UNIQUENESS OF CROWN DIAGRAMS OF SMOOTH 4-MANIFOLDS

JONATHAN D. WILLIAMS

Abstract. Previously the author presented a new way to specify any smooth, closed oriented 4-manifold by an orientable surface decorated with simple closed curves. These curves are cyclically indexed, and each curve has a unique transverse intersection with the next. These are called crown diagrams, and each comes from a certain type of map from the 4-manifold to the 2-sphere. Given such a map, the corresponding crown diagram is defined up to diffeomorphism and isotopy of individual curves. The aim of this paper is to give a uniqueness theorem stating that, for maps within any fixed homotopy class, crown diagrams are unique up to four moves: stabilization, handleslide, multislide, and shift.

Contents

1. Introduction

This decidedly old-school paper is concerned with a certain class of maps from an arbitrary smooth closed orientable 4-manifold M to the 2-sphere, introduced as *simplified purely wrinkled fibrations* in [W1] and more briefly called *simple wrinkled fibrations* in [BH]. In this paper, the maps will be called crown maps for short, following a suggestion of R. Kirby. Under mild conditions, such a map defines a *crown diagram* (Σ, Γ) of M (previously called a *surface diagram*), in which Σ is a closed, orientable surface decorated with a collection Γ of simple closed curves. A crown diagram specifies M up to diffeomorphism. Section 2 describes four moves on crown diagrams for a fixed 4-manifold and explains how these moves come from certain model homotopies of the fibration map. Section 3 proves that, for a fixed homotopy class of maps $M \to S^2$, any pair of crown maps are the endpoints of some sequence of these model homotopies. Though the moves of [W1] are sufficient to

relate even maps in different homotopy classes, it is an open question to determine the diagrammatic relation between crown diagrams coming from different homotopy classes of maps $M \to S^2$ (see [KMT] for discussion of the collection of homotopy classes for a fixed 4-manifold).

In this paper, the notation M_I for $I \times M$ where $I = [0,1]$ will be convenient. Similarly, the notation M_t will denote the level set (t, M), and for a homotopy α, the notation α_t will refer to the map $\alpha|_{M_t}$. Abusing notation even further, symbols like $M_{\{t_0,t_1\}}$ will denote $M_{t_0} \cup M_{t_1}$. Here is the main result.

Theorem 1.1. Suppose $\alpha_0, \alpha_1 \colon M \to S^2$ are two homotopic crown maps with crown diagrams (Σ^i, Γ^i), $i = 0, 1$. Then there is a finite sequence of crown diagrams, beginning with (Σ^0, Γ^0) and ending with (Σ^1, Γ^1), obtained by performing stabilizations, handleslides, shifts, multislides, and their inverses.

The remainder of this section gives background on the origins of this work, the fundamentals of crown diagrams and the maps that induce them, and an overview of the proof of Theorem 1.1.

1.1. **Background.** The subject of crown maps arose from the study of broken Lefschetz fibrations on smooth 4-manifolds, which were first introduced in [ADK] to generalize the correspondence between Lefschetz fibrations and symplectic 4-manifolds up to blowup (see [GS] for further details on symplectic structures and Lefschetz fibrations). Roughly, the result of [ADK] was that if (M, ω) is a *near-symplectic* 4-manifold, then a suitable blowup of (M, ω) has a smooth map to the 2-sphere, called a *broken Lefschetz fibration*, such that ω restricts to a volume form on any fiber that does not intersect its vanishing locus. The vanishing locus of a near-symplectic form is always a smooth 1-submanifold of M. Discussed below, broken Lefschetz fibrations are a generalization of Lefschetz fibrations in which the critical locus is allowed to contain a 1-submanifold of critical points with a particular local model (see Equation 1); when the fibration corresponds to a near-symplectic form ω, this critical 1-submanifold coincides with its vanishing locus; see for example [B1, L1]. Though mild, this generalization greatly increases the collection of 4-manifolds that admit such a fibration structure: it is known that every smooth, orientable 4-manifold admits a broken Lefschetz fibration (though it may not correspond to any near-symplectic form); see for example [AK, B2, L1].

The study of broken Lefschetz fibrations is partly motivated by the effort to understand the Seiberg-Witten invariants of a smooth 4-manifold M geometrically. When b^{2+} is positive, the Seiberg-Witten invariants at their most basic level define a map from $H^2(M; \mathbb{Z})$ to the integers, defined as an algebraic count of solutions to a nonlinear elliptic pair of partial differential equations on M [M]. In a 1996 paper, Taubes showed that, for symplectic 4-manifolds, solutions to the Seiberg-Witten equations correspond to pseudoholomorphic curves which contribute to a special Gromov invariant he defined, called Gr. Pseudoholomorphic curves are submanifolds of two real dimensions, possibly with singularites, that are singled out by the chosen symplectic structure and other auxiliary data (these are choices which are later shown to not affect the values of Gr) [T1].

Related efforts have revolved around equipping a given manifold with structures that resemble surface bundles, namely Morse functions in three dimensions and (broken) Lefschetz fibrations in four dimensions, then considering associated

spaces of pseudoholomorphic curves. In 2003 Donaldson and Smith defined a *standard surface count* for symplectic 4-manifolds, counting pseudoholomorphic curves which are sections of a fiber bundle associated to a Lefschetz fibration [DS]. Soon after, Usher showed that the standard surface count is equivalent to Gr, and thus the Seiberg-Witten invariant [U]. Though this equivalence is only known to hold for a suitably chosen Lefschetz fibration whose existence is equivalent to the existence of a symplectic form, it gave a promising inroad to generalization: broken Lefschetz fibrations offer a way to continue their approach into nonsymplectic territory, further guided by an existence result of [T2] stating that if a near-symplectic 4-manifold has nonvanishing Seiberg-Witten invariant, then there is a pseudoholomorphic curve with boundary given by the vanishing locus of ω. In 2007, Perutz defined a generalization of the standard surface count for near-symplectic broken Lefschetz fibrations, called the *Lagrangian matching invariant*, that fits in nicely with natural expectations in numerous ways. For example, there are formal similarities with ideas in [T2], vanishing theorems and calculations in special cases that coincide with those for the Seiberg-Witten invariant. However, it remains to show that it is a smooth invariant of the underlying 4-manifold (not just the isomorphism class of its chosen fibration structure), and it is generally difficult to compute. A large amount of the motivation for studying crown diagrams is the possibility that they will prove useful in addressing such issues.

1.2. **Critical points.** Let M be a smooth closed 4-manifold. Any generic smooth map $M \to S^2$ resembles a surface bundle, except one allows the presence of a one-dimensional critical locus that is a union of *fold* and *cusp* points. The fold locus is an embedded smooth 1-submanifold of M with the following local model at each point:

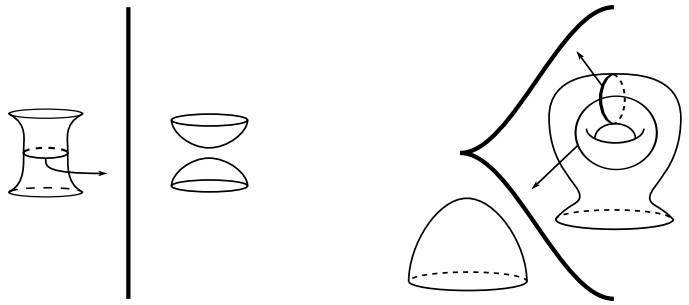
$$(x_1, x_2, x_3, x_4) \mapsto (x_1, x_2^2 + x_3^2 \pm x_4^2). \tag{1}$$

When the sign above is negative, it is known variously as an *indefinite fold*, *round singularity*, or *broken singularity* depending on context. Figure 1a is a picture of the target space, schematically depicting the fibration structure. In that figure, like the one to its right, what appears is the target disk of a map $D^4 \to D^2$, with bold arcs representing the image of the critical locus. A surface is pictured in the region of regular values that have that surface as their preimage, and tracing point preimages above a horizontal arc from left to right gives the foliation of $\mathbb{R}^3$ by hyperboloids, first one-sheeted, then two-sheeted, with a double cone above the fold point. The circle (drawn on the cylinder to the left) that shrinks to the cone point is called the *round vanishing cycle* for that arc in the context of broken Lefschetz fibrations. When dealing exclusively with Morse 2-functions as this paper does, there are no Lefschetz vanishing cycles, so the term *vanishing cycle* will suffice.

Each cusp point is a common endpoint of two open arcs of fold points, with the local model

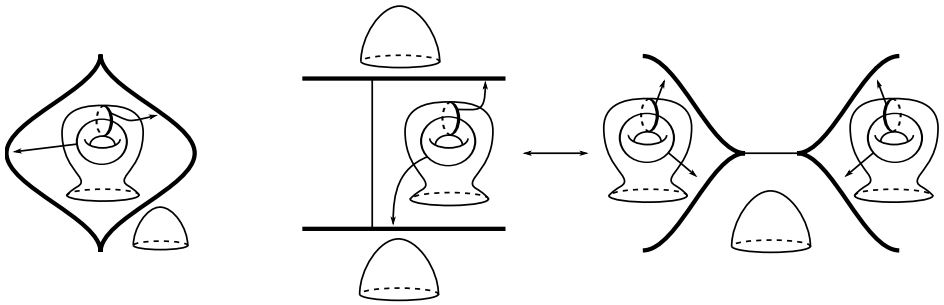
$$(x_1, x_2, x_3, x_4) \mapsto (x_1, x_2^3 - 3x_1x_2 + x_3^2 \pm x_4^2). \tag{2}$$

When the sign above is negative, it is called an *indefinite cusp* and is adjacent to two indefinite fold arcs as in Figure 1b: here two fold arcs meet at a cusp point, for which the two vanishing cycles must transversely intersect at a unique point in the fiber.



a. Fold. b. Cusp.

FIGURE 1. Critical points of Morse 2-functions.



a. Birth. b. Merge.

c. Flip.

FIGURE 2. Deformations of Morse 2-functions.

1.3. **Deformations of Morse 2-functions.** This paper is focused on homotopies of maps $M \to S^2$ which fail to be Morse 2-functions at finitely many values of the homotopy parameter t. Near such values of t, these homotopies have local models called *birth, merge* and *flip*. Using the same visual language of Figure 1, their effects on a Morse 2-function appear in Figure 2. Here follow the definitions from [W1, Section 2.4].

Definition 1.2. The *birth* move begins with a submersion and introduces a circle of critical points. It has the local model

$$(t, x_1, x_2, x_3, x_4) \to (t, x_1, x_2^3 + 3(x_1^2 - t)x_2 + x_3^2 - x_4^2).$$

The *merge* move results in the connect sum between two critical arcs, and is given by

$$(t, x_1, x_2, x_3, x_4) \mapsto (t, x_1, x_2^3 + 3(t - x_1^2)x_2 + x_3^2 - x_4^2).$$

The *flip* move introduces a double point in the critical image, and is given by

$$(t, x_1, x_2, x_3, x_4) \mapsto (t, x_1, x_2^4 + x_2^2 t + x_1 x_2 + x_3^2 - x_4^2).$$

The conditions under which these moves and their inverses can be performed are spelled out in [L1] and [W1, Section 2.4], where these local models were introduced as moves on what were then called *purely wrinkled fibrations* but are now more commonly called indefinite Morse 2-functions. An additional type of homotopy, called *isotopy* in [L1] and *2-parameter crossing* in this paper and [GK1], is one in which the critical image moves around in the base after the fashion of the Reidemeister moves. These are considered in detail in [W2].

Definition 1.3. For Morse 2-functions α_0, α_1, a *deformation* $\alpha = \alpha_t$, $t \in [0,1]$ is a homotopy realized by a sequence of births, merges, flips and 2-parameter crossings.

1.4. **Crown diagrams.** For the purposes of this paper, a broken Lefschetz fibration is simply a smooth map from a 4-manifold M to a surface F (usually the sphere or the disk) whose critical locus is a collection of Lefschetz critical points and indefinite folds, and a Morse 2-function is a stable map $M \to F$ whose critical locus is free of definite folds. In other words, the critical set of a Morse 2-function is a union of indefinite cusps and indefinite folds. For details on stability and the critical loci of stable maps, see [W2], and perhaps more importantly [L1], which relates the results of [Wa] concerning stability of families of real-valued functions to the world of Morse 2-functions.

Definition 1.4. By Corollary 1 of [W1], any continuous map $M \to S^2$ is homotopic to a Morse 2-function f whose critical locus is either empty or is a connected, smooth 1-submanifold of M mapped injectively by f into S^2. Such a map is called a *crown map*.

After an arbitrarily small perturbation to turn the map into a Morse 2-function, the rest of the homotopy is a given by a sequence of moves chosen from a list that initially appeared in the context of 4-manifold topology in [L1]. Such a map $f\colon M \to S^2$ has an orientable genus g surface as regular fiber over one of the disks that constitute $S^2 \smallsetminus f(\operatorname{crit} f)$, and the regular fiber over the other disk has genus $g-1$; choosing a regular value p in the higher-genus side and reference arcs from p to the various indefinite fold arcs, one obtains a collection of simple closed curves in Σ_g by recording the round vanishing cycles in a counter-clockwise direction going around the cusped critical circle. For this reason, the curves are relatively indexed by $\mathbb{Z}/k\mathbb{Z}$, where k is the number of fold arcs.

Definition 1.5. Assuming $g \geq 3$, a *crown diagram* (Σ_g, Γ) of M is the higher-genus fiber Σ_g of a crown map, decorated with the relatively $\mathbb{Z}/k\mathbb{Z}$-indexed collection $\Gamma = (\gamma_1, \ldots, \gamma_k)$ of round vanishing cycles, up to orientation-preserving diffeomorphism.

An observation by Paul Melvin is that composing with the antipodal map reverses the cyclic ordering of the vanishing cycles as measured in this way, while preserving M. For a fixed orientation of M, this reverses the preimage orientation of Σ.

Remark 1.6. The requirement that $g \geq 3$ (which can always be satisfied by applying the modification discussed in Section 2.2.1 below) is necessary for the following reason: After using (Σ_g, Γ) to form the fibration consisting of the preimage of a neighborhood of the critical image, if $g = 1$ or $g = 2$ there are various ways to close off the higher- or lower-genus sides with a copy of $\Sigma_{g-1} \times D^2$, according to the elements of $\pi_1(\mathrm{Diff}(S^2)) \cong \mathbb{Z}_2$ and $\pi_1(\mathrm{Diff}(T^2)) \cong \mathbb{Z}^2$, where $\mathrm{Diff}(\Sigma)$ is the group of orientation-preserving diffeomorphisms of Σ (see [ADK, B1, H] for explicit examples). Since $\mathrm{Diff}(\Sigma_g)$ is simply connected for $g \geq 2$, a crown diagram as defined specifies the total space of the fibration up to orientation-preserving diffeomorphism.

Remark 1.7. The local model for the cusp requires the two round vanishing cycles to transversely intersect at a unique point in the fiber; for this reason, consecutive elements of Γ must intersect in this way. Also, each element of Γ must be an embedded circle in Σ_g. These two facts could be called an *intersection condition* for crown diagrams, and it is natural to wonder if (Σ, Γ) is a crown diagram if it satisfies the intersection condition. Unfortunately this is too much to ask: This remark discusses the conditions under which (Σ, Γ) specifies a smooth 4-manifold. Just as the vanishing cycles of a Lefschetz fibration over the sphere must yield a trivial monodromy map, there is an associated monodromy map for (Σ, Γ) that is trivial if and only if the pair specifies a closed 4-manifold.

Given the pair (Σ, Γ) satisfying the intersection condition, construct a crown map over the closed unit disk (like Figure 3) in which the vanishing cycles read $\gamma_1, \dots, \gamma_k$ in a counter-clockwise direction. Then perform unsinking moves on all the cusps ([L1, Figure 1]) to obtain a broken Lefschetz fibration over the disk whose round vanishing cycle is γ_1 and whose Lefschetz vanishing cycles read $c_i = t_{\gamma_i}(\gamma_{i+1})$, $i \in \mathbb{Z}/k\mathbb{Z}$, so that t_{c_i} is the unique positive Dehn twist sending γ_i to γ_{i+1} (see [L1, W2] about sinking and unsinking). Let Σ' be the surface with two marked points (p, q) obtained from Σ by replacing a tubular neighborhood of γ_1 with two disks. The mapping class $\mu = [t_{c_k} \circ \cdots \circ t_{c_1}] \in MCG(\Sigma)$ of these Dehn twists preserves the isotopy class of the unoriented circle γ_1 by construction, and so one could interpret μ as an element of $MCG(\Sigma')(p, q)$, the group of orientation-preserving diffeomorphisms of Σ' that preserve its marked points as a set, and then forget about the marked points, taking it as an element of $MCG(\Sigma')$. Denoting the subgroup of elements that fix the unoriented circle γ_1 up to isotopy by $MCG(\Sigma)(\gamma_1)$, this amounts to a homomorphism

$$\Phi_{\gamma_1} \colon MCG(\Sigma)(\gamma_1) \to MCG(\Sigma')(p, q) \to MCG(\Sigma')$$

that was first hinted at in [ADK], mentioned explicitly in [B1] and then studied in more detail in [H, Be]. It is known that the fibration of Figure 3 can be capped off by a copy of $\Sigma' \times D^2$ to obtain a crown map exactly when $\mu \in \ker(\Phi_{\gamma_1})$. This offers a meaningful contrast between crown diagrams and Kirby diagrams of 4-manifolds: To check that the handlebody given by a framed and dotted link in S^3 closes off with 3- and 4-handles can be a difficult problem, because one must exhibit a diffeomorphism between the boundary of that handlebody and $\#_n(S^1 \times S^2)$ for some n. The corresponding task for crown diagrams is tedious at worst and large parts of it can be done by a computer, using Alexander's method (see [FM]) to check that $\mu \in \ker(\Phi_{\gamma_1})$.

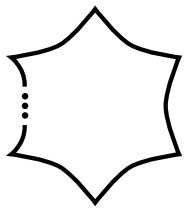


FIGURE 3. A Morse 2-function over the disk. The higher genus side of the fold is the inside of the circle.

Section 2 contains a list of discrete moves that may be used to modify any crown diagram. These moves preserve the diffeomorphism type of the 4-manifold specified by the crown diagram because they come from the endpoints of *deformations* of the fibration map. By the main result of [W1] (and also [GK1] with the added value of a connected fibers result for deformations between maps with connected fibers), one may modify any homotopy α, whose endpoints are required to be Morse 2-functions, to be a deformation. These were first described in [L1], using the term *deformation of purely wrinkled fibrations*. Such maps are the main focus of the paper.

To keep the distinction clear, the terms *handleslide*, *stabilization*, etc. will always refer to the move on crown diagrams, while the terms *handleslide deformation* and *stabilization deformation* will refer to the corresponding homotopies of fibration maps. Section 2 describes the so-called *model deformations* that induce the moves on crown diagrams. This is important important for two reasons. First, the correspondence between model deformations and moves on crown diagrams shows that performing a move does not change the diffeomorphism type of the 4-manifold described by the diagram, because each move is shown to be the result of deforming the map that specifies the original diagram. Second, the proof of Theorem 1.1 does not make use of the moves themselves: The argument is that there exists a way to change a deformation into a sequence of model deformations, which are linked to diagrammatic moves in Section 2.

Remark 1.8 (Surgered crown diagrams)**.** For each of the moves in Theorem 1.1, the highest-genus region of a base diagram momentarily becomes disconnected. Initially, reference fibers chosen over various points in the higher-genus region are decorated with identical sets of vanishing cycles up to isotopy. Depending on what goes on while there are two higher-genus components, there can be some variation in how the reference fibers over points in each component become identified when the components reunite, and one way to keep track of how these fibers are identified is to use a common reference fiber in a lower genus region. This viewpoint is used to great effect in [BH, H].

Consider in the initial crown map a reference fiber Σ_g over a point just to the lower-genus side of a fold arc γ, along with a short *reference path* across γ into its higher-genus side. As one traces fibers above this path, two points p, p' become identified in a surgery that increases the genus of the fiber by one as in Figure 1a. Then the endpoint of this path gives a reference fiber Σ_{g+1} to which one can add the vanishing cycles for folds bounding that higher-genus region. Pulling this picture back across γ, all the vanishing cycles descend to circles in Σ_g except those that

intersect the vanishing cycle of γ; these instead appear as *vanishing arcs* whose endpoints are p and p'. The ends of these arcs come equipped with a bijection β between those at p and those at p' by the way they pair up on either side of the vanishing cycle of γ in Σ_{g+1}. Further, β must reverse the order of the ends going around p and p' in the sense that if the ends $e_1, \dots, e_n$ are numbered clockwise around p, then the ends $\beta(e_1), \dots, \beta(e_n)$ are numbered counter-clockwise going around p'. For this reason, the union of vanishing arcs obtained by picking a point on one vanishing arc, following it to one end e, continuing at $\beta(e)$ and on until returning to the chosen point will be called a vanishing cycle just like the other simple closed curves. In this way, the crown diagram can be recovered from a reference fiber in the lower-genus region, along with a chosen path into the higher-genus region.

Taking a family of such reference fibers and paths (with fixed endpoints) that travels along the critical circle of a crown map yields a one-parameter family of such diagrams, and this paragraph describes how such a family evolves when a reference path passes a cusp (it is helpful to imagine the family of arcs in Figure 1b). As the family of paths approaches a cusp, there is a vanishing arc ν whose ends correspond under β coming from the vanishing cycle just past the cusp. This arc shrinks to a point where p and p' momentarily meet, so that the diagram whose path intersects the cusp itself has only one distinguished point. Passing the cusp, p and p' separate again. Those vanishing cycles that intersected ν form the new collection of vanishing arcs, while those ends that came from vanishing cycles disjoint from ν (and were brought together with p and p') remain identified, becoming arcs that pass between p and p', intersecting γ, which appears as a short arc running between p and p'.

1.5. **Overview of the argument.** The first step is to show that the moves on crown diagrams correspond to particular model deformations. Section 2 describes the moves and supplies these deformations along with arguments linking them to the moves. Importantly, each model deformation (call a given one α) has a specific pattern for three aspects of its critical locus:

- the Morse function T given by projecting $\operatorname{crit}\alpha$ to the t parameter,
- the restriction of T to the cusp locus of α,
- the stratified immersion $\operatorname{crit}\alpha \to [0,1] \times S^2$.

The map $\alpha|_{\operatorname{crit}\alpha}$ is a *stratified* immersion because it has a nonempty critical locus consisting of the cusp and swallowtail points, while its restriction to each of these strata is an immersion. The surface $\operatorname{crit}\alpha$ is orientable, with an orientation at every fold point coming from the t direction and the coorientation of the critical image at any time t coming from the direction of decreasing fiber genus.

The strategy to prove Theorem 1.1 stems from the observation that if the three aspects of a deformation above are shared with that of a model deformation, then it is a model deformation. The strategy is to modify $\alpha\colon M_{[0,1]} \to S^2_{[0,1]}$ until there is a partition of $[0,1]$ for which α follows one of the four patterns on each member.

After the sections that outline the model deformations, Sections 3.1–3.3 collect notation and some tools that appear in the proof of the main Theorem 1.1. The required modification of α appears in Sections 3.4 and 3.5.

Section 3.4 addresses the embedded surface $\operatorname{crit}\alpha$ itself in a series of lemmas to the effect that one may assume all fibers are connected with genus at least two, $\operatorname{crit}\alpha$ is connected, has at most two components at any value of t, and that T has

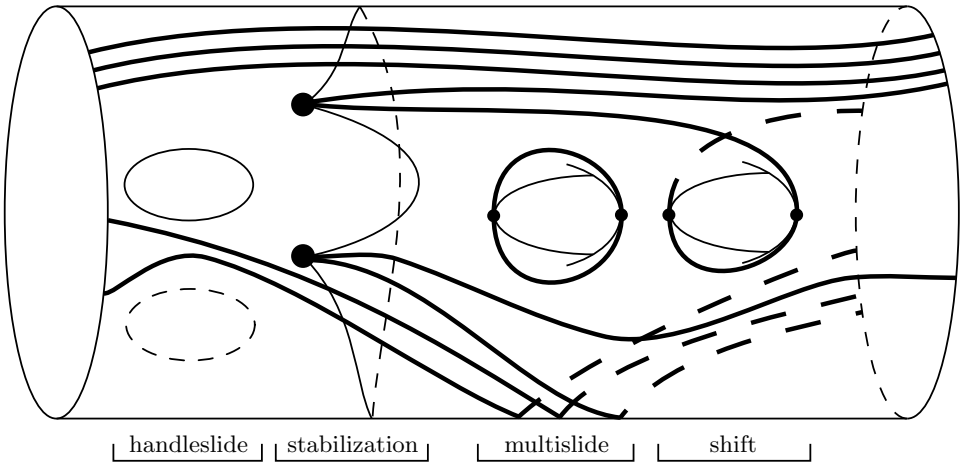


FIGURE 4. The decorated critical surface of a deformation. The decorations here correspond to the four moves of Theorem 1.1.

no canceling Morse critical points. Trading precision for intuition, one could say that applying these lemmas results in a one-parameter family of maps whose critical circle remains connected, except sometimes it splits in two for a short interval in t. Going further, the last of these lemmas, Lemma 3.7, uses tools from Section 3.2 to modify $\operatorname{crit}\alpha$ around each interval where the circle splits to distinguish each splitting as coming from a multislide or a shift deformation. Assuming $\operatorname{crit}\alpha$ has undergone the above modifications, Section 3.5 modifies the stratified immersion $\operatorname{crit}\alpha \to [0,1]\times S^2$. Decorating the critical surface with so-called *immersion arcs*, there are several cases one must examine in order to ensure that they precisely follow the prescriptions given by the model deformations. What remains is the required deformation. To give some idea of how these decorations appear, Figure 4 depicts $\operatorname{crit}\alpha$, where α is a sequence of the model deformations for handleslide, stabilization, multislide, and finally shift (each is underlined below the figure). The t coordinate increases to the right, and the critical points of T are marked out by small dots while swallowtail points (where flips occur) are large dots. The bold arcs are swept out by cusp points while the other arcs are mapped to each other in pairs by α. All these phenomena are discussed in more detail below. One quirk of these figures is that each cylinder is cut out of $\operatorname{crit}\alpha$ at an angle and drawn with a slight bend to eliminate artifacts of parallax: It looks like a straight right cylinder on the paper, yet each pair of points in each slice α_t appears vertically aligned. This makes it easier to identify which arcs have the same image in S^2_I using the *simultaneity condition* defined below.

Figure 5 exhibits some of the behavior that is ruled out for sequences of model deformations. Two highlights are the existence of a canceling pair of T-critical points at the top (Lemma 3.6) and an intersection triple of immersion arcs. In the absence of cusps, such an intersection appears as a Reidemeister-III fold crossing in S^2, and there are no instances of such a move in the model deformations. Another issue in Figure 5 is that none of the swallowtail points occur as part of a stabilization as in Figure 4. The remedy for such a deformation would be to apply Lemma 3.6 for the canceling critical points (see Figure 27 for the heart of the

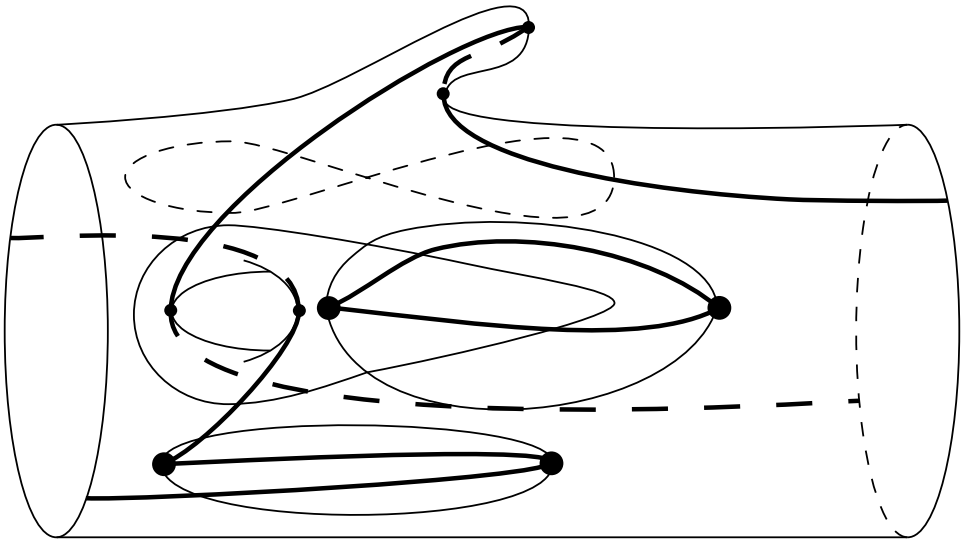


FIGURE 5. A sampling of what can go wrong in a deformation.

argument), and to include the swallowtails into stabilization deformations. As for the intersection triple, Section 3.5 contains lemmas that allow one to include the arcs into handleslide or stabilization deformations. One of the more subtle parts of the paper is concerned with the immersion arc that encircles the pair of index one critical points of T: Section 3.5.1 is dedicated to eliminating this phenomenon.

Acknowledgments. The author would like to thank Denis Auroux, David Gay, Kenta Hayano and Rob Kirby for their exceedingly helpful input during the preparation of this work, which was partly supported by NSF grant DMS-0635607. The author would also like to thank the reviewer(s) for their extraordinary fortitude.

2. Moves and model deformations

2.1. **Handleslide.**

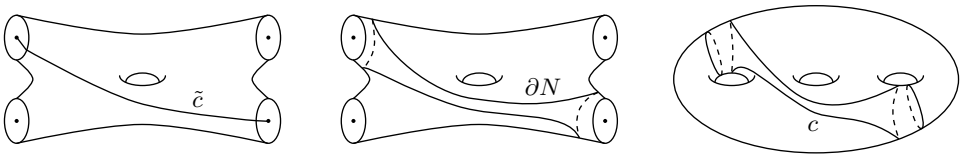
2.1.1. *Handleslide move.* The first move comes from an application of Lemma 3.8 and Theorem 3.9 of [H]. To perform a handleslide on (Σ, Γ), the first step is to choose a pair of disjoint vanishing cycles, say γ_1 and γ_n, such that the result of surgery on $\gamma_1 \cup \gamma_n$ (replacing a neighborhood of each of γ_1 and γ_n with a pair of disks) is a connected surface of genus at least 2. Next, choose a product of three Dehn twists $\phi = t_{\gamma_1}^{-1} \circ t_{\gamma_n}^{-1} \circ t_c \in \operatorname{Diff} \Sigma$, where c is constructed immediately below. Then the new crown diagram is (Σ, Γ'), where

$$\Gamma' = (\gamma_1, \phi(\gamma_2), \dots, \phi(\gamma_{n-1}), \gamma_n, \dots, \gamma_k).$$

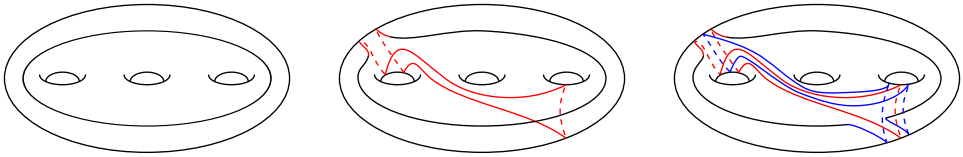
Since the move can only affect $\gamma_2, \dots, \gamma_{n-1}$, these are called the *nonstationary set*, and the *stationary set* comprises the rest of the circles. These terms will also be useful for the multislide and shift moves.

The curve c comes from the following construction. First, replace neighborhoods of γ_1 and γ_n with pairs of marked disks (D_i, p_i), (D_i', p_i'), $i \in \{1, n\}$, and let

$$h\colon \Sigma_{g-2} \smallsetminus \{p_1, p_1', p_n, p_n'\} \to \Sigma_g \smallsetminus (\gamma_1 \cup \gamma_n)$$



a. Constructing a curve c for the handleslide move. In the right side, c is the middle curve, while γ_1 and γ_n are interchangeably the other two.



b. An example of how the choices above affect the curve on the left by a sequence of connect sums with γ_1 and γ_n.

FIGURE 6

be the resulting diffeomorphism relating the surfaces on the left and right sides of Figure 6a (minus their respective subsets). Then connect one of $\{p_1, p_1'\}$ to one of $\{p_n, p_n'\}$ by any embedded arc $\tilde{c}$ (an example of such an arc is on the left side of Figure 6a) and let N be the closure of a neighborhood of $\tilde{c}$ containing the disks that contain the endpoints of $\tilde{c}$. The boundary of N appears in the middle of Figure 6a. Then $c = h(\partial N)$ is a simple closed curve in $\Sigma_g \smallsetminus (\gamma_1 \cup \gamma_n)$.

It may help the reader to observe that the monodromy ϕ, viewed in the disk N with two marked points, simply revolves the points around each other without rotating their tangent spaces. Then map then becomes the identity if we forget either one of the marked points. From there, it is straightforward to see how the given projection can change the monodromy by ϕ while being invisible over each of the lower genus regions.

Section 2.1.1 ends with a short digression on why this is called a handleslide move. In a Heegaard diagram for a 3-manifold, the attaching circles can be considered vanishing cycles for critical points of an associated Morse function; see e.g. [L2] for an application of this viewpoint. The deformation underlying the handleslide move is similar to that of the 3-dimensional handleslide in the sense that both can be seen as inducing push maps in surgered diagrams: In the 3-dimensional case, one chooses whether the handleslide occurs among α or β circles, or in the language of this paper which circles form the nonstationary set (suppose it is the α circles). Then one replaces, say, α_1 with two marked disks (that is, one moves the reference fiber just past the critical point that gave α_1) and applies a push map to the other α circles. Each time one of these circles crosses a marked point, it is recorded in the Heegaard diagram as a handleslide over α_1. Indeed, the map ϕ above is constructed as such a lift of a push map in the proof of [H, Lemma 3.8]. As an example, Figure 6b shows how the nonstationary curve γ_i undergoes a sequence of connect sums with parallel copies of γ_1 or γ_n according to the choices made in Figure 6a. Another way to check that any handleslide modifies a nonstationary curve by a sequence of connect sums with γ_1 and γ_n is to perform the relevant

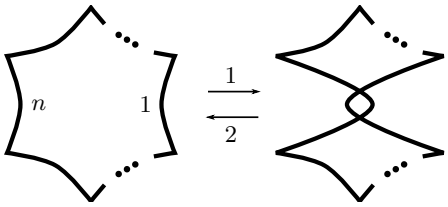


FIGURE 7. Base diagrams for a handleslide deformation. Steps 1 and 2 are both Reidemeister-II fold crossings.

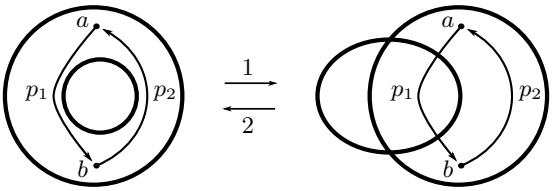


FIGURE 8. A reproduction of [H, Figure 5], in which the circle γ is given as a concatenation of two paths p_1 and p_2. The leftmost lune after step 1 corresponds to the central lune in Figure 7.

Dehn twists on short arcs representing intersections of $h^{-1}(\gamma_i)$ with $\tilde{c}$ (each gives two intersections between γ_i and c), or of γ_i with γ_1 or γ_n (each of which gives one intersection of γ_i with c and one intersection of γ_i with γ_1 or γ_n).

2.1.2. *Handleslide deformation.* The model deformation for the handleslide move is the R_2-*move* studied in [H].

Definition 2.1. A *local handleslide deformation* consists of a consecutive pair of Reidemeister-II fold crossings. In the first crossing, the higher-genus side of a fold arc approaches the higher-genus side of another, and two intersections form. In the second Reidemeister-II fold crossing, the newly-formed lune of regular values contracts, canceling the two intersections. A *handleslide deformation* is a local handleslide deformation beginning with a crown map as in Figure 7, throughout which all fibers are connected and have genus at least 2.

A deformation such as the one in Figure 7 has disconnected fibers in the central lune if $\Sigma \setminus (\gamma_1 \cup \gamma_n)$ is disconnected. According to [GK1, Theorem 1.2], the initial deformation connecting α_0 to α_1 can be assumed to have connected fibers. Going further, Lemma 3.3 asserts there is a deformation whose fibers all have genus at least 2. The modifications in Section 3 preserve this condition. In [H, Theorem 3.8], Hayano did essentially all of the work to prove that any handleslide deformation induces a handleslide move in a crown diagram. In [H, Theorem 3.9] he proved that any handleslide move is induced by some handleslide deformation.

The fibration that appears in [H, Figure 5] is not a crown map, so it remains to clarify why Hayano's argument applies to handleslide deformations. In Figure 8, the path p_2 and some chosen horizontal distribution specifies a diffeomorphism $\phi_2\colon F_b \to F_a$ between the fibers above the endpoints a and b. This diffeomorphism remains unchanged through Hayano's homotopy because the chosen horizontal distribution over p_2 does not change. On the other hand, Hayano shows that his homotopy

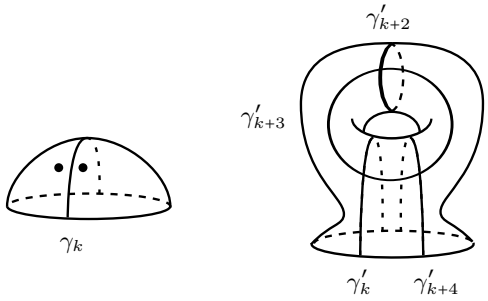


FIGURE 9. Stabilizing a crown diagram near a point in γ_k. The circle γ'_{k+1} has a more subtle description.

changes the monodromy along $p_1 * p_2$ precisely by composition with a map ϕ as in Section 2.1.1. For this reason, the map $\phi_1 \colon F_a \to F_b$ becomes $\phi(\phi_1)$ as a result of his homotopy. Finally, in Figure 7, p_1 gives the identification between higher-genus fibers in regions above and below the central lune in the right side of the figure, so the vanishing cycles for fold arcs contained in the upper region, as they appear in F_a, change by application of a map ϕ after a handleslide deformation.

Remark 2.2. Hayano's construction of the monodromy φ_γ in [H, Lemma 3.8] as a product of three Dehn twists for a non-separating pair (γ_1, γ_n) is the same as what appears in [H, Theorem 3.7] for a pair which separates the fiber into surfaces of genus at least 2, so there is an immediately analogous version of handleslide for separating pairs. On the other hand, his version of the move is not necessary for the uniqueness Theorem 1.1.

2.2. **Stabilization.**

2.2.1. *Stabilization move.* When $\operatorname{crit} \alpha_0$ is nonempty, the stabilization move comes from a lightly disguised generalization of a homotopy of Baykur that first appeared in Figures 5 of [B2] (where it is called "flip and slip") and 11 of [L1]. Hayano in [H] gave a careful description of this move; most of how it affects a crown diagram is shown in Figure 9. On the left of Figure 9 is a small neighborhood of a point in a vanishing cycle, for which the numbering has been chosen so that it is the "last" vanishing cycle of $\Gamma = (\gamma_1, \dots, \gamma_k)$. According to [H, Theorem 6.5], the stabilization deformation results in a diagram (Σ_{g+1}, Γ'), with Σ_{g+1} and the first k entries of Γ' obtained from (Σ_g, Γ) by replacing a pair of small disks as shown at the left with a cylinder. Three of the last four entries of $\Gamma' = (\gamma'_1, \dots, \gamma'_k, \gamma'_{k+1}, \dots, \gamma'_{k+4})$ are as in the figure, with γ'_{k+4} parallel to γ'_k. The remaining circle γ'_{k+1} is given by $\mu' \circ \phi^{-1}(\gamma'_{k+3})$, where $\phi \in MCG(\Sigma_{g+1})$ is any element such that $\Phi_{\gamma'_k}(\phi) = id \in MCG(\Sigma_g)$ and

$$\Phi_{\gamma'_{k+2}}(\phi) = t_{t_{\gamma_k}(\gamma_1)} \circ t_{t_{\gamma_{k-1}}(\gamma_k)} \circ \cdots \circ t_{t_{\gamma_1}(\gamma_2)}.$$

See Remark 1.7 for a discussion of Φ. The problem of how to systematically produce a mapping class ϕ given Γ and a choice of γ_k is largely open, though numerous examples appear in [H].

One special case of stabilization is available when $\operatorname{crit} \alpha_0$ is empty. For crown diagrams that come from surface bundles over the sphere, that is, blank crown diagrams, performing a birth move as in [W1, Section 2.4.2] increases the genus

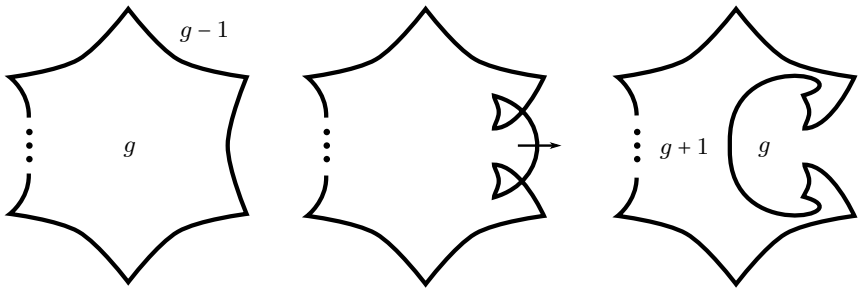


FIGURE 10. The stabilization deformation involves two flipping moves that occur on the same fold arc, eventually followed by an R_2 deformation that cancels the resulting intersections. The arrow indicates a fold arc that passes across the back of the sphere, completely sweeping over the lower-genus side.

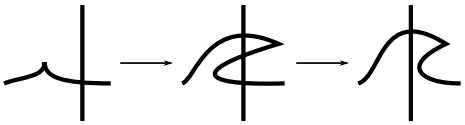


FIGURE 11. Cusps might cross fold crossings during an R_2 deformation by a pair of 2-parameter crossings of this type, or its reverse. Here the depicted critical image is contained in a small disk that contains no other critical values.

of the crown diagram by one and introduces two simple closed curves whose only requirement is that they intersect at a unique point. In [BH], but not in this paper, it is part of their definition of *surface diagram* that there are at least two circles.

2.2.2. *Stabilization deformation.* Figure 10 is a sequence of base diagrams for the stabilization deformation when $\operatorname{crit}\alpha_0$ is nonempty, corresponding to the move from Section 2.2.1, studied in detail in [H]. Here, two *flipping moves* from [L1] have been applied to the same fold arc, then the part of the fold arc between the two loops sweeps across the lower-genus side of the fibration to cancel the intersection points, crossing each cusp point (other than the four that appeared in the flips) exactly once.

The part of the deformation following the flips is what will be called an R_2 *deformation*, which is defined here using the language of [GK1, Section 2, item (4b)]:

Definition 2.3. An R_2 deformation is a *Reidemeister-II fold crossing* (or its reverse), slightly generalized to allow cusps to cross into or out of the boundary of the central lune, with each entry or exit as in Figure 11. An R_2 deformation is also allowed to begin or end with a cusp-fold crossing.

As depicted in [H, Figure 5], the first and second fold crossings of what that paper names an R_2*-move* are both examples of what this paper calls an R_2 deformation. Like in Figure 4, depictions of the deformation in Figure 11 will be suppressed from

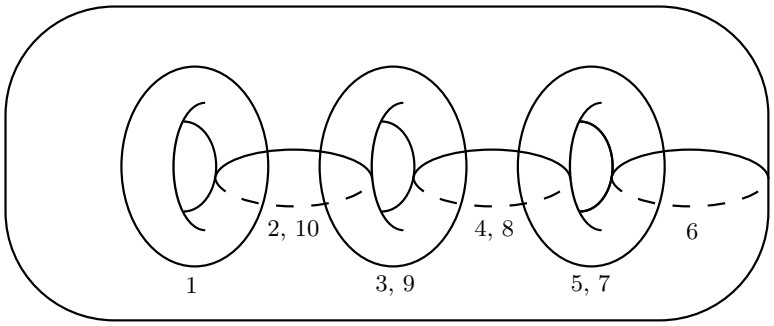


FIGURE 12. A crown diagram for $S^2 \times S^2$ which comes from stabilizing the empty genus-0 diagram for the projection $S^2 \times S^2 \to S^2$ three times.

pictures of the critical set. They should be understood to occur at any crossing between a cusp arc and an immersion arc.

Proposition 2.4. Given any fold arc in a crown map, there is a stabilization deformation whose initial flips occur on that fold arc.

Proof. First perform two flips on the chosen fold arc. By repeated application of [W2, Propositions 2.5(2) and 2.7], it is possible to repeatedly apply deformations as in Figure 11 to move all cusps into the two loops introduced by the flips. At this point it is possible to cancel the two fold crossings by [W2, Proposition 2.5(3)]. □

Example 2.5. Now that the term *stabilization* has appeared, it makes sense to give the following example of a crown diagram. Figure 12 is a crown diagram for $S^2 \times S^2$, with the circles numbered in the order they would appear going counterclockwise around the image of a 10-cusped critical circle. The corresponding crown map is homotopic to the sphere bundle over the sphere that projects onto either of the factors of $S^2 \times S^2$. Beginning with that map, perform a birth move and perform two stabilization deformations to obtain Figure 12. It is curious that the other S^2 bundle over S^2 with total space diffeomorphic to $\mathbb{CP}^2 \# \overline{\mathbb{CP}}^2$ also begins with a base diagram that has a sphere fiber with no vanishing cycles, and that one may perform the "same" three moves; however, a crown diagram specifies the total space up to diffeomorphism, so the resulting vanishing cycles must be different from those in Figure 12. For this reason, stabilization (as an operation that takes (Σ, Γ) as input) is not well-defined if the genus of Σ is less than 2.

Hayano [H] was able to sufficiently understand this phenomenon to convert the fibration of S^4 originally appearing in [ADK] (whose fibers are genus 0 and 1, with one cusp-free fold circle between) into one with arbitrarily high genus to yield a family of crown diagrams of S^4. Starting with one of these diagrams, one could apply [Be, Lemma 5.1] to get a diagram of $\mathbb{CP}^2 \# \overline{\mathbb{CP}}^2$.

2.3. **Multislide.**

2.3.1. *Multislide move.* Independently found by Denis Auroux and Rob Kirby, then carefully explained with proof in [BH], this move comes from a deformation in which the critical locus becomes momentarily disconnected. To perform the multislide, one first finds a non-consecutive pair of vanishing cycles that intersect transversely

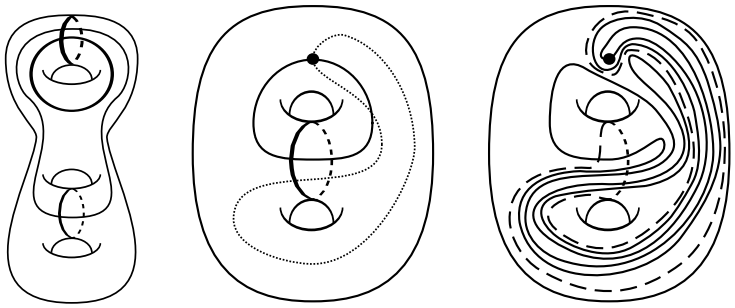


FIGURE 13. An example of multislide (omitting the last step, a connect sum with T^2 at the dot).

at a unique point in Σ, say γ_1 and γ_n (the two bold curves in Figure 13a). One also chooses one of the subsets $\{\gamma_2, \dots, \gamma_{n-1}\}$ or $\{\gamma_{n+1}, \dots, \gamma_k\}$ on which to perform the move, calling it the nonstationary set (two of which appear as the other two curves in Figure 13a). A small tubular neighborhood of the chosen pair $\gamma_1, \gamma_n \subset \Sigma_g$ is a punctured torus T. The multislide proceeds as a one-parameter family of diffeomorphisms of $(\Sigma_{g-1}, \gamma_2, \dots, \gamma_{n-1})$ in which T travels along an arbitrary circle α based at a point $p \in \partial T$, returning to where it started and dragging the nonstationary set along with it (but leaving all the other vanishing cycles $\gamma_n, \dots, \gamma_1$ unchanged). In other words, if one were to replace T with the point p as in Figure 13b, the modification of the nonstationary set would be to apply the point-pushing map of the Birman exact sequence of mapping class groups along the circle α (pushing along the dotted circle in Figure 13b results in Figure 13c), then replace T so that the vanishing cycles glue back together in the obvious way. For details on the point push map, see for example section 4.2 of [FM]. A choice of framing also allows the nonstationary set to be affected by a power of a Δ-twist, which is in the language of [BH] denotes the square root of a positive or negative Dehn twist along ∂T; see Section 6.2 and the proof of Lemma 5.2 of that paper. To explain the terminology, it is as if a handleslide has been performed in which a collection of vanishing cycles slides over a disk spanning ∂T (followed by half Dehn twists along ∂T).

2.3.2. *Multislide deformation.* The multislide deformation, corresponding to the move in Section 2.3.1, is depicted in Figure 14. To begin, choose a pair of fold arcs whose vanishing cycles happen to intersect transversely at a unique point in the fiber (they do not have to be consecutive). In the left side of Figure 14, a vertical arc joins the chosen pair, signifying a *fold merge* that results in the right side of the figure.

Remark 2.6. This move is called a *merge* in [L1] and [W1], and an *unmerge* in [GK1]; the terms *fold merge* for when two folds meet and *cusp merge* for when two cusps meet appeared in [BH], will be used in this paper, and seems likely to enjoy future consensus.

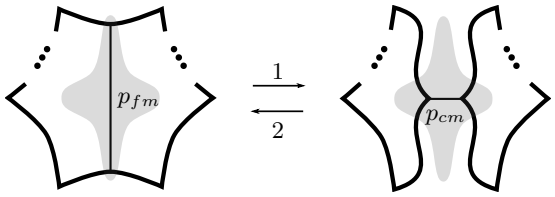


FIGURE 14. The multislide deformation consists of a fold merge followed by a cusp merge. The local version is supported on a tubular neighborhood of the union of the fold merge path p_{fm} and the cusp merge path p_{cm}, the images of which are shown, with the image of the support shaded.

The inside of each of the two circles now has Σ as its regular fiber, decorated with vanishing cycles obtained from Γ by deleting those whose fold arcs became contained in the other circle as a result of the fold merge (this is where the partition $\Gamma \smallsetminus \{\gamma_1, \gamma_n\} = \{\gamma_2, \dots, \gamma_{n-1}\} \cup \{\gamma_{n+1}, \dots, \gamma_k\}$ comes from). Before any other moves such as cusp-fold crossings, flips, or R_2 deformations occur, the deformation concludes with a cusp merge in which the two circles reunite along the two newly-formed cusps.

Definition 2.7. A *local multislide deformation* is a generalization of a multislide deformation satisfying the following two conditions:

(i) There is an embedding of a fibration structure of the type depicted over the shaded $D^2_{[0,1]}$ in Figure 14 into the fibration structure over $S^2_{[0,1]}$ given by α.
(ii) Let $\iota \subset \operatorname{crit} \alpha$ denote the collection of points where $\alpha|_{\operatorname{crit} \alpha}$ fails to be injective. As suggested in Figure 14, $\alpha(\iota)$ is bounded away from the one-parameter family of shaded disks in $S^2_{[0,1]}$ depicted in Figure 14.

The merge points of a local multislide deformation form a *multislide pair*.

To explain the corresponding modification of crown diagrams, consider the deformation as a map $[0,1] \times M \to [0,1] \times S^2$ that restricts to the identity map in the first factors. The critical image is a properly embedded twice-punctured torus in $[0,1] \times S^2$. If one chooses two reference points in the left side of the figure such that one reference point splits into each circle, initially their crown diagrams are isotopic. However, the identification between them is modified by the time the two circles rejoin: note that the cusps that form and then disappear trace out a circle in $[0,1] \times M$. Projecting out the homotopy parameter yields a circle in M, and without loss of generality one of the pair of cusps, say the right, remains stationary throughout its life, while the other traces out the circle as the homotopy progresses (the resulting crown diagram will then be the one obtained using a reference point that stays in the right circle through the deformation). To say it a different way, the horizontal arc in the middle of Figure 14 can be taken as the image of the path to be taken by the left cusp. This so-called *cusp merge path* (less precisely referred to as a *joining curve* in [Lev, W2], α in [L1], and elsewhere unnamed) is framed, is everywhere transverse to the fiber, and has the two cusps as its endpoints.

Remark 2.8. All cusp merge paths will be denoted p_{cm} in this paper (p_{fm} for fold merge paths), and note that once a cusp merge path is defined, its sub-paths

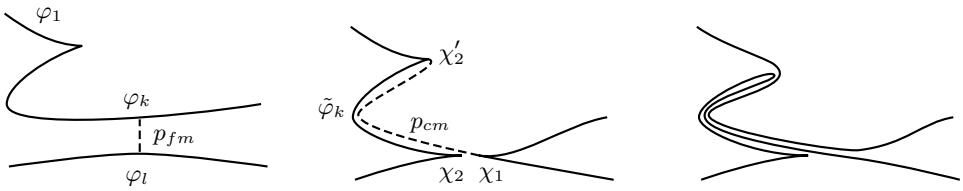


FIGURE 15. The shift deformation consists of a fold merge followed by a cusp merge. The dotted arcs are the image of a pair of merge paths in M; the left for a fold merge and the right for a cusp merge.

at later t values as the two cusps approach each other also serve as cusp merge paths, and similar for fold merge paths, so that really a cusp merge or fold merge involves a one-parameter family of merge paths $[p_{cm}]_t$, though such pedantry is only necessary in a few parts of the paper.

At the moment t_0 when the two cusps form, the endpoints of $(p_{cm})_{t_0}$ coincide with the merge point, so $(p_{cm})_{t_0}$ is a circle in M. The fibers containing points on the interior of $(p_{cm})_t$ serve as Σ', with the disk mentioned above coming from a fiberwise neighborhood of $(p_{cm})_t$. The circle $(p_{cm})_{t_0}$ projects to a circle in a fiber Σ' along which this disk travels, inducing the above-mentioned isotopy of the vanishing cycles in the left circle, which in turn induces an element of the mapping class group of Σ that only applies to those vanishing cycles coming from the left circle.

Choosing the left cusp instead of the right as the stationary cusp (and the left circle's reference point as the ending reference point) results in performing exactly the reverse modification to $\gamma_{n+1}, \dots, \gamma_k$, moving along the circular path in the opposite direction and applying oppositely oriented Dehn twists along the boundary of the disk. For completeness, it should be mentioned that the surgered crown diagram explanation for the multislide involves a reference path connecting the two newly-formed cusps throughout their existence; the details are left to the reader.

2.4. **Shift.** This move could be seen to embody the possible variations on the ordering of the vanishing cycles in a crown diagram, though it does not simply re-index the elements of Γ. It was studied carefully in [BH], and for the reader's convenience a summary follows. Because of the wording of Proposition 2.9, it is convenient to first outline the shift deformation. As with the multislide, one identifies a pair $\gamma_l, \gamma_k \in \{\gamma_1, \dots, \gamma_k\}$, along with a nonstationary subset $\{\gamma_{l+1}, \dots, \gamma_{k-1}\}$.

In base diagrams, the shift deformation is similar to that of the multislide in Figure 14, in which a fold merge occurs between the fold arcs φ_l and φ_k corresponding to γ_l and γ_k, but instead of reuniting the two circles by a cusp merge between the two newly-formed cusps, the circles reunite by a cusp merge between one of the newly-formed cusps (χ_1 in Figure 15) and the cusp between φ_1 and φ_k. Such a deformation is called a *generalized shift deformation* in [BH]. The following theorem, the terminology has been modified to match that of the current paper.

Proposition 2.9 ([BH], Proposition 6.3)**.** Let $w\colon X \to D^2$ be a crown map with crown diagram $(\Sigma; \gamma_1, \dots, \gamma_k)$ such that for some $1 < l < k$ the curves γ_l and γ_k intersect at one point.

(1) If $w'\colon X \to D^2$ is obtained from w by a generalized shift deformation whose initial fold merge is applied around the two fold arcs with vanishing cycles γ_l and γ_k, then the crown diagram of w' is given by

$$(\Sigma, \gamma_1, \dots, \gamma_l, \gamma_k, \chi(\gamma_{l+1}), \dots, \chi(\gamma_{k-1}))\,, \tag{3}$$

where $\chi \in \mathrm{Mod}(\Sigma)$ is a mapping class that satisfies the following properties:
   (a) $\chi t_{\gamma_k}^{-1} t_{\gamma_l}^{-1} \in \operatorname{Ker} \Phi_{\gamma_k}$,
   (b) $\chi t_{\gamma_1}^{-1} t_{\gamma_k}^{-1} \in \operatorname{Ker} \Phi_{\gamma_1}$.

(2) For any $\chi \in \mathrm{Mod}(\Sigma)$ satisfying the conditions (a) and (b), there exists a generalized shift deformation from w to a crown map whose crown diagram is given by Expression 3.

The above modification simplifies somewhat with the following assumption about the deformation. Suppose that, once the initial fold merge occurs, there is a path traced out by one of the resulting cusps, with the other cusp remaining stationary in M. Slightly after the initial fold merge, the endpoint of this path (whose interior consists of regular points at all times) travels along the fold arc φ_k to the cusp between φ_k and φ_1, then serves as the cusp merge path $p_{cm}\colon [0,1] \to M_{1/2}$ by which the two critical circles reunite. In other words, there is a smooth embedded disk in M whose boundary c is naturally partitioned into three arcs as follows: two arcs are cusp merge paths between the cusps involved in the shift, and the third edge consists of fold points in φ_k. In [BH, Remark 6.5] it is explained that there exist precisely two such disks, up to isotopy relative to the boundary, yielding a possibly distinct pair of shift moves on crown diagrams, and Behrens and Hayano describe how to construct the mapping class χ in this case. For such crown diagram moves or deformations this paper and [BH] use the term *shift* or *shift deformation*. The deformation $M_{[0,1]} \to S^2_{[0,1]}$ depicted in Figure 15 is supported in an open ball containing the cusp between φ_1 and φ_k and an open arc from φ_l, and restricts to a product map $\alpha_0 \times \mathrm{id}_{[0,1]}$ on the rest of the fibration. The following definition singles out the merge points of such a deformation.

Definition 2.10. A *local shift deformation* is a generalization of a shift deformation satisfying the following two conditions:

(i) There is an embedding of the fibration structure over $D^2_{[0,1]}$ corresponding to the deformation in Figure 15 into the fibration structure over $S^2_{[0,1]}$ given by α.

(ii) Let $\iota \subset \operatorname{crit} \alpha$ denote the collection of points where $\alpha|_{\operatorname{crit}\alpha}$ fails to be injective. As in [BH] and as suggested in Figure 15, $\alpha(\iota)$ is bounded away from the one-parameter family of disks in $S^2_{[0,1]}$ depicted in Figure 15.

Similar to a multislide pair, a *shift pair* is the pair of merge points in $\operatorname{crit} \alpha$ at either end of a local generalized shift deformation. In the proof of Lemma 3.7 there will appear pairs that satisfy Condition (i) but not (ii); these will be called *shift pair candidates*. In other words, a shift pair candidate is a shift pair with the possibility of fold crossings coming from fold arcs in Figure 15.

Like a local multislide deformation, a local shift deformation differs from the shift deformation that appears in [BH] only in one way: In this paper, a local shift deformation might be applied to a fibration with immersed fold arcs or disconnected critical locus, while a shift deformation is a local shift deformation whose endpoints are crown maps. Condition (ii) is nontrivial: it proscribes, for example, embeddings

such that $\alpha|_{\tilde{\varphi}_k}$ is not injective, or embeddings such that the image of χ_1 crosses folds as it travels along p_{cm}.

Remark 2.11. The shift deformation contrasts to the multislide deformation, in which the circle in M traced by one cusp (while the other remains stationary) may not bound a disk. Suppose we are given a deformation that corresponds to a shift, except that the cusp merge path p_{cm} and the path λ traced by χ_1 from m_1 to $p_{cm}(0)$ gives a circle

$$c = p_{cm} \cup \tilde{\varphi}_k \cup \lambda \subset M_{1/2}$$

that does not bound a disk. One may decompose this into a multislide deformation followed by a shift deformation as follows. Beginning at a value $t = t_0$ immediately after the initial fold merge, perform the following deformation: Cusp merge χ_1 and χ_2 using a cusp merge path that causes χ_1 to trace out a circle which is homotopic to c, and then immediately perform the reverse of that cusp merge. From there, the deformation continues as it did originally, beginning with the map α_{t_0}. In the critical surface, it looks like the result of performing a self-connect sum along a pair of points contained in the cusp arcs emanating from the initial merge point. Now the deformation consists of a multislide followed by a shift. There is clearly an analogous modification in the local case.

Definition 2.12. A deformation that consists of a deformation followed precisely by its reverse is called a *detour*.

As in Remark 2.11, any detour beginning with the Morse 2-function α_{t_0} clearly can be inserted into α at $t = t_0$ to get a new deformation α' with the same endpoints, and this is a tool that will be used repeatedly. It is easy to see that α and α' are homotopic: A detour is just a loop in the space of maps $M \to S^2$ given by an arc followed by its reverse, which is a nullhomotopic loop based at α_{t_0}.

3. Proof of the main theorem

Previous sections connected model deformations with moves on crown diagrams. The rest of the paper is concerned with the part of the proof of Theorem 1.1 that converts a given deformation into a sequence of model deformations.

3.1. **Some notation.** Maps $M_I \to S^2_I$ will always be smooth and will always be the identity on the first factor (that is, they are smooth homotopies). Finally, we routinely conflate various maps with the fibration structures they induce. The critical surface of a deformation is a smooth 2-submanifold of $M_{[0,1]}$ corresponding to the moves in [W1, Section 2.4] (also paraphrased in Section 1.3), where there appears proof of the statements of the rest of this paragraph. Since α is a deformation, the function $T = t|_{\operatorname{crit}\alpha}$ that projects $\operatorname{crit}\alpha$ to the t-coordinate is Morse and its critical points are those at which birth and merge moves occur. The further restriction $T|_{\overline{\chi}}$ to the closure of the cusp locus $\chi \subset \operatorname{crit}\alpha$ is also a Morse function on a smooth 1-manifold, whose boundary consists of the cusp points in α_0 and α_1. Since we are assuming α_0 and α_1 are crown maps, there is one boundary circle in each M_i, possibly dotted with cusps, and α_i is injective on each. Examining the local models for birth, merge and flip, it becomes clear that the index zero and two critical points of T are precisely where births and their inverses occur, while the merges occur precisely at the index one critical points of T. Call these points birth and merge points, regardless of whether they correspond to births and merges or

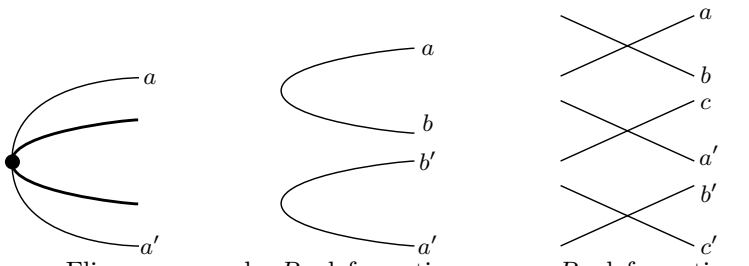
their inverses. The critical points of T at birth and merge points are all cusps. The critical points of $T|_{\overline{\chi}}$ which are not birth or merge points are all swallowtail points, where flipping moves occur in the deformation.

As in Figures 4 and 5, the critical manifold of a deformation can be depicted with various decorations: the immersion locus ι consisting of paired immersion arcs; the cusp locus χ, which is a smoothly embedded 1-submanifold of $\operatorname{crit}\alpha$; and the collection of swallowtail points, which coincides with $\overline{\chi} \smallsetminus \chi$.

Definition 3.1. For a double point x in the critical image of α, let $\{p, p'\} = \alpha^{-1}(x) \cap \operatorname{crit}\alpha$. Then the *counterpart* of p is p'. Similarly, an arc or a circle in ι can have a counterpart, defined as the arc or circle in ι with the same image under α. For any $A \subset \iota$, the counterpart of A will be denoted A'.

Though such a decorated depiction of $\operatorname{crit}\alpha$ is a picture of a smooth two-dimensional submanifold of the 5-manifold M_I, it is only a recording of the one-parameter family of stratified immersions $\alpha(\operatorname{crit}\alpha)$ and vanishing cycles, so it more properly should be considered a depiction of a one-parameter family of base diagrams. When a decorated critical surface is a concatenation of decorated critical surfaces coming from the four model deformations (as for example in Figure 4), there is a unique one-parameter family of corresponding base diagrams. For this reason, modifying the decorated critical surface of a deformation until it is such a concatenation is a valid strategy for proving Theorem 1.1. On the other hand, more complicated decorated critical surfaces may not uniquely specify a family of base diagrams. An example due to an anonymous referee is that there could be two concentric circles in the critical image, and the outside circle undergoes a local handleslide deformation in such a way that the inside circle could lie in one of two regions of regular values during the time the crossings exist. What is required for this paper is that at least one corresponding deformation exists. It is important to note that a base diagram does not in general specify the total space up to diffeomorphism; see for example the discussion around [GK2, Figure 8]. Similarly, a decorated critical surface is a typically lossy tool for recording a deformation, though it turns out to be sufficient for proving Theorem 1.1.

It follows from Lemma 1 of [W1] that, as t increases or decreases, there are only two ways for crossings to form in the critical image of a deformation: flips and R_2 deformations. In the critical surface, a flipping move is encoded by a pair of immersion arcs nested around a pair of cusp arcs whose common endpoint is a swallowtail point as in Figure 16a. The R_2 deformation appears as two immersion arcs ι_1 and ι_2 that appear as in Figure 16b. An important feature of deformations, which will be called property s_2, is that the maps $T|_{\iota_i}$ each have a single common critical value for each R_2 deformation; in other words, the arcs ι_i are tangent to the same level set M_t, and such tangencies always come in pairs, one pair for each R_2 deformation. Finally, there may be Reidemeister-III fold crossings in which cusps may lie in the initial triangle (call these R_3 deformations for short), whose immersion loci appear in $\operatorname{crit}\alpha$ as in Figure 16c. These also come with a simultaneity condition which will be called property s_3: the three intersection points between immersion arcs in the critical surface must all have the same t-coordinate. Since α_0 and α_1 are injective on their critical circles, the closure of the immersion locus is a union of circles; let c denote one of these (recall that swallowtail points are included in the immersion locus).



a. Flip. b. R_2 deformation. c. R_3 deformation.

FIGURE 16. Critical and immersion loci of α. In such diagrams t increases to the right, cusp arcs are bold, and pairs of fainter arcs are mapped to each other by α (components may be flipped or reordered vertically depending on α). The dot is a swallowtail point.

The constructions used in the proof of Theorem 1.1 are mostly recorded as modifications of $\operatorname{crit}\alpha$ and its stratified immersion into S^2_I, and it is crucial to know that those changes actually produce deformations. Given a decorated critical surface, there are at least three conditions that must be satisfied by the result:

(1) The critical points of T must be precisely the birth and merge points of α, and $\operatorname{crit}(T|_{\overline{\chi}})$ must be the disjoint union of $\operatorname{crit} T$ with all the swallowtail points.
(2) The immersion locus must come from an immersion of a surface into S^2_I (possibly with corners coming from cusps) and follow one of the local models in Figure 16 at each tangency with any M_t.
(3) The vanishing cycles around birth, merge, swallowtail and immersion points must be valid according to:
   (a) The local models for birth, merge, and flip;
   (b) The disjointness of vanishing sets as discussed in [W2, Definition 2.1];
   (c) The identifications prescribed by each pair of immersion points (p, p'): The vanishing cycle of an immersion point p must be the vanishing cycle of the fold containing p'.

It is clear that the critical surface of any deformation satisfies these three conditions. In the other direction, it is a subtle question to determine whether a decorated critical surface actually comes from a deformation. In the constructions of this paper, verifying that modifications of a decorated critical surface actually yield new deformations is typically verified on an ad hoc basis by exhibiting corresponding base diagrams where necessary, using results from [W2], and using other arguments that do not appeal to the properties of any particular decorated critical surface. On the other hand, the results of Section 2 that connect the moves on crown diagrams with model deformations show that a deformation is given by a sequence of model deformations if and only if its decorated critical surface is a concatenation of the decorated critical surfaces of model deformations. The approach is to modify the decorated critical surface until the sequence of moves can be read off like in Figure 4, though the precise implementation of each move as a modification of vanishing cycles is lost.

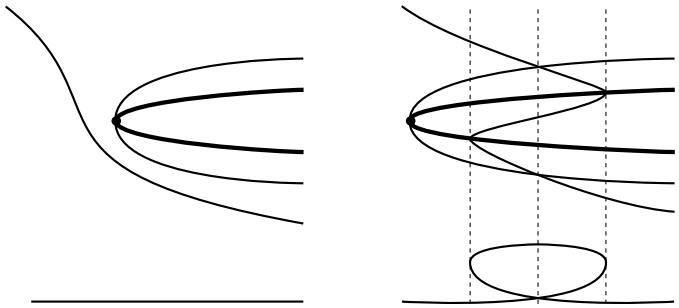


FIGURE 17. Figure for Lemma 3.2. A flipping move is pushed backward in t so that its swallowtail point moves across an immersion arc.

3.2. **Splicing cusp arcs.** This section gives a few tools which will be important for gaining some control over how the cusp locus is embedded in $\operatorname{crit}\alpha$. The section begins with an observation about how one can push swallowtails around in a decorated critical surface.

Lemma 3.2. By homotopy of α, it is possible to move the swallowtail point of a flipping move backward in t (or an inverse flipping move forward in t) across any adjacent immersion arc. The result appears in Figure 17.

Proof. In such a homotopy, the initial deformation is just a flipping move occurring on a fold arc A near an intersection of A with another fold arc B. The decorated critical surface for such a deformation appears on the left side of Figure 17. Following the picture on the right, the terminal deformation is also a flip on a point of A, but on the other side of the crossing. Then the fold arc B moves across the loop formed by the flip, undergoing a pair of cusp-fold crossings and a Reidemeister-III fold crossing.

The local model for a swallowtail has a particular fibration structure

$$f_{[-\varepsilon,\varepsilon]}\colon B^4_{[-\varepsilon,\varepsilon]} \to B^2_{[-\varepsilon,\varepsilon]}$$

for any $\epsilon > 0$, and the required one-parameter family of deformations comes from simply choosing an appropriate extension of this fibered neighborhood by a ball of regular points of α (for example, take the union of $B^4_{[-\varepsilon,\varepsilon]}$ with a neighborhood of an arc that is parameterized by t, disjoint from $\operatorname{crit}\alpha$ except for its terminal point which is the swallowtail), then within this larger ball making a one-parameter family of coordinate changes that sends the swallowtail point backward in t. □

The first tool for modifying the critical locus is a local one that introduces two swallowtail points into a cusp arc, as shown in Figure 18. As usual, the two dots are swallowtail points: the dot on the left corresponds to a flipping move, the one to the right an inverse flip. The bold arcs are cusps, and there is a fainter circle of immersion points, where vertically aligned points are mapped to each other by α.

Figures 19 and 20 give base diagrams for the right side of Figure 18. At first, there is a cusp as in Figure 19a; a flip occurs, giving Figure 19b; the two upper vanishing cycles in the reference fiber follow directly from the local model for flips as described in [L1, W2, GK1], while the lower two come from the fold arcs on

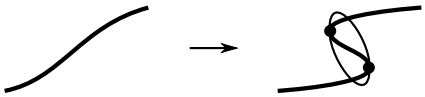
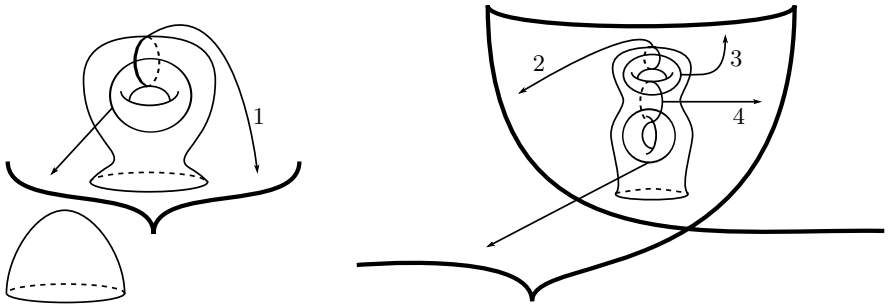
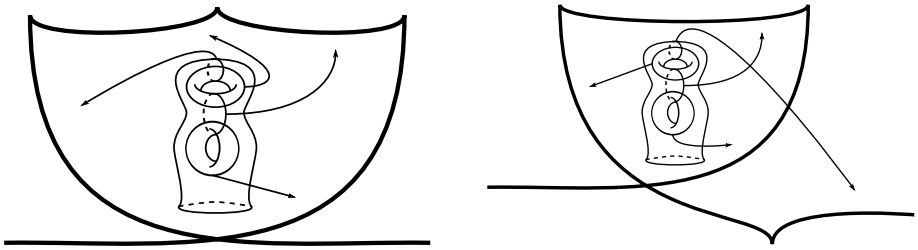


FIGURE 18. A pair of swallowtail points inserted into a cusp arc.



a. b.

FIGURE 19. A pair of swallowtails introduced to a cusp arc, part one.



a. b.

FIGURE 20. A pair of swallowtails introduced to a cusp arc, part two.

either side of the original cusp. To obtain Figure 20a, the cusp at the lower left moves into the higher-genus region, with the vanishing cycles unchanged. Next, the cusp at the top left of Figure 20a moves into the lower-genus region resulting in Figure 20b, at which point the loop can be shrunk away by an inverse flipping move, and the upper two vanishing cycles survive.

Following a suggestion from an anonymous referee, it may be helpful in Figure 19b to use a reference point that lies just above the loop formed by the flipping move, and reference paths that cross the upper fold arc. With those choices, vanishing cycle 3 appears as two points in the fiber and vanishing cycles 2 and 4 more clearly appear to come from vanishing cycle 1. In this way, it becomes clear that any intersections of distant vanishing cycles with vanishing cycle 1 can be moved freely between vanishing cycles 2 and 4.

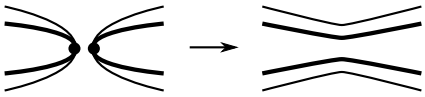


FIGURE 21. Canceling a pair of swallowtails.

FIGURE 22. Moving a swallowtail into a neighboring region of fold points. The bold arcs are cusps and the thin ones are immersion arcs.

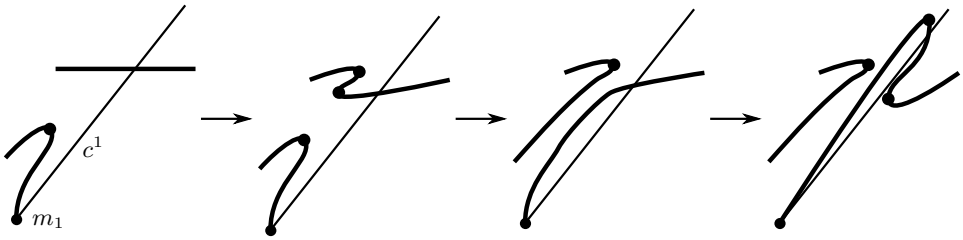


FIGURE 23. How to string a cusp arc coming from the merge point m_1 along the arc c^1 when c^1 crosses another cusp arc (this notation appears in the proof of Lemma 3.7). The immersion locus is omitted.

It is also possible to cancel two swallowtails that are connected by an arc of fold points as in Figure 21. In this situation, the base diagrams for the left side have an inverse flip, only for a flipping move to later occur on the same fold arc. Taking advantage of the local model for flips,

$$(t, x_1, x_2, x_3, x_4) \mapsto \left(t, x_1, x_2^4 + x_2^2 t + x_1 x_2 \pm x_3^2 \pm x_4^2\right) \tag{4}$$

in which a flipping move occurs at $t = 0$, the left side of the figure has a local model obtained by replacing t with $t^2 + \epsilon$ for some small $\epsilon < 0$; increasing ϵ to a positive value gives the right side. The gluing between the two local models of flips implicit in this construction comes from a normal framing to the fold locus along the arc of fold points connecting the two swallowtail points. One application allows one to change the fold arc upon which any flip occurs in a deformation, as in Figure 22. Combining Figure 22 with another instance of Figure 18, it is possible to arrange for a cusp arc to follow any path in $\operatorname{crit}\alpha$ parameterized by its t coordinates, as in Figure 23.

3.3. **Switching between cusp and fold merges.** In some places it will be necessary to ensure a cusp merge occurs at a particular merge point $m \in \operatorname{crit}\alpha$, or that a fold merge occurs at m. For instance, in a multislide deformation the first merge point is a fold merge and the second is a cusp merge. To address this issue we use a trick originally due to Denis Auroux, enabling one to switch between a fold merge and cusp merge in the presence of a flipping move, as follows.

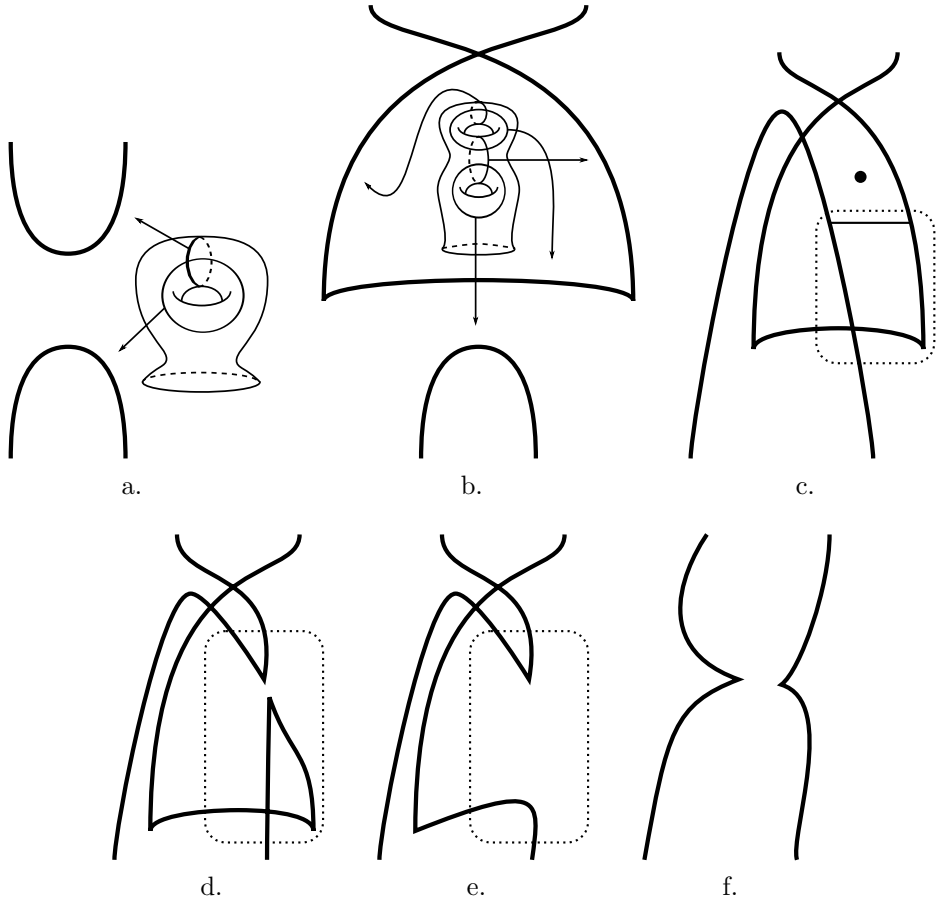


FIGURE 24. Putting a fold merge into position for reversal.

Suppose a fold merge occurs at m, and we wish to modify α so that a cusp merge occurs instead. The other case, converting a cusp merge into a fold merge, is precisely the same modification with t reversed. The first step is to put m into a situation to which Auroux's trick applies. For this one may introduce a pair of swallowtails along its cusp arc qualitatively like in Figure 18, except the two swallowtails lie before and after the merge point with respect to t; the precise modification appears in Figure 24. The way to interpret this figure is as a replacement for the fold merge deformation, whose base diagrams are given by Figure 24a followed immediately by Figure 24f; for fiber genus at least 2 (which will follow from Lemma 3.3 wherever it is used), the validity of the replacement follows from the validity of the intervening base diagrams and the fact that the modifications therein occur relative to the fibration above the boundary of the target disk. The modification proceeds with a flipping move followed by an R_2 deformation to obtain Figure 24c; the vanishing cycles follow directly from the local model for the flip and the R_2 deformation is valid by [W2, Proposition 2.7]. The decorated reference fiber in Figure 24b can be transferred to be the reference fiber above the dot in

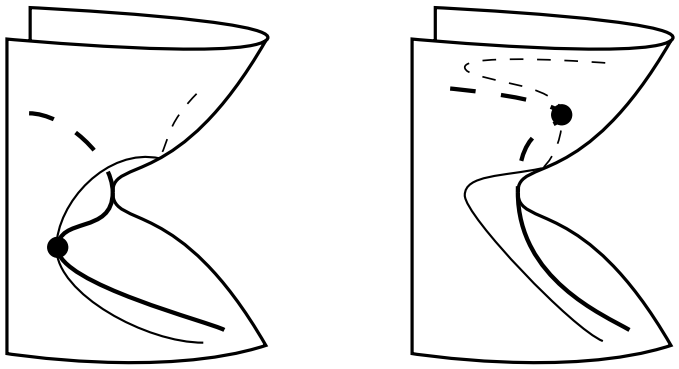
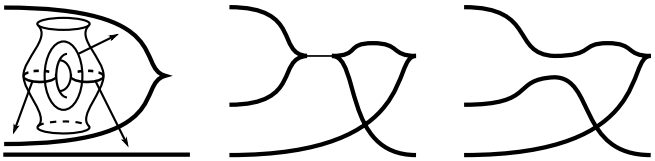


FIGURE 25. Switching between cusp merge (left) and fold merge (right).

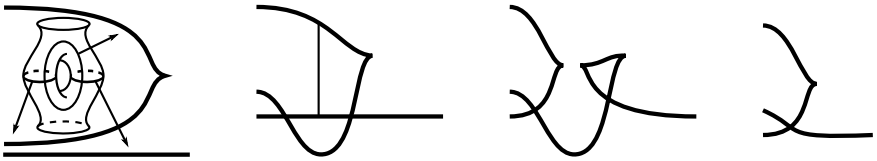
Figure 24c, with arrows pointing at the same fold arcs as before. In other words, the fold arcs of Figure 24c *inherit* their vanishing cycles from Figure 24b and we will use that term repeatedly in such contexts. Performing the indicated fold merge to obtain Figure 24d, the next steps are to cancel the top two intersections with an R_2 deformation (which is valid by an application of [W2, Proposition 2.7] followed by [W2, Proposition 2.5(2)]) and to close off the lower loop by an inverse flip, which can be seen to be valid by drawing in the three relevant vanishing cycles inherited from those of Figure 24b. The deformation above the three dotted rectangles now contains a fold merge that can be replaced by one containing a cusp merge using Auroux's trick, which follows.

The left side of Figure 25 is a depiction of the saddle-shaped critical surface in M_I above the dotted rectangles in Figure 24 (for t increasing left to right, it corresponds to the reversed order 24e, 24d, 24c because it is easier to depict and validate the deformations this way). As usual, the bold arcs are cusp arcs, and in each figure m lies at the saddle point of the surface, with a flipping move occurring at the dot. The fainter arcs are the immersion locus. It is is not difficult to deduce the base diagrams specified by the left side of Figure 25 by considering vertical slices of the figure from left to right: the progression for the left side appears in Figure 26a, which is a copy of what happened in the dotted rectangles from before. Here a fold arc experiences a flip, then one of the resulting cusps merges with another preexisting cusp.

The right side of Figure 25 has base diagrams given by Figure 26b, which will replace those of Figure 26a in the dotted rectangles. Again, following vertical slices from left to right, one may deduce that the sequence of moves must begin with an R_2 deformation between a fold arc and another fold arc containing a cusp. The deformation concludes with a fold merge and an inverse flip as appears in Figure 26b. For fiber genus at least 2, which will be the case according to Lemma 3.3 below, the validity of these moves and the intended substitution follow entirely from the vanishing cycles in the initial base diagrams in Figure 26, which are inherited from the base diagrams of Figure 24. In Figure 26b the initial R_2 deformation is valid because the vanishing cycles involved are disjoint. The following fold merge and inverse flip are straightforward to verify using the vanishing cycles inherited from the initial base diagram. To restate the prescription for turning a fold merge into a



a. The reverse of this deformation appears in the sequence of three dotted rectangles in Figure 24.



b. The reverse of this deformation can replace the sequence of three dotted rectangles in Figure 24.

FIGURE 26. Base diagrams for the left and right side of Figure 25, respectively.

cusp merge, one first substitutes the reverse of Figure 26b into the dotted rectangles of Figure 24, and then substitutes the resulting deformation for a neighborhood of the merge point m.

3.4. **Modification of the critical manifold.** Here we begin to use the tools from above to prove Theorem 1.1; it may be helpful to refer to the flowchart in Section 4.

As discussed in Section 1.5, the argument begins with a given deformation $\alpha\colon M_I \to S^2_I$ whose endpoints are maps that induce crown diagrams. Provided the endpoints are homotopic, a deformation between them was proven to exist in [W1, Theorem 1], but we need something more: by a theorem of Gay and Kirby, one may assume the fibers of α are connected. Taking their arguments slightly further, the following lemma is the starting point for our modifications. From this point on, α should be taken to be the deformation that results from applying any preceding lemmas.

Lemma 3.3. Let α_0 and α_1 be homotopic crown maps, both of which have lower fiber genus at least 2. Then there is a deformation α from α_0 to α_1 such that every regular fiber is connected and has genus at least 2.

Proof. By [GK1, Theorem 1.2], One may assume α has connected fibers. It remains to ensure those fibers have genus at least 2.

Here is an argument whose idea of inserting many birth detours is due to David Gay. The set of regular values of α is naturally partitioned into path components, each labeled with the genus of any one of its regular fibers. Given a path component P whose genus is minimal among all of these path components, break the closure of P into a stack of sausages; that is, a collection of closed 3-balls $B^1, \dots, B^n$ such that if $1 \le i < j \le n$, then $B^i \cap B^j$ is either empty or a 2-disk common to their boundaries, and such that $B^i \cap S^2_t$ is connected and simply connected for all i and for all t. The following is a technique to increase the fiber genus in a neighborhood

of any B^i, say B^1, in a way that allows the same technique to be used on what remains of the stack.

The boundary of P is a union of polygons in the critical image. These polygons meet at 1-dimensional seams wherever there are crossings in the base diagrams of ∂P and at a finite set of corners, at each of which three seams meet at a Reidemeister-III fold crossing. If ∂B^1 intersects ∂P but has no seams, then ∂B^1 also contains no corners, and there is a restriction of the fibration over B^1 given by α to an honest fiber bundle whose fiber is the regular fiber over B^1 with one puncture. This comes from choosing a horizontal distribution for which the vanishing set in each regular fiber over B^1 coming from critical points over ∂B^1 is a unique pair of points, choosing an embedded arc connecting those points, and removing the closed arc from the fiber (note that the minimal genus of P causes P to lie on the lower-genus side of its boundary). We remove the arc because the two points of a vanishing set approach each other as the end of a reference path approaches a cusp, so the arc allows a uniform choice of restricted fiber F. Similarly, if ∂B^1 intersects any positive number of seams in ∂P but has no corners, there is a fiber bundle whose fiber is twice punctured, and three times if there is a positive number of corners.

In all three cases, the base of the bundle can be extended just past ∂B^1, where a puncture can be interpreted as coming from removing a neighborhood of a vanishing cycle, then removing an arc connecting the two boundary circles that flows smoothly to the arcs in the lower-genus fibers. The disjointness of vanishing cycles needed for this to work near a seam or corner comes from the existence of the relevant seam or corner.

In all three cases, the resulting bundle over a tubular neighborhood of B^1 is trivial,

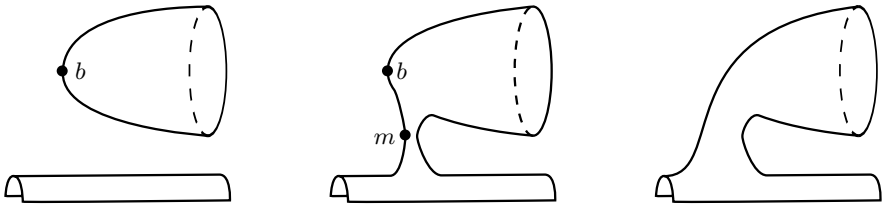
so it has a section, and that section has a neighborhood U of regular points. Then there is a deformation supported on U consisting of a birth followed by its reverse, such that B_1 is contained in the open region R^1 enclosed by the resulting critical sphere in S^2_I. With this accomplished, the genus of fibers near B^1 has increased by one.

Now, supposing the above modification of α has been applied for $B^1, \dots, B^k$, the same argument applies inductively to $B^{k+1} \setminus \cup_{i=1}^k R^i$. Finally, this modification can be applied repeatedly to each minimal-genus region whose genus is less than 2, until all path components of regular values have genus at least 2. □

Remark 3.4. The modifications that appear from now on do not decrease fiber genus or introduce disconnected fibers in the sense that they do not increase $\max_{p \in S^2_I} \{\chi(\alpha^{-1}(p))\}$. This is achieved by not performing genus-decreasing finger moves or R_3 deformations in which the triangle is on the higher-genus side of all three fold arcs (see [W2, Propositions 2.5(1a) and 2.9], respectively), and by similarly restricting modifications to the immersion locus (see Lemma 3.19).

Lemma 3.5. Let α be a deformation given by Lemma 3.3. Then there is a deformation with the same endpoints whose critical locus is connected.

Proof. The following argument is reminiscent of Theorem 6.1 of [L1], though variations of it have appeared from time to time, going as far back as [Lev]. Denote by A the path component of $\operatorname{crit} \alpha$ whose image under T contains the minimum value of T, and suppose there is a path component $B \subset \operatorname{crit} \alpha$ distinct from A. Then B must



a. A birth occurs. Cusp arcs run along the profile of the upper disk.

b. A self-connect sum in $\operatorname{crit}\alpha$.

c. Perturb to eliminate b and m.

FIGURE 27. Canceling critical points of T in the proof of Lemma 3.6.

have an index 0 T-critical point because $\operatorname{crit}\alpha_0$ is connected, and by a homotopy of α one may simply push that point backward in t until $T(A)\cap T(B)$ contains an open interval; choose t_0 in that interval. Here follows a short deformation f whose endpoints are α_{t_0}.

If A has a cusp at $t = t_0$ then there exists a curve $v\colon [0,1]\to M$ from that cusp to one of the two cusps formed in the birth by which B appeared, transverse to the fibers, whose intersection with $\operatorname{crit}\alpha$ at t_0 is precisely the two cusps at its endpoints. Restricting the fibration to a neighborhood of v, by a small perturbation near its endpoints it is possible to arrange for v to specify a cusp merge as in the right side of Figure 4 of [W1]. Starting with the fibration α_{t_0}, insert the detour given by performing the cusp merge and immediately its reverse. If there is no such cusp, begin with a flip on a fold arc of A to introduce two cusps in A, perform the merges mentioned above, then end with the inverse flip that returns the fibration to α_{t_0}. This is the deformation f. It is clear that $\operatorname{crit} f$ is a cylinder with two T-critical points, one at each merge point. The endpoints of f agree with α_{t_0}, so that it is possible to insert this deformation into α to get a new deformation, resulting in the connect sum of A and B. Repeating this process a finite number of times decreases the number of components of $\operatorname{crit}\alpha$ to one. □

Lemma 3.6. Let α be the deformation resulting from Lemma 3.5. Then there is a deformation with the same endpoints satisfying exactly one of the following conditions:

- If $\operatorname{crit}\alpha_0$ and $\operatorname{crit}\alpha_1$ are nonempty, then $\operatorname{crit} T$ consists entirely of merge points.
- If exactly one of $\operatorname{crit}\alpha_0$ and $\operatorname{crit}\alpha_1$ is nonempty, then $\operatorname{crit} T$ has merge points and a single birth point.

Note that if both endpoints have empty critical locus, then the corresponding crown diagrams have the same genus, and $\Gamma = \varnothing$ for both diagrams, so Theorem 1.1 automatically holds.

Proof. First assume $\operatorname{crit}\alpha_0$ and $\operatorname{crit}\alpha_1$ are nonempty, so that the goal, in other words, is to remove all the birth points. By Lemma 3.5, $\operatorname{crit}\alpha$ is connected so that $\operatorname{crit}\alpha$ is nonempty just before any birth move.

Suppose a birth occurs, with birth point $b \in M_t$, and choose small $\epsilon > 0$. As suggested by Figure 27, and just like in the proof of Lemma 3.5, perform a self-connect sum of $\operatorname{crit}\alpha$, where the initial cusp merge has its merge point $m \in M_{t+\epsilon}$, and occurs between one of the cusps that emanate from the birth point b, and a cusp point that does not lie within the newly introduced critical circle coming from b. Choosing ϵ sufficiently small, the deformation in a neighborhood of the cusp arc connecting b and m appears in base diagrams as a birth near a cusp point, immediately followed by a cusp merge between that cusp and one of the cusps in the newly introduced critical circle C. Now it is not hard to see that there is a homotopy of such a deformation in which m moves toward the level set M_t, canceling b as in Figure 27c. This completes the argument for removing birth points from α, resulting in a new deformation; applying the same argument with t reversed removes the inverse birth points in a symmetric manner. □

Recall the Definitions 2.7 and 2.10 for *multislide pair* and *shift pair*, respectively. Hereafter, the term *merge pair* refers to a pair of merge points that forms either a shift pair or a multislide pair.

Lemma 3.7. The deformation α resulting from Lemma 3.6 can be modified so that every index one critical point of $T\colon \operatorname{crit}\alpha \to I$ is part of a merge pair.

Proof. Denote the first merge point of the deformation α at $t = t_1$ by $m_1 \in \operatorname{crit}\alpha_{t_1}$. Because $\operatorname{crit}\alpha_t$ is connected for $t \le t_1$, $\operatorname{crit}\alpha_t$ has two components for t slightly larger than t_1. Since $\operatorname{crit}\alpha_1$ is connected, eventually all components much reunite at some merge point (called $m_2 \in \operatorname{crit}\alpha_{t_2}$ below) and the first step (speaking as if there were never more than two components) is to define a pair of paths between m_1 and m_2, one in each component. More precisely, choose two smooth curves $c^i_{[t_1,t_2]} \subset \operatorname{crit}\alpha_{[t_1,t_2]}$, $i = 1, 2$, satisfying the following conditions:

- $c^i_{t_1} = m_1$ and $c^i_{t_2} = m_2$ for $i = 1, 2$ and for some $m_2 \in \operatorname{crit}\alpha_{t_2}$, where $t_1 < t_2$.
- c^1_t and c^2_t lie in distinct path components of $\operatorname{crit}\alpha_t$ for $t \in (t_1, t_2)$.

The above conditions imply that the point m_2 is the first merge point at which the path components of $\operatorname{crit}\alpha_t$ containing c^i_t become reunited (as above, there exists such m_2 because $\operatorname{crit}\alpha_1$ is connected). The idea of the proof is to first use the splicing deformation to introduce a cusp arc which is in some sense as isotopic as possible to the circle $c = c^1 \cup c^2$, then use it in a surgery to shrink c to size, which connects those path components at each t except for a short interval containing t_1. The next step is to insert detours containing pairs of merge points near m_1 which allow the immersion locus to be organized such that the result is a sequence of merge pairs starting with m_1. The result then follows inductively, applying the same construction for the next merge point to occur after this sequence.

Recall that for m_1, m_2 to form a merge pair, it is necessary for a fold merge to occur at m_1 and a cusp merge at m_2. If this is not the case for some m_i, use Auroux's method from Section 3.3 to make it so, and construct c^1, c^2 as above.

This paragraph arranges for m_1 and m_2 to be contained in a single cusp arc as required for any merge pair, or if that is not easily possible, arranges for the cusp arcs containing m_1 and m_2 to be adjacent in the part of $\operatorname{crit}\alpha_{(t_1,t_2)}$ containing the arcs c^1 and c^2. In other words, for $i = 1, 2$ and for $t \in (t_1, t_2)$ there is an arc of fold points in M_t which contains c^i_t and connects the two cusp points contained in the cusp arcs which contain m_1 and m_2. If necessary, perturb c^i to be transverse in

$\operatorname{crit}\alpha$ to the cusp and swallowtail locus, so that c^i consists of fold points except for transverse intersections of the interior of c^i with the cusp locus. The modification of Figure 23 allows one to arrange that the two cusp arcs coming from m_1 can follow c^1 and c^2 (respectively) to approach m_2, and for each i, a final application of Figure 22 involving the cusp arcs containing m_1 and m_2 causes those merge points to lie within a single cusp arc or to be adjacent, depending on which of the two cusp arcs coming from each of the two swallowtails contain m_1 and m_2 (this can be understood by imagining m_1 lies at the upper right in Figure 22, and m_2 lies at either the left or right of the lower cusp arc). In either case, let $\gamma \subset \operatorname{crit}\alpha_{[t_1,t_2]}$ denote the cusp arc containing m_2.

After applying these modifications to the cusp locus, the next step is to choose a positive $\epsilon \ll t_2 - t_1$ and a cusp merge path $p_{cm}\colon [0,1] \to M_{t_1+\epsilon}$ between the cusps that reunite at m_2, as follows. The cusp arc $\gamma \subset M_{[t_1+\epsilon,t_2]}$ from the last paragraph has both endpoints in $M_{t_1+\epsilon}$, and $T|_\gamma$ has exactly one critical point, the merge point m_2. Let $\tilde{p}_{cm}$ denote the projection of γ to $M_{t_1+\epsilon}$. Now, $\tilde{p}_{cm} \cup \gamma$ is the boundary of a smooth embedded disk $D \subset M_{[t_1+\epsilon,t_2]}$, whose intersections with the level sets M_t form a one-parameter family of embedded arcs which is transverse to the fibers (TD is spanned by $\frac{\partial}{\partial t}$ and a lift to M_t of a nonzero tangent vector to the image of each path in S^2_t). Further, since $\dim D + \dim \operatorname{crit}\alpha < \dim M_{[0,1]}$, D can be chosen disjoint from $\operatorname{crit}\alpha$ except along the boundary arc that coincides with γ. After possibly perturbing D near the part of its boundary that coincides with γ, $D_{[t_1+\epsilon,t_2-\epsilon]}$ becomes a one-parameter family of cusp merge paths (each with an unspecified framing in the sense of [BH, Section 3.2.2]); set p_{cm} to be a parameterization of the arc $D_{t_1+\epsilon}$. There is a framing of the arc $D_{t_2-\epsilon}$ that makes it a cusp merge path for m_2, and transporting this framing back across D to lie in $M_{t_1+\epsilon}$ also turns p_{cm} into a framed cusp merge path between the endpoints of γ. This is the unique framing such that, if one were to insert the detour given by performing the cusp merge according to p_{cm} (producing a merge point m_1' in, say, $M_{t_1+2\epsilon}$) and then immediately its reverse (a fold merge producing the merge point $m_1'' \in M_{t_1+4\epsilon}$), in a neighborhood of $D \cap M_{[t_1+3\epsilon,t_2]}$ the deformation would be given by the concatenation of local models for the fold merge at m_1'' followed precisely by its reverse at m_2. The fibration structure near $D \cap M_{[t_1+3\epsilon,t_2]}$ could then be replaced by one whose base diagrams all consist of a pair of parallel fold arcs, leaving the merge points m_1 and m_1' and no others. For those who are not satisfied with the claim that the fibration structure near $D \cap M_{[t_1+3\epsilon,t_2]}$ is itself a detour, and so can be replaced as claimed, the replacement can be explicitly achieved using a homotopy of α supported on a fibered neighborhood of D, parameterized by $s \in [0,1]$, modeled on

$$H_{s,t}\colon (s,t,x_1,x_2,x_3,x_4) \mapsto (t, x_1, x_2^3 + 3((1-s)(1-2t^2) - s - x_1^2)x_2 + x_3^2 - x_4^2). \quad (5)$$

This model comes from beginning with the fold merge deformation

$$FM_t\colon (t,x_1,x_2,x_3,x_4) \mapsto (t, x_2^3 + 3(t - x_1^2)x_2 + x_3^2 - x_4^2),\ t \in [-1,1]$$

[W1, Equation 8], doubling it along its terminal fibration FM_1 by replacing the t in the coordinate

$$x_2^3 + 3(t - x_1^2)x_2 + x_3^2 - x_4^2$$

with $1 - 2t^2$ to obtain

$$x_2^3 + 3((1-2t^2) - x_1^2)x_2 + x_3^2 - x_4^2,$$

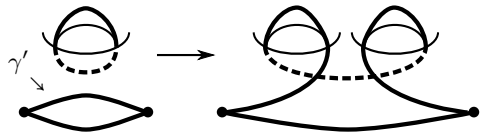


FIGURE 28. Turning a multislide pair candidate into two shift pair candidates. The dots are swallowtails and the bold arcs are cusps.

then interpolating the resulting deformation from $H_{0,t} = FM_{1-2t^2}$ toward the deformation $H_{1,t} = FM_{-1}$ by scaling the timelike parameter $1-2t^2$ by $1-s$. It is important to note that the presence of immersion arcs intersecting c^1 or c^2 does not affect the validity of this construction, because this is a modification of the fibration structure of M_I induced by α only near D.

Choosing ϵ small enough, there are no birth, merge or flipping moves in the interval $(t_1, t_1+2\epsilon)$ between the two merge points m_1, m_1' because they are supported in arbitrarily small balls in M_I that can be assumed disjoint from M_{t_1}. However, Condition (ii) for multislides or (ii) for shifts is not automatically satisfied: using for instance the labels from Figure 15, there may be cusp-fold crossings involving χ_1 as it travels along its cusp merge path p_{cm} (which could even have self-intersections in its image in S^2), or fold crossings involving $\tilde{\varphi}_k$ that exist throughout the interval $[t_1, t_1+2\epsilon]$. Both of these are forbidden in Definitions 2.7 and 2.10. One way to address this is to convert (m_1, m_1') into a sequence of shift pair candidates (recall the end of Definition 2.10), then give an algorithm for achieving Condition (ii) for these candidates. Once this is done, Condition (i) is easily verified.

Suppose m_1 and m_1' are contained in a circle of cusp points like at the top left of Figure 28. To turn the pair into two pairs of merge points, each of which is a shift pair candidate, insert a detour given by a flip at $t = t_1 - \epsilon$, followed by its reverse, so that there appears a small loop just to the side of either of the cusps that forms at m_1, then immediately perform its reverse. This creates a pair of cusp arcs in $\operatorname{crit}\alpha$ that meet in a pair of swallowtails. Use Lemma 3.2 to extend the local model for the inverse flip forward in t past m_1', resulting in the left side of Figure 28. Now there is a cusp arc γ' (whose endpoints are the newly introduced swallowtails) adjacent in the fold locus to the circle containing m_1 and m_1'. Using an arbitrary cusp merge path, perform a self-connect sum in $\operatorname{crit}\alpha$ between γ' and the cusp circle containing m_1 and m_1' using the same kind of detour as for Figure 27b (such a cusp merge path exists by [W2, Proposition 2.7]); the result is a pair of shift pair candidates.

Now suppose it is not the case that m_1 and m_1' are connected by a cusp arc, so that the cusp merge path p_{cm} connects cusps χ_1', χ_2' that are adjacent to the cusp points χ_1 and χ_2 (respectively) emanating from m_1 (left side of Figure 29). Modify this as above by performing a self connect sum in $\operatorname{crit}\alpha$ according to a cusp merge path between, say, χ_1 and χ_2'. This also results in a pair of shift pair candidates, and this concludes the list of cases for how the cusp arcs containing m_1 and m_1' lie within the fold locus for $t \in [t_1, t_1+2\epsilon]$.

We may now assume without loss of generality that the pair of merge points under consideration, m_1 and m_1', form a shift pair candidate, and now is a good time to state the notation, using when possible the labels from Figure 15. There is a fold merge at $m_1 \in M_{t_1}$ between fold arcs φ_k and φ_l forming two cusp points

FIGURE 29. Taking a connect sum between points in the two solid cusp arcs produces a pair of shift pair candidates.

χ_1 and χ_2, and a cusp merge at $m_1' \in M_{t_1+2\epsilon}$ according to an arbitrary cusp merge path $p_{cm}\colon [0,1] \to M_{t_1+\epsilon}$ which travels from χ_1 to χ_2', where χ_2 and χ_2' bound $\tilde{\varphi}_k$. Finally, let $p_{fm}\colon [0,1] \to M_{t_1-\epsilon}$ be the fold merge path for m_1, oriented so that $p_{fm}(1) \in \varphi_k$. The rest of the proof applies to any such pair of merge points, so it suffices to consider only this pair.

By definition, the shift pair candidate generally fails to be a shift pair because the image of the fold arc $\tilde{\varphi}_k$ (and possibly the cusp merge path p_{cm}) may cross itself or other fold arcs. The list below summarizes an algorithm to address these issues. The algorithm produces many shift pair candidates, successively more similar to a sequence of shift pairs. At each step, there are fold arcs, cusps, and merge paths that play analogous roles to those in previous steps. The following argument refers to all of these objects using the labels from Figure 15 to streamline the notation, but is careful to keep track of necessary distinctions. The result is that m_1 is the first merge point in a sequence of local shift deformations interspersed with many 2-parameter crossings.

(1) By a homotopy of α, shorten p_{cm} so that the cusp χ_2' between φ_1 and φ_k lies close to χ_1 (that is, pointing into the same region of regular values) as soon as it appears. This causes the new $\tilde{\varphi}_k$ to possibly have fold crossings and causes the new p_{cm} to be embedded and disjoint from the critical image of α.
(2) Break the deformation into a sequence of shift pair candidates that do not have self-crossings in their respective $\tilde{\varphi}_k$ fold arcs, but may have fold crossings between their respective $\tilde{\varphi}_k$ and other fold arcs. Do this in a particular way to allow item (4) below.
(3) For a shift pair candidate resulting from item (2), modify α by a homotopy so that $\tilde{\varphi}_k$ is free of all fold crossings, converting it to a shift pair.
(4) Immediately after the sequence of shift pairs from item (3), perform a sequence of 2-parameter crossings to return α to the endpoint of the original shift pair candidate deformation.

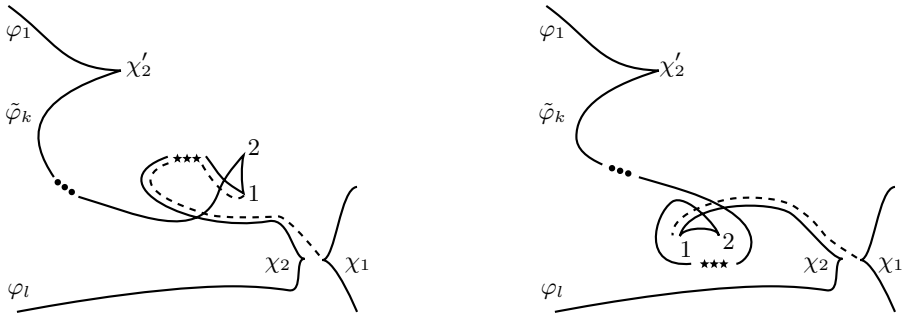
Item (1) is an extension of the modification involving D and $H_{s,t}$ above. Alternatively, but less rigorously, one may view it as pushing χ_2' along most of the reverse of p_{cm} before the t-coordinate t_1 of the merge point m_1, so that χ_1 is pointing into the same region of regular values as χ_2' as soon as it appears. Recall that the construction of the shift pair candidate resulted in a cusp merge in which χ_1 traveled along a cusp merge path p_{cm} toward the stationary cusp χ_2' (recall the labels from Figure 15). Insert a detour supported near p_{cm} in which χ_2' travels along the reverse of p_{cm} (as specified by the framing of p_{cm} as a joining curve), then returns to where it started. Now delay the return of χ_2' so that χ_1 trails it closely during its own traversal of p_{cm}. The part of p_{cm} that remains between the two cusps as they move together along p_{cm} is a one-parameter family of joining curves, so it traces out a disk in $M_{[0,1]}$ that serves an analogous role to D above, to

which the deformation $H_{s,t}$ may be applied. The fold merge at m_1 now produces cusps χ_1, χ_2 that point into the same region of regular points as χ_2'. The cusp-fold crossings originally undergone by χ_1 are now cusp-fold crossings undergone by χ_2' at t-values less than that of m_1.

Item (2) involves inserting a flip near each self-intersection of $\tilde{\varphi}_k$, and using its two cusps to break a candidate into three candidates, thereby breaking an immersed $\tilde{\varphi}_k$ into a collection of fold arcs, each serving as an embedded $\tilde{\varphi}_k$ for a member of a sequence of shift candidates. This addresses self-crossings in $\tilde{\varphi}_k$; the remaining crossings are treated in Item (3). Because of Item (1), assume without loss of generality that $\alpha(p_{cm})$ is a short embedded arc of regular values in $S^2_{t_1+\epsilon}$ connecting the cusp points $\alpha(\chi_1), \alpha(\chi_2')$. To begin, modify the fibration near each self-crossing of $\tilde{\varphi}_k$ according to two cases that correspond to the two types of self-crossings available to a fold arc. These cases appear in Figure 30: Each self-crossing in $\tilde{\varphi}_k$ receives an additional loop coming from a flip at $t = t_1 - \epsilon$ at the location suggested by the figure, which later goes away by an inverse flip at $t = t_1 + 3\epsilon$ (as usual, one adds these flips by inserting a flip immediately followed by its inverse and extending the local model for each move forward and backward in t by Lemma 3.2). In either case, instead of traveling along p_{cm}, the cusp χ_1 will first undergo cusp-fold crossings that exist by [W2, Proposition 2.7] to lie just to the higher genus side of cusp 1, following a path whose image is the dotted line in the figure, and which is transverse to the fibers and disjoint from $\operatorname{crit}\alpha$. Perform the detour consisting of a cusp merge between χ_1 and cusp 1, followed by its inverse, so that the fold that is parallel to the dotted line in Figure 30 now serves as the fold arc $\tilde{\varphi}_k$ for the first of three candidates to be produced by the modification. Similarly, send χ_1 along a path that follows the fold arc between cusps 1 and 2, and perform the same kind of detour with χ_1 and cusp 2. This happens all along the critical arc until χ_1 cusp merges with the cusp at the end of φ_1: χ_1 follows along $\tilde{\varphi}_k$, performing cusp-then-fold-merge moves at each cusp it encounters. This adds two candidates to the deformation for each self-intersection of the original $\tilde{\varphi}_k$. If any flips are inserted in this item, then the fold arcs trailing behind χ_1 will be different from that indicated by the short embedded path p_{cm} (and this is the only difference between the ending maps, because the merges were detours). Item (4) addresses this. Before that, however, Item (3) turns all of these candidates into shift pairs.

Item (3) addresses the fold crossings in $\tilde{\varphi}_k$, which by this stage are not self-crossings. Assuming items 1 and 2 have been applied, the given shift pair candidate has a cusp merge path whose image closely follows φ_k as in the dotted path of Figure 15, with the possible addition of other critical arcs crossing φ_k. In particular, Item (3) applies to the first and third candidates of any triplet produced in Item (2), while the second candidate needs no modification and is already a shift pair. During a t interval before the fold merge, the idea is to push any crossings in φ_k so that they do not that end up in $\tilde{\varphi}_k$ once the fold merge occurs, then perform the local shift deformation, then return those crossings to where they started. The merges and the movements of crossings have disjoint supports called U_{me} and U_{cr}, respectively.

The first step is to define U_{me} and U_{cr}. The set U_{me} will be the support of the shift candidate (*me* stands for *merge*), while U_{cr} will be the support of a homotopy described below that pushes the fold crossings away from the part of φ_k that will become $\tilde{\varphi}_k$ (*cr* stands for *crossings*). The second step is to prove that such a homotopy exists, and that one may arrange for U_{me} and U_{cr} to be disjoint, so that



Figures for Item (2) in the proof of Lemma 3.7 giving the placement of a flip. The text of Item (4) refers to the star ellipsis in each.

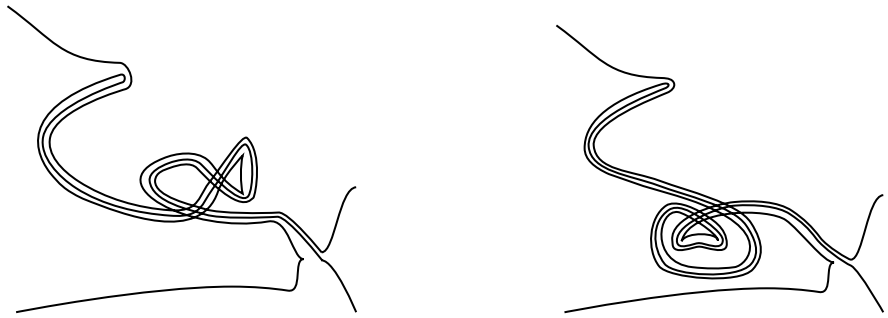


FIGURE 30. The fibrations which result from the added merge pairs indicated in the two figures above, assuming $\tilde{\varphi}_k$ had just one self intersection. Each is not the original fibration $\alpha_{t_1+2\epsilon}$; the discrepancy is addressed by Item (4).

the proposed deformation, a local generalized shift deformation occurring between some 2-parameter crossings, has the same endpoints as (and thus can be substituted for) the candidate we started with.

To set notation, pick $\epsilon > 0$ and call the merge points of the candidate $m_1 \in M_{t_1}$, $m_1' \in M_{t_1+2\epsilon}$, using the same labels as in Figure 15. To precisely define U_{me}, it helps to refer to Figure 31. The initial fold merge is supported on a tubular neighborhood of the fold merge path $p_{fm}\colon [0,1] \to M_{t_1-\epsilon}$ between φ_k and φ_l, and the subsequent cusp merge is supported on a tubular neighborhood of the cusp merge path $p_{cm}\colon [0,1] \to M_{t_1+\epsilon}$ from the resulting cusp on the right, χ_1, to the cusp between φ_1 and $\tilde{\varphi}_k$ at $p_{cm}(1)$. Now, as in Remark 2.8, p_{fm} specifies a family of merge paths $[p_{fm}(s)]_{[t_1-\epsilon,t_1)}$. Thus, $[p_{fm}(1)]_{t_1-\epsilon}$ is connected to $[p_{cm}(0)]_{t_1+\epsilon}$ by the union of $[p_{fm}(1)]_{[t_1-\epsilon,t_1]}$ and a cusp arc $\gamma \subset M_{[t_1,t_1+\epsilon]}$. Denote the projection $\pi_t\colon M_I \to M_t$ and the concatenation of the four paths by P:

$$P = [p_{fm}([0,1])]_{t_1-\epsilon} * [p_{fm}(1)]_{[t_1-\epsilon,t_1]} * \gamma * [p_{cm}([0,1])]_{t_1+\epsilon} \subset M_{[t_1-\epsilon,t_1+\epsilon]}.$$

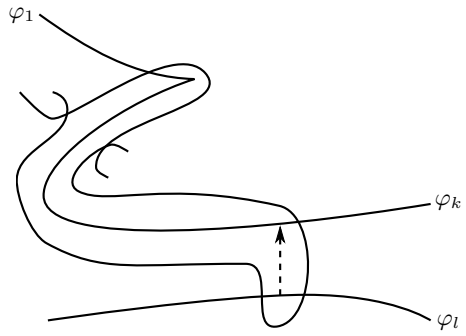


FIGURE 31. Applying Item (3) to push two fold crossings off the relevant part of φ_k. The dotted line is $\alpha(p_{fm})$.

Denote a tubular neighborhood of this path in M_{t_1} by $\nu\pi_{t_1}(P)$, and define

$$U_{me} = \left\{(t,m) \in M_{[t_1-\epsilon,t_1+2\epsilon]} : \pi_{t_1}(m) \in \nu\pi_{t_1}(P)\right\}.$$

This is essentially the product of $[t_1-\epsilon, t_1+\epsilon]$ with a neighborhood of $\pi_{t_1}(\gamma)$, slightly extended to contain part of φ_l.

To define U_{cr}, it is necessary to describe the deformation of which it is the support. Begin by traveling downward along φ_k away from the cusp χ_2' in Figure 31. Consider the first fold crossing one encounters between φ_k and some fold arc A. If χ_2' is not on the lower-genus side of A, move on to the next crossing. Otherwise, push the crossing between A and φ_k upward along φ_k past χ_2'. More precisely, there is a deformation as in Figure 11 in which A is the vertical fold arc and the fold arcs φ_1, φ_k are respectively to the left and right of the cusp, in effect pushing the crossing off what will be $\tilde{\varphi}_k$ to lie in φ_1. The initial cusp-fold crossing exists by [W2, Proposition 2.7] and the subsequent R_2 deformation exists by [W2, Proposition 2.5(3)]. The intersection point has now been moved to lie in φ_1, as shown in Figure 31. This procedure can be repeated for all crossings oriented like A: Use [W2, Proposition 2.5(1a)] to perform finger moves across any other arcs that might intersect φ_k, and then use [W2, Proposition 2.9] to perform R_3 deformations, to move each crossing up over the cusp. Now move the remaining crossings (for which χ_2' is on their higher-genus sides) downward away from χ_2', just past $\alpha(p_{fm}(1))$, then perform finger moves of the offending fold arcs downward along the image of $\alpha(p_{fm})$ to cross $\alpha(p_{fm}(0))$. An illustration of what would result from pushing the one of each of the two types of intersections out of the way appears in Figure 31. The support of the finger move pushing a fold arc down to intersect φ_l is a neighborhood of a pair of disks which can be chosen disjoint from a neighborhood of p_{fm}, and the disjointness condition of [W2, Definition 2.1] and [W2, Theorem 2.4] implies that the support of the rest of this deformation is disjoint from the part of φ_k whose image runs parallel to $\alpha(p_{cm})$. For this reason, it is possible to arrange for Condition (ii) in the definition of local shift deformation by first pushing off the intersections, then performing the pair of merge moves, and then reversing the deformation that pushed off the intersections. Once this is achieved, Condition (i) is verified by simply examining the base diagrams in the

image of U_{me} just after pushing off the intersections. In this way, the initial shift candidate is broken into a sequence of local shift deformations.

The last step, Item (4), takes as input the fibration that results from the sequence of local shift deformations coming from Item (3) and sends the pair of fold arcs that trailed behind χ_1 (which may have many loops as in Figures 30c and 30d) back to coincide with the fold arcs that trailed behind χ_1 at the end of Item (1) (which follow a crossing-free path from where the initial fold merge occurred to where the final cusp merge occurred). Once Item (4) is applied, the fibration structure just after the final cusp merge agrees with the fibration structure at the analogous t value before the algorithm began (that is, immediately after the shift candidate). The algorithm is defined to begin with the same fibration as the beginning of the given shift candidate, so the result can be substituted into α to satisfy the requirements of Lemma 3.7.

Consider $\tilde{\varphi}_k$, before the flips are introduced by Item (2), as a path oriented from χ_2 to χ_2'. The two fold points that map to the first self-intersection of $\tilde{\varphi}_k$ bound a fold arc that maps to an oriented based loop; call it $\ell\colon [0,1] \to S^2$. Within that loop, the first self-intersection to occur in $\ell|_{(0,1)}$ singles out a smaller loop within ℓ, and so on, until one reaches an innermost loop. Item (2) puts a flip loop nearby, for which the picture is just like Figures 30a and 30b, though the fold arc is not interrupted by a star ellipsis.

Considered only as paths in S^2, there is the obvious Reidemeister-II homotopy canceling crossings in an innermost loop in a configuration like Figure 30c and there is the homotopy depicted in Figure 32 for an innermost loop like in Figure 30d). Thus, if P is the path in S^2 taken by the image of χ_1 and Q is the straight path of regular values from χ_1 to χ_2' coming from Item (1), the flips suggested by Figure 30 double the number of crossings in $\tilde{\varphi}_k$ in such a way P is homotopic to Q by a smooth homotopy f_v that consists only of Reidemeister-II and Reidemeister-III moves. The immediate goal is to show that this homotopy leads to a corresponding sequence of Reidemeister-II and Reidemeister-III fold crossings sending the pair of fold arcs that are parallel to P to be parallel to Q. Instead of relying on the disjointness condition of [W2, Definition 2.1], which is used to construct a ball of regular points on which a proposed deformation is supported, it is possible to construct that support directly. Let t_e be a t-value just after the final cusp merge between χ_1 and χ_2'. The homotopy f_v from P to Q is a map $D^2 \to S^2$ which has a lift $\tilde{f}\colon D^2 \to M_{t_e}$, that is, $\alpha_{t_e} \circ \tilde{f} = f$, which is generic in the sense that $\tilde{f}$ is injective except for a finite number of transverse self-crossings and $\tilde{f} \cap \operatorname{crit} \alpha = \varnothing$.

Now the path taken by χ_1 in M could have been chosen to be the corresponding boundary arc of $\tilde{f}(\partial D^2)$ (recall there was a detour taken by χ_1 toward each cusp introduced by the flips, but these to not show up in the fibration α_{t_e} and may be ignored). Then the required deformation of α_{t_e} is supported on a neighborhood of $\tilde{f}$. The self-intersections in $\tilde{f}$ are no problem because generically the preimage of an intersection in D^2 occupies different v-level sets of the homotopy f_v.

Once these movements are applied, the deformation consists of a sequence of local generalized shift deformations instead of the single shift candidate given in Item (1), followed by a sequence of 2-parameter crossings. Because the cusp and fold merges that occurred between the first and last merge points were constructed as detours, the result is a fibration that agrees with the ending of the candidate yielded by Item (1), except that even though the pair of fold arcs that trailed behind

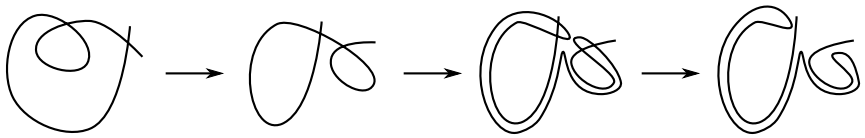


FIGURE 32. A schematic for resolving intersections coming from Figure 30b. Each arc represents the image of two parallel fold arcs that trailed behind χ_1. This movement occurs in the interval between one candidate pair and the next.

χ_1 parallel to Q have the same image as the original pair resulting from Item (1), they may not be isotopic to the original pair in M (relative to their endpoints). This is easily rectified by adding one more local multislide deformation whose initial fold merge is between these fold arcs. □

3.5. **Modification of the stratified immersion of the critical manifold.** At this point, α has some of the characteristics of the required deformation: The Morse function $T\colon \operatorname{crit}\alpha \to I$ and the part of the cusp locus containing its critical points are as required; however, the base diagrams are still disorganized because of self-intersections in the critical image. The object of this section is to modify the stratified immersion $\operatorname{crit}\alpha \to S^2_I$ until the remaining double points all come from the model deformations for the moves in Theorem 1.1.

3.5.1. *Removing ι from the merge pairs.* Here begins the process to arrange ι into a form taken by the model deformations. The first step is to reduce to considering a single immersed critical cylinder, rather than a pair of cylinders or a higher-genus surface, by causing ι to inhabit a union of t-intervals which is disjoint from that inhabited by the merge pairs. Recall from Sections 2.3 and 2.4 that a merge pair whose immersion locus is empty is either a shift or multislide pair.

Lemma 3.8. There is a homotopy of α that causes all merge pairs to occur within multislide or shift deformations.

Proof. A merge pair fails to belong to a shift or multislide deformation because the critical image has crossings when the perge pair appears. The first goal is to give a method to remove these crossings by a deformation whose starting point is just before the merge pair. The second goal is to prove that the support of this deformation can be made disjoint from the support of the merge pair, thereby allowing it to be inserted as the first half of a detour, with the merge pair occurring during a short pause in the detour at its halfway point, so that the merge pair belongs to a shift or multislide deformation. Since there is nothing to do if there are no crossings in the critical image just before the merge pair, assume there is at least one crossing.

This paragraph and the next are devoted to finding a suitable loop or bigon in the critical image for crossing cancellation. To begin, the endpoints of the fold merge path are also the endpoints of a pair of critical arcs whose union is $\operatorname{crit}\alpha$, and if α fails to be injective on one of these (call it A), let x_1 denote a crossing in $\alpha(A)$. The case in which α is injective on both critical arcs is dealt with in the next paragraph. Let $a(x_1)$ denote the loop in $\alpha(A)$ based at x_1. If $a(x_1)$ has self-crossings, let x_2 be one of those and let $a(x_2)$ be the loop contained in $a(x_1)$

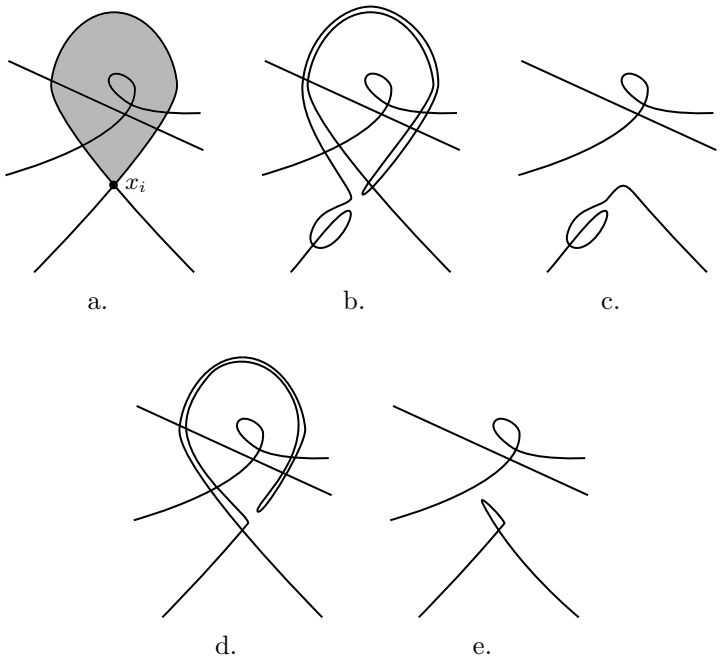


FIGURE 33. The cancellation movement for the case in which the shaded region D lies on the higher-genus side of $a(x_i)$ (a–b–c), or on the lower-genus side of $a(x_i)$ (a–d–e). Two other fold arcs are included to show how the movement doubles and then cancels the crossings in $a(x_i)$, shortening the resulting finger in a counter-clockwise movement. The lower crossings in Figures 33c and 33e are removed as in Figure 34.

FIGURE 34. Eliminating a genus-decreasing loop in the critical image by performing a flipping move followed by an R_2 deformation which exists by [W2, Proposition 2.5(3)]. The arrow indicates decreasing fiber genus.

based at x_2. Repeating this process, there is an i such that $a(x_i)$ is free of self-crossings, because $\alpha(A)$ had finitely many self-crossings. For later reference, let D be the component of $S^2 \smallsetminus a(x_i)$ which is shaded in Figure 33a. This figure gives an overview of how the cancellation proceeds for $a(x_i)$, which will be explained below.

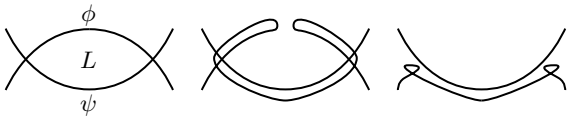
If α is injective on each of the two critical arcs F, G connecting the two ends of the fold merge path in M, then the critical image contains a bigon with one side in $\alpha(F)$ and the other side in $\alpha(G)$, found as follows. Begin by following $\alpha(F)$ starting at either of its endpoints to the second crossing encountered, and

observe both crossings must be crossings with $\alpha(G)$. The path in $\alpha(F)$ from the first crossing to the second crossing is one side of the bigon B_1. At the second crossing, there is one direction along $\alpha(G)$ that reaches the first crossing without passing an endpoint of the merge path. This path in $\alpha(G)$ is the other side of B_1 (the crossings have opposite orientations as required for a bigon because the first side was constructed to be free of any other crossings with $\alpha(G)$). Let L_1 denote the open lune in S^2 bounded by B_1. Since the endpoints of F and G lie outside L_1, every fold arc within L_1 is contained in $\alpha(F)$ or contained in $\alpha(G)$, and so is free of self-crossings and can only intersect one side of B_1. Thus, if L_1 contains critical values, then there is an open lune $L_2 \subsetneq L_1$ bounded by a bigon B_2, again with one side in $\alpha(F)$ and the other side in $\alpha(G)$. Proceeding in this way, there is a j such that the open lune L_j contains no critical values.

By the construction of $a(x_i)$ and B_j, each has a neighborhood which is disjoint from the image of the support of the merge pair (recall this is either the shaded region in Figure 14 or the entire region depicted in Figure 15). The crossing cancellation will be a genus-increasing movement of $a(x_i)$ or one side of B_j, and the deformation corresponding to this movement will be shown to have support which is disjoint from that of the merge pair.

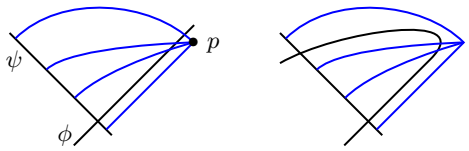
The rest of the argument makes liberal use of [W2, Proposition 2.11], in which the goal is to push the image of an arbitrarily short open fold arc ϕ of a Morse 2-function $f = \alpha_t$ for some t over a lune L in the base. This paragraph gives an additional argument for that proposition; such a digression seems merited because of the importance of the result for this paper and historical uncertainty about the general existence of such movements. The proof of the lemma itself continues in the next paragraph. The first step is to construct a section σ of f over L such that $\sigma \cap \operatorname{crit} f = \phi$. Choose a finite open cover of L by disks which are small enough so that their pairwise intersections contain no more than one fold arc without crossings, or exactly two embedded fold arcs which meet at one crossing or one cusp point (this is to guarantee that the union of all vanishing cycles of critical arcs in the overlap, depicted in any fiber in the overlap, are nonseparating). Since ϕ and the disks are both arbitrarily small, this cover may be chosen so that $\alpha(\phi)$, and no other critical value, is contained in exactly one of these disks. Over this disk, define σ as one of the two fiberwise points forming the vanishing set of ϕ for some choice of connection over L. The rest of σ is defined inductively: Assuming σ has been defined over a disk $D \subset L$ given by a union of elements of the cover, take an arbitrary section s of regular points over an adjacent disk, which exists as long as the disks in the cover are small enough. Since the complement of the vanishing sets in each fiber of all folds in the overlap is connected, there is a perturbation of s causing it to coincide with the part of σ already defined over the overlap. After finitely many such extensions, σ is defined over L as required. Let U denote a tubular neighborhood of σ whose intersection with $\operatorname{crit} f$ is ϕ. Identify U with the normal bundle $N(\sigma)$ in M, and choose a section σ' of $N(\sigma)$ whose vanishing set is ϕ. Then there is a connection for $f|_U$ such that the vanishing set of ϕ is $\sigma \cup \sigma'$. Any extension of this connection to the preimage of L will then exhibit the required disjointness condition for the proposed movement of $f(\phi)$.

In order to describe the cancellation of crossings for B_j and for $a(x_i)$, it is necessary to explain how to cancel the crossings at the corners of a bigon whose central open lune L has no critical values. When L is on the lower-genus side of



a. b. c.

FIGURE 35. Cancellation when the lune L is on the higher-genus side of the critical arc ϕ, which moves across the back of S^2 to obtain Figure 35b. The two small loops are contracted off the other side ψ to obtain 35c, then their crossings are resolved as in Figure 34.



a. b.

FIGURE 36. In (a), a depiction of the left crossing in Figure 35a and in (b) the narrow bigon on the top-left of Figure 35b, with a family of reference paths starting at p.

both of its sides, the cancellation exists by [W2, Proposition 2.5(3)]. Otherwise, the lune lies on the higher-genus side of at least one of its sides, such as the arc ϕ in Figure 35a.

Figure 35b comes from Figure 35a by pushing a small arc of fold points in ϕ around the back of S^2 so that it closely follows ψ, and the corresponding deformation exists by [W2, Proposition 2.11]. Note ϕ and ψ may have cusps, but the movement which results in Figure 35b is chosen so the cusps of ψ are excluded from the arcs that cross the two fingers in ϕ. One of these arcs is also shown in Figure 36b as the lower-left side of the bigon. Figure 35c comes from shrinking the two loops in ϕ, and if L lies on the lower-genus side of ψ then this is an application of [W2, Proposition 2.5(3)]. Otherwise, the movement that produced Figure 35b must be chosen with care to satisfy the matching requirement of [W2, Proposition 2.5(2)], which is that if one were to measure the vanishing cycle of ψ using the top and bottom reference paths in Figure 36b, then the two measurements must yield isotopic vanishing cycles in the fiber over p. In that case, the desired R_2 deformation exists. The following paragraph explains why this is possible.

Figure 36a shows a neighborhood of the left crossing in Figure 35a with a family $r(s)$ of reference paths from a point p to the image of the fold arc ψ. Having chosen a connection on the fibration, for each s there is a circle in the reference fiber that flows to ψ, and for those reference paths that cross the image of ϕ, there is also a pair of points that flow to ϕ. The connection may be chosen so that, within the preimage of this neighborhood, these circles and points are constant with respect to s. A movement of ϕ corresponds to a choice of connection according to [W2,

Remark 2.2], and the movement of ϕ corresponding to such a choice of connection, to obtain Figure 35b, produces a bigon in Figure 36b which satisfies the matching condition. This paragraph also holds for the crossing at the right of Figure 35a, with Figures 36a and 36b reflected horizontally. For this reason, the loops in Figure 35c may be contracted, regardless of the orientation of ψ. Note that this argument also completes the proof that the crossings at the corners of the bigon B_j above may be canceled.

It remains to verify the movements of Figure 33. The deformations corresponding to the movements to obtain Figures 33b and 33d starting with Figure 33a exist by [W2, Proposition 2.11]. The subsequent movements, shortening the resulting fingers counter-clockwise, can be understood as a sequence chosen from two movements: either an R_2 deformation using [W2, Proposition 2.5(3)] or the movement in Figure 35, depending on the orientation of each fold arc crossing $a(x_i)$ in Figure 33a. The lower crossing in Figures 33c and 33e may then be removed using the modification of Figure 34.

Using either a lune L_j or a loop $a(x_i)$ as above, one may decrease the number of crossings in the critical image, and since there are finitely many crossings, this may be repeated until there are no crossings. Such a sequence of movements, followed by its reverse, is a detour one may insert just before the merge pair, and the insertion of a detour can be achieved by a homotopy of α.

The goal is now to show that the second half of the above detour may be delayed until after the merge pair has occurred. Every movement in the first half of the detour is realized as a movement of the image of a critical arc ϕ or ϕ_0 in a genus-increasing direction over a lune Δ in S^2, from one side to the other side (in this paragraph, call this arc ϕ). This movement is parameterized in the sense that Δ is foliated by a family of paths connecting its two sides, and each point of $\alpha(\phi)$ moves along the leaf containing it. Having chosen a connection $\mathcal{H}$ on M, the support of such a movement is a tubular neighborhood of the pair of points in each fiber over Δ that flows to ϕ along the leaf to which that fiber is mapped. These points coincide along ϕ and along leaves that contain a cusp of ϕ. For this reason, the support of the fold crossing deformation is an immersed 2-disk in M. The support of the merge pair is a tubular neighborhood of a circle in M. In the case of a multislide, this circle is traced by the two cusps that momentarily appear as the endpoints of p_{cm} (recall Figure 14) and for a shift it is the union of $\tilde{\phi}_k$ and the path traced by the cusp χ_1 in Figure 15. In either case, the circle associated to the merge pair and the immersed disk associated to the fold crossing deformation, and thus its reverse, are generically disjoint in M (that is, such disjointness is realized by a slight perturbation of $\mathcal{H}$). For this reason, the second half of the detour can be moved forward in t by a homotopy of α to occur after the merge pair. □

3.5.2. *Immersion loci of deformations.* It will be necessary to collect some tools and terminology before the next step in the modification, which is to remove all crossings in ι. To avoid clutter in the exposition, it will be convenient to include swallowtail points in the immersion locus throughout the rest of the paper. The author would like to thank the anonymous referee for suggesting case (3) of the following lemma, and its proof.

Lemma 3.9. Suppose α is a deformation whose critical locus is connected at each value of t, and let $c \subset \operatorname{crit} \alpha$ be an immersion circle. Then:

(1) If c is one of a pair of circles which are mapped to each other by α, then c contains no swallowtail points.
(2) If c has a swallowtail point, then it has exactly two such points, and α identifies the two arcs between them.
(3) If c has no swallowtail points, then c is one of a pair as in (1).

Proof. For (1), if c contained a swallowtail point, one could choose a point on c and mark its counterpart on the other circle. One could then follow the one-parameter family of double points along the immersion arcs until one of them enters a neighborhood of the swallowtail point as in Figure 16a, and here we come across a one-parameter family of triple points (or higher), the third arc coming from the other side of the immersion arc passing through the swallowtail point, which cannot be simplified by a small perturbation. This is a contradiction because in a deformation the only triple points are those coming from Reidemeister type three moves, which result in isolated triple points. Thus, c is free of swallowtail points.

For (2), because triple points are isolated, the maximum number of immersion circles in $\operatorname{crit}\alpha$ with common image is two. For this reason, if c is not one of a pair of circles, then $\alpha|_c$ must be two-to-one outside a finite set of points. Orient c and trace a point p and its counterpart p' on c in opposite directions. According to Figure 16b, when p passes the point at which a Reidemeister-II fold crossing occurs, so does p', in a way that preserves the condition that the two points are counterparts. For this reason, one may trace along c until p meets p'. An immersion point can only meet its counterpart at a swallowtail, according to the local models in Figure 16, so this completes the proof.

For (3), if there is no swallowtail point, then $\alpha|_c$ is locally two-to-one and has no branch points (since the only source of branch points in $\alpha|_\iota$ is swallowtails), so we must rule out the case where $\alpha|_c$ is a double cover followed by an immersion into S^2_I. Since $\operatorname{crit}\alpha$ is orientable, c has a nowhere-zero normal vector field v in $\operatorname{crit}\alpha$. Since c has no swallowtails, α is an immersion near c, after smoothing the critical image near cusp arcs, so v has nonzero image in S^2_I. This is locally a pair of linearly independent normal vector fields along $\alpha(c)$ (tangent to the two sheets of the critical image). Globally, however, the vectors must interchange when we go once around $\alpha(c)$ (since they both come from v upstairs). This means the normal bundle of $\alpha(c)$ is nonorientable, contradicting orientability of S^2_I. □

Definition 3.10. Lemma 3.9 suggests the partition of the set circles of ι into two subsets, the set of immersion pairs and the set of immersion singles: An *immersion pair* is a pair of circles which have the same image, and an *immersion single* is a circle of immersion points containing a pair of swallowtail points.

It is time to further classify immersion circles and the critical points in their complement.

Definition 3.11. Each point $p \in \operatorname{crit}\alpha \smallsetminus \iota$ has an associated genus $g(p)$ as follows. Choose a regular fiber F_p just to the higher genus side of $\alpha(p)$. Define *the genus at* p as the genus of F_p.

Definition 3.12. Fix a short path in $\operatorname{crit}\alpha$, parameterized by t, that transversely crosses an immersion circle at the immersion circle's minimal t-value. Where the path crosses the immersion circle, the genus of the regular fiber increases or decreases by 1.

- If the genus decreases, call the circle *genus-decreasing*. If it increases, call it *genus-increasing*.
- Immersion pairs composed of genus-increasing (resp. decreasing) circles are called *genus-increasing pairs* (resp. *genus-decreasing pairs*).
- If the genus increases at one member of an immersion pair but decreases at the other, call the pair *mixed*.
- Disregarding the global aspects of immersion circles, there are notions of genus-increasing, genus-decreasing, and mixed R_2 deformations in which two fold crossings form with increasing t. Similarly, there are mixed and genus-increasing cusp-fold crossings. This explains the modifiers *genus-increasing, genus-decreasing*, or *mixed* for an R_2 deformation that forms a pair of fold crossings.

Remark 3.13. Some examples of this terminology: The Reidemeister-II fold crossing that initiates a handleslide deformation is genus-decreasing, and the disk bounded by the loop in the bottom right of Figure 17 lies on the higher-genus side of the immersion arc that forms its boundary. In other words, the loop appears by a genus-increasing R_2 deformation.

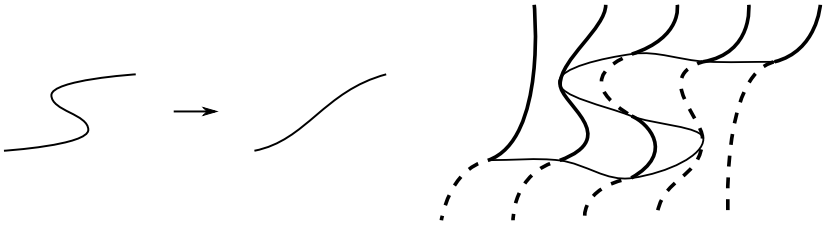
On a related note, it will be useful to *trace orientations* near any crossing triple in ι as follows. Suppose in Figure 16c that a is oriented so that the genus is decreasing as one crosses a in the increasing t direction. This actually comes from the orientation of the strip of fold points passing through the page along a, so that the genus decreases as one travels along b across a with increasing t. The orientation of this strip also causes the genus to decrease traveling along b', so that c' gets an orientation: The genus decreases traveling across c' with increasing t.

3.5.3. *Tools for eliminating crossings in ι.* A basic tool is to push the critical image around in S^2_I according to an ambient isotopy; an example appears in Figure 37. The following Lemma states conditions under which such movements preserve the condition that α is a deformation.

Lemma 3.14 (Ambient isotopy lemma)**.** Suppose i^s is an ambient isotopy of S^2_I with $s \in [0,1]$ which does not introduce or eliminate tangencies between the critical image C of α and the level sets S^2_t. Then there is a one-parameter family of homotopies α^s such that $\alpha^0 = \alpha$, $\alpha^s(\operatorname{crit}\alpha^s) = i^s(C)$ for all $s \in [0,1]$, and α^1 is a deformation. In particular, the critical image of α^1 is $i^1(C)$.

Proof. The claim is that $\alpha^s = i^s \circ \alpha$, with time parameter coming from composing with the projection $S^2_I \to I$, satisfies the conclusion of the lemma. Since i^0 is the identity, $\alpha = \alpha^0$. The requirement that i^s does not introduce or eliminate tangencies between the image of the fold or cusp locus and the level sets S^2_t guarantees that for each fixed t, α^s is a deformation consisting of flipping moves (for when i^s pushes a swallowtail across a t-level set) and 2-parameter crossings (for when i^s pushes an intersection in the critical image across a t-level set), both of which are known to preserve the total space of the fibration. Thus, for each fixed s, α^s is a one-parameter family of endpoints of deformations of maps $M \to S^2$, so it is itself a 1-parameter family of maps $M \to S^2$, that is, a homotopy. □

Here follows a particular application of Lemma 3.14. It appears explicitly because of its fundamental nature, and to illustrate the seemingly nontrivial modifications to sequences of base diagrams such an isotopy can produce.



a. A basic simplifying move in ι. The arcs' counterparts have the same appearance.

b. The critical image in S^2_I before simplifying.

FIGURE 37. Depictions of a certain behavior of ι for Corollary 3.15. In both images, the page is part of $\operatorname{crit}\alpha$ or its image.

Corollary 3.15. Pairs of critical points of $T|_\iota$ as in the left side of Figure 37a (in which no other immersion arcs or critical points are allowed to be present in the pictured region of $\operatorname{crit}\alpha$) can be eliminated by a homotopy of α to obtain the right side of the figure. The same modification is available to the vertical and horizontal reflections of the figure.

Proof. Up to reflections of Figure 37a, the deformation corresponding to Figure 37a has base diagrams given by Figure 11, with the cusp replaced by a fold point. The required homotopy of α is realized by an ambient isotopy in S^2_I that gives a one-parameter family of homotopies (by Lemma 3.14) in which the t-value of the first R_2 deformation (that is, the t-value at which the images of the fold arcs become tangent) approaches that of the second R_2 deformation; see Figure 33b for a depiction of the critical image of the deformation. In that figure, one of the fold arcs sweeps out the plane given by the page, while the other fold arc sweeps out a surface with stripes given by five T-level sets in $\operatorname{crit}\alpha$. The stripes become dotted lines where they dive below the page. The remaining S-shaped curve in Figure 33b is where they intersect. □

Here is a similar application that finds immediate use; its proof is essentially the same as that of Corollary 3.15. Figure 38 depicts the two ways in which a cusp arc may cross an immersion arc. In that figure, either all arrows point into the lower-genus side of each fold or immersion arc, or they all point into the higher-genus side of each fold or immersion arc. Note that Figure 38a consists of mixed R_2 deformations and Figure 38b consists of either genus-increasing R_2 deformations or genus-decreasing R_2 deformations (according to the interpretation of the arrows).

Corollary 3.16. By Proposition 3.14, one may convert between Figures 38a and 38b by a homotopy of α.

Before adding to the list of moves on deformations, note there is generally not a Reidemeister-II move between a given cusp arc and immersion arc. For example, the single coming from a stabilization deformation has a particular form laid out in Section 2.2.2, and canceling the fold crossings coming from two flips using some other R_2 deformation does not necessarily yield a stabilization move, as detailed in

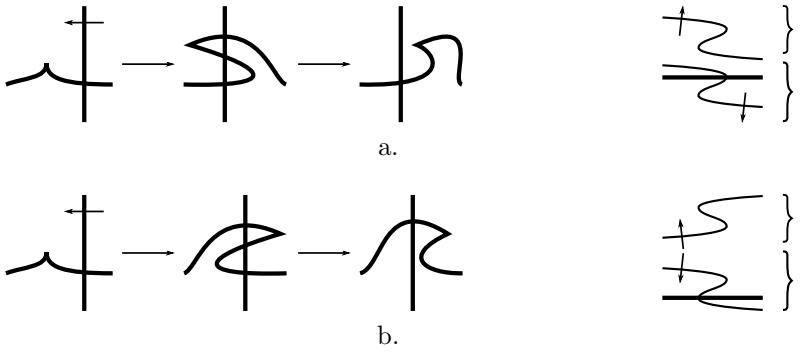


FIGURE 38. Base diagrams for the two versions of Figure 11, with corresponding decorated critical sets to the right. As is common with these pictures, each bracketed piece can be flipped vertically (also flipping its arrow), and each pair of bracketed pieces can switch places, and still represent the same deformation. The arrows give the correspondence between orientations in the pictures, and may be all marked "1" or all marked "2" as in [GK1].

the proof of Lemma 3.26: It could yield a stabilization occurring within a sequence of handleslides. Nevertheless, immersion arcs enjoy some freedom of movement.

Lemma 3.17. Immersion arcs can move around cusp arcs in at least three ways:

(1a) Suppose there is a cusp arc lying on the lower-genus side of an immersion arc, with no other decorations between them. Then one may perform a finger move between the two arcs, introducing two intersection points between them.

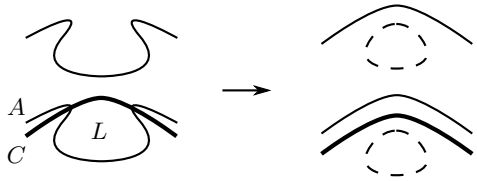
(1b) Suppose there is a lune L in $\operatorname{crit}\alpha$ such that:
- One side of L is a cusp arc.
- The other side is an immersion arc.
- The interior of L has no cusp arcs or immersion arcs.

Then, possibly after an application of Corollary 3.16, there is a homotopy of α whose endpoints appear in Figure 39a.

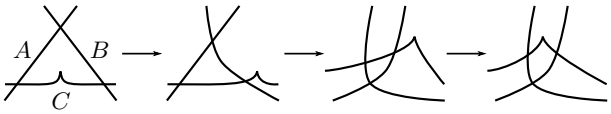
(2) Suppose there is a triangle Δ in $\operatorname{crit}\alpha$ such that two sides are immersion arcs A and B, the third side is a cusp arc C, and $\Delta \cap (\iota \cup \chi) = A \cup B \cup C$. Then there is a homotopy of α realizing a Reidemeister-III type move as indicated by Δ.

Proof. For item (1a), the modification is simply a detour in which the cusp moves into the higher-genus side of a nearby crossing as in Figure 38, then returns.

For item (1b), the base diagrams of Figure 38a, followed by the reverse, is the sequence of base diagrams on the left side of Figure 39a, followed by the reverse (except the orientation arrows in that figure should be ignored). The proposed homotopy would cancel the pair of cusp-fold crossings that occur, leaving behind an immersion pair. It is not hard to verify the claim: The pair of cusp-fold crossings is a homotopy supported in a 4-ball neighborhood of the cusp, and the proposed



a. Moving an immersion arc A past a cusp arc C in Lemma 3.17(1b). A similar pinching movement occurs with the counterpart of A, producing a new immersion pair.



b. An example deformation corresponding to Δ in Lemma 3.17(2).

FIGURE 39. Figures for Lemma 3.17.

homotopy of α consists of retracting that neighborhood until its image in S^2 is disjoint from the fold arc it crosses.

For item (2), the deformation corresponding to the triangle $\Delta \subset \operatorname{crit}\alpha$ begins with a triangle in the critical image, consisting of arcs A, B and C, with a cusp lying on side C (see Figure 39b for one such configuration). The cusp crosses side A out of the triangle, then there is a Reidemeister-III fold crossing, and then the cusp crosses fold arc B back into the triangle. The proposed modification would result in the cusp crossing side B first, then side A. The midpoint of the proposed homotopy would have the cusp crossing A and B simultaneously at their own point of intersection. The disjointness of the vanishing set of C from those of A and B that is required for the existence of these two deformations (and similarly for the intervening deformations) follows from the existence of all the intersections in the initial deformation (in verifying this, it helps to reverse t if necessary so that the cusp in the initial base diagram points into the triangle like in Figure 39b). □

According to Lemma 3.8, each critical twice-punctured torus associated to a merge pair is mapped by injectively by α into S^2_I, while the typical intervening critical cylinder between shift pairs has nonempty immersion locus ι. For this reason, α_t is as required in the conclusion of Theorem 1.1 near those values of t for which $\operatorname{crit}\alpha_t$ is disconnected, and it suffices to restrict attention to the intervening critical cylinders, where the goal is to first arrange for ι to be an embedded 1-submanifold of $\operatorname{crit}\alpha$.

In light of Lemma 3.17, which does not create new R_3 deformations and so should simply be understood to be used as needed, the cusp locus does not show up in the arguments until Lemma 3.25. The model deformations from Section 2 do not involve R_3 deformations, and this property is important enough to have a name.

Definition 3.18. An immersion circle is *isolated* if it has no crossings with itself and no crossings with any other immersion circles.

The following lemma is the main tool for removing the R_3 deformations from α.

Lemma 3.19. Let α be a deformation which is free of merge, birth and flipping moves. Suppose L is a lune in $\operatorname{crit}\alpha \smallsetminus \iota$ and Δ is a path component in $\operatorname{crit}\alpha \smallsetminus \iota$ which is a disk with three sides, where L and Δ lie on the lower-genus side of each of their sides, and none of their sides contain a swallowtail point.

(1) (Pinching lemma) As in Figure 40, suppose a, b are points in immersion arcs such that:
- $T(a) = T(b)$
- $\alpha(a) \neq \alpha(b)$
- There is a smooth arc $f\colon [0,1] \to \operatorname{crit}\alpha_{T(a)}$ which connects a to b and is otherwise disjoint from ι,
- The image of f lies on the lower-genus side of both immersion arcs.
- There is another arc f' connecting the counterparts a', b' of a and b with the analogous properties.

Then there is a homotopy of α, which this paper calls a *pinching move*, realizing the modification that appears in that figure.

(2) (Finger move lemma) Suppose an immersion arc c lies parallel and adjacent to another immersion arc d on its lower-genus side, and that d is not the counterpart of c. Then it is possible to perform what this paper calls a *finger move* in ι, pushing a small part of c across d. See Figure 41.

(3) (Bigon move lemma) Suppose L is a lune in $\operatorname{crit}\alpha$, bounded by immersion arcs, lying on the lower-genus side of both of its sides. Then there is a homotopy of α causing L to appear as in Figures 44a or 44b, at which point L can be contracted, yielding Figure 44c, or contracted by three simultaneous Reidemeister-II type moves in ι, respectively. (both are called a *bigon move* in this paper).

(4) (Triangle move lemma) Suppose Δ is a 2-simplex in $\operatorname{crit}\alpha$, bounded by immersion arcs, lying on the lower-genus side of its three sides. Then there is a homotopy of α causing Δ to appear as in Figure 45, and there exists a further homotopy of α as in that figure which this paper calls a *triangle move*; see Figure 45.

(5) (Pass-over move lemma) Suppose there is an isolated immersion circle on the lower-genus side of an immersion arc. Then, by a homotopy of α, that arc may pass over the circle as in Figures 46a-46c, resulting in a new deformation with the same endpoints.

It is straightforward to verify that all movements of critical points in Lemma 3.19 correspond to movements of critical values toward their lower-genus side, which is in other words a genus-increasing movement. For this reason, any application of Lemma 3.19 preserves the conditions of Remark 3.4, that all fibers are connected and have genus at least 2.

Proof. Item (1) is a simple application of [W2, Proposition 2.5(3)]. The curves $\alpha \circ f$ and $\alpha \circ f'$ comprise the relevant bigon in $S^2_{T(a)}$, and the modification is to insert a detour given by that fold crossing, followed by its reverse.

Item (2) follows quickly from considering Figure 41a, which gives base diagrams for the first half of a detour whose initial finger move exists by [W2, Proposition 2.5(1a)], and whose initial R_3 deformation exists by [W2, Proposition 2.9].

Before proving Items (3) and (4), here is an explanation for why it is possible, and therefore sufficient, to arrange for Δ and L to appear as shown in Figures 45

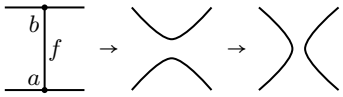
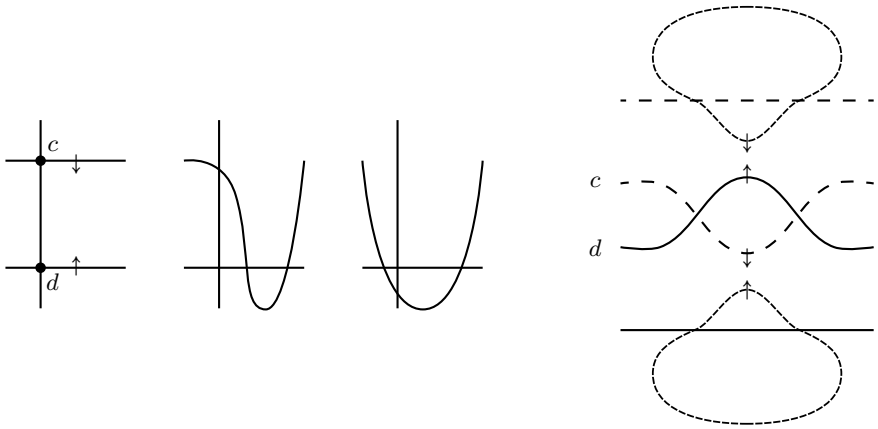


FIGURE 40. Modification of ι in Lemma 3.19(1). Two immersion arcs undergo connect sum according to the reference path f. The same figure with labels a', b', f' depicts what happens to the counterparts of these immersion arcs.



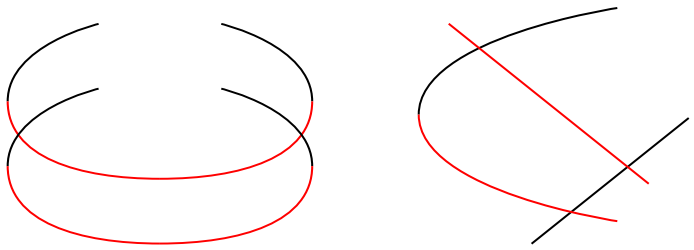
a. Base diagrams for the first half of the detour.

b. A depiction of ι.

FIGURE 41. Performing a finger move in ι for Lemma 3.19(2).

and 44. First, the corners in those figures have pairwise distinct t coordinates, so their corners come from three (respectively two) R_3 deformations. This is easily achieved by subdividing Δ or L as necessary using the Finger Move Lemma, Item (2). Second, no two points in the sides of L or Δ are counterparts (see Figure 42). This is automatically true, because both would require a mixed R_2 deformation in which the T-maximal points of the associated immersion pair (say, in $\operatorname{crit}\alpha_{t_0}$) are connected by an arc in $\operatorname{crit}\alpha_{t_0} \smallsetminus \iota$ (that is, there is no immersion arc "between" the T-minimal points of the pair), which is easily seen to be impossible by considering the base diagrams for such a deformation. Third, the sides of Δ and L are depicted without T-critical points, as may be arranged by applying Corollary 3.15 in the case that one corner of L is at the maximum value of $T|_{\partial L}$ or $T|_{\partial\Delta}$ and another corner is at the minimum. In the situation where a corner is oriented in $\operatorname{crit}\alpha$ in a way that forces T-critical points, one or two applications of Lemma 3.14 as in Figure 43 turns it to allow the sides of L or Δ to have regular T projections. Finally, Lemma 3.17(1b) and (2) allow all cusps be removed from L or Δ.

For Item (3), the lune L and the counterparts of its sides are shown in Figures 44a or 44b, according to whether the base diagrams show a fold arc moving across two different crossings, which would necessarily form the corners of a bigon formed by two fold arcs, or a fold arc moving across a crossing and back. The orientations



a. b.

FIGURE 42. Figures for Lemma 3.19, paragraph above the proof of Item (3), explaining why there is no such thing as a bigon (a) or a lune (b) in which there is a point on one side whose counterpart is on another side.

of the sides of L are as shown in those figures by hypothesis, and the eight other arrows in Figure 44 are deduced from tracing orientations as in Remark 3.13. In the base diagrams for Figure 44a, the bigon of critical values over which the fold arc passes is on the lower-genus side of both of its sides, so it can be removed by a Reidemeister-II fold crossing by a detour as in the proof of Lemma 3.3, where the crossings are canceled before the fold arc begins its movement over the bigon, then re-form after the fold arc has completed its movement. The resulting formation in ι is shown in Figure 44c.

In the base diagrams for Figure 44b, the central triangle initially lies on the higher-genus side of all three fold arcs. Choosing a reference fiber outside of the central triangle, but adjacent to any one of its sides, it is clear that the isotopy classes of the vanishing cycles for the two other sides are unchanged by this deformation, as is the isotopy class for the pair of points corresponding to the remaining side, because the fibers are connected. For this reason, the deformation can indeed be interpreted as a detour, whose removal by homotopy of α corresponds to the three Reidemeister-II type moves in ι suggested by the three bigons.

For Item (4), first recall that Δ can be assumed to appear as depicted in Figure 45 by the paragraph preceding the proof of Item (3). Initially, α is the sequence of Reidemeister-III fold crossings $a_1, a_3, 2_3$ in Figure 45. The alternative deformation given by the Reidemeister-III fold crossings $b_1, \ldots, b_5$ exist by the following argument. Because of the orientations of the fold arcs, b_1, b_2 and b_3 exist by [W2, Proposition 2.9]. Move b_4 is the same as a_1, and b_5 exists because it is the reverse of b_5^{-1}, which exists by [W2, Proposition 2.9]. Alternatively, b_5 can be interpreted as the inverse of b_1. The fact that these moves exist is enough to demonstrate the existence of the movement of $\operatorname{crit}\alpha$ in S_I^2 indicated by the top of the figure, in which a strip of critical values bulges to the left across $\alpha(\Delta)$. In turn, this shows the move b_i can be performed in such a way that the two endpoints of the original and alternative deformations coincide as maps to S^2.

For Item (5), there are two ways to interpret the proposed movement in ι. One way is to start with Figure 46a, use Lemma 3.14 to obtain Figure 46b, then use a pinching move to obtain Figure 46c. Another way, using the base diagrams, is to observe that the map in Figure 46g can be obtained from the map in Figure 46d

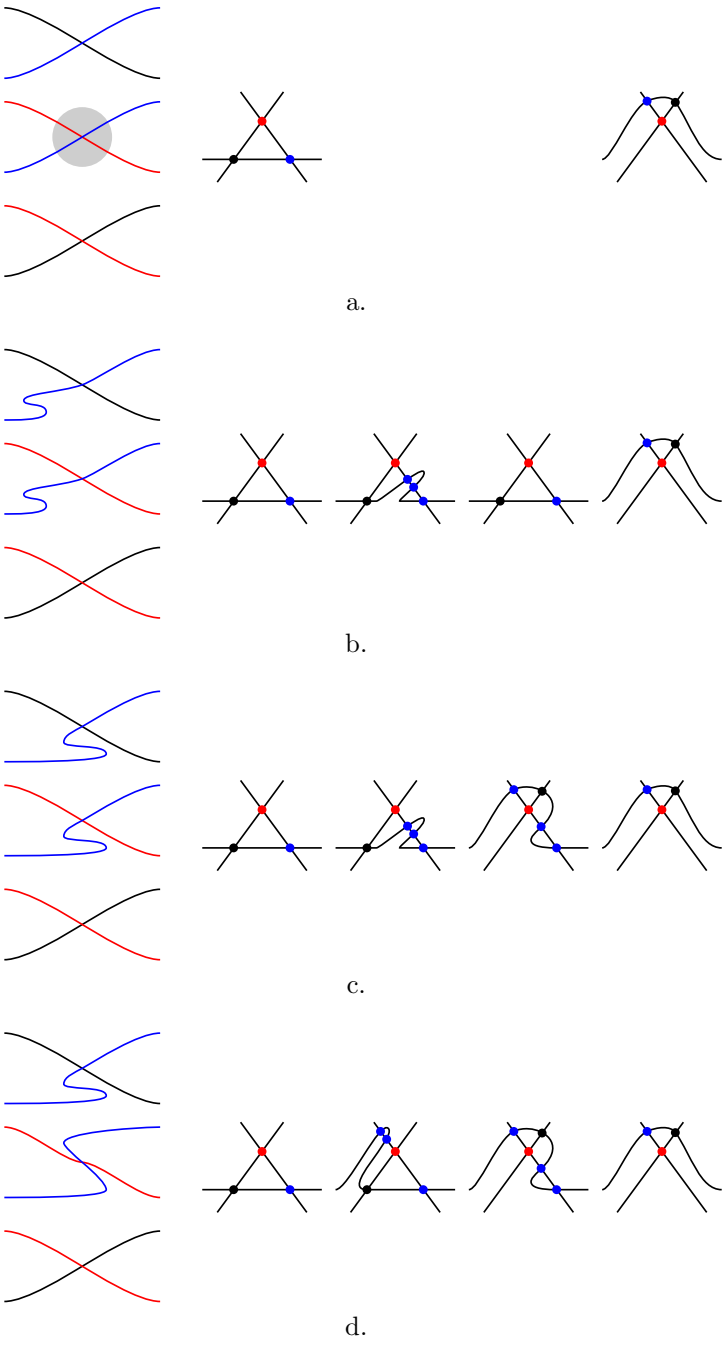


FIGURE 43. Rotating a corner of L or Δ to appear as in Figure 44 or 45 as part of a bigon move (3) or triangle move (4). This corner may initially occupy any one of the four shaded quadrants in 43a. This is a family of deformations, changing near the crossing in ι, whose immersion locus is at left, with corresponding base diagrams at right.

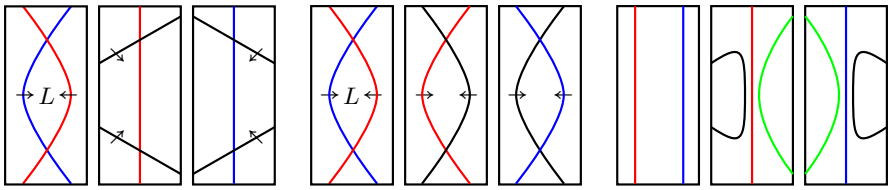


FIGURE 44. Figures for the Bigon Lemma 3.19(3). All arrows in this figure indicate decreasing fiber genus. The first two figures are the two possibilities for the immersion arcs not forming sides of L and the last figure is the configuration after shrinking L in the first figure.

by either pushing a bit of the vertical fold arc either upward or downward past the upper horizontal fold arc. Since the endpoints of these deformations are the same map, one might as well insert the latter deformation as a detour whose halfway point has a pause during which the isolated single occurs. □

3.5.4. *Elimination of crossings in ι.* The argument for the main result of this section, Lemma 3.23, comes in two steps, beginning with any deformation α whose endpoints are crown maps and whose critical set is connected for all t. It suffices to limit attention to this kind of deformation because of Lemma 3.8, which implies $T(\iota)$ is contained in a disjoint union of t-intervals, on each of which α restricts to such a deformation. For the rest of Section 3.5.4, α is assumed to be such a deformation.

The first step is to make α *ready* by a process which may increase the number of crossings in ι. The second step is to decrease the number of crossings in a way that preserves readiness.

Definition 3.20. Let $\iota' \subset \iota$ be the collection of non-isolated immersion circles, let R be the collection of path components of $\operatorname{crit}\alpha \smallsetminus \iota$ other than those containing $\operatorname{crit}\alpha_0$ and $\operatorname{crit}\alpha_1$, let $\{r_1, \dots, r_n\} \subset R$ and let $r = \cup_i \overline{r}_i$. Call r a *clearable region* if $\iota' \cap \partial r \neq \varnothing$ and r lies on the lower-genus side of each of its boundary arcs. The deformation α is *ready* if α has no clearable regions.

Clearable regions are going to be the ones to which Lemma 3.19 is applied to make α ready. To help with understanding the definition, observe that a clearable region may contain isolated immersion circles in its interior, and isolated genus-decreasing pairs are allowed in a deformation which is ready.

The Pass-over Move Lemma 3.19(5) is the main reason isolated pairs can be ignored in the course of verifying that a deformation is ready or causing a deformation to be ready: Essentially, immersion arcs can pass over isolated circles in a genus-increasing direction without introducing clearable regions.

Lemma 3.21. Suppose α is a deformation with no merge moves, no birth moves and no swallowtails. Then α is homotopic to a deformation which is ready.

Proof. The first step is to show that any clearable region contains a *local minimum of g*, that is, a path component of $\operatorname{crit}\alpha \smallsetminus \iota'$ which lies on the lower-genus side of

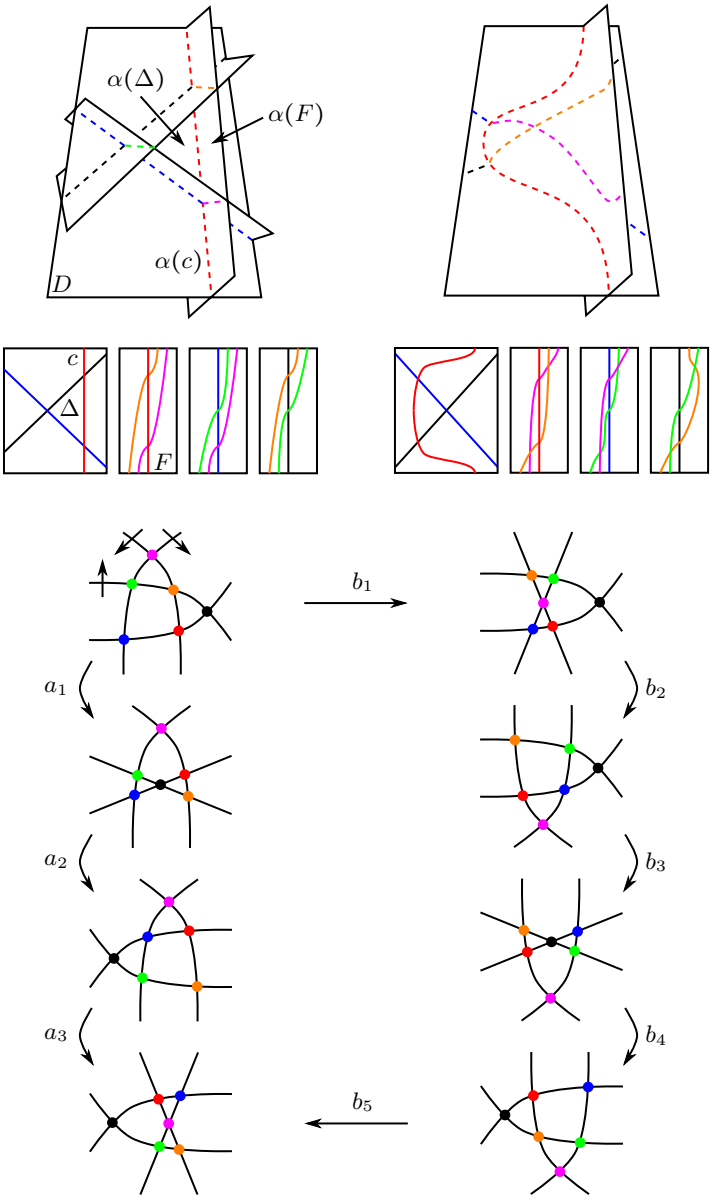


FIGURE 45. Figure for the Triangle Move Lemma 3.19(4). Top: critical image. Middle: crit α decorated with ι. Bottom: the corresponding base diagrams (a_i for top left, b_i for top right). Unlike other depictions of ι, t ranges vertically here. The strip $\alpha(F)$ is transverse to $\alpha(\Delta)$. At top right, $\alpha(F)$ bulges to the left, having crossed $\alpha(\Delta)$. Parts of the critical image are omitted from the top to make the image more understandable. Finally, the new crossings on the front and back side of the bubble are connected by the green immersion arc, which is drawn slightly tilted toward the viewer for clarity. The real deformation would have it tilted away, so that the new crossings on front and back are the first and last to occur, respectively.

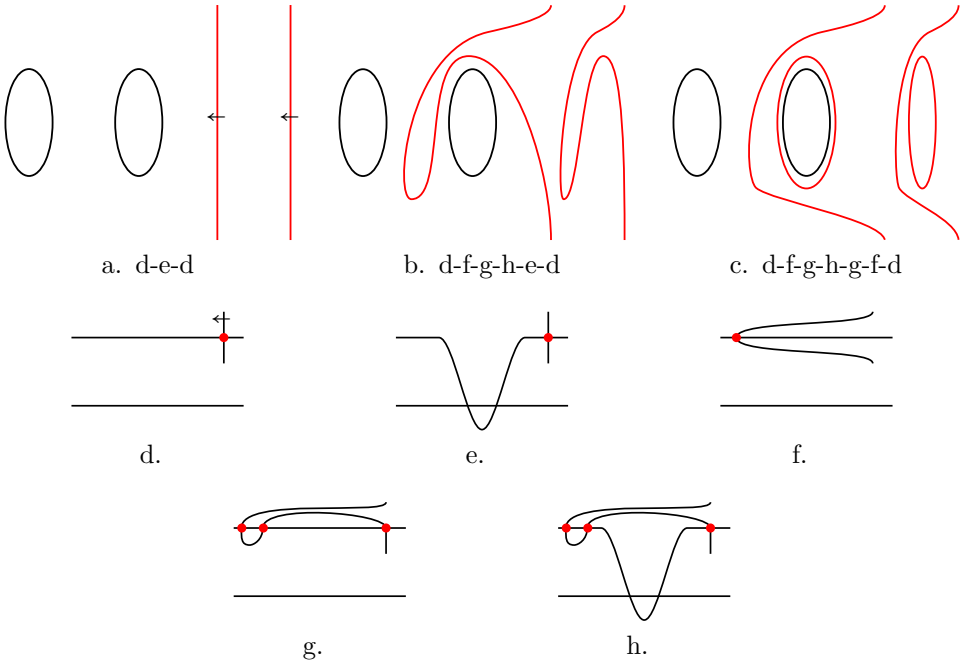


FIGURE 46. Figures for the Pass-over Move Lemma 3.19(5). Figures (a-c) have t ranging vertically; they show an immersion arc passing over an isolated immersion circle, creating a genus-decreasing pair in the process. Their captions denote their corresponding sequences of base diagrams, shown in Figures (d-h). The movement h-e in Figure 46b is a cancellation of the left two dotted crossings.

each of its boundary arcs (note this term is somewhat of a misnomer because it uses ι' instead of ι, so a local minimum may contain isolated circles, and thus more than one g-value, and a clearable region is generally a union of local minima). To prove this, start with a clearable region r and consider a sequence of path components (r_i) of $r \smallsetminus \iota'$ such that:

(1) The boundary of r_1 contains a boundary arc of r.
(2) For each i, r_i shares a boundary arc with r_{i-1}, and r_i is on the lower-genus side of that boundary arc.
(3) The sequence (r_i) is maximal in the sense that, if an entry r_i is not a local minimum for g, then r_i is not the last term of the sequence.

Then (r_i) is a finite sequence whose last term is a local minimum for g.

For this reason, if there is a clearable region r, the number of local minima in r may be decreased by the following steps:

(1) Find a local minimum m of g within r
(2) Use the over-pass move, Lemma 3.19(5), to remove the isolated circles from m, which is now a single path component of crit $\alpha \smallsetminus \iota$ which lies on the lower-genus side of each of its sides.

(3) Subdivide m into lunes and triangular regions by finger moves (Lemma 3.19(2)).
(4) Use bigon and triangle moves (Lemma 3.19(3) and (4)) on those regions to eliminate the resulting local minima.

This paragraph is the explanation for why this list of steps actually decreases the number of local minima. It boils down to the fact that the moves in Lemma 3.19 are genus-increasing movements of all of their involved immersion arcs. Step (1) does not modify ι. Step (2) only creates isolated genus-decreasing circles, and such circles are not considered in the definition of *local minimum*. Step (3) breaks m into a collection of local minima of g, but no other local minima of g are introduced by the finger moves because the immersion circles created by a finger move are genus-increasing. The bigon and triangle moves in step (4) similarly do not introduce local minima of g because all immersion arcs in those moves move toward each other's lower-genus sides. For this reason, the number of local minima contained in r may be repeatedly decreased until r is no longer a clearable region, and repeating this process, the number of clearable regions can be decreased, eventually making α ready. □

The following notion will be useful in the proof of Lemma 3.23.

Definition 3.22. A *pseudostabilization* is any deformation $\alpha\colon M_{[t_1,t_2]} \to S^2_{[t_1,t_2]}$ satisfying the following criteria:

- The map α_{t_i} is a crown map for $i = 1, 2$.
- There are no birth or merge points.
- The immersion locus of α consists of an isolated immersion single which is not nullhomotopic in $\operatorname{crit}\alpha$.

A stabilization deformation is a pseudostabilization, the term *pseudostabilization deformation* is redundant, and there is no *pseudostabilization move* on crown diagrams defined in this paper (see Lemma 3.26). Basically, a pseudostabilization is a stabilization deformation whose immersion circle is allowed to have more than four critical points in its restriction of the projection T to the t parameter, and may arbitrarily intersect the cusp locus within the limits imposed by coming from a deformation.

Lemma 3.23. Any deformation resulting from Lemma 3.8 is homotopic fixing endpoints to a deformation which has no R_3 deformations.

Proof. We begin with a way to ensure all singles are isolated. Using Lemma 3.2, if there is a swallowtail corresponding to a flipping move, it may be pushed back to lie in $\operatorname{crit}\alpha_0$. Figure 47a shows a way to then convert it into a stabilization deformation, then a genus-decreasing R_2 deformation, and then a reverse flipping move. The base diagrams are as follows: Immediately after any flipping move applied to a crown map, insert the detour whose first half is the rest of the stabilization. The second half of the detour then only has one reverse flipping move, the one in Figure 47b. Similarly, inverse flipping moves may be postponed until α_1, and then converted to have ι appear as in the mirror image (in t) of Figure 47b. Apply this to every swallowtail in α, and in the next paragraph these swallowtails will be canceled in pairs. Later in the proof, it will be useful to similarly convert any genus-increasing R_2 deformation applied to a crown map; its pair of decorated critical surfaces are in Figure 48. The base diagrams are slightly more complicated, so it helps to refer

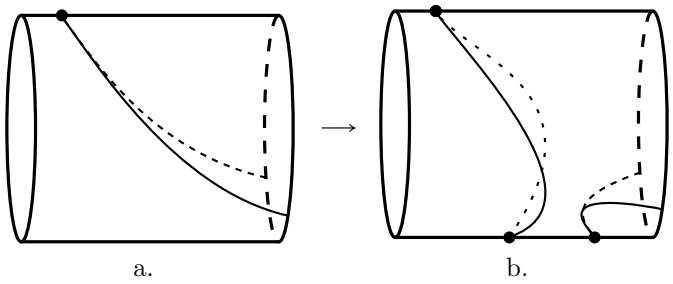


FIGURE 47. Converting a single.

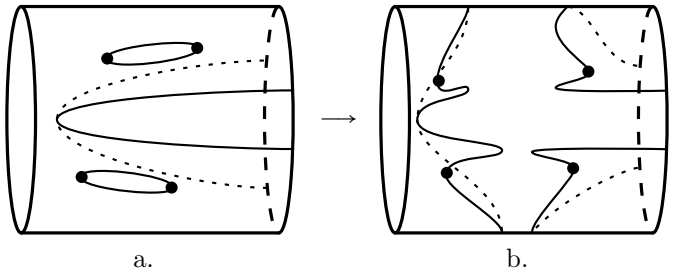


FIGURE 48. Converting a pair.

Decorated critical surfaces which depict the conversion of genus-increasing immersion arcs in the proof of Lemma 3.23. In each case, the beginning of the deformation under consideration is understood to move forward past the pseudostabilization to the "halfway point" where crit α is embedded.

to Figure 49. In that figure, cusps are shown to give the reader some perspective

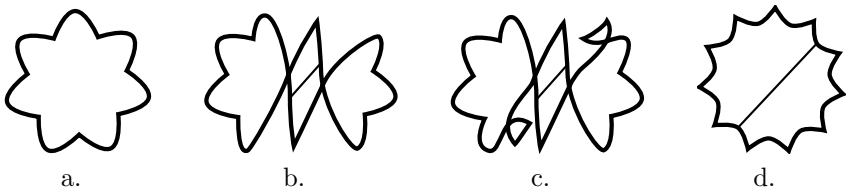


FIGURE 49. Base diagrams for the detour of Figure 48b. The deformation proceeds a-b-c-d-c-b.

on how such a deformation would appear; indeed, these modifications are valid regardless of the placement of cusps, and what matters is the placement of the swallowtails in relation to ι. Figures 49a and 49b depict the initial genus-increasing R_2 deformation (recall that ι_0 is empty). Immediately, as depicted in Figure 49c, two flips occur, followed by a pair of genus-increasing R_2 deformations to obtain

Figure 49d, valid by [W2, Proposition 2.5(3)]. The deformation then reverses itself back to Figure 49b and α resumes as it did before.

Since the flipping move is the only way to change the parity of the number of crossings in the critical image, and that parity is 0 for both α_0 and α_1, the number of flips has the same parity the number of inverse flips (note at this point the flips are clustered near $t = 1$ and the inverse flips near $t = 0$). For some integer $n > 0$, suppose there are $2n$ more flips than reverse flips (reverse the t parameter for this paragraph, if necessary, to make $n \geq 0$). Let S_n denote a detour consisting of n stabilizations followed by the reverse of those n stabilizations, and insert S_n at $t = 0$. Since the first half of S_n satisfies the conclusion of the present lemma, we may omit the first half of S_n from the deformation under consideration, setting α_0 to be the halfway point of S_n. Now there is a collection of inverse flipping moves near $t = 0$ and a collection of the same number of flipping moves near $t = 1$. Choose a bijection β between these two sets of swallowtails, and choose a swallowtail σ near $t = 0$. Using Lemma 3.2 and the trick from Figure 22, push σ forward in t and possibly across cusp arcs to lie adjacent to $\beta(\sigma)$ as in Figure 21. Cancel the two swallowtails as in that figure. Repeat for all the other swallowtails near $t = 0$. Now α begins with a sequence of stabilization deformations, then there is a sequence of fold crossings (but no merges, births, or swallowtails), then a sequence of inverse stabilization deformations. Since the stabilizations satisfy the conclusion of the lemma, omit them from the deformation under consideration: Without loss of generality $\operatorname{crit}\alpha_t$ is connected for all t and is free of swallowtail points. By Lemma 3.9, ι consists entirely of immersion pairs. Now apply Lemma 3.21 to make α ready.

Since α is ready, if an immersion circle c is nullhomotopic, then either c is isolated, and the disk bounded by c in $\operatorname{crit}\alpha$ is disjoint from ι', or c is genus-increasing. This is due to the fact from the proof of Lemma 3.21 that if the disk is not disjoint from ι', then it contains a sequence of regions terminating in a clearable region. Another immediate consequence is that if the first immersion pair to appear has a genus-decreasing circle c, then that circle is not nullhomotopic. We now turn to non-nullhomotopic circles.

For any immersion circle c, let the term *subloop* denote a proper sub-arc of c whose endpoints lie at a single self-intersection of c. Viewing $\operatorname{crit}\alpha$ as $S^1 \times [0,1]$, if a subloop of c projects to $\operatorname{crit}\alpha_0$ with degree ± 1, call it *primitive*. A primitive subloop ℓ is *genus-decreasing* if a transverse path from $\operatorname{crit}\alpha_0$ to $\operatorname{crit}\alpha_1$ crosses ℓ in a genus-decreasing direction more times than it crosses ℓ in a genus-increasing direction (an analogously define *genus-increasing primitive subloop*). If c is genus-decreasing, then any primitive loop ℓ_{first} containing the T-minimal point of c is genus-decreasing, because otherwise c contains a nullhomotopic genus-decreasing subloop, to which the previous paragraph applies with *immersion circle* c replaced by *subloop of* c. If we additionally assume c and its counterpart contain the minimum of $T|_\iota$, and that ℓ_{first} is genus-decreasing, then all primitive subloops in ι must also be genus-decreasing, since otherwise there would be a region bounded by arcs in ℓ_{first} and a genus-increasing primitive subloop, which would contain a sequence of regions terminating in a clearable region.

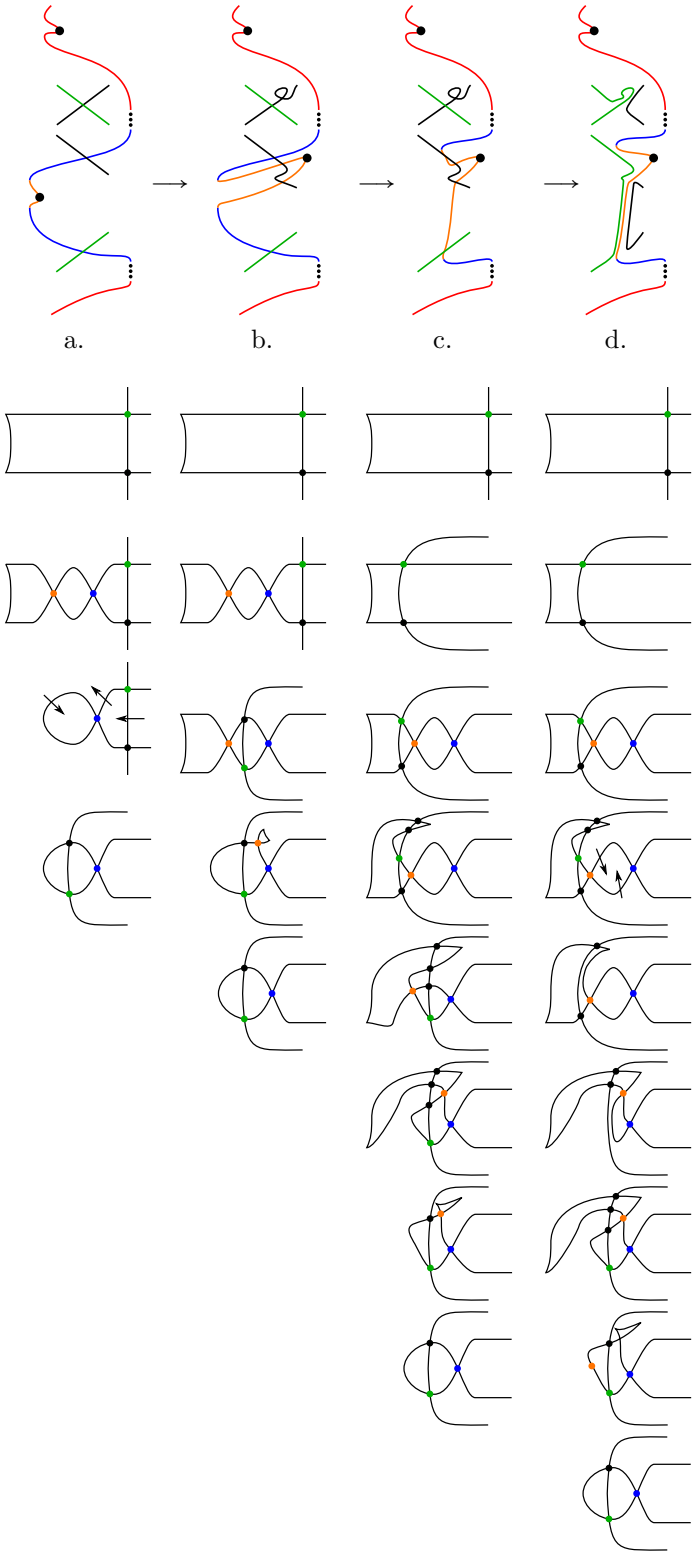
Order the primitive subloops according to their minimal T-values for genus-increasing primitive subloops and maximal T-values for genus-decreasing primitive subloops. The previous two paragraphs show that if α is ready, then all nullhomotopic subloops are genus-increasing and the primitive subloops occur as a

sequence of genus-increasing primitive subloops followed by a sequence of genus-decreasing primitive subloops. With this understood, the first immersion pair to appear is genus-increasing (if all primitive subloops are genus-decreasing, reverse the t-parameter). Now apply the conversion of Figure 48 to that pair, resulting in Figure 50a. In that figure, the short immersion arcs crossing the newly-created single are the those of the first crossing to occur in the single, possibly after applying the modification of Figure 43, and the topmost crossing might be on the other (lower-genus) side of the single, which is something that does not affect the validity of the following modifications. The bigon in Figure 50b, formed by applying Lemma 3.2, is clearable by an application of the Bigon Move Lemma 3.19(3) as shown in Figures 50c-50d. Below each depiction of ι is the corresponding sequence of base diagrams, with t increasing from top to bottom in each column. This procedure may be repeated for the next crossing in the single, with readiness guaranteeing that the orientations of immersion arcs remain as required for repeated use of bigon moves, until the single is isolated.

An alternative method for isolating the single appears in Figure 51. That figure seems a bit simpler, but boils down to the same basic idea of including an R_3 deformation into a pair of R_3 deformations which come from a fold arc moving across a bigon. In both cases, cancelcancellation of the pair of R_3 deformations is achieved by inserting a detour which causes the bigon to be absent during the time when the fold arc would have passed into its interior. The movements in Figure 51 exist by the following argument. Consider the base diagram at the bottom right of that figure. The bigon lies on the lower-genus side of both of its sides, so the Reidemeister-II fold crossing that contracts that bigon exists by [W2, Proposition 2.5(3)]. A flipping move applied to the left side of that bigon is supported in an arbitrarily small ball neighborhood of a point of the fold arc mapping to that side, so the bigon cancellation can be supported disjointly from that ball. For this reason, in all of the base diagrams of Figure 51, the vertical arc may be pushed to the left to be disjoint from the other arc by a deformation.

In order to show that this process may be repeated, it is necessary to consider the possibility that there is an immersion circle on the lower-genus side of the (now isolated) single with its counterpart on the higher-genus side of the single. These circles of this pair are both nullhomotopic because property s_2 requires such a pair to inhabit the same interval in t as the single, but the pair is also disjoint from the single. But if there are such pairs, then the first such pair is easily seen to be a mixed pair, considering the base diagrams for the appearance of the single followed by an R_2 deformation between a pair of critical arcs on the higher- and lower-genus sides of the single. The genus-decreasing circle in that mixed pair must be isolated, and the disk it bounds must contain only isolated circles, because otherwise α is not ready. For this reason, the immersion circles of any non-isolated pair are not separated by the single, and can be moved in the t-direction away from the single using Lemma 3.14 to occupy t-intervals which are disjoint from the t-interval occupied by the single, so that ι' is contained in t-intervale which do not contain any singles, and the isolation process may be repeated. In this way, all immersion circles become isolated. □

3.5.5. *Organizing the embedded immersion locus.* At this point, we have shown that α may be modified to consist of a sequence of shift deformations, multislide deformations, pseudostabilizations and Reidemeister-II fold crossings. This last



e. Base diagrams for Figures 50a-50d.

FIGURE 50. One method for removing crossings in the proof of Lemma 3.23.

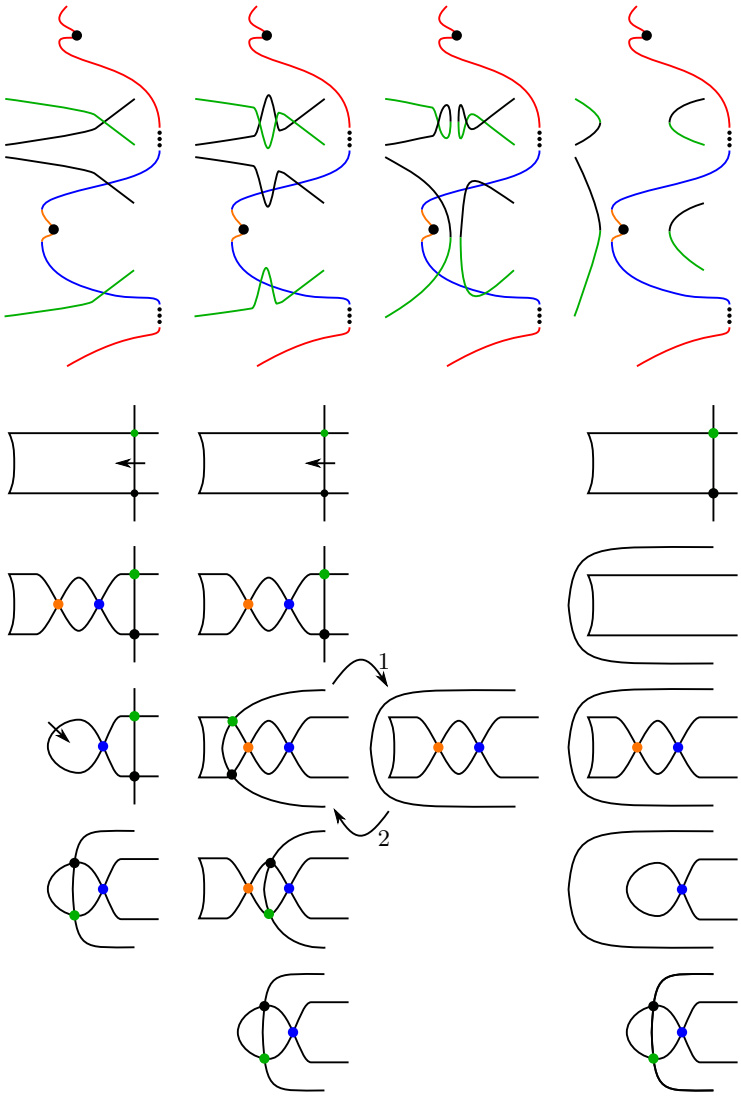


FIGURE 51. Another method for removing crossings. Unnumbered arrows indicate decreasing fiber genus. The third column is the same as the second, with an added detour as indicated by arrows 1 and 2 (this corresponds to a pinching move in ι). Essentially, all movements are valid because the bigon at the bottom of each column can be eliminated.

section converts the fold crossings into a sequence of stabilization deformations and handleslide deformations.

Definition 3.24. An isolated immersion circle c is *nested* if it or its counterpart bounds a disk in $\operatorname{crit}\alpha$ whose interior contains an immersion circle d (for short, just say c or c' contains d). In this situation, the inner circle d is also said to be nested.

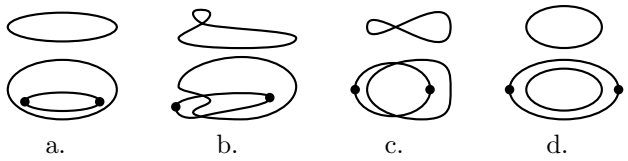


FIGURE 52. Everting a non-nested genus-decreasing circle for Lemma 3.25. This changes the pair into a mixed pair, because the top circle becomes genus-increasing. Everting the lower circle makes the pair genus-increasing.

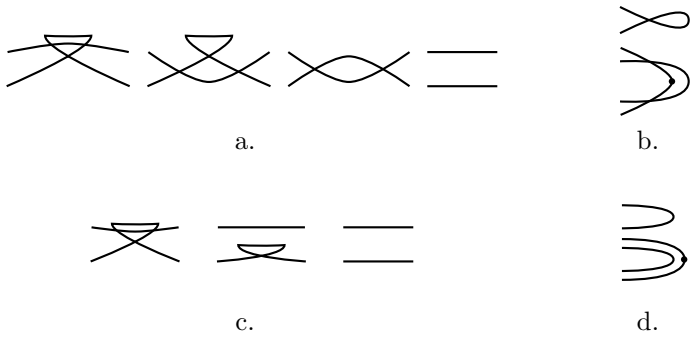


FIGURE 53. Base diagrams for part of Figures 52c and 52d. The upper fold arc may have either orientation.

Lemma 3.25. Suppose ι consists of isolated circles in $\operatorname{crit}\alpha \approx S^1_{[a.b]}$. Then, by a homotopy of α, it is possible to arrange for all immersion circles to come from pseudostabilizations.

Proof. The first step is to explain how to *evert* any isolated non-nested genus-decreasing circle. First come the details for when that circle is one of a genus-decreasing pair.

The moves in Figure 52 proceed from left to right, and the genus-decreasing circle which is everted is the top circle. These movements are perhaps familiar:

- 52a comes from performing and immediately reversing a flipping move on one side of the bigon of critical values.
- 52b is an instance of Lemma 3.2.
- 52c is an application of Lemma 3.14, and
- 52d is the same trick as employed in the rightmost column of Figure 51, where a fold arc is pushed off another before an inverse flipping move. For clarity's sake, Figure 53 has the relevant base diagrams.

Here is why the same movements are available if the lower circle in Figure 52 is genus-increasing; that is, it is possible to evert the genus-decreasing circle of an isolated non-nested mixed pair. In this situation, the upper fold arc in Figures 53a and 53c has its higher-genus side on top, so that the R_2 deformation in Figure 53c is mixed. This deformation exists because the existence of the Reidemeister-II

fold crossing in Figure 53a shows that the required matching condition for [W2, Proposition 2.5(4)] is satisfied.

The first modification is to pinch all the immersion pairs to be disjoint from the cusp locus: Given a cusp arc which has intersections with ι, that set of crossings is partitioned into pairs which cancel up to isotopy of ι because at this point the pairs are all isolated and therefore nullhomotopic (if one of the circles of a pair had a primitive subloop, then that circle would intersect its counterpart). Because the circles are isolated, there is an *innermost* pair of intersections in the sense that the cusp arc connecting them is otherwise disjoint from ι. Use Lemmas 3.14 and 3.17(1b) to cancel that pair of intersections. Proceed inductively until the cusp locus is disjoint from all immersion pairs. Going further, if any circle contains two circles which bound disjoint disks in $\operatorname{crit}\alpha$, start with the innermost circle c containing a and b and pinch c so that c becomes two circles, one containing a and the other containing b. Repeating this, any collection of nested circles is a collection of concentric circles. The next goal is to arrange that each such collection has its genus-decreasing circles inside of all of its genus-increasing circles. The main tool is Figure 54, which uses two finger moves, then the ambient isotopy lemma, then two bigon moves (Lemmas 3.19(2), 3.14 and 3.19(3), respectively) to expand a circle a to contain a circle b. Let such a movement be called *expanding* a genus-increasing circle. Observe that such a movement applies to the situation in which a' is not contained in b'; in that case, it is not hard to see that in the base diagrams there forms a bigon, then the two sides of the bigon meet again, forming a smaller bigon, and such a pair of immersion pairs is eligible for a pinching move, Lemma 3.19(1). This pinching move unifies the circles into an isolated genus-decreasing pair.

With this understood, it is possible to cause every family of nested circles concentric, with the genus-increasing circles containing the genus-decreasing circles. Once this is achieved, call a circle *innermost* if neither it nor its counterpart contain another circle. Then evert (as in Figure 52) an innermost genus-decreasing circle and repeatedly expand it to contain all of the genus-decreasing circles in its family. Repeating this, every nullhomotopic circle eventually becomes genus-increasing. This also introduces nullhomotopic genus-increasing singles, one for each eversion. Convert every such single and every genus-increasing pair as in Figures 47 and 48, this time working with outermost circles. This completes the proof. □

Lemma 3.26. By a homotopy of α it is possible to turn any pseudostabilization into a deformation consisting of one stabilization and a collection of handleslides.

Proof. Given the immersion single d coming from a pseudostabilization, use Corollary 3.15 repeatedly until $T|_d$ has four extrema: Two coming from swallowtails s_1, $s_2 \in d$ and the other two coming from a Reidemeister-II fold crossing. Next, use Lemma 3.17 to remove all intersections between one component of $d \setminus \{s_1, s_2\}$ and the cusp locus (first push the swallowtails to lie in the same strip of fold points, then pinch along the cusps as in the proof of Lemma 3.25. This converts d into a stabilization single and pinches off a number of non-nested genus-decreasing pairs. Because their images lie in distinct regions of $S^2_{[0,1]}$, these pairs may be separated into distinct intervals in t using Lemma 3.14. □

Proof of Theorem 1.1. The results of Section 2 that give the correspondence between model deformations and moves on crown diagrams show that a deformation consists of a sequence of model deformations exactly when its decorated critical

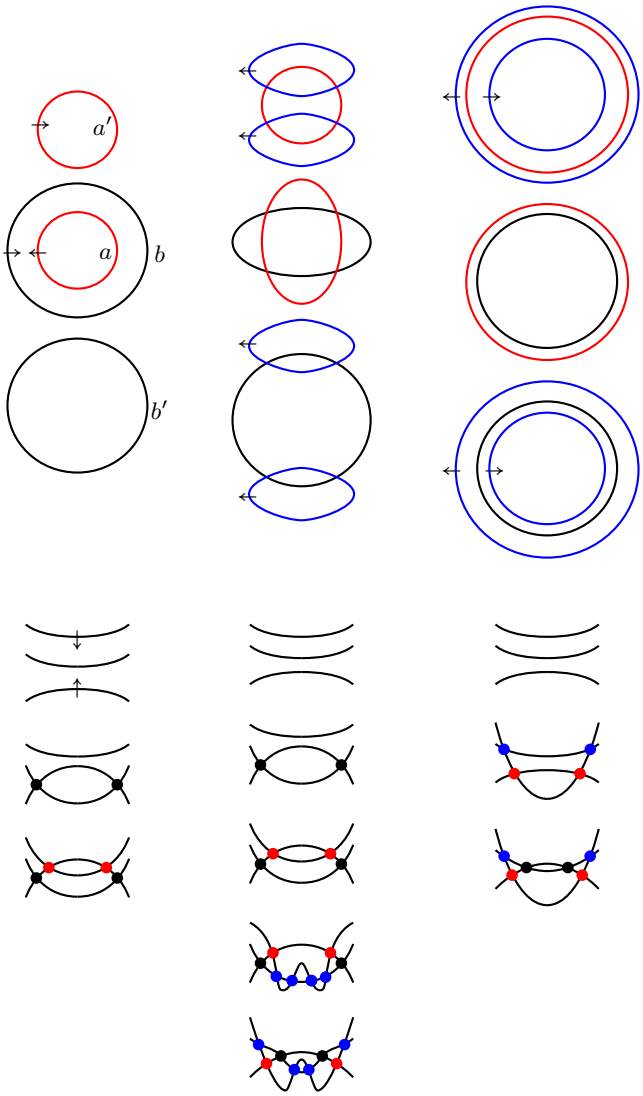
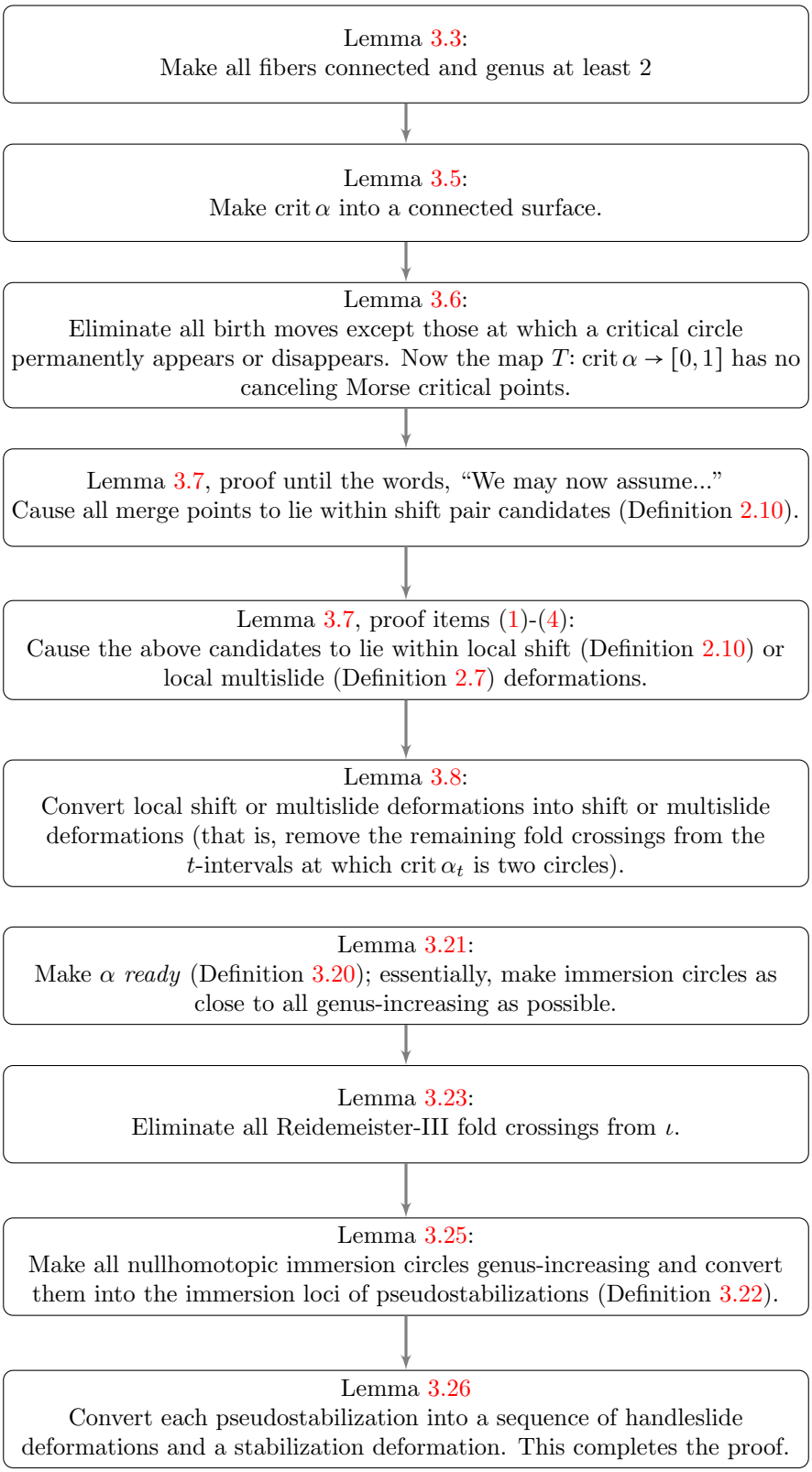


FIGURE 54. Expanding a genus-increasing circle a to contain a genus-decreasing circle b. In the depictions of ι along the top, t increases to the right. In sequences of base diagrams below, t goes from top to bottom and back to the top. All arrows indicate decreasing fiber genus. Observe this modification occurs arbitrarily close to a, a', b, b'. If a' was inside b' the circles would be eligible for a pinching move, Lemma 3.19(1).

surface is a concatenation of decorated critical surfaces whose decorations are characteristic of model deformations. The lemmas of Section 3 yield a deformation whose decorated critical surface is such a concatenation. □

4. Flowchart for the modification of α

Here is a flowchart which gives an overview of the proof of Theorem 1.1.


Lemma 3.3:
Make all fibers connected and genus at least 2
Lemma 3.5:
Make crit α into a connected surface.
Lemma 3.6:
Eliminate all birth moves except those at which a critical circle permanently appears or disappears. Now the map $T\colon \operatorname{crit}\alpha \to [0,1]$ has no canceling Morse critical points.
Lemma 3.7, proof until the words, "We may now assume..."
Cause all merge points to lie within shift pair candidates (Definition 2.10).
Lemma 3.7, proof items (1)-(4):
Cause the above candidates to lie within local shift (Definition 2.10) or local multislide (Definition 2.7) deformations.
Lemma 3.8:
Convert local shift or multislide deformations into shift or multislide deformations (that is, remove the remaining fold crossings from the t-intervals at which $\operatorname{crit}\alpha_t$ is two circles).
Lemma 3.21:
Make α *ready* (Definition 3.20); essentially, make immersion circles as close to all genus-increasing as possible.
Lemma 3.23:
Eliminate all Reidemeister-III fold crossings from ι.
Lemma 3.25:
Make all nullhomotopic immersion circles genus-increasing and convert them into the immersion loci of pseudostabilizations (Definition 3.22).
Lemma 3.26
Convert each pseudostabilization into a sequence of handleslide deformations and a stabilization deformation. This completes the proof.

References

[AK] S. Akbulut and Ç. Karakurt, Every 4-manifold is BLF, *J. Gökova Geom. Topol.* **2** (2008) 40-82.

[ADK] D. Auroux, S. Donaldson, and L. Katzarkov, Singular Lefschetz pencils, *Geom. Topol.* **9** (2005), 1043-1114. doi:10.2140/gt.2005.9.1043

[B1] R. I. Baykur, Topology of broken Lefschetz fibrations and near-symplectic 4-manifolds, *Pac. J. Math* **240** No. 2 (2009), 201-230. doi:10.2140/pjm.2009.240.201

[B2] R. I. Baykur, Existence of broken Lefschetz fibrations, *Int. Math. Res. Notices* **2008** (2008). doi:10.1093/imrn/rnn101

[Be] S. Behrens, On 4-manifolds, folds and cusps, *Pac. J. Math* **264**(2) (2013), 257-306. doi:10.2140/pjm.2013.264.257

[BH] S. Behrens and K. Hayano, Elimination of cusps in dimension 4 and its applications, *Proc. London Math Soc.* **113** (5), 674-724. doi:10.1112/plms/pdw042

[DS] S. K. Donaldson and I. Smith, Lefschetz pencils and the canonical class for symplectic 4-manifolds, *Topology* **42** (2003), 743-785. doi:10.1016/S0040-9383(02)00024-1

[FM] B. Farb and D. Margalit, A Primer on Mapping Class Groups, Princeton Mathematical Series **49**, Princeton University Press, Princeton, NJ 2011.

[GK1] D. Gay and R. Kirby, Indefinite Morse 2-functions; broken fibrations and generalizations, Geom. Topol. **19** (2015), 2465–2534. DOI: 10.2140/gt.2015.19.2465

[GK2] D. Gay and R. Kirby, Fiber connected, indefinite Morse 2-functions on connected n-manifolds, *Proc. Nat. Ac. Sci.* **108**, no. 20 (2011), 8122-8125.

[GS] R. Gompf and A. Stipsicz, 4-manifolds and Kirby Calculus, Graduate Studies in Math. **20**, Amer. Math. Soc., Providence, RI 1999.

[H] K. Hayano, Modification rule of monodromies in R_2-move, Alg. Geom. Topol. **14** (2014), 2181–2222. doi: 10.2140/agt.2014.14.2181

[KMT] R. Kirby, P. Melvin, and P. Teichner, Cohomotopy sets of 4-manifolds, *Geom. Topol. Monographs* **18** (2012), 161-190. doi:10.2140/gtm.2012.18.161

[L1] Y. Lekili, Wrinkled fibrations on near-symplectic manifolds, *Geom. Topol.* **13** (2009), 277-318. doi:10.2140/gt.2009.13.277

[L2] Y. Lekili, Heegaard–Floer homology of broken fibrations over the circle, Adv. Math. **244** (2013), 268–302. doi:10.1016/j.aim.2013.05.013

[Lev] H. Levine, Elimination of cusps, *Topology* **3**, Suppl. 2 (1965), 263-296. doi:10.1016/0040-9383(65)90078-9

[M] J. Morgan, The Seiberg-Witten equations and applications to the topology of smooth four-manifolds, Mathematical Notes **44** (1996), Princeton, NJ: Princeton University Press, pp. viii+128.

[T1] C. Taubes, Counting pseudo-holomorphic submanifolds in dimension 4, *J. Diff. Geom.* **44** no. 4 (1996), 818-893. MR1438194

[T2] C. Taubes, Seiberg-Witten invariants and pseudo-holomorphic subvarieties for self-dual, harmonic 2-forms, *Geom. Topol.* **3** (1999), 167-210. doi:10.2140/gt.1999.3.167

[U] M. Usher, The Gromov invariant and the Donaldson-Smith standard surface count, *Geom. Topol.* **8** (2004), 565-610. doi:10.2140/gt.2004.8.565

[Wa] G. Wassermann, Stability of unfoldings in space and time, *Acta. Math.*, **135** (1975), 58-128. doi:10.1007/BF02392016

[W1] J. Williams, The h-principle for broken Lefschetz fibrations, *Geom. Topol.* **14** no. 2 (2010), 1015-1061. doi:10.2140/gt.2010.14.1015

[W2] J. Williams, Existence of 2-parameter crossings, with applications, *Geom Dedicata* (2019).

Department of Mathematical Sciences, Binghamton University
Email address: jdw.math@gmail.com