphisms and the annulus conjecture, *Annals of Math.* 89 (1969), 575–582.

[R] T. Radó, Über den Begriff der Riemannschen Fläche, *Acta Sci. Math. Szeged.* 2 (1925), 101–121.

[S] L. Siebenmann, The Osgood–Schoenflies theorem revisited, *Russian Math. Surveys* 60 (2005), 645–672.