Math. Res. Lett.
Volume 25, Number 6, 1827–1836, 2018

Dehn's Lemma for immersed loops

Michael Freedman and Martin Scharlemann

Suppose δ is a generic immersed closed curve in the boundary of a 3-manifold M and δ is null-homotopic in M. Then δ can be displaced by a height function in a collar of the boundary of M so that the resulting simple closed curve in the collar bounds a disk in M.

We know five theorems whose conclusion is the existence of an embedded disk, perhaps with additional structure, in some larger space. Each introduced an influential technique and had broad consequences. They are: (1) - 1913 Boundary continuity of the Riemann mapping to Jordan domains (Carathéodory [1], Osgood and Taylor [2]); Carathéodory's proof introduced "external length"; applications to quasifuschian groups. (2) - 1944, the Whitney disk (with appropriate normal frame extension) [3]; applications to Whitney embedding theorem, h-cobordism theorem. (3) - 1957 Dehn's Lemma - loop theorem (Papakyriakopoulos [4]); correctly treated triple points; applications: hierarchy in Haken manifolds, Thurston's geometrization theorem. (4) - 1982 Disk embedding theorem (Freedman [5]); used decomposition theory to identify Casson handles; application: topological classification of simply-connected 4-manifolds. (5) - Existence of Pseudoholomorphic disks (Gromov [6]); brought Kahler manifold techniques into symplectic context; applications: the nonsqueezing theorem, Seiberg-Witten invariants, quantum cohomology. This paper is a comment on (3); we prove a simply stated extension of Dehn's Lemma.

Let M be a 3-manifold with boundary ∂M, and $\delta : S^1 \looparrowright \partial M$ be a generic immersion, where $S^1 := [0, 2\pi]/_{0 \equiv 2\pi}$. By generic, we mean that δ has only simple crossings. We say that δ' is δ "displaced by a height function f" if $\delta'(\theta) = (\delta(\theta), f(\theta))$, where $f : S^1 \to (0, \epsilon)$ is a Morse function and $[0, \epsilon]$ is a normal collar coordinate on ∂M into M. We call a simple closed curve (scc) $\alpha \subset \mathrm{int}(M)$ unknotted iff α bounds an embedded disk $\bar{\alpha} : D^2 \hookrightarrow \mathrm{int}(M)$.

Theorem 1. *Let $\delta : S^1 \looparrowright \partial M$ be a generic immersed loop so that the composition into M is null homotopic. There is a height function $f : S^1 \to (0, \epsilon)$ so that $\delta' = (\delta, f) : S^1 \hookrightarrow \mathrm{int}(M)$ is unknotted.*

The theorem readily implies two familiar facts:

1) Dehn's Lemma: For any scc $\delta \subset \partial M$ which is null homotopic in M there is a properly embedded disk $(D, \partial D) \subset (M, \partial M)$ with ∂D parameterizing δ.

2) Given any knot diagram there is always a way to rechoose the crossings to produce an unknot. (This is the case when $(M, \partial M) \cong (B^3, \partial B^3)$ is a 3-ball.)

Regarding 1: A short argument connects the special case of the theorem where δ is one-to-one to Dehn's Lemma. Let C be the annular collar joining δ to δ' and D a disk with boundary δ'. If C and D are in general position initially there may be arcs of intersection, but a perturbation of D near $\partial D = \delta'$ starting from an outermost arc ensures that C and D intersect only in sccs contained in $\text{int}(C)$ and $\text{int}(D)$. Let $\sigma \subset D$ be an innermost circle of intersection bounding a subdisk $\overline{\sigma} \in D$. σ may be paired with either an essential or inessential scc in C. Perform disk exchanges to modify C until either $C \cup D$ is an embedded disk or some innermost σ is paired to an essential scc in C. In this case a final cut and glue operation yields an embedded disk with boundary δ.

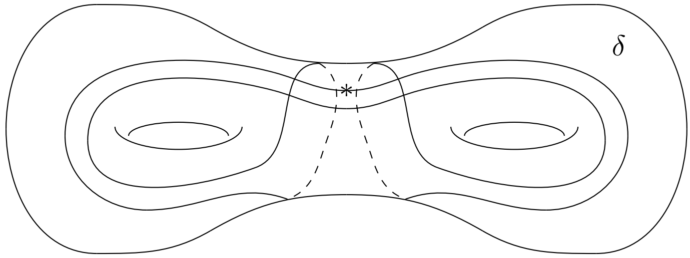
Regarding 2: We should note a subtlety. The height function produced in the theorem may, in general, be more complicated than the familiar height function which solves the knot diagram problem (see Figure 1b). For knot diagrams the unknotting function may be taken to be any function with a unique local maximum and unique local minimum.

Actually Theorem 1 is a corollary of a stronger Theorem 0, better adapted to the required induction.

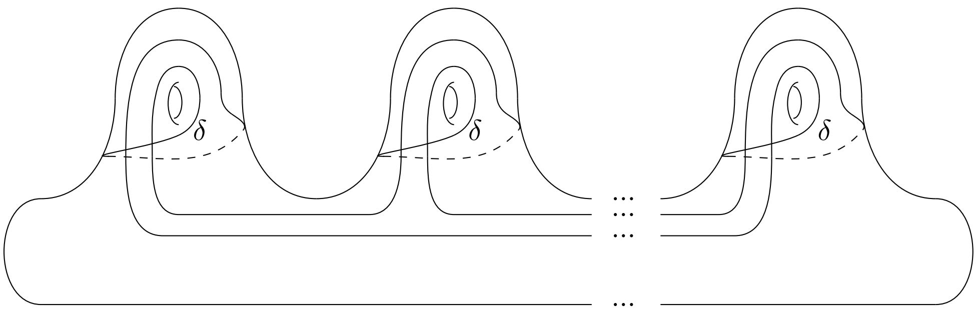
Theorem 0. *Let $\delta : S^1 \looparrowright \partial M$ be a generic immersed loop with base point $*$ ($*$ is assumed disjoint from multiple points of δ) whose composition into M is null homotopic. There is a height function $f : S^1 \to [0, \epsilon)$, $f(*) = 0$, $f(S^1 \backslash *) \subset (0, \epsilon)$, so that $\delta' = \{(\delta(\theta), f(\theta)\}$ bounds an embedded disk $\Delta \subset M$ with $\Delta \cap \partial M = *$. We say δ' collapses to $*$, and call Δ a "lollypop" for δ'.*

Proof. No methods post dating Papakyriakopoulos [4] are required. He would have found this proof rather easy to understand and perhaps to generate. First we build a tower.

δ is the pointed, immersed loop, D the general position null homotopy bounding it, and N the regular neighborhood of D in M. Subscript will indicate height in the tower.



(a) This example of a loop on the boundary of a genus 2 handlebody shows that the height function f, constructed in the proof of Theorem 0, cannot always have a unique local maximum, the familiar form for unknotting knot diagrams on the 2-sphere. The unknotting function f, relative to the indicated base point $*$ with $f(*) = 0$, must have at least two local maxima.



handlebody genus $= m + 1$

(b) For this absolute example, the unknotting function f produced by Theorem 1 must have at least m local minima (and maxima)

Figure 1

To build the tower we should ask about the Z_2 Betti number $b_1(N) :=$ rank$(H_1(N; Z_2))$. If $b_1(N) = 0$, there is no tower. For homological reasons ∂N is a disjoint union of 2-spheres and δ is contained in one of these: $\delta \looparrowright S \subset \partial N$.

If $b_1(N) > 0$ choose a 2-fold cover $\widetilde{N}_0 \to N_0 := N$ and choose a lift $l_1 : D \looparrowright \widetilde{N}_0$ and let $N_1 = \text{neib}(l_1(D)) \subset \widetilde{N}_0$. If $b_1(N_1) = 0$ then N_1 is the top of the tower. If $b_1(N_1) > 0$ continue and find a 2-fold cover $\widetilde{N}_1 \to N_1$, and lift $l_2 : D \looparrowright \widetilde{N}_1$ and set $N_2 = \text{neib}(l_2(D)) \subset \widetilde{N}_1$. Again if $b_1(N_2) = 0$ then N_2 is the top of the tower. If not, proceed to construct $l_3 \dots l_n$ and $N_3 \dots N_n$ until $b_1(N_n) = 0$. A simple complexity argument, where the complexity can be the number of simplicies identified by l_k for a fixed triangulation of D making all l_k simplicial, shows that the tower is indeed finite.

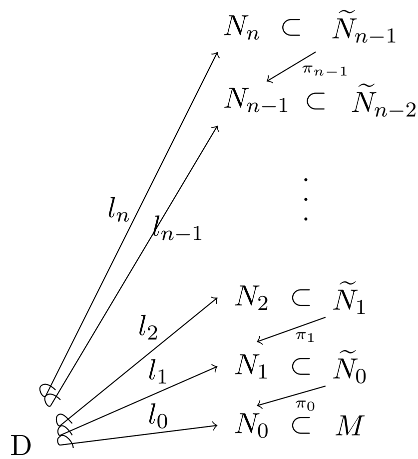


Figure 2. The tower

Observation: $\delta_n := l_n(\partial D)$ can be unknotted by a suitable resolution of its crossings (i.e. a normal function f as in Theorem 1). As in Theorem 0, given for any base point $*$ (chosen away from crossings) we can resolve crossings $\delta_n \to \delta'_n$ and produce an embedded disk Δ with $\partial\Delta = \delta'_n$ and $\Delta \cap \partial N_n = *$.

Explanation of δ'_n and its null-isotopy. One might expect to choose $\delta'_n = \{\delta_n(\theta), f(\theta)\}$, $f(*) = 0$, $f(\theta \neq *) > 0$, where f has a unique local maximum and a unique local minimum. However, this solution does not, in general, push down the tower. We prefer to give a second solution. There are two cases. If l_n is one-to-one $\delta_n = l_n(\partial D)$ bounds a hemisphere $E_1 \subset S \subset \partial N_n$. Pushing the disk E_1 normally toward interior N_n gives a disk $(\Delta, *) \subset (N_n, \partial N_n)$ bounding δ'_n.

Now assume δ_n is not one-to-one. Let $\alpha \subset \delta_n$, not containing $*$, be a subarc so that $l_n(\alpha)$ is a scc $\subset \partial N$. Let E_1 be one of the two disks in ∂N_n bounded by $l_n(\alpha)$. Begin to resolve the crossings of δ by following this rule: $l_n(\alpha)$ lies above $l_n(\beta)$, where β is the complementary arc, $\delta = \alpha \cup \beta$ and $\alpha \cap \beta = \partial\alpha = \partial\beta$. Also the two endpoints of α are at a crossing: resolve this crossing arbitrarily.

Now an isotopy across the disk E_1 bounding α effectively erases the loop α. If the simplified diagram is a scc we may continue the unknotting using a hemisphere 2-cell $E_2 \subset S, \partial E_2 = l_n(\beta)$ making no further crossing choices.

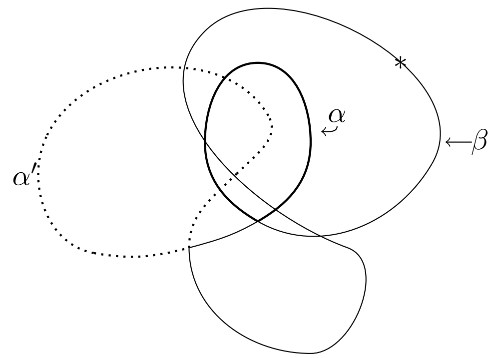


Figure 3

If the simplified diagram is still singular choose another arc $\alpha' \subset \beta$ whose image is again a scc not containing $*$ and proceed as before. Continuing in this way, guided by a sequence of embedded 2-cells, say $\{E_1, \ldots, E_j\}$, all crossings are eventually assigned so that the diagram resolution is unknotted with $*$ remaining in the final 2-cell E_j. The cells $E_1, \ldots, E_j$ determine a sequence of isotopies $\mathcal{I}_1, \ldots, \mathcal{I}_j$ so that the composition $\mathcal{I}_j \circ \cdots \circ \mathcal{I}_1$ shrinks δ'_n toward the base point $*$. Note: the 2-cells E_i will not generally have disjoint interiors. □

This solution will now be pushed down the tower

Dissection of disks with double arcs: Let E be a properly embedded disk $(E, \partial E) \subset (M^3, \partial M)$ in a 3-manifold and $\pi : (M, \partial M) \to (P, \partial P)$ a 2-fold cover with covering translation t. Assume E and tE are in generic position and meet only in double arcs and double loops. Double loops are easily removed by an innermost circle argument and will not be discussed further. E is assumed to have a base point $* \subset \partial E$. Call $\pi(E) = F$ and $\pi(*) = *$ in an abuse of notation.

We now describe how to form an ordered list of embedded disks $F_1, \ldots,$ $F_k \subset (P, \partial)$ from pieces (some used several times) of F. Constructing $\{F_1,$ $\ldots, F_k\}$ will dictate "crossing choices" for $\partial F := \gamma$, which yield γ', bounding a lollypop. Each F_i has base point $*_i$ with $*_k = * \subset \partial F$. Furthermore, once $\{F_1, \ldots, F_k\}$ are constructed we may view them as the instructions for an isotopy $\mathcal{I} = \mathcal{I}_k \circ \cdots \circ \mathcal{I}_1$, as above, collapsing γ' toward $*$.

Let β_1 be an outermost arc cutting off an outermost disk $D_1 \subset F$ not containing $*$. Let β'_1 be the partner arc of β_1 and D'_1 be the subdisk of F

cut off by β_1' which also does not contain $*$. There are two cases, shown in Figure 4.

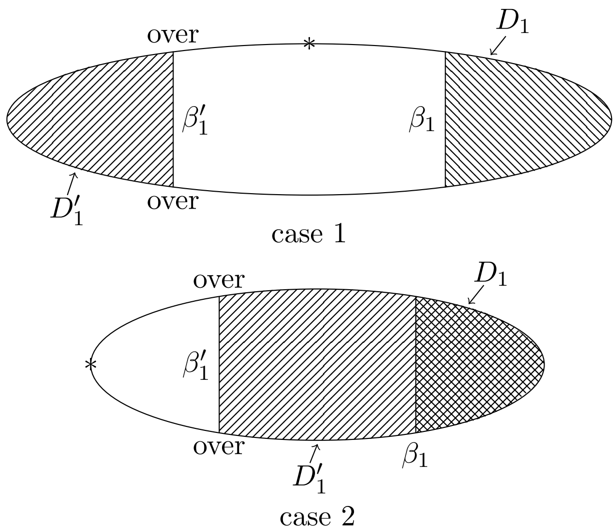
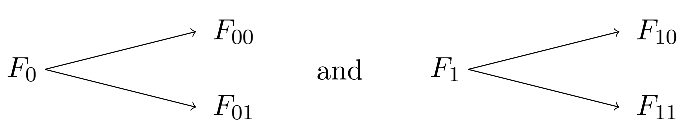


Figure 4

In both cases $D_1 \cup D_1'$ can be perturbed into a proper map of a disk F_0. There is a residual general position proper map with only double arc singularities of a disk formed from $D_1 \cup (F \backslash D_1')$. In both cases, label this map F_1. Because D_1 is outermost, both F_0 and F_1 have fewer double arcs than F.

We need to discuss crossing choices (resolutions) and base point choices. The end points on $\partial D_1'$, and its crossings with <u>other</u> segments of ∂F_1, are deemed <u>overcrossings</u> (in both cases). F_0 is provided a base point $*_0$ in its copy of D_1 and F_1 retains $*_1 := *$ as its base point.

Now unless $F_0(F_1)$ is embedded find an outermost arc cutting off an outermost disk $D_\bullet$, with $D_\bullet'$ disjoint from $*_0(*_1)$ and sharing a double arc with $D_\bullet$, <u>not</u> containing $*_0(*_1)$ and as above dissect:



to obtain general position proper pointed maps F_{00}, F_{01}, F_{10}, and F_{11} of disks with only double arc singularities. We call such maps <u>good</u> maps. In each case follow the preceding rule for resolving the crossings of $\partial D'_\bullet$ as overcrossings.

Continuing in this way a dyadic tree of good maps is obtained. The leaves of this tree are called great maps as they have the additional property of being embeddings; their imbedded images are called great disks. (No disjointness has been constructed or assumed for these great disks.) The leaves are now linearly ordered by the base 2 numerical value of their subscripts considered as decimals. By the time we reach the leaves all crossings have been resolved. These great disks, monotonically reindexed, become the ordered list $\{F_1, \ldots, F_k\}$, built from pieces of F, which we sought. We call this list a great sequence (for ∂F) guiding a collapse of $(\partial F)'$ to $*$. Note that the great sequence uniquely defines the crossing resolution $(\partial F)'$ of ∂F. The simple expedient of successively declaring $\partial D'_\bullet$ to over-cross other segments of the boundary has given us the well-defined crossing choices and thus defines $(\partial F)'$. The isotopy $\mathcal{I}$ shrinks $(\partial F)'$ toward $*$.

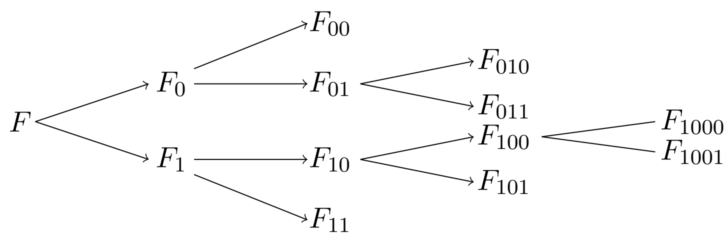


Figure 5. Sample tree with great maps as leaves

To summarize the Dissection section:

The tree, as constructed above, of good disks terminating in great disks is a dissection of F. The same tree, according to our convention, determines crossing choices $\gamma' = (\partial F)'$ for $\gamma = \partial F$. The great sequence $\{F_1, \ldots, F_k\}$ (the leaves of the dissection) are said to guide a sequence of isotopies $\mathcal{I}_k \circ \cdots \circ \mathcal{I}_1 := \mathcal{I}$ with $\mathcal{I}_i$ supported near F_i, $1 \leq i \leq k$, so that $\mathcal{I}$ collapses γ' toward its base point.

Example. We illustrate (Figure 6) dissection, the crossing resolutions (for $F_1 \to F_{10}$ viewed from opposite ends of the double curve, i.e. from the exterior of M), and the collapsing isotopy with an example. Our crossing convention implies that middle sheet of the three sheets of the outermost D in $F_{\ldots 0} \amalg F_{\ldots 1}$ should be compressed slightly into interior (M) near $\partial\beta_1$ when interpreting the disk F_j, $1 \leq j \leq k$, which contain this sheet of D as guiding the isotopies $\mathcal{I}_j$.

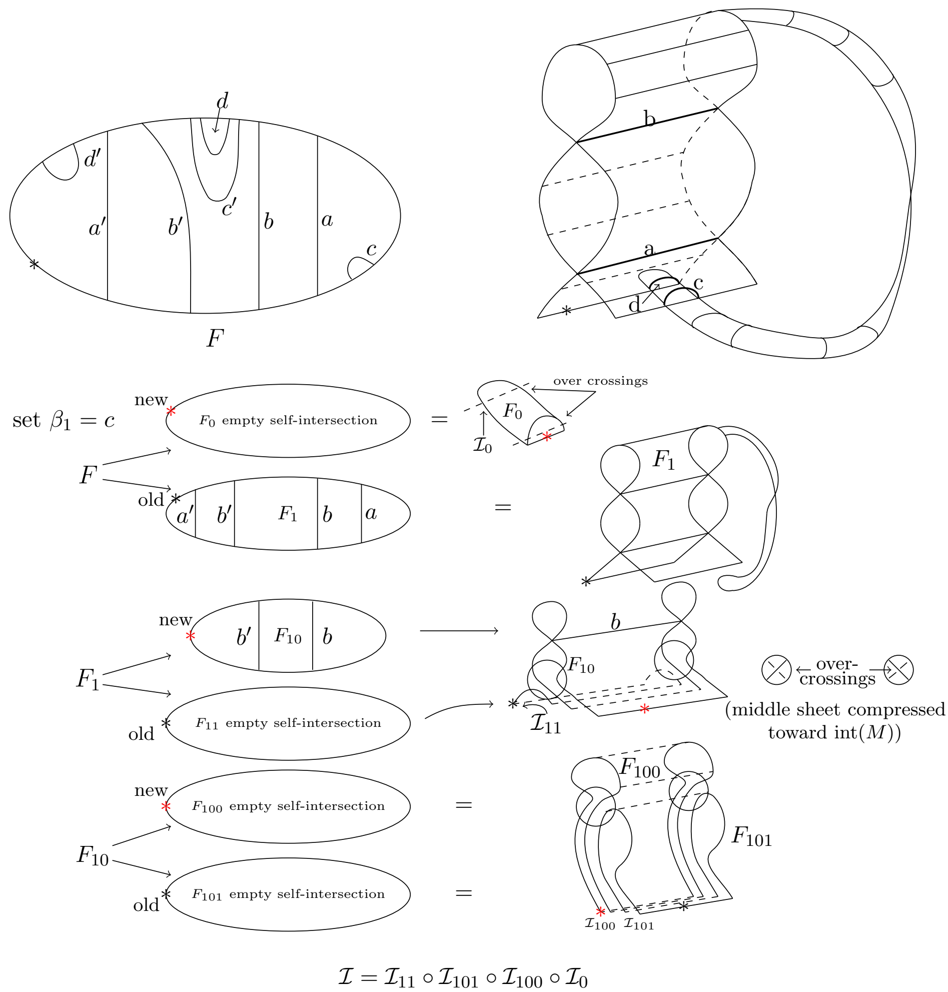


Figure 6

We now finish the proof of Theorem 0.

At the top of the tower the collapse of $((l_n(\partial D))', *)$ is guided by an initial sequence of properly embedded 2-cells $\{E_1, \ldots, E_j\}$, where the base point of each $E_1, \ldots, E_j$ is taken to be the self intersection point of its boundary, and with the original $* \subset E_j$ serving as the base point for the final E_j.

Now consider each of these cells E_i, $1 \leq i \leq j$, and its singular image F_i under the 2-fold cover $\pi_{n-1} : \widetilde{N}_{n-1} \to N_{n-1}$. Again we may remove double loops and only need consider double arcs. As above, <u>dissect</u> F_i into a sequence of great disks $\{F_{i,1}, \ldots, F_{i,k_i}\}$ and replace each E_i with that ordered sublist to obtain a great sequence:

$$\{F_{1,1}, \ldots, F_{1,k_1}, \ldots, F_{j,1}, \ldots, F_{j,k_j}\}$$

which guides the collapse of $(l_{n-1}(\partial D))'$ to $*$ (and defines the crossing resolution indicated by the "prime").

Now proceed all the way down the tower. At height m by induction assume a great sequence $\{G_1, \ldots, G_p\}$ for $(l_m(\partial D))'$. Project each G_i, $1 \leq i \leq p$, by π_{m-1} to N_{m-1} and construct, as above, a great sequence for each $(\pi_{m-1}(\partial G_i))$. The great sequence for $(l_{m-1}(\partial D))$ is obtained by concatenating these p great sequences. The great sequence for $(l_{m-1}(\partial D))$ is obtained by concatenating these p great sequences. The great sequence for $(l_{m-1}(\partial D))$ determines an isotopy compressing $(l_{m-1}(\partial D))'$ to its base point. It may be viewed as an 'elaboration' of the corresponding null isotopy, one stage higher, for $(l_m(\partial D))'$, which accounts for the double circles and arcs induced by the projection π_{m-1}. Finally, setting $m = 1$ we obtain a great sequence guiding the null isotopy which establishes Theorem 0. □

Note: There is a potentially exponential efficiency in describing a null isotopy as a long composition, as we have done in the proof of Theorem 0, compared with attempting a direct description of the bounding disk. The Fox-Artin-like unknot in Figure 7 makes this clear: any bounding disk must pass 2^k times over the rightmost protuberance. This motivated our proof strategy.

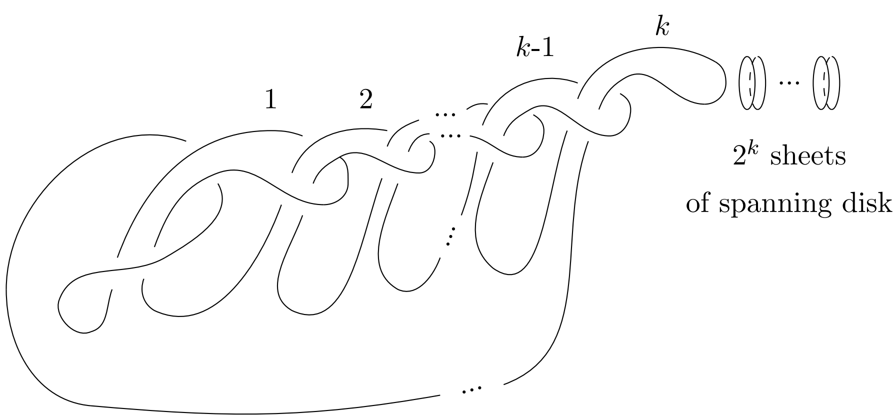


Figure 7

References

[1] C. Carathéodory, *Über die gegenseitige Beziehung der Ränder der Konformen Abbidung der Inneren einer Jordanschen Kurve auf einen Kreis*, Ann. Math. **73** (1913), 305–320.

[2] W. Osgood and E. Taylor, *Conformal transformations on the boundaries of their regions of definition*, Trans. Amer. Math. Soc. **14** (1913), 277–298.

[3] H. Whitney, *The self-intersections of smooth n-manifolds in $2n$-space*, Ann. Math. **45** (1944), 247–293.

[4] C. Papakyriapoulos, *On Dehn's lemma and the asphericity of knots*, Ann. Math. **66** (1957), no. 1, 126.

[5] M. Freedman, *The topology of four-dimensional manifolds*, J. Differential Geom. **17** (1982), no. 3, 357–453.

[6] M. Gromov, *Psuedo holomorphic curves in symplectic manifolds*, Invent. Math. **82** (1985), 307–347.

Microsoft Research, Station Q, and Department of Mathematics
University of California, Santa Barbara
Santa Barbara, CA 93106, USA
E-mail address: mfreedman@math.ucsb.edu

Department of Mathematics, University of California, Santa Barbara
Santa Barbara, CA 93106, USA
E-mail address: mgscharl@math.ucsb.edu

Received August 3, 2017