ELIMINATION OF CUSPS IN DIMENSION 4
AND ITS APPLICATIONS

STEFAN BEHRENS AND KENTA HAYANO

ABSTRACT. Several new combinatorial descriptions of closed 4–manifolds have recently been introduced in the study of smooth maps from 4–manifolds to surfaces. These descriptions consist of simple closed curves in a closed, orientable surface and these curves appear as so called *vanishing sets* of corresponding maps. In the present paper we focus on homotopies canceling pairs of cusps so called *cusp merges*. We first discuss the classification problem of such homotopies, showing that there is a one-to-one correspondence between the set of homotopy classes of cusp merges canceling a given pair of cusps and the set of homotopy classes of suitably decorated curves between the cusps. Using our classification, we further give a complete description of the behavior of vanishing sets under cusp merges in terms of mapping class groups of surfaces. As an application, we discuss the uniqueness of *surface diagrams*, which are combinatorial descriptions of 4–manifolds due to Williams, and give new examples of surface diagrams related with Lefschetz fibrations and Heegaard diagrams.

1. INTRODUCTION

The study of maps from 4–manifolds to surfaces has received considerable attention in the last years. Motivated by Perutz's generalization [Per07, Per08] of the Donaldson-Smith invariant [DS03] for Lefschetz fibrations to an invariant of broken Lefschetz fibrations introduced in [ADK05], Lekili [Lek09] discussed homotopies between so called wrinkled fibrations, which are stable maps on a 4–manifold with only indefinite folds and cusps. Lekili's results were later improved by Williams [Wil10] and Gay-Kirby [GK11b, GK11a]. The results guarantee that for any two homotopic wrinkled fibrations there exists a homotopy such that all but finitely many maps in the homotopy are wrinkled fibrations and near the exceptional maps the homotopy has one of essentially six possible local models. While it is easy to see that basic homotopies change the critical value sets as in Figure 2, for a map with one of the critical value configurations shown in Figure 2 it is not always possible to realize the corresponding modification of critical values by a homotopy. In order to guarantee the existence of such a homotopy one has to investigate what we call *vanishing sets* in fibers of the map, which are reminiscent of ascending and descending manifolds in Morse theory (this problem is addressed in [Wil14], for example). Moreover, even if a homotopy can be found, it is not enough to understand the critical values but one should also keep track of how the vanishing sets behave throughout the homotopy. On a related note, various existence results for maps with special properties have been obtained. In particular, Williams's *simple wrinkled fibrations* [Wil10] can be used to obtain combinatorial descriptions of closed 4–manifolds known as *surface diagrams*. These diagrams consist of collections of simple closed curves in surfaces arising as vanishing sets of simple wrinkled fibrations. (Other combinatorial descriptions of 4–manifolds related to the idea of vanishing sets can be obtained from Baykur's *simplified broken Lefschetz fibrations* [Bay09] and Gay and Kirby's *trisections* [GK12].) Williams also discussed the uniqueness of simple wrinkled fibrations up to homotopy, introducing four basic homotopies which he called *handleslides*,

stabilizations, *multislides* and *shifts* [Wil11b]. Our main motivation was to understand how these homotopies affect the associated surface diagrams. However, our results are applicable in more general contexts.

Given a wrinkled fibration, the general strategy is to fix one reference fiber over each connected component of the regular value set and to collect the vanishing sets associated to neighboring critical values in those fibers. If one wants to understand how the vanishing sets change in homotopies, the crucial problem is that two formerly disconnected regions of regular values can be joined. As a result, after the homotopy there is one region that contains two previously unrelated fibers with their vanishing sets. However, the two fibers can now be related by parallel transport so that the two collections of vanishing sets appear in a single fiber again. But it turns out that the way they appear relative to each other usually depends on the homotopy. This phenomenon can be caused by two kinds of homotopies known as R_2–*moves* and *cusp merges*. The former have been studied in [Hay14] and the latter are the main focus of the present paper.

It is interesting to note that cusp merge homotopies have already been studied 50 years ago in Levine's work [Lev65] on the elimination of cusps. It is well known that if the source manifold of a stable map to a surface is closed, then the number of cusps in this map is finite and has the same parity as the Euler characteristic of the source. Using cusp merges Levine showed that the Euler characteristic is the only obstruction for the elimination of cusps: he proved that any stable map to a surface is homotopic to a map with at most one cusp. In some sense, part of the present paper is a natural extension of [Lev65], whence the title. For the elimination of cusps it was enough to understand *when* a pair of cusps can be eliminated. However, as explained in the previous paragraph, the recent developments of the topology of 4–manifolds have led us to study *how many ways* there are to eliminate a given pair of cusps in order to understand the behavior of vanishing sets. An analogous situation occurred in Morse theory: in order to prove the h– and s–cobordism theorems it was enough to understand *when* two critical points can be canceled, but Cerf's approach to the pseudo-isotopy problem made it necessary to study – among many other things – the ambiguities in the cancellation procedure (see [Cer70, HW73]).

Statement of results and outline. We first address the classification problem of cusp merges for a stable map on a 4–manifold which cancel a given pair of indefinite cusps. For this purpose we will define a notion of *elementary cusp merges* and characterize them up to homotopy by their so called *framed joining curves*, which are certain (suitably decorated) curves that connect the cusps. We will prove the following result in Section 3 to which we also refer for precise definitions. The proof relies on some technical results that we outsource into Appendices I and II.

Theorem A. *Let $f\colon X \to B$ be a stable map and $p, q \in X$ a pair of indefinite cusps. Denote by $\mathfrak{CM}_0(f; p, q)$ and $\mathbf{\Lambda}(f; p, q)$ the spaces of elementary cusp merge homotopies and framed joining curves for p and q. Then there are canonical bijections between $\pi_0\big(\mathfrak{CM}_0(f; p, q)\big)$ and $\pi_0\big(\mathbf{\Lambda}(f; p, q)\big)$.*

Building on Theorem A we go on to clarify how cusp merges between indefinite cusps affect vanishing sets. As explained, the problem comes from the fact that the two regions in the left and the right sides of Figure 1(a) are joined as shown in Figure 1(b) and the two fibers Σ_1 and Σ_2 over p_1 and p_2 can be identified by parallel transport along the dotted path in Figure 1(b). As a consequence, the two collections of vanishing sets in Σ_1 and Σ_2 derived from the initial map can be considered in only one of these fibers after the homotopy. We can thus understand the behavior of vanishing sets in cusp merges once we know which diffeomorphisms can appear as parallel transport. For the statement of our result, recall that for
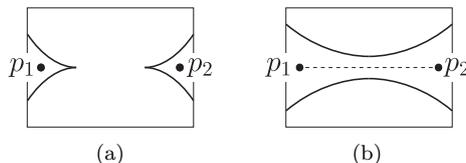
FIGURE 1. Critical value sets in a cusp merge.

a non-separating simple closed curve $c \subset \Sigma$ in a closed, orientable surface there is a so called *surgery homomorphism* $\Phi_c \colon \mathrm{Mod}(\Sigma)(c) \to \mathrm{Mod}(\Sigma'_c)$ where $\mathrm{Mod}(\Sigma)(c)$ consists of all mapping classes that fix the isotopy class of c and Σ'_c is the surface obtained by surgery on c (see Section 5.2 for more information).

Theorem B. *The subset of $\pi_0(\mathrm{Diff}(\Sigma_1, \Sigma_2))$ obtained from merging cusps along a fixed arc in B has a free and transitive action of either of the groups*

$$\mathcal{K}(c_i, d_i) = \ker \Phi_{c_i} \cap \ker \Phi_{d_i} \subset \mathrm{Mod}(\Sigma_i), \quad i = 1, 2$$

where $c_i, d_i \subset \Sigma_i$ are simple closed curves related to the vanishing sets of the cusps.

We will also describe a set of generators for $\mathcal{K}(c_i, d_i)$ in Lemma 5.4.

In Section 6 we discuss some applications of Theorem B to the theory of surface diagrams. Section 6.2 is concerned with the uniqueness problem for surface diagrams. More precisely, we use Theorem B to study how surface diagrams change in two types of homotopies introduced in [Wil11b], the so called *multislides* and *shifts*. The analogous problem for the remaining homotopies (handleslides and stabilizations) was studied by the second author in [Hay14]. Finally, in Section 6.3 we discuss how to obtain surface diagrams from some constructions of simple wrinkled fibrations involving cusp merges. More precisely, we show how to construct surface diagrams for total spaces of Lefschetz fibrations from the knowledge of the (Lefschetz) vanishing cycles, and for products of 3–manifolds with the circle from a given Heegaard diagram. As concrete examples, we obtain new surface diagrams for S^4 and $S^1 \times S^3$ (see Figures 10 and 15(b)). These have the interesting property that they are *not* related to the previously known diagrams given in [Hay14] by the moves discussed in Section 6.2 because the corresponding simple wrinkled fibrations are not homotopic.

The observant reader will have noticed that we have neglected to mention Sections 2 and 4 so far. These contain necessary background material for Sections 3 and 5 that might also be of independent interest. Although the main purpose of Section 2 is to introduce terminology and notation used in the subsequent sections, we go a bit further. The recent developments in the study of stable maps on 4–manifolds have heavily relied on folklore facts in singularity theory. However, the authors found it extremely difficult to find references including complete proofs of these results. For this reason, we take the chance to review these facts with outlines of proofs. In the remaining Section 4 we introduce a notion of *connections* for general smooth maps generalizing the usual concept for fiber bundles. We thereby provide a conceptual framework for discussing vanishing sets and parallel transport which has previously been done in an ad hoc fashion (if at all).

2. GENERALITIES ABOUT MAPS TO SURFACES

The purpose of this section is twofold. First and foremost, we will introduce necessary terminology and notation that will be used throughout and explain the necessary background for our results. Since the proofs of some technical results

require more sophisticated notions from singularity theory (such as stability of map-germs and their unfoldings), our review will be a little more extensive than one might expect. Second, we take the chance to give a survey some "generic" properties of maps from 4–manifolds to surfaces and homotopies between them from the perspective of singularity theory. Specifically, we want to address the following "well known facts" that have been used in a number of papers in the context of wrinkled fibrations and Morse 2–functions.

(1) "Generic" maps from 4–manifolds to surfaces have only folds and cusp singularities and if a critical value is covered by more than one critical point, then it is covered by two folds whose images meet transversely in the target.

(2) In "generic" homotopies of maps from 4–manifolds to surfaces all but finitely many maps have the above structure. While passing through the exceptional maps one of six phenomena occurs in which the critical image changes as shown in Figure 2.

While the first statement indeed follows from rather standard (albeit not completely trivial) results in singularity theory, the second one is a little more complicated. In fact, not much literature is available about global phenomena in families of maps. What both statements have in common is that it is surprisingly difficult to find concrete references in the literature. We are aware of only one source, namely [Chi77], but unfortunately it is very hard to find. So as a third purpose, we aim to provide a more readily available reference.

To be clear, we do not claim any originality. Theorem 2.1 can be proved immediately using Mather's criterion of stability in [Mat70], while Theorems 2.5 and 2.6 are contained in [Chi77], although our proofs are slightly different. We are aware that what we write will be cryptic to some and at the same time obvious to others. However, we hope that some low dimensional topologists who are working with surface valued maps will find our exposition helpful, either as a reference or to obtain some idea about the inner workings of the "black box" known as singularity theory.

2.1. **Basic notation and terminology.** Throughout we assume that all manifolds and maps are smooth. As a precautionary measure we also assume that all manifolds are connected and oriented, and that sources of maps are closed unless otherwise noted. We reserve the letters X and B for 4–manifolds and surfaces, respectively. In more general situations we follow the tradition in singularity theory and consider maps between manifolds N and P of dimensions n and p; in this case we always assume that $n \geq p$.

Maps and homotopies. Let $C^\infty(N, P)$ be the set of smooth maps from N to P. Throughout the paper we endow this set with the *Whitney C^∞ topology*[1]; for its definition and basic properties we refer to [GG73, II.§3]. As usual in differential topology, objects of interest are studied up to diffeomorphism. In this spirit, two maps $f \in C^\infty(N, P)$ and $f' \in C^\infty(N', P')$ are called *(right-left) equivalent* if there exist diffeomorphisms $\phi : N \to N'$ and $\psi : P \to P'$ which satisfy $f' \circ \phi = \psi \circ f$. By a *homotopy* of maps we mean a smooth map $F: J \times N \to P$ where J is some fixed interval; we will mostly use $J = [-1, 1]$. We frequently think of homotopies as 1–parameter families of maps $F = (f_s)_{s \in J}$ where $f_s: N \to P$ is defined by $f_s(p) = F(s, p)$. We will mostly consider homotopies with fixed initial map which we simply denote by the lower case letter f. Another useful way to think of homotopies is to consider the map $\widetilde{F}: J \times N \longrightarrow J \times P$ defined by $\widetilde{F}(s, p) = (s, f_s(p))$. We call $\widetilde{F}$ the *unfolding* associated to F. Of course, the objects F, (f_s), and $\widetilde{F}$ all

---

[1]Since we are assuming that X is closed, the Whitney topology agrees with the topology of uniform convergence of all partial derivatives and is generally rather well-behaved. However, for non-compact X the story is more complicated.

contain the same information but each perspective has its advantages. A natural notion of *equivalence* for homotopies $F \in C^{\infty}(J \times N, P)$ and $F' \in C^{\infty}(J \times N', P')$ is given by commutative diagrams of the form

$$
\begin{array}{ccccc}
J \times N & \xrightarrow{\widetilde{F}} & J \times P & \xrightarrow{\mathrm{pr}_J} & J \\
\downarrow{\scriptstyle\Phi} & & \downarrow{\scriptstyle\Psi} & & \downarrow{\scriptstyle\theta} \\
J \times N' & \xrightarrow{\widetilde{F'}} & J \times P' & \xrightarrow{\mathrm{pr}_J} & J
\end{array}
\tag{1}
$$

where Φ, Ψ, and θ are diffeomorphisms and pr_J denotes the projections onto J. We say that F is *constant* if $\widetilde{F} = \mathrm{id}_J \times f$ or, equivalently, $f_s = f$ for all $s \in J$. We call F *trivial* if it is equivalent to the constant homotopy at the initial map of F. In this case, we can assume that $\theta = \mathrm{id}_J$ so that we can consider Φ and Ψ as families of diffeomorphisms $(\phi_s) \in \mathrm{Diff}(N)$ and $(\psi_s) \in \mathrm{Diff}(P)$. Moreover, they satisfy $f_s = \psi_s \circ f \circ \phi_s^{-1}$ and we can assume that both families of diffeomorphisms start from the identity map; we will refer to homotopies of this form as *isotopies* of f.[2]

Critical points. Recall that a *critical point* of a smooth map $f \colon N \to P$ is a point $x \in N$ where the derivative df_x fails to have maximal rank. The number $p - \mathrm{rk}(df_x)$ is called the *corank* of f at x. The image of a critical point is called a *critical value*. We denote the sets of critical points and critical values by $\mathcal{C}(f) \subset N$ and $\mathcal{D}(f) = f(\mathcal{C}(f)) \subset P$, respectively, and refer to $\mathcal{C}(f)$ as the *critical locus* and to $\mathcal{D}(f)$ as the *critical image* or *discriminant*. Points in the complements of $\mathcal{C}(f)$ and $\mathcal{D}(f)$ are called *regular* points and values, respectively.

Germs, singularities, and local models. In order to study local structure of maps, it is convenient to use the language of germs. Let $f \colon N \to P$ be a smooth map and let $S \subset N$ be a subset. Recall that the *germ* of f at S is the equivalence class of maps that are defined in a neighborhood of S and agree with f on a possibly smaller neighborhood. The most important special cases are when S is a single point or a finite set of N such that $f(S)$ is a point in P. In these situations we speak of *mono-germs* and *multi-germs*. It is customary to denote the germ of f at S by $f \colon (N, S) \to (P, f(S))$ which should not be confused with the notation for maps of pairs. Two germs $f \colon (N, S) \to (P, f(S))$ and $f' \colon (N', S') \to (P', f(S'))$ are called *equivalent* if there are germs of diffeomorphisms $\phi \colon (N, S) \to (N', S')$ and $\psi \colon (P, f(S)) \to (P', f(S'))$ which satisfy $f' \circ \phi = \psi \circ f$. A mono-germ $f \colon (N, x) \to (P, y)$ is *singular* if x is a critical point of some (hence any) representative. With this understood, a *singularity* is an equivalence class of singular mono-germs. More generally, a *multi-singularity* is an equivalence class of multi-germs such that each of its mono-germs is singular. Obviously, by suitable choices of coordinates all (multi-)singularities become equivalent to (finite collections of) singular mono-germs $(\mathbb{R}^n, 0) \to (\mathbb{R}^p, 0)$. The set of germs $(\mathbb{R}^n, 0) \to (\mathbb{R}^p, 0)$ is usually denoted by $\mathcal{E}(n, p)$. If a map $\mu \colon \mathbb{R}^n \to \mathbb{R}^p$ with $\mu(0) = 0$ represents a given singularity, we will refer to μ as a *local model* for that singularity.

2.2. **Maps from 4–manifolds to surfaces.** The first task is to give a precise version of the statement about the generic structure of maps from a 4–manifold X to a surface B. We first define the relevant multi-singularities in terms of local models (which are maps from $\mathbb{R}^4$ to $\mathbb{R}^2$). Here, and throughout the rest of the

---

[2]The reader should be warned that "isotopy" has a slightly different meaning in [Lek09] and [Wil10], there it refers to a homotopy that stays within a special class of maps. Such homotopies are usually not related to isotopies of X and B.

paper, we will denote coordinates on the source by (t, x, y, z) and on the target by (u, v). The first pair of models describes the *fold singularities*

$$\bigl(t,\, x, y, z\bigr) \mapsto \bigl(\, t,\, x^2 + y^2 \pm z^2\bigr) \tag{2}$$

which can be thought of as a trivial family of 3–dimensional Morse singularities of index 1 or 0. In the index 1 case, that is, when the sign is negative, we speak of *indefinite* folds while the index 0 folds are called *definite*. If we superimpose two fold singularities in such a way that their discriminants intersect transversely, we obtain a multi-singularity which we call a *transverse double fold*. Next there are the (*definite* and *indefinite*) *cusp singularities*

$$\bigl(t,\, x, y, z\bigr) \mapsto \bigl(\, t,\, x^3 + 3tx + y^2 \pm z^2\bigr). \tag{3}$$

With these definitions (together with the topology of $C^\infty(X, B)$) in hand, we can make the statement (1) in page 4 precise as follows:

Theorem 2.1. *Let X be a closed 4–manifold and B a surface. Denote by $\mathcal{S}^0(X, B)$ the set of smooth maps $X \to B$ whose only multi-singularities are folds, cusps and transverse double folds. Then $\mathcal{S}^0(X, B)$ is open and dense in $C^\infty(X, B)$.*

As we will explain, $\mathcal{S}^0(X, B)$ agrees with the set of *stable* maps which are classical objects in singularity theory. Since we are deliberately taking the point of view of singularity theory, we will henceforth use this terminology. We note, however, that these maps have gained some popularity under the name of *Morse 2–functions* in the low dimensional topology community through the work of Gay and Kirby [GK11a, GK11b, GK12]. We now go on to explain how Theorem 2.1 follows from the vastly more general theory of stable maps.

2.2.1. *A glance at stability theory.* In this interlude we first consider smooth maps between manifolds N and P of arbitrary dimensions which we write as a pair (n, p). As before, we assume that N is closed and equip $C^\infty(N, P)$ with the Whitney C^∞ topology. We begin with the central definition.

Definition 2.2 (Stable maps)**.** A map is $f\colon N \to P$ is called *stable* if there is an open neighborhood $\mathcal{U}_f$ of f in $C^\infty(N, P)$ such that each $g \in \mathcal{U}_f$ is equivalent to f, that is, there are $(\phi_g, \psi_g) \in \mathrm{Diff}(N) \times \mathrm{Diff}(P)$ such that $g = \psi_g \circ f \circ \phi_g^{-1}$.

The group $\mathrm{Diff}(N) \times \mathrm{Diff}(P)$ is commonly denoted by $\mathcal{A} = \mathcal{A}(N, P)$ in this context. Stability of smooth maps was carefully studied in a highly influential series of papers by Mather (including [Mat70, Mat71]) who attributes the above definition to Thom. The theory has long matured and excellent textbook references are available, for example [GG73, Mar82, Gib79]. Note that stability is an open condition by definition. Moreover, a deep theorem of Mather states that stable maps are also dense in $C^\infty(N, P)$ provided that the dimensions (n, p) lie in the so called range of "nice dimensions" which includes $(n, 2)$ for all n (see [GG73, VI.§6] or [Mat71]). So in order to prove Theorem 2.1 it is enough to show that the $\mathcal{S}^0(X, B)$, which was defined by restricting the allowed multi-germs, agrees with the set of stable maps in dimensions $(4, 2)$. This is done in two steps: the first is a characterization of stable maps in terms of their multi-germs, and the second is the classification of the multi-germs that appear in stable maps.

Observe that $\mathcal{A}$ acts on $C^\infty(N, P)$ and f is stable if and only if its $\mathcal{A}$–orbit is open. Intuitively, one should think of the finite dimensional situation where a Lie group acts properly on a manifold. In this case, the tangent spaces of orbits provide valuable information about the manifold. Even though there are no completely satisfactory manifold structures on $C^\infty(N, P)$ and $\mathcal{A}$, there are natural candidates for their tangent spaces. The main theorem of [Mat70] states that f is stable if and only if the (formal) tangent space to its $\mathcal{A}$–orbit has codimension 0 in the tangent

space to $C^\infty(N, P)$; this is known as *infinitesimal stability* and can be interpreted as the surjectivity of the differential of the map $\mathcal{A} \to C^\infty(N, P)$ given by evaluation of the action at f at the identity (see [GG73, III.§1]). In general, the codimension of $T_f \mathcal{A} f$ in $T_f C^\infty(N, P)$ is called the *codimension of f*.

A similar discussion applies to germs. A group of diffeomorphism germs, usually also denoted by $\mathcal{A}$, acts on the set of map germs and all relevant objects have (formal) tangent spaces; for more details see [Gib79] or [Wal80], for example. This allows to define the *codimension* of a multi-germ $f \colon (N, S) \to (P, y)$ as the codimension of the tangent space to its orbit in its full tangent space, and the codimension 0 multi-germs are called *(infinitesimally) stable*.[3] The connection to the notions of codimension for maps and multi-germs is given by a formula of Mata-Lorenzo [ML89] which expresses the codimension of a map as the sum of codimensions of all its multi-germs. In particular, a map is stable if and only if all its multi-germs are stable; the latter condition is also known as *local stability*.[4]

We end this digression with a word of warning. One should be aware that the compactness assumption on N is crucial for the whole discussion. In fact, the relations of stability with its infinitesimal and local versions become delicate issues when the source is not compact. For state of the art accounts on these matters we refer the interested reader to [dPW95] and [dPV01].

2.2.2. *The proof of Theorem 2.1 (Sketch).* Now let us go back to the situation of Theorem 2.1 where $(n, p) = (4, 2)$. The theory outlined in the previous section reduces the problem to showing that the stable multi-singularities in dimensions $(4, 2)$ are exactly the folds, cusps and transverse double folds. As a first step, one can show directly using the definition that these germs are infinitesimally stable, which is an instructive exercise. In the other direction, another exercise in the definitions shows that all stable (mono-)singularities in dimensions $(4, 2)$ have corank 1. Then the normal forms for corank 1 singularities obtained by Morin [Mor65] show that we are only dealing with folds and cusps. As a last step, we have to discuss multi-singularities. It is a basic fact that for any stable multi-singularity $f \colon (N, S) \to (P, y)$ we have $|S| \le p$. In particular, for $p = 2$ any critical value is covered by at most two critical points. We leave it to the interested reader to show that (a) a bi-singularity that involves a cusp cannot be stable, and (b) that the discriminants of a stable bi-singularity consisting of two folds must intersect transversely. Although some of the verifications that we have left out are tedious, they are all elementary.

2.3. **Homotopies of maps from 4–manifolds to surfaces.** The purpose of this section is briefly reviewing generic properties of homotopies from 4–manifolds to surfaces, in particular making statement (2) on page 4 precise. As in the previous section these properties are described in terms of local models. We thus first introduce another equivalence relation of germs appropriate for homotopies. Let $S \subset \mathbb{R}^n$ be a finite set and $f \colon (\mathbb{R}^n, S) \to (\mathbb{R}^p, f(S))$ a germ. A germ $F \colon (\mathbb{R}^{n+k}, \{0\} \times S) \to (\mathbb{R}^{p+k}, \{0\} \times f(S))$ is called an (*k*–parameter) *unfolding* of f if the restriction $F|_{\{0\} \times \mathbb{R}^n}$ is equal to f and $pr_{\mathbb{R}^k} \circ F = pr_{\mathbb{R}^k}$, where $pr_{\mathbb{R}^k} \colon (\mathbb{R}^{m+k}, \{0\} \times T) \to (\mathbb{R}^k, 0)$ is the germ of the projection onto the former components. Two *k*–parameter unfoldings F_0 and F_1 of f are said to be *k–equivalent* if there exist germs of diffeomorphisms $\Phi \colon (\mathbb{R}^{n+k}, \{0\} \times S) \to (\mathbb{R}^{n+k}, \{0\} \times S)$, $\varphi \colon (\mathbb{R}^{p+k}, \{0\} \times f(S)) \to (\mathbb{R}^{p+k}, \{0\} \times f(S))$ and $\phi \colon (\mathbb{R}^k, 0) \to (\mathbb{R}^k, 0)$ such that the following diagram

---

[3]In fact, there are two notions of codimension related to the $\mathcal{A}$–action on germs. We are interested in the $\mathcal{A}_e$–codimension which is denoted by $d_e(f, \mathcal{A})$ in [Wal80].

[4]The equivalence of stability and local stability was already proved in [Mat70, p.314].

commutes:

$$
\begin{array}{ccccc}
(\mathbb{R}^{n+k}, \{0\} \times S) & \xrightarrow{F_0} & (\mathbb{R}^{p+k}, \{0\} \times f(S)) & \xrightarrow{pr_{\mathbb{R}^k}} & (\mathbb{R}^k, 0) \\
{\scriptstyle \Phi} \downarrow & & {\scriptstyle \varphi} \downarrow & & {\scriptstyle \phi} \downarrow \\
(\mathbb{R}^{n+k}, \{0\} \times S) & \xrightarrow{F_1} & (\mathbb{R}^{p+k}, \{0\} \times f(S)) & \xrightarrow{pr_{\mathbb{R}^k}} & (\mathbb{R}^k, 0).
\end{array}
\tag{4}
$$

An unfolding $F : (\mathbb{R}^{n+k}, \{0\} \times S) \to (\mathbb{R}^{p+k}, \{0\} \times f(S))$ of f is said to be *trivial* if F is k–equivalent to the constant unfolding $\mathrm{id}_{\mathbb{R}^k} \times f$.

Let $F : (\mathbb{R}^{n+k}, \{0\} \times S) \to (\mathbb{R}^{p+k}, \{0\} \times f(S))$ be an unfolding of a germ $f : (\mathbb{R}^n, S) \to (\mathbb{R}^p, f(S))$ and $h : (\mathbb{R}^l, 0) \to (\mathbb{R}^k, 0)$ a germ. We can obtain an l–parameter unfolding $h^* F$ of f as follows:

$$
h^* F(s, x) = (s, pr_{\mathbb{R}^p} \circ \hat{F}(\hat{h}(s), x)),
$$

where $\hat{F}$ and $\hat{h}$ are representatives of F and h, respectively. This unfolding is called the *pull-back* of F by h. An unfolding $F : (\mathbb{R}^{n+k}, \{0\} \times S) \to (\mathbb{R}^{p+k}, \{0\} \times f(S))$ of $f : (\mathbb{R}^n, S) \to (\mathbb{R}^p, f(S))$ is said to be *versal* if any l–parameter unfolding of f is l–equivalent to some pull-back of F for any l. The connection to the theory outlined in Section 2.2.1 is given by the following standard result.

Proposition 2.3 ([Mar82, p.189])**.** *Let $f : (\mathbb{R}^n, S) \to (\mathbb{R}^p, 0)$ be a multi-germ.*

  (i) *f is stable if and only if it is versal when considered as an unfolding of itself. In particular, any unfolding of a stable multi-germ is trivial.*
  (ii) *More generally, the codimension of f agrees with the minimal number of parameters needed to obtain a versal unfolding of f.*

Remark 2.4. As an immediate consequence of Proposition 2.3 (i) is that whenever we have a normal form for a given stable multi-germ, we get a parametric normal form for free (that is, if the singularity appears embedded in a family of maps, then the whole family is locally equivalent to the constant family of the normal form). Moreover, Proposition 2.3 (i) globalizes to the statement that any family of stable maps is trivial (see [GG73, V.§2]).

For the purpose of understanding homotopies of maps, the significance of the concept of versal unfoldings is illustrated by the following result:

Theorem 2.5 ([Chi77])**.** *Let X be a 4–manifold, B a surface and J some interval. Let $\mathcal{H}^0(X, B)$ be the set of homotopies $F : J \times X \to B$ with the following properties:*

  (a) *For any $t \in J$ and finite subset $S \subset X$ the germ of F at $\{t\} \times S$ is a versal unfolding of the germ of f_t at S.*
  (b) *Each map f_t has codimension at most 1 (that is, it contains at most one multi-germ of codimension 1).*

Then $\mathcal{H}^0(X, B)$ is dense in $C^\infty(J \times X, B)$ with respect to the Whitney topology.

In analogy with the situation of maps, we call the elements of $\mathcal{H}^0(X, B)$ *stable homotopies*. In fact, this terminology is justified by the work of Chincaro [Chi77] who develops a theory of stability for families of maps. As in the situation of maps, this immediately implies that $\mathcal{H}^0(X, B)$ is also open in $C^\infty(J \times X, B)$. Before giving an outline of a proof of Theorem 2.5, we quickly review the classification of versal 1–parameter unfoldings of germs from $\mathbb{R}^4$ to $\mathbb{R}^2$. We begin with general remarks on versal unfoldings. First, it is easy to see that $\tilde{F} : (\mathbb{R}^{n+k}, \{0\} \times S) \to (\mathbb{R}^{p+k}, \{0\} \times f(S))$ is versal if and only if the restriction $\tilde{F} : (\mathbb{R}^{n+k}, \{0\} \times (S \cap f^{-1}(q))) \to (\mathbb{R}^{p+k}, (0, q))$ is versal for any $q \in f(S)$. Thus we can assume that $f(S)$ consists of a single point, say $0 \in \mathbb{R}^p$, without loss of generality. Second, a multi-germ
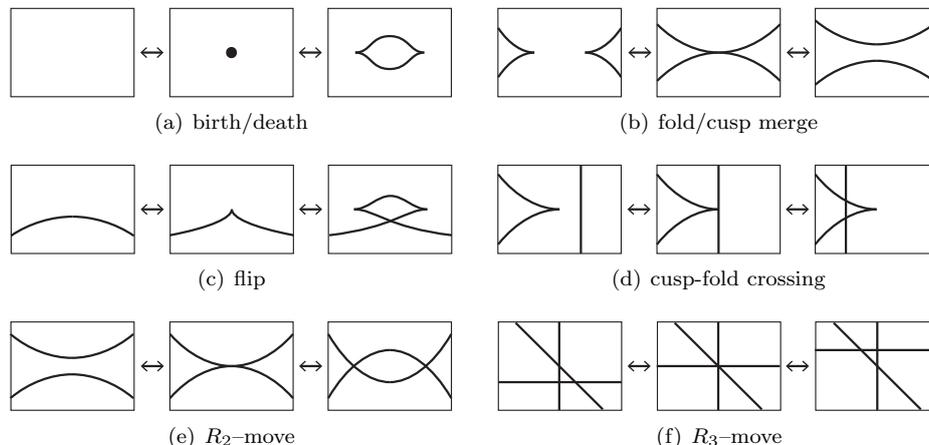
(a) birth/death   (b) fold/cusp merge

(c) flip   (d) cusp-fold crossing

(e) R_2–move   (f) R_3–move

FIGURE 2. The discriminants of versal unfoldings.

$f : (\mathbb{R}^n, S) \to (\mathbb{R}^p, 0)$ has a versal 1–parameter unfolding if and only if it has codimension at most one (in the sense of Section 2.2.1, see Proposition 2.3). Furthermore, in this case the versal unfolding of f is unique up to equivalence. In particular, we can classify 1–parameter versal unfoldings once we obtain the classification of codimension–1 germs (and these versal unfoldings). Lastly, since the codimension of $f : (\mathbb{R}^n, S) \to (\mathbb{R}^p, 0)$ is equal to that of $f \sqcup s : (\mathbb{R}^n \sqcup \mathbb{R}^n, S \sqcup \{0\}) \to (\mathbb{R}^p, 0)$, where $s : (\mathbb{R}^n, 0) \to (\mathbb{R}^p, 0)$ is a submersion-germ, in what follows we only deal with singular germs.

We now turn our attention to the specific dimension pair $(n, p) = (4, 2)$. The germs with codimension 0 are nothing but stable germs given in Section 2.2. As for germs with codimension 1, one can obtain a list of mono-germs using [RR91, Lemma 1.1] together with the classification of mono-germs between planes with small $\mathcal{A}$–codimensions due to Rieger [Rie87]. The codimension–1 multi-germs can be determined using an algorithm due to Cooper, Mond and Wik Atique [CMWA02, Theorem 5.22 and Remark 5.23]. We eventually obtain the following list of 1–parameter versal unfoldings of germs from $\mathbb{R}^4$ to $\mathbb{R}^2$:

Theorem 2.6. *Any versal 1–parameter unfolding of a germ $f : (\mathbb{R}^4, S) \to (\mathbb{R}^2, 0)$ is 1–equivalent to either a trivial unfolding of a stable germ or one of the six types of unfoldings of germs with codimension 1, whose discriminants are shown in Figure 2.*

The former three germs in Figure 2, named birth/death, fold/cusp merge and flip, are mono-germs, while the other ones are multi-germs. We can find local models of the three mono-germs in [Lek09], for example. As for the multi-germs, one can easily obtain local models of them following an algorithm in [CMWA02].

Proof of Theorem 2.5 (Sketch). As usual in differential topology, density is proved using some form of transversality. In this case, (a) can be rephrased as a transversality condition on some multi-jet extension of F relative to the parameter and the density follows from standard methods (see [Wal09, 2.1, 2.2][5] or [Chi77, III.4.]). Similarly, one can use transversality to show that the maps f_s generically avoid all multi-singularities of codimension at least two which proves the density of (b). Roughly, one has to show that the unions of orbits of multi-singularities of higher codimension have codimension at least two. More precisely, this boils down to

---

[5]Wall discusses $\mathcal{K}_e$–versal unfoldings, but a similar statement holds for $\mathcal{A}_e$–versal unfoldings.

estimating the codimension of algebraic sets $_s\Sigma \subset {_sJ^r}(4,2)$ for $s \leq 3$ and sufficiently large r as in [Mat70], where $_s\Sigma$ is the union of $\mathcal{A}$–orbits of s–fold r-jets of multi-singularities with codimension at least two. This calculation is not hard but tedious, thus for reasons of brevity we leave the details to the really interested reader. $\qquad\square$

2.3.1. *Stability of unfoldings.* Since we need to deal with families of homotopies in Section 3, we introduce a notion of unfoldings for unfoldings of germs and equivalence relation for them, which is a generalization of (r, s)–*equivalence* for unfoldings of function germs introduced by Wassermann [Was75].

Let $F : (\mathbb{R}^{n+k}, 0) \to (\mathbb{R}^{p+k}, 0)$ be a k–parameter unfolding of a germ f. An unfolding $\mathbf{F} : (\mathbb{R}^{n+k+l}, 0) \to (\mathbb{R}^{p+k+l}, 0)$ of f is called an (l–parameter) *unfolding* of F if the restriction $\mathbf{F}|_{\{0\}\times\mathbb{R}^{n+k}}$ is equal to F. Two l–parameter unfoldings $\mathbf{F}_0, \mathbf{F}_1 : (\mathbb{R}^{n+k+l}, 0) \to (\mathbb{R}^{p+k+l}, 0)$ of a k–parameter unfolding F of f are said to be (k, l)–*equivalent* if there exist germs of diffeomorphisms $\boldsymbol{\Phi}$, Φ, φ and ϕ on $(\mathbb{R}^{n+k+l}, 0)$, $(\mathbb{R}^{p+k+l}, 0)$, $(\mathbb{R}^{k+l}, 0)$ and $(\mathbb{R}^l, 0)$, respectively, such that the following diagram commutes:

$$
\begin{array}{ccccccc}
(\mathbb{R}^{n+k+l}, 0) & \xrightarrow{\ \mathbf{F}_0\ } & (\mathbb{R}^{p+k+l}, 0) & \xrightarrow{\ pr_{\mathbb{R}^{k+l}}\ } & (\mathbb{R}^{k+l}, 0) & \xrightarrow{\ pr_{\mathbb{R}^l}\ } & (\mathbb{R}^l, 0) \\
{\scriptstyle\boldsymbol{\Phi}}\downarrow & & {\scriptstyle\Phi}\downarrow & & {\scriptstyle\varphi}\downarrow & & {\scriptstyle\phi}\downarrow \\
(\mathbb{R}^{n+k+l}, 0) & \xrightarrow{\ \mathbf{F}_1\ } & (\mathbb{R}^{p+k+l}, 0) & \xrightarrow{\ pr_{\mathbb{R}^{k+l}}\ } & (\mathbb{R}^{k+l}, 0) & \xrightarrow{\ pr_{\mathbb{R}^l}\ } & (\mathbb{R}^l, 0).
\end{array}
\tag{5}
$$

An unfolding $\mathbf{F}$ of an unfolding F is said to be *trivial* if $\mathbf{F}$ is (k, l)–equivalent to the constant unfolding $\mathrm{id} \times F$. In this paper we need the following theorem:

Theorem 2.7 ([Mar82]). *If F is a versal unfolding of a germ f, then every unfolding of F is trivial as an unfolding of F (and thus also as an unfolding of f).*

Remark 2.8. Theorem 2.7 *does not* follow directly from uniqueness of versal unfoldings (i.e. [Mar82, p.190, Theorem 1.2]). Indeed, the uniqueness of versal unfoldings only guarantees that any l–parameter unfolding $\mathbf{F}$ of a k–parameter versal unfolding F of f is $(k+l)$–equivalent to the constant unfolding of F. However, the construction of diffeomorphisms in [Mar82, Ch. XIV] gives rise to (k, l)–equivalence between $\mathbf{F}$ and the constant unfolding of F.

Remark 2.9. Theorem 2.7 provides an extension of the parameterized normal forms for stable germs in families mentioned in Remark 2.4. Whenever a versal unfolding which has a normal form is embedded in a higher dimensional family of maps, then there is a normal form for the whole family given as a trivial product with the original normal form.

2.4. **Wrinkled fibrations.** A map $f \in \mathcal{S}^0(X, B)$ is called a *wrinkled fibration*[6] if f does not have definite folds and all fibers are connected. Note that a wrinkled fibration cannot have any definite cusps since these require definite folds. It is easy to see that wrinkled fibrations are open maps, in particular they are surjective. In the case $B = S^2$ a wrinkled fibration $f : X \to S^2$ over S^2 is said to be *simple* if the critical set $\mathcal{C}(f)$ is connected, the restriction $f|_{\mathcal{C}(f)}$ is injective, and f has at least one cusp.

As for the existence of wrinkled fibrations, Saeki [Sae06] first proved that any smooth map $f : X \to S^2$ is homotopic to a wrinkled fibration, and later Gay and Kirby [GK11a] generalized this result: they showed that $f : X \to B$ is homotopic to a wrinkled fibration if and only if $f_*\pi_1(X, x)$ has finite index in $\pi_1(B, f(x))$. They further proved that any two homotopic wrinkled fibrations can be connected

---

[6]Wrinkled fibrations are also known as *fiber-connected, indefinite Morse 2–functions* [GK11a].

by a stable homotopy f_t such that each fiber of f_t is connected for all t and f_t is a wrinkled fibration for all but finitely many values of t. The latter statement was also proved by Williams [Wil10] for $B = S^2$ who also shows that any map $f : X \to S^2$ is homotopic to a simple wrinkled fibration. Moreover, using this fact he introduced *surface diagrams* which describe closed 4–manifolds by sequences of simple closed curves in closed surfaces. The simple closed curves in such a diagram represent *vanishing cycles* of indefinite folds in a simple wrinkled fibration, which reflects configuration of singularities in the fibration. We will discuss vanishing cycles in detail in Section 4 and surface diagrams will be studied in Section 6.

3. Elimination of Cusps I: Merge Homotopies and Joining Curves

In this section we will prove our first main result which we restate for convenience.

Theorem A. *Let $f : X \to B$ be a stable map and $p, q \in X$ a pair of indefinite cusps. Denote by $\mathfrak{CM}_0(f; p, q)$ and $\mathbf{\Lambda}(f; p, q)$ the spaces of elementary cusp merge homotopies and framed joining curves for p and q. Then there are canonical bijections between $\pi_0\big(\mathfrak{CM}_0(f; p, q)\big)$ and $\pi_0\big(\mathbf{\Lambda}(f; p, q)\big)$.*

We begin by giving precise definitions of all involved objects. Recall from Section 2.1 that we consider $C^\infty(J \times X, B)$ where $J = [-1, 1]$ as the space of homotopies of maps from X to B and that we can think of a homotopy F as a 1–parameter family of maps $(f_s : X \to B)_{s \in J}$.[7] We will be interested in following special class of homotopies.

Definition 3.1 (Merge homotopies). A stable homotopy $F = (f_s)$ is called a *merge homotopy*, or simply a *merge*, if the following two conditions are satisfied:

(a) All but one multi-germs that appear in the maps f_s are stable except for a single beak-to-beak point in f_{s_0} for some $s_0 \neq \pm 1$.

(b) The numbers of cusps of f_{-1} and f_1 differ by two.

More precisely, we call F a *cusp merge* if the number of cusps decreases, and a *fold merge* otherwise. We denote by $\mathfrak{CM}$ and $\mathfrak{FM}$ for the subspaces of $C^\infty(J \times X, B)$ formed by the cusp and fold merges. Moreover, we write $\mathfrak{CM}(f)$ and $\mathfrak{CM}(f, g)$ for the cusp merges starting with f or going from f to g, respectively.

In what follows we will mostly be concerned with cusp merges and only briefly comment on fold merges in Section 5.4. Recall from Section 2.3 for any $F \in \mathfrak{CM}$ the passage of through the beak-to-beak point is governed by the *cusp merge model*

$$\mu_s(t, x, y, z) = \big(t, x^3 - 3(t^2 + s)x + y^2 \pm z^2\big), \quad s \in \mathbb{R}. \tag{6}$$

A direct calculation shows that the critical locus of μ_s is cut out by the equations $x^2 - t^2 = s$ and $y = z = 0$ and mostly consists of folds except for pairs of cusps for $s < 0$ located at $(\pm\sqrt{-s}, 0, 0, 0)$ which approach a beak-to-beak point in the origin for $s = 0$. In particular, for $s > 0$ all critical points are folds. Note that for $s \in [-1, 0]$, the cusps and the beak-to-beak trace out the line segment

$$L_0 = \{(\tau, 0, 0, 0) \mid \tau \in [-1, 1]\} \subset \mathbb{R}^4 \tag{7}$$

which will soon play an important role.

---

[7] Since we are assuming that X is compact, the map $s \mapsto f_s$ is continuous in the Whitney C^∞ topology. This is not true for non-compact X! In the non-compact situation, continuity is equivalent to the homotopy being constant outside of some compact subset of X.

3.1. **Elementary cusp merge homotopies.** While the cusp merge model gives a precise picture of cusp merge homotopies near their beak-to-beak singularities, we do not have any control over the behavior further away. And although all other multi-singularities are stable, they might move in complicated patterns. It is therefore more convenient to work with a special class of cusp merge homotopies whose effect is purely local.

Definition 3.2 (Elementary cusp merges). A cusp merge $F \in \mathfrak{CM}(f)$ is called *elementary* if there are open subsets $U \subset V \subset X \times [-1, \varepsilon]$ (for sufficiently small $\varepsilon > 0$) such that

  (i) the image of V under the projection to X is contained in a compact set K diffeomorphic to the 4–ball and F is constant in $X \setminus K$,
 (ii) there are coordinates on U and $\widetilde{F}(U)$, called *merge coordinates*, in which $\widetilde{F}$ coincides with the unfolding associated with the model map μ_s in (6),
(iii) there are diffeomorphisms Φ of $X \times [-1, \varepsilon]$ and Ψ of $B \times [-1, \varepsilon]$ with support in V and $\widetilde{F}(V)$, respecitvely, both preserving the parameter levels, such that on $\left(X \times [-1, \varepsilon] \right) \setminus U$ we have $\widetilde{F} = \Psi \circ (f_{-1} \times \mathrm{id}) \circ \Phi$.

We write $\mathfrak{CM}_0$, $\mathfrak{CM}_0(f)$ and $\mathfrak{CM}_0(f, g)$ for the spaces of elementary cusp merges with or without fixed endpoints.

In order to justify the terminology, we observe that an arbitrary cusp merge can be deformed into an isotopy followed by an elementary cusp merge and another isotopy. Recall that an isotopy is a homotopy of the form $f_s = \psi_s \circ f \circ \phi_s^{-1}$ where $(\phi_s, \psi_s) \in \mathrm{Diff}(X) \times \mathrm{Diff}(B)$ with $\phi_{-1} = \mathrm{id}_X$ and $\psi_{-1} = \mathrm{id}_B$; let $\mathfrak{Iso}(f, g)$ be the space of isotopies from f to g.

Lemma 3.3. *For any cusp merge $F \in \mathfrak{CM}(f, g)$ there are stable maps f', g' as well as $\mathcal{I} \in \mathfrak{Iso}(f, f')$, $F' \in \mathfrak{CM}_0(f', g')$, and $\mathcal{J} \in \mathfrak{Iso}(g', g)$ such that F can be deformed within $\mathfrak{CM}(f, g)$ to the concatenation $\mathcal{I} * F' * \mathcal{J}$.*

In principle, this result could be proved directly using similar methods as in Appendix I. However, since this would be rather lengthy and technical and Lemma 3.3 only serves as a motivation, we content ourselves with sketching a proof using some general results on stratifications due to Cerf [Cer70].

Proof of Lemma 3.3 (sketch). Let us write $\mathcal{S}^k$ for the subspace of $C^\infty(X, B)$ given by the maps of codimension k (as defined in Section 2.2.1). This is completely analogous to the stratifications on spaces of real valued functions used by Cerf [Cer70]. Conveniently, many of Cerf's general results about stratifications [Cer70, I.§1–3] carry over to the surface valued setting. In particular, the results about maps of codimension at most one in Sections 2.2 and 2.3 show that we can apply the "elementary path lemma"[8] [Cer70, p.20] to the stratification of $\mathcal{S}^0 \cup \mathcal{S}^1$ by the sets $(\mathcal{S}^0, \mathcal{S}^1)$. This reduces Lemma 3.3 to the following problem: for any $f \in \mathcal{S}^1$ with a beak-to-beak singularity we have to find an elementary cusp merge homotopy $F = (f_s)_{s \in J}$ with $f_0 = f$. But such a homotopy can easily be obtained using the normal form for beak-to-beak singularities, which allows to identify f with μ_0 near its beak-to-beak, and extending f to a homotopy using a truncated version of the cusp merge model satisfying $\mu_s = \mu_0$ outside of a neighborhood of the origin which can be obtained as in Example I.3.   □

3.2. **Joining curves.** We now focus on the second ingredient in Theorem A. Let $F = (f_s) \in \mathfrak{CM}_0(f)$ be an elementary cusp merge. If we choose merge coordinates for F, we can transfer the line segment $L_0 \subset \mathbb{R}^4$ defined in (7) to an

---

[8]Named "lemme des chemins élémentaires" by Cerf.

embedded arc $L_F \subset X$ with endpoints on two cusps of $f = f_{-1}$, say $p, q \in \mathcal{C}(f)$, which are eliminated by F. If we want to emphasize which cusps are eliminated we will use the more refined notation $\mathfrak{CM}_0(f; p, q)$. Since L_0 can be understood as the trace of cusps in the merge model, we see that L_F is filled out by cusps of the map f_s for $s < 0$ and the beak-to-beak of f_0. Since the positions of the critical points of the map f_s are clearly independent of the choice of merge coordinates, so is L_F. We will call L_F the *(unparameterized) joining curve* of F. Alternatively, we could describe L_F without reference to a model as follows. According to Remark 2.4, any cusp $c \in \mathcal{C}(f)$ will begin to evolve smoothly in X to a family of cusps $c_s \in \mathcal{C}(f_s)$ as the parameter s increases from -1 onward. For $s < 0$ no problems occur since each f_s is stable, but for $s = 0$ two things can happen. In either case, the curve c_s, $s < 0$, has a limit $c_0 \in \mathcal{C}(f_0)$. However, c_0 can either be a cusp, in which case c_s will continue to be a cusp for all s, or c_0 is the beak-to-beak point of f_0 and the curve cannot be continued for $s > 0$. The latter occurs for exactly the cusps $p, q \in \mathcal{C}(f)$ and L_F is traced out by the curves p_s and q_s for $s \in [-1, 0]$. In fact, looking at the model again, we see that

$$\lambda_F \colon [-1, 1] \longrightarrow X, \quad \lambda_F(\tau) = \begin{cases} p_{-\tau^2}, & \text{for } \tau \leq 0 \\ q_{-\tau^2}, & \text{for } \tau \geq 0 \end{cases} \tag{8}$$

constitutes a smooth parametrization of L_F; we therefore call λ_F the *parameterized joining curve* of F. Note that the definition of λ_F implicitly involves the non-canonical choice of p as the first cusp. We could just as well have chosen q and we would have obtained the reversed curve $\lambda_F^{\mathrm{rev}}(\tau) = \lambda_F(-\tau)$. However, this choice turns out to be irrelevant, see Remark 3.11.

Soon we will see that λ_F (or even L_F) almost contain enough information to reconstruct F up to deformation within $\mathfrak{CM}_0(f; p, q)$. What is missing is a suitable notion of framing. For simplicity we will only focus on *indefinite cusp merges*, by which we mean cusp merges involving indefinite cusps, since these are the most important from the point of view of wrinkled fibrations and broken Lefschetz fibrations. However, our arguments can be modified to handle general cusp merges, see Section 3.6. In the indefinite case, the framing will take the form of the line field spanned by the ∂_y vector field in some merge coordinates. Unfortunately, the independence of the framing of the choice of coordinates is not as easy to see as in the case of L_F or λ_F. We therefore embark on a small digression before we continue the study of elementary cusp merges.

3.2.1. *Tangent spaces of indefinite cusps and beak-to-beaks.* Suppose first that a map $f \colon X \to B$ has an indefinite cusp at $p \in X$. For concreteness, we fix *cusp coordinates* (t, x, y, z) around p and (u, v) around $f(p)$ such that f is given by $(t, x^3 + 3tx + y^2 - z^2)$. However, we will try to keep the discussion as intrinsic as possible. Let $K_p = \ker(df_p)$ and $Q_p = \mathrm{coker}(df_p)$. Our first observation is that the image of df_p, which is 1–dimensional, has a preferred orientation induced by the orientations of X and B. Intuitively, this is determined by the direction in which the cuspidal tip of the discriminant points in B. In the coordinates this direction agrees with the ray $D_p \subset T_{f(p)}B$ spanned by the coordinate vector field ∂_u. Moreover, the results in [Lev65, (4.2)] show that D_p is independent of the choice of coordinates and therefore well-defined. We will refer to D_p as the *direction of the cusp*. Following Levine, we say that a tangent vector $v \in T_p X$ *points downward* or *upward* at p if $df_p(v) \in D_p \setminus \{0\}$ or $df_p(-v) \in D_p \setminus \{0\}$, respectively. Next we note that the orientation of B together with the direction of the cusp induces an orientation on Q_p, which is also 1–dimensional, by extending a non-zero vector in D_p to an oriented basis of $T_{f(p)}B$. We therefore have a notion of positivity in Q_p. In coordinates, we can identify Q_p with the span of ∂_v. Lastly, we consider

the *intrinsic second derivative* of f which gives a symmetric bilinear map

$$\delta_p^2 f \colon K_p \times K_p \longrightarrow Q_p$$

and can be considered as a generalization of the Hessian for functions. In fact, in coordinates K_p is spanned by ∂_x, ∂_y, and ∂_z and in this basis $\delta_p^2 f$ appears as the Hessian of the function $x^3 + y^2 - z^2$ using the identification $Q_p \cong \mathbb{R}\partial_v$. We can therefore consider $\delta_p^2 f$ as a symmetric bilinear form which provides a decomposition of K_p into three disjoint sectors K_p^+, K_p^-, and K_p^0 defined as the set of vectors $v \in K_p$ for which $\delta_p^2 f(v,v)$ is positive, negative, and zero, respectively. From the coordinate description of $\delta_p^2 f$ we see that $\partial_x \in K_p^0$, $\partial_y \in K_p^+$, and $\partial_z \in K_p^-$. Moreover, K_p^+ and K_p^- each have two connected components which are convex and interchanged by the multiplication with any negative number. In particular, the image of $K_p^\pm$ in $\mathbb{P}(K_p)$, henceforth denoted by $\mathbb{P}(K_p^\pm)$, is contractible.

The above discussion of $\delta_p^2 f$ goes through verbatim if f has a beak-to-beak at p and thus has the local model $(t, x^3 - 3t^2 x + y^2 - z^2)$. However, this time there is no preferred orientation for Q_p but $\delta_p^2 f$ is still defined and singles out a well defined null sector K_p^0 whose complement in K_p has four contractible components which are paired by comparing the sign of $\delta_p^2 f$ with respect to a fixed orientation of Q_p.

3.2.2. *Abstract joining curves and framings.* We now give an abstract definition of joining curves which is essentially due to Levine [Lev65, (4.4)]. However some adjustments were necessary since we are working with stable maps while Levine allows arbitrary multi-singularities made of folds and cusps. Throughout this subsection we consider a stable map $f \colon X \to B$ and a pair of cusps $p, q \in \mathcal{C}(f)$.

Definition 3.4 (Joining curves). A curve $\lambda \colon [-1,1] \to X$ with $\lambda(-1) = p$ and $\lambda(1) = q$ is called a *joining curve* for f (from p to q) if
  (i) both λ and $f \circ \lambda$ are smooth embeddings,
 (ii) λ points downward at p and upward at q, and
(iii) the interior of λ does not meet any singular fibers.
The joining curves for f form a subspace of $C^\infty([-1,1], X)$ which we denote by $\Lambda(f)$, and by $\Lambda(f; p, q)$ indicates fixed cusps at the endpoints.

It is easy to see that $\Lambda(f; p, q)$ is non-empty if and only if $f(p)$ and $f(q)$ can be connected by an arc of regular values, say $R \subset B$, and p and q are in the same connected component of $f^{-1}(R)$. Moreover, Levine shows that if $\Lambda(f; p, q)$ is non-empty and if p and q form a "matching pair" (see [Lev65, p.284]), then they can be eliminated by a cusp merge homotopy. We will not explain the matching pair condition, we only observe that it is satisfied for any two indefinite cusps so that the existence of a joining curve guarantees that the pair can be eliminated.

Next we have to discuss a notion of framings for joining curves which are slightly different for definite and indefinite cusps. Again, we only treat the indefinite case and refer to Section 3.6 for some remarks about other cases. Consider the subspace of TX given by $\ker(df) = \cup_p \ker(df|_p)$. Although $\ker(df)$ is not a vector bundle over X it is still meaningful to speak of sections and pullbacks. Moreover, since the fibers are vector spaces (albeit of varying ranks) we can take the fiber-wise projectivization which we denote by $\boldsymbol{X} = \mathbb{P}(\ker(df)) \subset \mathbb{P}(TX)$ and we let $\pi \colon \boldsymbol{X} \to X$ be the obvious projection.

Definition 3.5 (Framed joining curves). Let $\lambda \in \Lambda(f; p, q)$ be a joining curve. A *framing* for λ is a map $\boldsymbol{\lambda} \colon J \to \boldsymbol{X}$ such that $\pi \circ \boldsymbol{\lambda} = \lambda$ and $\boldsymbol{\lambda}(\pm 1) \in \mathbb{P}\big(K_{\lambda(\pm 1)}^\pm\big)$. Any map $\boldsymbol{\lambda} \colon J \to \mathbb{P}(\ker(df))$ with the latter property and $\pi \circ \boldsymbol{\lambda} \in \Lambda(f)$ is called a *framed joining curve*. We denote by $\boldsymbol{\Lambda}(f; p, q)$ and $\boldsymbol{\Lambda}(f)$ the spaces of framed joining curves with or without fixed endpoints.

In what follows it will be understood that whenever we discuss a framed joining curve $\boldsymbol{\lambda}$, the underlying joining curve is denoted by $\lambda = \pi \circ \boldsymbol{\lambda}$.

Remark 3.6. With the arguments in Section 5 in mind, we take a moment to unravel Definition 3.5. Let $\lambda \in \Lambda(f)$ be a joining curve and let Σ_τ be the fiber of f containing $\lambda(\tau)$, that is, $\Sigma_\tau = f^{-1}\big(f(\lambda(\tau))\big)$. For $\tau \neq \pm 1$, Σ_τ is a smooth submanifold of X whose tangent bundle is the restriction of $\ker(df)$ to Σ_τ. Therefore a framing of λ amounts to the choice of lines in $T_{\lambda(\tau)}\Sigma_\tau$ with limits in the negative sector K_p^- over the negative end $p = \lambda(-1)$ and in K_q^+ over $q = \lambda(1)$.

3.3. From $\mathfrak{CM}_0(f)$ to $\Lambda(f)$. Now that we have proper definitions for the objects in the statement of Theorem A, we have to understand how they are related. A first step in this direction was already taken in the beginning of Section 3.2 where we discussed the parameterized and unparameterized joining curves λ_F and L_F of an elementary cusp merge homotopy $F \in \mathfrak{CM}_0(f)$. We leave the following easy verifications to the reader.

Lemma 3.7. *The curve λ_F defined in* (8) *is a joining curve for f in the sense of Definition 3.4. Moreover, in merge coordinates for F the line field spanned by ∂_z constitutes a framing for λ_F.*

As mentioned before, there is no canonical choice of merge coordinates and therefore no canonical framing. However, we we can get around this issue as follows.

Lemma 3.8. *There is a connected space of preferred framings of λ_F induced by F which contains the framings coming from merge coordinates. In particular, an elementary cusp merge homotopy $F \in \mathfrak{CM}_0(f; p, q)$ gives rise to a well defined element $\boldsymbol{\lambda}_F \in \pi_0\big(\Lambda(f; p, q)\big)$.*

Proof of Lemma 3.8. For simplicity, we write $\lambda = \lambda_F$. As usual, we consider F as a family (f_s) of maps. The discussion of the tangent spaces of cusps and beak-to-beaks in Section 3.2.1 results in a collection of 3–dimensional vector spaces

$$K_\tau = \ker\big(d(f_{-|\tau|})_{\lambda(\tau)}\big) \subset T_{\lambda(\tau)}X. \tag{9}$$

In merge coordinates these are spanned by the vector fields ∂_x, ∂_y and ∂_z and therefore fit together to a smooth vector sub-bundle $K \subset \lambda^* TX$. The crucial observation is that for $\tau \neq 0$ the sectors $K_\tau^\pm$ and $K_{-\tau}^\mp$ defined by the cusps of $f_{-|\tau|}$ as in Section 3.2.1 limit to the same sector of K_0 as τ approaches zero from either side. This can again be checked in the model. The change of signs is caused by the fact that λ points downward at one cusp and upward at the other. As a consequence, we obtain a sub-bundle $\widetilde{\mathcal{F}}_\lambda$ of the projectivization $\mathbb{P}(K)$ such that for $\tau > 0$ the fibers of $\widetilde{\mathcal{F}}_\lambda$ over τ and $-\tau$ respectively agree with $\mathbb{P}(K_\tau^+)$ and $\mathbb{P}(K_{-\tau}^-)$. Note that $\widetilde{\mathcal{F}}_\lambda$ has contractible fibers. In order to make the connection to framings we consider the intersection of K with the kernel of df_{-1} along λ. First we observe that K and $\lambda^* \ker(df_{-1})$ agree over the end-points of λ. Over the interior, another local calculation shows that $\ker(df_{-1})$ is spanned by ∂_y and ∂_z along λ, thus $\lambda^* \ker(df_{-1})$ forms a subset of K. Finally, after passing to projectivizations we arrive at a subset

$$\mathcal{F}_\lambda = \widetilde{\mathcal{F}}_\lambda \cap \mathbb{P}(\lambda^* \ker(df_{-1})) \subset \mathbb{P}(K).$$

The proof is finished by the following consequences of the (admittedly somewhat convoluted) definitions and the arguments in Section 3.2.1: the map $\mathcal{F}_\lambda \to [-1, 1]$ has contractible fibers, sections of $\mathcal{F}_\lambda$ can be considered as framings of λ, and any choice of merge coordinates give rise to a section of $\mathcal{F}_\lambda$ via ∂_z. $\qquad\square$

We arrive at the main conclusion of the present subsection.

Proposition 3.9. *The assignment $F \mapsto \boldsymbol{\lambda}_F$ descends to a map*

$$\pi_0\big(\mathfrak{CM}_0(f;p,q)\big) \longrightarrow \pi_0\big(\boldsymbol{\Lambda}(f;p,q)\big).$$

Proof. We have to show that the $\boldsymbol{\lambda}_F$ does not change in smooth families. So let $\boldsymbol{F} = (F_r)_{r \in [0,1]}$ be a smooth 1–parameter family where $F_r \in \mathfrak{CM}_0(f;p,q)$ and let λ_r be the joining curve of F_r. Since each F_r is stable as a homotopy, it follows from the Remark 2.9 that the family $(\lambda_r)_{r \in [0,1]}$ depends smoothly on r. Moreover, the discussion of framings in the proof of Lemma 3.8 is only notationally more complicated in the presence of a parameter. $\square$

3.4. **From $\boldsymbol{\Lambda}(f)$ to $\mathfrak{CM}_0(f)$.** So far we have been able to extract framed joining curves from elementary cusp merge homotopies. Now we try to go back and construct homotopies with prescribed joining curves and framings. As mentioned earlier, the construction is essentially due to Levine [Lev65].

Lemma 3.10 (Parametric Levine construction). *Let $\boldsymbol{\lambda}_r \in \boldsymbol{\Lambda}(f;p,q)$ be a smooth family of framed joining curves. Then there exists a smooth family $F_r \in \mathfrak{CM}_0(f;p,q)$ of elementary cusp merges such that $\boldsymbol{\lambda}_r$ is the framed joining curve of F_r.*

Proof of Lemma 3.10. We first consider a single framed joining curve $\boldsymbol{\lambda}$, that is, the case of a constant homotopy. Let $L \subset X$ denote the image of the underlying joining curve λ and let $L_0 \subset \mathbb{R}^4$ as in (7). The proof has two main steps.

> **Step 1:** Identify f with the initial map μ_{-1} of the cusp merge model (6) in neighborhoods of L and L_0.
>
> **Step 2:** For any neighborhood νL_0 of L_0 in $\mathbb{R}^4$ construct an elementary cusp merge homotopy with initial map μ_{-1} which is constant outside of νL_0.

Once this is done, we can promote the homotopy constructed in the second step to an elementary cusp merge homotopy from f using the identification obtained in the first step in the obvious way. Let us call this homotopy

$$F_{\boldsymbol{\lambda}} \in \mathfrak{CM}_0(f;p,q).$$

Step 1 is treated carefully in [Lev65, (4.6),(4.8)] to which we refer for more details. The basic observation is that the map $\mu_{-1}\colon \mathbb{R}^4 \to \mathbb{R}^2$ can be divided into three parts: an inner part where $|t| < 1$ consisting of a tubular neighborhood of L_0 fibered over a tubular neighborhood of $\mu_{-1}(L_0)$, and two outer parts where $|t| \geq 1 - \epsilon$ which are both easily matched with the cusp model. So the idea is to implant these structures into a neighborhood of L in X. To begin with, we choose cusp coordinates around the points p and q. We can assume that the ∂_z vector fields of both coordinate patches take values in the tangent lines specified by $\boldsymbol{\lambda}$ near the ends of L. Let us write $\partial_z^{(p)}$ and $\partial_z^{(q)}$ in order to distinguish between the two. Next we use a partition of unity argument to extend $\partial_z^{(p)}$ to a vector field ν_z on X which takes values in $\boldsymbol{\lambda}$ along L and agrees with $\pm\partial_z^{(q)}$ in the other coordinate patch. If the sign is negative, then we change the coordinates near q by the symmetry $(t,x,y,z) \mapsto (t,x,-y,-z)$. After this is done, we can also find vector fields ν_x and ν_y on X that agree with the ∂_x and ∂_y fields in the coordinate patches and are point-wise linearly independent near L. Lastly, we choose a vector field T on X that agrees with $\partial_t^{(p)}$ near p and with $-\partial_t^{(q)}$ near q and has L as a flow line. (In particular, they form an honest *normal framing* for L as a submanifold of X.) If the extensions T, ν_x, ν_y, and ν_z were chosen carefully enough, then we can use their flows to connect the two cusp coordinate systems to a single coordinate patch containing L in which f appears as μ_{-1} and L corresponds to L_0. At this point, we already have constructed enough of $F_{\boldsymbol{\lambda}}$ to form its joining curve whose image agrees with L by construction, and possibly after rescaling the coordinates in the t–direction we can assume that the parametrization agrees with λ as well.

As for Step 2, we refer to Example I.3. We can obtain an elementary cusp merge homotopy with the desired property by taking sufficiently small K_ε and a family $S_{s,t}$ sufficiently close to $(t, 0, 0, 0)$ as in Example I.3 so that the support of the resulting homotopy is contained in the given neighborhood νL_0.

Lastly, in order to treat the case of non-constant homotopies of joining curves, it is enough to have a parametric version of the first step. But a closer investigation of Levine's arguments shows that the essential ingredients are the normal form for a cusp and the tubular neighborhood theorem which both have parametric versions. $\qquad\square$

Remark 3.11. It turns out that the Levine construction applied to $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}(f; p, q)$ and its reverse $\boldsymbol{\lambda}^{\mathrm{rev}}(\tau) = \boldsymbol{\lambda}(-\tau)$, which lies in $\boldsymbol{\Lambda}(f; q, p)$ give rise to the same cusp merge homotopy. The reason is that the map μ_{-1} admits a symmetry that interchanges the two cusps at $(\pm 1, 0, 0, 0)$. For example, one can take the linear diffeomorphisms $(t, x, y, z) \mapsto (-t, -x, z, y)$ on $\mathbb{R}^4$ and $(u, v) \mapsto (-u, -v)$ on $\mathbb{R}^2$.

Remark 3.12. Note that Levine takes a slightly different approach to the second step in the above proof (see [Lev65, (4.9)]). Unfortunately, his version does not lead to elementary cusp merge homotopies in the sense of Definition 3.2, as it results in a slightly more complicated model than (μ_s) from (6). Since we found the model (μ_s) much more convenient for computations, we were forced to develop the machinery in Appendix I.

The main goal of this section is to establish the following.

Proposition 3.13. *The parametric Levine construction induces a map*

$$\pi_0\big(\boldsymbol{\Lambda}(f; p, q)\big) \longrightarrow \pi_0\big(\mathfrak{CM}_0(f; p, q)\big)$$

Proof. The Levine construction involves several choices and we have to make sure that for each $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}(f; p, q)$ we get a well defined element of $\pi_0\big(\mathfrak{CM}_0(f; p, q)\big)$. Let $F, F' \in \mathfrak{CM}_0(f; p, q)$ be two elementary cusp merges resulting from different choices in the Levine construction. The ambiguities in the first step of the construction affect the behavior of the homotopies near the image of $\boldsymbol{\lambda}$, while the ambiguity in the second step concerns the extension to all of X. The first ambiguities can be eliminated using our results about the symmetries of cusps in Proposition II.5 combined with the standard fact that a parameterized tubular neighborhood of a submanifold is determined by a normal framing. These results show that we can deform F' so that it agrees with F near the image of $\boldsymbol{\lambda}$ through homotopies arising from Levine constructions. Instead of dealing with the second type ambiguities directly, we observe that the desired result follows from Lemma 3.14 below. $\qquad\square$

Lemma 3.14. *Suppose that $F, F' \in \mathfrak{CM}_0(f)$ agree near their joining curves. Then they can be deformed into each other within $\mathfrak{CM}_0(f)$.*

Proof. Let $L \subset X$ be the image of the common joining curves of F and F'. By assumption we can find nested neighborhoods $U \subset U'$ of L such that F and F' are constant in U' and U is contained in a common set of merge coordinates of F and F'. We take other neighborhoods $V \subset U$ and $V' \supset U'$ of L and U', respectively. For convenience, we assume that $U, f(U), V$ and $f(V)$ are all open balls with compact closures and smooth boundaries. As we construct 1–parameter families of diffeomorphisms in the proof of Theorem I.1 we can take codimension 0 compact submanifolds $W \subset [-1, \varepsilon] \times V$ and $K \subset f(V)$ which contain $[-1, \varepsilon] \times L$ and $f(L)$, respectively, an open neighborhood $W' \subset [-1, \varepsilon] \times V$ of $\partial W \setminus (\{-1, \varepsilon\} \times V)$, level preserving diffeomorphisms Φ, Φ' of $[-1, \varepsilon] \times X$ and Ψ, Ψ' of $[-1, \varepsilon] \times B$ such that

- $\widetilde{F} = \Psi \circ (\mathrm{id} \times f) \circ \Phi^{-1}$ and $\widetilde{F}' = \Psi' \circ (\mathrm{id} \times f) \circ \Phi'^{-1}$ on $W' \cup W^c$,

- Supports of Φ and Φ' are contained in $V' \times [-1, \varepsilon]$, while those of Ψ and Ψ' are contained in $[-1, \varepsilon] \times \nu \partial K$ where $\nu \partial K$ is a neighborhood of ∂K,
- $\Phi = \Phi'$ on $V \times [-1, \varepsilon]$ and $\Psi = \Psi'$ on $f(V) \times [-1, \varepsilon]$.

We can consider the composition $(\Phi^{-1}\Phi', \Psi^{-1}\Psi')$ as a map from $[-1, \varepsilon]$ into the space of pairs $(\phi, \psi) \in \mathrm{Diff}(X) \times \mathrm{Diff}(B)$ such that $\phi|_V$ and $\psi|_{f(V)}$ are the identity maps and $(\psi \circ f \circ \phi^{-1})|_{X \setminus V} = f|_{X \setminus V}$ that sends -1 to $(\mathrm{id}_X, \mathrm{id}_B)$. In other words, $(\Phi^{-1}\Phi', \Psi^{-1}\Psi')$ corresponds to a path in the space just mentioned based at $(\mathrm{id}_X, \mathrm{id}_B)$. Since we do not have to keep the endpoint of this path fixed, we can deform it to the constant path. But this means that we can find a 1–parameter family of pairs (Φ_r, Ψ_r), $r \in [0, 1]$, that connects (Φ, Ψ) to (Φ', Ψ') with the properties that Φ_r and Ψ_r are independent of r in $[-1, \varepsilon] \times V$ and $[-1, \varepsilon] \times f(V)$ and the support of Φ_r (resp. Ψ_r) is contained in $V' \times [-1, \varepsilon]$ (resp. $\nu \partial K \times [-1, \varepsilon]$). Using this we obtain a 1–parameter family of homotopies F_r with unfoldings

$$\widetilde{F}_r = \begin{cases} \Psi_r \circ (\mathrm{id} \times f) \circ \Phi_r^{-1} & \text{on } W' \cup W^c \\ \widetilde{F} = \widetilde{F}' & \text{on } W \end{cases}$$

which provides a deformation of $F = F_0$ into $F' = F_1$. It remains to show that each F_r is an elementary cusp merge homotopy. But this is clear from the construction. Indeed, on W and $\widetilde{F}(W)$ the common merge coordinates for F and F' can be used for each F_r. Moreover, Φ_r and Ψ_r, which are supported on $V' \times [-1, \varepsilon]$ and $\nu \partial K \times [-1, \varepsilon]$, satisfy the second condition in the definition of elementary cusp merges.   □

3.5. **The proof of Theorem A.** According to Propositions 3.9 and 3.13, we have natural maps between the sets $\pi_0\big(\mathfrak{CM}_0(f)\big)$ and $\pi_0\big(\Lambda(f)\big)$. We will now show that they are mutually inverse. We briefly recall the constructions:

- Given $\boldsymbol{\lambda} \in \Lambda(f)$ we can find coordinates around the image of λ in which f is represented by μ_{-1} and $\boldsymbol{\lambda}$ corresponds λ_0, that is, $\lambda_0(\tau) = (\tau, 0, 0, 0) \in \mathbb{R}^4$ framed by the lines spanned by ∂_y. A compactly supported version of the merge model then gives rise to $F_{\boldsymbol{\lambda}} \in \mathfrak{CM}_0(f)$.
- Given $F \in \mathfrak{CM}_0(f)$ we choose merge coordinates for f and obtain $\boldsymbol{\lambda}_F \in \Lambda(f)$ by pulling back $\boldsymbol{\lambda}_0$ to X.

From Lemma 3.10 we already know that $\boldsymbol{\lambda}_{F_{\boldsymbol{\lambda}}}$ is equal to $\boldsymbol{\lambda}$. So we only have to show that $F_{\boldsymbol{\lambda}_F}$ and F represent the same element of $\pi_0\big(\mathfrak{CM}_0(f)\big)$. But this follows from the same arguments as in the proof of Proposition 3.13. Indeed, since $F_{\boldsymbol{\lambda}_F}$ and F have the same framed joining curve, we can deform them to agree near the joining curve and then apply Lemma 3.14. This finishes the proof of Theorem A.

3.6. **More general cusp merges.** While we have stated and proved Theorem A only for the case of indefinite cusp merges in dimensions $(4, 2)$, our arguments can easily be modified to obtain an analogous result for general cusp merges in dimensions $(n, 2)$ with $n \geq 2$. Essentially the only difference arises in the discussion of framings of joining curves. Indeed, the general cusp merge model in dimensions $(n, 2)$ takes the form

$$(s; t, x, y_1, \ldots, y_k, z_1, \ldots, z_l) \mapsto (t, x^3 - 3(t^2 + s)x + \sum_i y_i^2 - \sum_i z_i^2)$$

where $k + l = n - 2$ and the framing has to keep track of $\partial_{y_1}, \ldots, \partial_{y_k}, \partial_{z_1}, \ldots, \partial_{z_l}$. With an appropriate notion of framings, the parametric Levine construction and all other ingredients in the proof of Theorem A are available in the general case.

4. Parallel Transport and Vanishing Sets

Before we can address Theorem B we have to develop a theory of parallel transport. We start in the setting of general smooth maps and then specialize to wrinkled fibrations on 4–manifolds. Besides the relevance of Theorem B, our goal is to provide a general framework for discussing vanishing sets of wrinkled fibrations which, as far as we know, has not been available.

4.1. **Connections for smooth maps.** For the moment, we consider an arbitrary smooth map $f\colon N \to P$ between connected manifolds of dimensions $n \geq p$. We need a tool to compare different fibers of f. The following definition should be thought of as a hybrid between connections in fiber bundles and gradients of (Morse) functions.

Definition 4.1 (Connections). A *connection* for a smooth map $f\colon N \to P$ is a subset $\mathcal{H} \subset TN$ obtained as the point wise orthogonal complement of $\ker(df)$ with respect to some Riemannian metric on N.

Observe that if f is a fiber bundle, then we recover the usual notion of connection, while if f is a real-valued function, then a connection is simply the point wise span of the gradient of f with respect to some metric. Recall that the central idea in the construction of parallel transport in fiber bundles is to consider lifts of curves in P to curves in X via f. We will follow a similar procedure; for simplicity, we restrict our attention to a special class of curves in P.

Definition 4.2 (Reference arcs and $\mathcal{H}$–lifts). Let $f\colon N \to P$ be a smooth map.
(a) A *reference arc* for f is an embedded arc $\gamma\colon [0,1] \to P$ such that $\gamma(0)$ and $\gamma(1)$ are regular values of f.
(b) Let $\mathcal{H}$ be a connection for f. A map $\tilde{\gamma}\colon O \to N$ is called an $\mathcal{H}$–*lift* of a curve $\gamma\colon [0,1] \to P$ if $O \subset [0,1]$ is a relatively open subinterval, $f \circ \tilde{\gamma} = \gamma|_O$, and the velocity vectors of $\tilde{\gamma}$ are contained in $\mathcal{H}$.

We will occasionally blur the notational distinction between a reference arc γ and its image in P. For example, we usually write $f^{-1}(\gamma)$ instead of $f^{-1}(\gamma([0,1]))$. Moreover, we will often abbreviate the fibers over γ by $\Sigma_t = f^{-1}(\gamma(t))$. The following technical result is the key ingredient for the parallel transport construction.

Proposition 4.3. *Let $f\colon N \to P$ be a proper map equipped with a connection $\mathcal{H}$. Furthermore, let $\gamma\colon [0,1] \to P$ be a reference arc for f and $p \in \Sigma_t$ a regular point for some fixed $t \in [0,1]$.*
 (i) There exists a unique $\mathcal{H}$–lift of γ, denoted by
$$\tilde{\gamma}_{t,p}^{\mathcal{H}}\colon O_{t,p} \longrightarrow N, \quad O_{t,p} \subset [0,1],$$
 where $O_{t,p}$ is a relatively open interval containing t such that $\tilde{\gamma}_{t,p}^{\mathcal{H}}(t) = p$ and all other $\mathcal{H}$–lifts of γ with this property are restrictions of $\tilde{\gamma}_{t,p}^{\mathcal{H}}$.
 (ii) If $\mathcal{H}$ depends smoothly on some auxiliary parameters, then so does $\tilde{\gamma}_{t,p}^{\mathcal{H}}$.
 (iii) If each fiber of f along γ contains at most finitely many critical points of f, then $\tilde{\gamma}_{t,p}^{\mathcal{H}}$ limits to a critical point on each open end of $O_{t,p}$.

Proof of Proposition 4.3. This is well known for submersions and the usual proofs of (i) and (ii) are easily adapted to take critical points into account, so we shall be brief. Since $f|_{M \setminus \mathcal{C}_f}$ is a submersion (albeit non-proper), the set $Y = f^{-1}(\gamma) \setminus \mathcal{C}_f$ is a non-compact smooth $(n - p + 1)$–manifold with boundary $\Sigma_0 \amalg \Sigma_1$. For $q \in \Sigma_t$ let $\Gamma(q) \in \mathcal{H}$ be the unique element that is mapped to the velocity vector $\dot{\gamma}(t)$ under df. It is easy to see that Γ is tangent to Y and can thus be considered as

a vector field on Y. The claims (i) and (ii) now follow from standard properties of the flow of Γ. In order to prove (iii), let $\tilde{\gamma}$ be some $\mathcal{H}$–lift of γ defined on an open interval $(a, b) \subset [0, 1]$. It suffices to show that $\tilde{\gamma}$ can be extended to the closed interval $[a, b]$. Since the arguments for a and b are virtually the same, we only focus on b. Let $\{b_n\} \subset (a, b)$ be a sequence that converges to b. Since $f^{-1}(\gamma)$ is compact and γ is continuous, the sequence $\tilde{\gamma}(b_n)$ has an accumulation point $\tilde{b} \in N$ with $f(\tilde{b}) = \gamma(b)$; in other words, we have $\tilde{b} \in \Sigma_b$. Note that if $\tilde{b}$ is a regular point of f, then $\tilde{\gamma}$ must be the restriction of the maximal lift $\tilde{\gamma}^{\mathcal{H}}_{b,\tilde{b}}$ which obviously extends $\tilde{\gamma}$ to b. On the other hand, if $\tilde{b}$ is a critical point of f, then the same reasoning provides an accumulation point of the sequence $\{\tilde{\gamma}(b_n)\}$ and we only have to show that it is unique. We argue by contradiction and assume that there are two or more accumulation points. The above arguments show that all accumulation points have to be critical points of f in the fiber Σ_b which are isolated by assumption. Let $U \subset \Sigma_b$ be the union of pairwise disjoint open neighborhoods U_i of all critical points in Σ_b. Since $\tilde{\gamma}$ must enter and leave some of U_i infinitely many times, we can find a subsequence of $\{c_k\} \subset \{b_n\}$ with $c_k \to b$ such that $\tilde{\gamma}(c_k) \in \Sigma_b \setminus U$. Now, $\Sigma_b \setminus U$ is also compact so that $\{\tilde{g}(c_k)\}$ must have an accumulation point. But this would have to be a regular point of f which is impossible. $\qquad\square$

Note that the finiteness assumption in (iii) of Proposition 4.3 is automatically satisfied if f is stable. In particular, the conclusion always holds for wrinkled fibrations.

4.2. **Vanishing sets and parallel transport.** We continue with the notation of Proposition 4.3. For a critical point $q \in f^{-1}(\gamma) \cap \mathcal{C}(f)$ we say that $\tilde{\gamma}^{\mathcal{H}}_{t,p}$ *runs into* q (or *emerges from* q) if its left (or right) limit is q. We can now define what we call the *vanishing sets* of the triple $(\gamma, \mathcal{H}; q)$ as

$$V_0(\gamma, \mathcal{H}; q) = \left\{ p \in \Sigma_0 \,\middle|\, \tilde{\gamma}^{\mathcal{H}}_{0,p} \text{ runs into } q \right\} \subset \Sigma_0$$
$$V_1(\gamma, \mathcal{H}; q) = \left\{ p \in \Sigma_1 \,\middle|\, \tilde{\gamma}^{\mathcal{H}}_{1,p} \text{ emerges from } q \right\} \subset \Sigma_1.$$

This is reminiscent of the theory stable and unstable manifolds of gradients of critical points of functions which is already a delicate subject for degenerate critical points (see Example 4.4 below). However, our situation can be even more complicated since $f^{-1}(\gamma)$ does not have to be a manifold. So we should be alarmed and proceed with caution. In particular, the following example shows that the vanishing sets can exhibit a delicate dependence on the choice of connection.

Example 4.4. In the case of a real-valued function the vanishing sets can be considered as the intersections of stable and unstable manifolds associated with critical points with level sets. In particular, for a reference arc whose preimage contains a single non-degenerate critical point we recover the well known ascending and descending spheres prominently featured in Morse theory and one can show that they are independent of the choice of connection up to ambient isotopies within the levels. However, for degenerate critical points the topological type of the stable and unstable manifolds can depend on the choice of metric, it can even jump in 1–parameter families of metrics (see [Tak71] for an example).

Continuing with the previous discussion, let us write $V_0 \subset \Sigma_0$ and $V_1 \subset \Sigma_1$ for the unions of $V_i(\gamma, \mathcal{H}; q)$ over all $q \in f^{-1}(\gamma) \cap \mathcal{C}(f)$. Of course, the unions V_i depend even more drastically on $\mathcal{H}$ than the vanishing sets alone as is easily seen by considering two critical points of a Morse function. Regardless, we can define the *parallel transport* along γ with respect to a fixed connection $\mathcal{H}$ by

$$\Pi^{\mathcal{H}}_\gamma \colon \Sigma_0 \setminus V_0 \to \Sigma_1 \setminus V_1, \qquad p \mapsto \tilde{\gamma}^{\mathcal{H}}_{0,p}(1).$$

Note that this is essentially the same construction as in the case of fiber bundles, the price to pay for the presence of critical points is that the parallel transport is only partially defined. The following observations are either obvious or follow from the standard theory of ordinary differential equations.

- $\Pi_\gamma^{\mathcal{H}}$ is a diffeomorphism.
- If two reference arcs γ and γ' have the same image, then they give rise to the same vanishing sets. Furthermore, $\Pi_\gamma^{\mathcal{H}}$ is equal to $\Pi_{\gamma'}^{\mathcal{H}}$.
- If γ is an arc of regular values, then $\Pi_\gamma^{\mathcal{H}} \colon \Sigma_0 \to \Sigma_1$ is everywhere defined and its isotopy class is independent of $\mathcal{H}$ and depends only on the homotopy class of γ relative to its endpoints in the connected component of $P \backslash f(\mathcal{C}(f))$ containing γ.
- The whole discussion generalizes immersed, piecewise smooth, and closed curves. For parallel transport along a closed curve of regular values we use the term *monodromy*.

Remark 4.5. As an addendum to the third statement, we note that if γ is an arc of regular values, then for any diffeomorphism $\phi \in \mathrm{Diff}(\Sigma)$ that is isotopic to the identity one can find a connection $\mathcal{H}'$ which agrees with $\mathcal{H}$ outside an arbitrarily small neighborhood of $f^{-1}(\gamma)$ such that $\Pi_\gamma^{\mathcal{H}'} = \Pi_\gamma^{\mathcal{H}} \circ \phi$. (This follows as in the proof of [Nic11, Lemma 2.28], for example.) Applied to regular parts of general reference arcs this provides useful flexibility for rearranging the vanishing sets.

4.3. **Vanishing Sets in Wrinkled Fibrations.** We now leave the general theory behind and focus on wrinkled fibrations on 4–manifolds as defined in Section 2.4. Recall that these are stable maps $f \colon X \to B$ from a closed 4–manifold X to a closed surface B which have no definite critical points and whose fibers are connected. The discriminant $\mathcal{D}(f) = f(\mathcal{C}(f))$ consists of finitely many cusp and double fold values, smooth arcs of simple fold values between them, and possibly also embedded circles of simple fold values. We refer to the connected components of $B \setminus \mathcal{D}(f)$ as the *regions* of f. The different fibers over a fixed region are all diffeomorphic and our goal is to understand the relation of fibers over different regions using parallel transport. Clearly, we can perturb an arbitrary reference arc so that it avoids the cusps and double folds and only intersects $\mathcal{D}(f)$ transversely in simple fold values. Moreover, any such reference arc can be subdivided into parts which contain only one critical value. We study these first.

4.3.1. *Parallel transport across folds.* A reference arc $\gamma \colon [0,1] \to B$ for a wrinkled fibration $f \colon X \to B$ is called a *fold reference arc* if it meets the discriminant of f in a single fold value and is transverse to the corresponding fold arc. The vanishing sets and parallel transport along a fold reference γ are easily understood. Indeed, it follows from the indefinite fold model that the preimage $f^{-1}(\gamma)$ is a smooth manifold and that $\gamma^{-1} \circ f \colon f^{-1}(\gamma) \to [0,1]$ is a Morse function with a single critical point of index 1 or 2 depending on the direction in which γ crosses the fold arc. For convenience, we write $\Sigma = \Sigma_0$ and $\Sigma' = \Sigma_1$ and we always assume that the index is 2 so that, according to Example 4.4, the vanishing sets with respect to any connection $\mathcal{H}$ are a simple closed curve $c \subset \Sigma$ and a pair of points $\{p, q\} \subset \Sigma'$. In analogy with the theory of Lefschetz fibrations c is usually called *vanishing cycle*.

Parallel transport along γ with respect to $\mathcal{H}$ gives a diffeomorphism

$$\Pi_\gamma^{\mathcal{H}} \colon \Sigma \setminus c \longrightarrow \Sigma' \setminus \{p, q\}$$

which can be considered as an identification of Σ' as the surface obtained from Σ by surgery on c as follows.[9] Note that the surgery of Σ on c can be identified with the

---

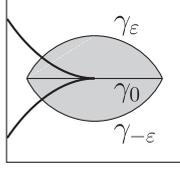[9] So Σ' is in some sense "derived" from Σ, whence the notation.

FIGURE 3. A cusp reference arc γ_0 embedded in a family γ_s of fold reference arcs for $s \neq 0$.

endpoint compactification of $\Sigma \setminus c$ while the endpoint compactification of $\Sigma' \setminus \{p, q\}$ is canonically identified with Σ'. Moreover, $\Pi_\gamma^{\mathcal{H}}$ extends to a diffeomorphism of the endpoint compactifications. As a consequence, we see that the vanishing cycle $c \subset \Sigma$ must be non-separating (otherwise Σ' would be disconnected, but fibers of wrinkled fibrations are by definition connected) and Σ' has genus one lower than Σ.

Note that the specific vanishing sets and parallel transport diffeomorphisms depend on both γ and $\mathcal{H}$. It is therefore important to understand this dependence.

Lemma 4.6. *Let $f \colon X \to B$ be a wrinkled fibration. For $s \in [0,1]$ we consider smooth families of connections $\mathcal{H}_s$ and fold reference arcs $\gamma_s \colon [0,1] \to B$ with common endpoints. Then the vanishing sets $c_s \subset \Sigma$ and $\{p_s, q_s\} \subset \Sigma'$ evolve by ambient isotopies. Moreover, all ambient isotopies of $c_0 \subset \Sigma$ and $\{p_0, q_0\} \subset \Sigma'$ can be realized by changing each $\mathcal{H}_s$ in an arbitrarily small neighborhood of $f^{-1}(\gamma_s)$.*

Proof. For brevity we denote the unit interval by $I = [0,1]$. For each $s \in I$ we have a smooth manifold $Y_s = f^{-1}(\gamma_s)$ and a Morse function $f_s = \gamma_s^{-1} \circ f \colon Y_s \to I$. We claim that the disjoint union $Y = \amalg_s(\{s\} \times Y_s)$ is a smooth submanifold of $I \times X$. To see this, we observe that Y can be expressed as the transverse preimage of $\amalg_s(\{s\} \times \gamma_s(I))$, which is a smooth submanifold of $I \times B$, under the product map $\mathrm{id}_I \times f$. Next we consider the map $F \colon Y \to I^2$ given by $F(s, y) = (f_s(y), s)$. Possibly after rescaling the reference arcs we can assume that $\gamma_s(\frac{1}{2})$ is the unique critical value along γ_s. It follows that the critical locus of F is an arc of indefinite folds which is mapped to $\gamma_I(\{\frac{1}{2}\})$. Another investigation of the fold model shows that F composed with the projection $pr_2 \colon I^2 \to I$ onto the second factor is a submersion. We can therefore integrate a suitable gradient vector field for $pr_2 \circ F$ to find a diffeomorphism $\phi \colon I \times Y_0 \to Y$ such that $F \circ \phi = \mathrm{id} \times f$. As a consequence, we get a family of diffeomorphisms $\phi_s \colon Y_0 \to Y_s$ satisfying $f_s \circ \phi_s = f_0$. Now let g_s be a family of Riemannian metrics on X that induce the connections $\mathcal{H}_s$. Restricting g_s to Y_s yields a gradient for f_s (with values in $\mathcal{H}_s$) which we pull back to a gradient v_s for $f_0 \colon Y_0 \to I$ via ϕ_s. According to Example 4.4, the vanishing sets of γ_s with respect to $\mathcal{H}_s$ are the same as the ascending and descending spheres of the unique critical points of f_0 with respect to v_s. This immediately shows that the vanishing sets evolve by ambient isotopies and for the realization of arbitrary isotopies we proceed as in Remark 4.5. $\square$

4.3.2. *Parallel transport through cusps.* While for many purposes it is enough to consider fold reference arcs, in some situations one is forced to consider reference arcs that interact with the critical values in more complicated ways. For example, in Section 5 we will have work with reference arcs that contain a single cusp value at which they are tangent to the direction of the cusp – we will refer to these as *cusp reference arcs*. Note that any cusp reference arc γ_0 can be embedded in a family $(\gamma_s)_{s \in [-\varepsilon, \varepsilon]}$ such that γ_s is a fold reference arc for $s \neq 0$ (see Figure 3). The following is common folklore in this situation:

(1) The vanishing sets of $(\gamma_0, \mathcal{H})$ are a single point $p \in \Sigma'$ and the union of two simple closed curves $c \cup d \subset \Sigma$ that intersect transversely in one point.

(2) As s approaches 0 from above the vanishing sets of $(\gamma_s, \mathcal{H})$ and $(\gamma_{-s}, \mathcal{H})$ converge to $c, d \subset \Sigma$ and $r \in \Sigma'$.

However, we have come to realize that this might be more subtle than expected. In fact, Example 4.4 indicates that (1) might actually fail for arbitrary connections. Since our arguments in Section 5 rely on (1) we include a proof for a reasonable class of connections which are "standard" near the cusps.

Lemma 4.7. *The conclusions (1) and (2) above are valid if for each cusp of f there are model coordinates in which $\mathcal{H}$ is induced by the standard metric on $\mathbb{R}^4$.*

Proof. We only give the details for (1) and note that (2) can be proved similarly. By the assumption on $\mathcal{H}$ it is enough to consider the indefinite fold model

$$f \colon \mathbb{R}^4 \to \mathbb{R}^2, \quad (t, x, y, z) \mapsto (t, x^3 + 3tx + y^2 - z^2)$$

and the reference arc $\gamma_0 \colon [-1, 1] \to \mathbb{R}^2$ given by $\gamma_0(\tau) = (\tau, 0)$. Our goal is to determine the vanishing sets $V_\pm$ of γ_0 in the fibers

$$\Sigma_\pm = f^{-1}(\pm 1, 0) = \{t = \pm 1\} \cap \{x^3 \pm 3x + y^2 - z^2 = 0\} \subset \mathbb{R}^4$$

with respect to the horizontal distribution $\mathcal{H}_0$ induced by standard Euclidean metric on $\mathbb{R}^4$. (It is easy to see that Σ_- is a once punctured torus while Σ_+ is diffeomorphic to the plane.) We fix points $p_\pm = (\pm 1, X_\pm, Y_\pm, Z_\pm)$ in $\Sigma_\pm$ and denote by $\tilde{\gamma}_0 \colon [-1, 1] \setminus \{0\} \to \mathbb{R}^4$ the unique $\mathcal{H}_0$–lift of γ_0 on the specified domain with $\tilde{\gamma}_0(\pm 1) = p_\pm$. Note that $V_\pm$ consists exactly of those $p_\pm$ for which $\tilde{\gamma}_0(\tau)$ converges to the origin as τ goes to zero from the left and the right. A direct computation shows that

$$\mathcal{H}_0 = \left\langle \tfrac{\partial}{\partial t}, 3(x^2 + t)\tfrac{\partial}{\partial x} + 2y\tfrac{\partial}{\partial y} - 2z\tfrac{\partial}{\partial z} \right\rangle$$

and that $\tilde{\gamma}_0$, which we henceforth write as $\tilde{\gamma}_0(\tau) = \big(\tau, X(\tau), Y(\tau), Z(\tau)\big)$, is the solution of the system of differential equations

$$\frac{dX}{d\tau} = -3X(X^2 + \tau)\big/\Xi, \quad \frac{dY}{d\tau} = -2XY\big/\Xi, \quad \frac{dZ}{d\tau} = 2XZ\big/\Xi \qquad (10)$$

where $\Xi(\tau) = 9(X^2 + \tau)^2 + 4Y^2 + 4Z^2$. There is one obvious solution given by $X = Y = Z \equiv 0$ which shows that $V_\pm$ contains the point $(\pm 1, 0, 0, 0)$. We next show that V_+ contains no other points. This follows from an inspection of (10) on the interval $(0, 1]$. The first equation shows that $X \equiv 0$ if and only if $X(\tau) = 0$ for some $\tau > 0$, and in that case the other equations force Y and Z to be constant. In particular, the only solution with $X_+ = 0$ that contributes to the vanishing set is the obvious one. On the other hand, if $X_+ \neq 0$, then X is nowhere zero; in fact, we have $|X| \geq |X_+|$. To see this, observe that if $X_+ > 0$, then the right hand side of the first equation in (10) is strictly negative everywhere. It follows that X is monotonically decreasing and therefore bounded below by X_+. The case $X_+ < 0$ is completely analogous. As a consequence, no solution with $X_+ \neq 0$ can contribute to V_+ which therefore only consists of the point $(1, 0, 0, 0)$, as claimed. It remains to determine V_- for which we restrict our attention to the interval $[-1, 0)$. Again, (10) shows that the only solution with $X_- = 0$ that contributes to V_- is the obvious one, and that if $X_- \neq 0$, then X is nowhere zero. Arguing similarly as above we can show that if $X_- > 0$, then $|Z| \geq |Z_-|$ and if $X_- < 0$, then $|Y| \geq |Y_-|$. To summarize, we have shown that V_- is contained in the union of $\Sigma_- \cap \{x \geq 0, z = 0\}$ and $\Sigma_- \cap \{x \leq 0, y = 0\}$ which is easily seen to be a pair of simple closed curves in Σ_- intersecting transversely in $(-1, 0, 0, 0)$. Finally, we

note that if $p_- \in \Sigma_- \cap \{x \geq 0,\, z = 0\}$, the limit of $\tilde{\gamma}$ must lie in the intersection $\Sigma_0 \cap \{x \geq 0,\, z = 0\}$ which contains only the origin. Similar arguments apply to $p_- \in \Sigma_- \cap \{x \leq 0,\, y = 0\}$ and we conclude that

$$V_- = \left(\Sigma_- \cap \{x \geq 0,\, z = 0\}\right) \cup \left(\Sigma_- \cap \{x \leq 0,\, y = 0\}\right)$$

has the desired structure. $\square$

Remark 4.8. We do not know whether the assumption on $\mathcal{H}$ in Lemma 4.7 is necessary but it seems likely that a proof for general $\mathcal{H}$ would have to involve more elaborate tools from the theory of dynamical systems.

4.3.3. *Vanishing cycles in simple wrinkled fibrations.* Using the techniques developed in this chapter we can prove another folklore fact about simple wrinkled fibrations. Let $w\colon X \to S^2$ be such a map. We parameterize the equator of S^2 by S^1 and denote the meridian through $\theta \in S^1$ by μ_θ. By a suitable reparametrization of S^2 we can assume that the critical values of w are arranged near the equator such that μ_θ is a fold reference arc for all but finitely many values of θ for which it is a cusp reference arc where the cusp points toward the south pole. Let $\theta_1, \ldots, \theta_l \in S^1$ be the exceptional values, cyclically ordered according to the orientation of S^1, and let Σ and Σ' be the fibers over the north and south pole, respectively.

Lemma 4.9. *If Σ has genus at least two, there is a connection for w such that the vanishing cycles of μ_θ are constant for each $\theta \in (\theta_i, \theta_{i+1})$, say $c_i \subset \Sigma$, and the union $c_{i-1} \cup c_i$ for $\theta = \theta_i$.*

The data $(\Sigma; c_1, \ldots, c_l)$ is known as the *surface diagram* of w and it was already observed in [Wil10] that $\mathfrak{S}$ contains enough information to recover w up to equivalence (also see [Beh13]). We will come back to these diagrams in Section 6.

Proof. Let $\mathcal{H}$ be an arbitrary connection for w that is standard near the cusps and let $c_i \cup d_i \subset \Sigma$ be the vanishing set for θ_i guaranteed by Lemma 4.7. As in Remark 4.5 and Lemma 4.6 we can modify $\mathcal{H}$ so that $d_i = c_{i-1}$ and that the vanishing cycles for θ near θ_i are constantly c_i on one side and c_{i-1} on the other. The claim now follows from Lemma 4.10 below. $\square$

Lemma 4.10. *Let $f\colon X \to B$ be a wrinkled fibration and let $(\gamma_s)_{s \in [0,1]}$ be a family of fold reference arcs with common endpoints such that the map $(s, t) \mapsto \gamma_s(t)$ is an embedding for $t \neq 0, 1$. Let $\mathcal{H}$ be a connection for f and let $c_s \subset \Sigma$ be the vanishing cycle of γ_s. If $c_0 = c_1$ and Σ has genus at least two, then $\mathcal{H}$ can be modified in $f^{-1}(\cup_{s \in (0,1)} \gamma_s)$ such that $c_s = c_0$ for all s.*

Proof of Lemma 4.10. This follows from a variation of the proof of Lemma 4.6 together the extra input that the space of homotopically non-trivial simple closed curves in a surface with negative Euler characteristic has simply connected components (see [Iva02, p.535ff]). We can therefore find a 2–parameter family of diffeomorphisms $\phi_{s,t} \in \mathrm{Diff}(\Sigma)$ such that $\phi_{s,0}(c_0) = c_s$ and $\phi_{s,1}(c_0) = c_0$ for all s. Finally, we can deform $\mathcal{H}$ to a 1–parameter family of connections $\mathcal{H}_t$ within $f^{-1}(\cup_{s \in (0,1)} \gamma_s)$ such that the vanishing cycle of $(\gamma_s, \mathcal{H}_t)$ is $\phi_{s,t}(c_0)$. $\square$

5. Elimination of Cusps II: A Mapping Class Group Interpretation

We can now prove our second main result which we first recall.

Theorem B. *The subset of $\pi_0(\mathrm{Diff}(\Sigma_1, \Sigma_2))$ obtained from merging cusps along a fixed arc in B has a free and transitive action of either of the groups*

$$\mathcal{K}(c_i, d_i) = \ker \Phi_{c_i} \cap \ker \Phi_{d_i} \subset \mathrm{Mod}(\Sigma_i), \quad i = 1, 2$$

where $c_i, d_i \subset \Sigma_i$ are simple closed curves related to the vanishing sets of the cusps.

We begin with a more precise description of the problem. Let $f\colon X \to B$ be a wrinkled fibration with a pair of (indefinite) cusps $p, q \in \mathcal{C}(f)$. We consider an elementary cusp merge $F = (f_s) \in \mathfrak{CM}_0(f; p, q)$ obtained by the Levine construction with a framed joining curve $\boldsymbol{\lambda}_F \in \boldsymbol{\Lambda}(f; p, q)$. It will be convenient to call the composition $\gamma_F = f \circ \lambda_F$ the *joining arc* of F. By a slight abuse of notation, we use the same notation for a slightly longer curve $\gamma_F \colon J_\varepsilon \to B$ where $J_\varepsilon = [-1-\varepsilon, 1+\varepsilon]$ whose ends run into the regions enclosed by the cusps of f. Since we can choose the support of F as an arbitrarily small neighborhood of the joining curve, we can assume that the fibers over the endpoints

$$\Sigma_1 = f_s^{-1}\big(\gamma_F(-1-\varepsilon)\big) \quad \text{and} \quad \Sigma_2 = f_s^{-1}\big(\gamma_F(1+\varepsilon)\big) \tag{11}$$

are independent of s. Observe that we can consider the extended version of γ_F as a reference arc for the various maps in the family (f_s). We will do this for the initial map $f = f_{-1}$ and a map f_δ where $\delta > 0$ is small enough so that (f_s) still evolves according to the cusp merge model. For f we find that γ_F is a concatenation of two cusp reference arcs for f in the sense of Section 4.3.2. So if we fix a connection $\mathcal{H}$ for f which is standard near the cusps, then according to Lemma 4.7 the vanishing sets of the nearest cusp of each endpoint of γ_F are the unions of simple closed curves $c_i, d_i \subset \Sigma_i$ that intersect transversely in one point. However, in the case of f_δ it is easy to see from the cusp merge model that γ_F is an arc of regular values. As a consequence, the parallel transport along γ_F with respect to any connection for f_F gives rise to a well defined isotopy class of diffeomorphisms

$$\varphi_F \in \pi_0\big(\mathrm{Diff}(\Sigma_1, \Sigma_2)\big).$$

In order to prove Theorem B we will establish a relation between the a priori unrelated objects φ_F, the framed joining curve $\boldsymbol{\lambda}_F$, and the vanishing cycles $c_i, d_i \subset \Sigma_i$. For that purpose, we also consider two fibers just outside of the cusps of f

$$\Sigma_1' = f^{-1}\big(\gamma_F(-1+\varepsilon)\big) \quad \text{and} \quad \Sigma_2' = f^{-1}\big(\gamma_F(1-\varepsilon)\big) \tag{12}$$

and observe that $\boldsymbol{\lambda}_F$ specifies tangent lines $L_i \in \mathbb{P}(T\Sigma_i')$. However, before we can continue this discussion we need to make some general considerations. We will return to the setup described above in Section 5.3.

5.1. **The difference of framed joining curves.** As above, let $f\colon X \to B$ be a wrinkled fibration with a pair of indefinite cusps $p, q \in \mathcal{C}(f)$. Observe that if $\boldsymbol{\Lambda}(f; p, q)$ is non-empty, then f maps the interior of the image of any joining curve from p to q into a connected component $R \subset B \setminus \mathcal{D}(f)$. We let $X_R = f^{-1}(R)$ and write $f_R\colon X_R \to R$ for the restriction of f to X_R. In addition, let $\boldsymbol{X}_R = \mathbb{P}(\ker(df_R))$ and let $\boldsymbol{f}_R\colon \boldsymbol{X}_R \to R$ be the composition of f_R and the projection $\pi\colon \boldsymbol{X}_R \to X_R$. It easily follows from the definitions that $\pi_0\big(\boldsymbol{\Lambda}(f; p, q)\big)$ admits a free and transitive action of $\pi_1(\boldsymbol{X}_R)$. For concreteness, let us fix $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}(f; p, q)$ and some fiber Σ' of f over the interior of the image of $\boldsymbol{f}_R \circ \boldsymbol{\lambda}$ in B. As mentioned before, $\boldsymbol{\lambda}$ gives rise to a tangent line $L \subset T_x\Sigma'$ which we take as a base point for both $\boldsymbol{X}_R$ and $\mathbb{P}(T\Sigma')$. Since R consists of regular values of f, both f_R and $\boldsymbol{f}_R$ are fiber bundles over R with fibers Σ' and $\mathbb{P}(T\Sigma')$, respectively. In particular, we can express $\pi_1(\boldsymbol{X}_R; L)$ as a semi-direct product of $\pi_1(\mathbb{P}(T\Sigma'); L)$ and $\pi_1(R; y)$ where $y = f(x)$. Thus, if we fix the $\pi_1(R)$ component, that is, if we only consider those $\boldsymbol{\lambda}' \in \boldsymbol{\Lambda}(f; p, q)$ with $\boldsymbol{f}_R \circ \boldsymbol{\lambda}' = \boldsymbol{f}_R \circ \boldsymbol{\lambda}$, then get an element

$$\delta(\boldsymbol{\lambda}, \boldsymbol{\lambda}') \in \pi_1(\mathbb{P}(T\Sigma'); L) \tag{13}$$

which measures the difference of the homotopy classes of $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$. It turns out that $\delta(\boldsymbol{\lambda}, \boldsymbol{\lambda}')$ has an interpretation in terms of mapping class groups that will be the key to the proof of Theorem B.

5.2. **A brief digression on mapping class groups.** In this section we will prove some abstract results about certain mapping class groups that are closely related to our problem. Let Σ' be a closed, orientable surface of genus g' and let $\mathrm{Mod}(\Sigma') = \pi_0(\mathrm{Diff}^+(\Sigma'))$ be its mapping class group. Moreover, let $L \subset T_x\Sigma'$ be a tangent line and let $\mathrm{Diff}^+(\Sigma'; L)$ consist of all $\phi \in \mathrm{Diff}^+(\Sigma')$ such that $d\phi(L) = L$. We consider the group $\mathrm{Mod}(\Sigma'; L) = \pi_0(\mathrm{Diff}^+(\Sigma'; L))$ and the "forgetful map"

$$F_L \colon \mathrm{Mod}(\Sigma'; L) \longrightarrow \mathrm{Mod}(\Sigma')$$

induced by the inclusion of $\mathrm{Diff}(\Sigma'; L)$ in $\mathrm{Diff}^+(\Sigma')$. We have the following analogue of the Birman exact sequence (see [FM11, Ch.4.2.3], for example).

Lemma 5.1 (Generalized Birman sequence)**.** *There is an exact sequence*

$$\cdots \to \pi_1(\mathrm{Diff}^+(\Sigma'), \mathrm{id}) \to \pi_1(\mathbb{P}(T\Sigma'), L) \xrightarrow{\mathrm{push}} \mathrm{Mod}(\Sigma'; L) \xrightarrow{F_L} \mathrm{Mod}(\Sigma') \to 1.$$

Moreover, $\pi_1(\mathrm{Diff}^+(\Sigma'), \mathrm{id})$ is trivial for $g' \geq 2$.

The map push comes from a variation of the usual point pushing construction which not only drags a point around but also keeps control of a tangent direction.

Proof. We consider the map $\mathrm{ev}_L \colon \mathrm{Diff}^+(\Sigma') \to \mathbb{P}(T\Sigma')$ defined by $\mathrm{ev}_L(\phi) = d\phi(L)$. An easy modifications of the arguments in [FM11, Ch.4.2.3] can be modified to show that ev_L is a fiber bundle with fiber $\mathrm{Diff}^+(\Sigma'; L)$. The desired exact sequence then follows from the long exact sequence of homotopy groups and the vanishing of $\pi_1(\mathrm{Diff}^+(\Sigma'), \mathrm{id})$ follows from the work of Earle and Eells [EE69]. $\square$

Now let Σ be another closed, orientable surface of genus $g = g'+1$ and let $c, d \subset \Sigma$ be a pair of simple closed curves that intersect transversely in a single point. We consider the subgroup $\mathrm{Mod}(\Sigma)(c, d) \subset \mathrm{Mod}(\Sigma)$ consisting of all mapping classes that have representatives $\phi \colon \Sigma \to \Sigma$ such that $\phi(c) = c$ and $\phi(d) = d$. In order to make the connection with the previous discussion, let $\Sigma'_{c,d}$ the surface given by the endpoint compactification of $\Sigma \setminus (c \cup d)$ with its distinguished endpoint $x \in \Sigma'_{c,d}$.

Lemma 5.2. *For any tangent line $L \subset T_x\Sigma'_{c,d}$ there is a canonical isomorphism*

$$\widetilde{\Psi}_L \colon \mathrm{Mod}(\Sigma'_{c,d}; L) \xrightarrow{\cong} \mathrm{Mod}(\Sigma)(c, d).$$

Proof. Let $T \subset \Sigma$ be a closed neighborhood of $c \cup d$ that is diffeomorphic to a one-holed torus and let Σ° be the closure of $\Sigma \setminus T$. Note that we can also consider Σ° as a subsurface of $\Sigma'_{c,d}$ where its complement is an open disk containing x whose closure we denote by $D \subset \Sigma'_{c,d}$. The two inclusions of Σ° induce homomorphisms of $\mathrm{Mod}(\Sigma^\circ)$ to $\mathrm{Mod}(\Sigma)(c, d)$ and $\mathrm{Mod}(\Sigma'_{c,d}; L)$.[10] It is easy to see that the groups $\mathrm{Mod}(T)(c, d)$ and $\mathrm{Mod}(D; L)$ are both infinite cyclic generated by $\Delta_{c,d} = (t_c t_d)^3$ and the right-handed half twist h_x around $x \in D$, respectively. Moreover, standard arguments yield short exact sequences

$$1 \to \left\langle (t_{\partial\Sigma^\circ}^{-1}, h_p^2) \right\rangle \to \mathrm{Mod}(\Sigma^\circ) \times \mathrm{Mod}(D; L) \to \mathrm{Mod}(\Sigma_{c,d}; L) \to 1$$

and

$$1 \to \left\langle (t_{\partial\Sigma^\circ}^{-1}, \Delta_{c,d}^2) \right\rangle \to \mathrm{Mod}(\Sigma^\circ) \times \mathrm{Mod}(T)(c, d) \to \mathrm{Mod}(\Sigma)(c, d) \to 1.$$

since $\Delta_{c,d}$ and h_p both square to a right-handed boundary parallel Dehn twist. The desired isomorphism Ψ_L is therefore induced by the identity map $\mathrm{Mod}(\Sigma^\circ)$ and the isomorphism of $\mathrm{Mod}(D; L)$ and $\mathrm{Mod}(T)(c, d)$. $\square$

---

[10]When discussing mappings class groups of surfaces with boundary we always assume that all diffeomorphisms restrict to the identity on the boundary.

By combining the maps from Lemmas 5.1 and 5.2 we obtain a homomorphism

$$\Psi_L = \widetilde{\Psi}_L \circ \mathrm{push} \colon \pi_1\big(\mathbb{P}(T\Sigma'_{c,d}); L\big) \longrightarrow \mathrm{Mod}(\Sigma) \tag{14}$$

that factors through the subgroup $\mathrm{Mod}(\Sigma)(c,d)$. Since $\widetilde{\Psi}_L$ is an isomorphism, it follows from Lemma 5.1 that Ψ_L is injective if Σ has genus at least three and its kernel can easily be computed from the exact sequence. More importantly, in the remainder of this section we will focus on the image of Ψ_L.

Let Σ'_c be the surfaces obtained by surgery on c. We have a subgroup $\mathrm{Mod}(\Sigma)(c)$ of $\mathrm{Mod}(\Sigma)$, defined exactly as $\mathrm{Mod}(\Sigma)(c,d)$ with the single difference that only one curve is preserved, which is the source of the so called *surgery homomorphism*

$$\Phi_c \colon \mathrm{Mod}(\Sigma)(c) \longrightarrow \mathrm{Mod}(\Sigma'_c)$$

obtained by choosing a representative of a given mapping class that fixes c set wise and extending it to the endpoint compactification. Of course, the same discussion applies to $d \subset \Sigma$. Surgery homomorphisms have appeared several times in the context of wrinkled fibrations and broken Lefschetz fibrations, see [Bay09, Beh13, Hay14].

Proposition 5.3. *The image of Ψ_L is given by the subgroup*

$$\mathcal{K}(c,d) = \ker \Phi_c \cap \ker \Phi_d \subset \mathrm{Mod}(\Sigma). \tag{15}$$

Proof. Let Σ° be as in the proof of Lemma 5.2. Observe that the three surfaces Σ'_c, Σ'_d and $\Sigma'_{c,d}$ are all obtained from Σ° by filling the boundary with a disk. Their mapping class groups are therefore canonically isomorphic to $\mathrm{Mod}(\Sigma^\circ)/\langle t_{\partial\Sigma^\circ}\rangle$. Using these canonical isomorphisms we obtain a diagram

$$
\begin{array}{ccc}
\mathrm{Mod}(\Sigma)(c) & \xrightarrow{\ \ \Phi_c\ \ } & \mathrm{Mod}(\Sigma'_c) \\[2pt]
{\scriptstyle\mathrm{incl}}\big\uparrow & & \big\downarrow{\scriptstyle\cong} \\[2pt]
\mathrm{Mod}(\Sigma)(c,d) \xleftarrow[\cong]{\ \widetilde{\Psi}_L\ } \mathrm{Mod}(\Sigma'_{c,d};L) & \xrightarrow{\ F_L\ } & \mathrm{Mod}(\Sigma'_{c,d}) \\[2pt]
{\scriptstyle\mathrm{incl}}\big\downarrow & & \big\uparrow{\scriptstyle\cong} \\[2pt]
\mathrm{Mod}(\Sigma)(d) & \xrightarrow{\ \ \Phi_d\ \ } & \mathrm{Mod}(\Sigma'_d)
\end{array}
$$

whose commutativity follows from the definitions. The diagram shows that $\widetilde{\Psi}_L^{-1}$ maps $\mathcal{K}(c,d)$ isomorphically onto $\ker F_L$ which agrees with $\mathrm{push}\big(\pi_1(\mathbb{P}(T\Sigma_{c,d}), L)\big)$ by Lemma 5.1. This proves the claim. $\qquad\square$

Observe that the definition of $\mathcal{K}(c,d)$ in (15) makes sense for pairs of non-separating with arbitrary intersection patterns. In the case where c and d are disjoint this group was studied by the second author who obtained a generating set in [Hay14, Theorem 3.4]. In our situation, where c and d intersect in a single point, we can also obtain a generating set. Again, let $\Sigma^\circ \subset \Sigma$ be the subsurface obtained by deleting a regular neighborhood of $c \cup d$. Let a be a simple closed curve in the interior of Σ° and let $\zeta \subset \Sigma^\circ$ be an embedded arc connecting a and $\partial\Sigma^\circ$ whose interior is disjoint from a. Then the band sum $a\#_\zeta \partial\Sigma^\circ$ is a simple closed curve in Σ and we define

$$\vartheta_{a,\zeta} := t_a\, t_{a\#_\zeta\partial\Sigma^\circ}^{-1} \in \mathrm{Mod}(\Sigma). \tag{16}$$

Lemma 5.4. *For $g(\Sigma) \geq 3$ the image of $\widetilde{\Psi}_L$ is generated by $\Delta_{c,d}$ and the elements $\vartheta_{a,\zeta}$ described in (16). For $g(\Sigma) \leq 2$ the image is generated by $\Delta_{c,d}$ which has infinite order for $g(\Sigma) = 2$ and order two for $g(\Sigma) = 1$.*

Proof. Note that $\pi_1(\mathbb{P}(T\Sigma'_{c,d}); L)$ is generated by a loop δ around $\mathbb{P}(T_x\Sigma'_{c,d}) \cong S^1$ and embedded loops in $\Sigma'_{c,d}$ based at x. The generalized point pushing map sends δ to the half-twist h_x or its inverse, depending on the orientation of δ. For the loops in $\Sigma'_{c,d}$ we get the usual point pushing map. Possibly after reversing δ we get $\Psi_L(\delta) = \Delta_{c,d}$ which is easily seen to have infinite order for $g(\Sigma) \geq 2$ and order two for $g(\Sigma) = 1$. For an embedded loop $\alpha\colon (S^1, 1) \to (\Sigma'_{c,d}, x)$ it is known that

$$\text{push}(\alpha) = t_{a_l}^{-1}\, t_{a_r} \in \text{Mod}(\Sigma'_{c,d}, x)$$

where $a_{l/r} \subset \Sigma^\circ$ are left and right push-offs of the image of α which can be chosen in the interior of Σ°. We can therefore also consider them as curves in Σ for which we use the notation $\tilde{a}_{l/r}$ to avoid confusion. Now, by definition we have

$$\widetilde{\Psi}_L\big(t_{a_l}^{-1}\, t_{a_r}\big) = t_{\tilde{a}_l}^{-1}\, t_{\tilde{a}_r} \in \text{Mod}(\Sigma).$$

and it is easy to see that $\tilde{a}_r$ can be obtained as a band sum of $\tilde{a}_l$ and $\partial\Sigma^\circ$ so that $\Psi_L(\alpha)$ has the form (16). It remains to show that $t_{\tilde{a}_l}^{-1}\, t_{\tilde{a}_r}$ is the trivial mapping class when $g(\Sigma) \leq 2$. But this is trivial for $g(\Sigma) = 1$ and for $g(\Sigma) = 2$ it is easy to see that $\tilde{a}_l$ and $\tilde{a}_r$ are isotopic in Σ°. $\square$

5.3. **Parallel transport before and after merging cusps.**
We now return to the setting described in the beginning of Section 5 and prove Theorem B. As before let $f\colon X \to B$ be a wrinkled fibration with two cusps $p, q \in \mathcal{C}(f)$ and let $R \subset B \setminus \mathcal{D}(f)$ be the region containing the images of joining curves for p and q. Given a fixed joining arc $\gamma\colon J \to B$ between $f(p)$ and $f(q)$ we denote by $\mathfrak{CM}_0^{[\gamma]}(f; p, q)$ the set of elementary cusp merges whose joining is homotopic in R to γ relative to its endpoints. Combining the considerations in Section 5.1 with Theorem A, we see that $\pi_0\big(\mathfrak{CM}_0^{[\gamma]}(f; p, q)\big)$ has a free and transitive action of $\pi_1(\mathbb{P}(T\Sigma'_i); L_i)$ with Σ'_i and L_i as in (12). Moreover, Section 5.2 already indicates how to connect the groups $\pi_1(\mathbb{P}(T\Sigma'_i); L_i)$ and $\mathcal{K}(c_i, d_i)$. The precise relation will be established by carefully choosing connections and investigating the resulting parallel transport diffeomorphisms.

Proof of Theorem B. We consider two elements F, G of $\mathfrak{CM}_0^{[\gamma]}(f; p, q)$. Up to deformations within $\mathfrak{CM}_0^{[\gamma]}(f; p, q)$ we can make a series of assumptions on F and G. Later it will become clear that they are irrelevant. First, we assume that the joining arcs γ_F and γ_G are not only homotopic to but actually agree with the fixed arc γ. Second, by Theorem A we can assume that both homotopies are obtained from the Levine construction applied to framed joining curves $\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_G \in \boldsymbol{\Lambda}(f; p, q)$ that agree near p and q. We can therefore choose the same extension of $\gamma\colon J_\varepsilon \to B$ and, since we can control the supports of F and G simultaneously, we can consider the same fibers Σ_i and Σ'_i as in (11) and (12) such that $\boldsymbol{\lambda}_F$ and $\boldsymbol{\lambda}_G$ span the same lines $L_i \subset T_{x_i}\Sigma'_i$. Next we choose connections for the maps f_s and g_s in the homotopies as follows. The proof of Lemma 3.10 shows that we can apply the Levine construction such that F and G are obtained by implanting the same truncated version $(\tilde{\mu}_s)$ of the merge model (μ_s) into f, and such that we obtain merge coordinates for both homotopies that agree near p and possibly differ near q by the involution $(t, x, y, z) \mapsto (t, x, -y, -z)$. Let U_F and U_G be the neighborhoods of the joining curves where F and G agree with $(\tilde{\mu}_s)$ which we can assume to have the same intersections with the preimages of the interior regions of the cusps of f. Moreover, by shortening γ we can assume that $U_F \cap \Sigma_i = U_G \cap \Sigma_i$ is non-empty. The Euclidean connections for the truncated model maps $\tilde{\mu}_s$ give rise to connections for the maps f_s and g_s near the images of λ_F and λ_G at least for those s where F and G evolve according to the cusp merge model. We extend these connections arbitrarily to the rest of X. For our purposes, we only need the resulting connections $\mathcal{H}'_F$

and $\mathcal{H}'_G$ for the initial map f as well as $\mathcal{H}_F$ and $\mathcal{H}_G$ for maps f_δ and g_δ where $\delta > 0$ is small enough. Observe that $\mathcal{H}'_F$ and $\mathcal{H}'_G$ agree near p and q so that both give rise to the same vanishing cycles $c_i, d_i \subset \Sigma_i$, assuming that the connections are standard near the cusps so that Lemma 4.7 applies. Clearly, this can be verified in the model. One readily checks that for $s < 0$ the maps $(t, x, y, t) \mapsto (\pm(s - t^2), x, y, z)$ and $(u, v) \mapsto (\pm(s - t^2), v)$ take the origin of the cusp model to the cusps of μ_s and also map the Euclidean connections of both maps into each other.

Using this setup we can now start the actual proof of Theorem B. Since the situation is symmetric in Σ_1 and Σ_2, we will only consider the case of Σ_1 . By taking parallel transport along γ we obtain diffeomorphisms

$$\widetilde{\varphi}'(F, G) = \left(\Pi^{\mathcal{H}'_G}_\gamma\right)^{-1} \circ \Pi^{\mathcal{H}'_F}_\gamma \in \mathrm{Diff}(\Sigma'_1)$$

$$\widetilde{\varphi}(F, G) = \left(\Pi^{\mathcal{H}_G}_\gamma\right)^{-1} \circ \Pi^{\mathcal{H}_F}_\gamma \in \mathrm{Diff}(\Sigma_1).$$

We are ultimately interested in the mapping class $\varphi(F, G) \in \mathrm{Mod}(\Sigma_1)$ represented by $\widetilde{\varphi}(F, G)$. However, we first show that $\widetilde{\varphi}'(F, G)$ preserves the tangent line $L_1 \in \mathbb{P}(T\Sigma'_1)$ and represents a mapping class

$$\varphi'(F, G) = \mathrm{push}\big(\delta(\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_G)\big) \in \mathrm{Mod}(\Sigma'_1; L_1) \tag{17}$$

with $\delta(\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_G) \in \pi_1(\mathbb{P}(T\Sigma'_1); L_1)$ as in Section 5.1. To see this we trivialize f over $\gamma(J'_\varepsilon)$ where $J'_\varepsilon = [-1 + \varepsilon, 1 - \varepsilon]$ using parallel transport with respect to $\mathcal{H}'_F$ so that we can identify f over $\gamma(J'_\varepsilon)$ with the projection $\Sigma'_1 \times J'_\varepsilon \to J'_\varepsilon$. Note that in this trivialization $\boldsymbol{\lambda}_F$ appears as the constant curve $\tau \mapsto (L_1, \tau) \in \mathbb{P}(T\Sigma'_1) \times J'_\varepsilon$ while $\boldsymbol{\lambda}_G$ takes the form $\tau \mapsto (\tau, \delta(\tau))$ where δ is a closed loop in $\mathbb{P}(T\Sigma'_1)$ representing $\delta(\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_G)$. Moreover, the parallel transport between the various fibers $\Sigma'_1 \times \{\tau\}$ give rise to an isotopy of Σ'_1 and it is easy to see that this isotopy realizes the mapping class $\mathrm{push}\big(\delta(\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_G)\big)$. Next we show that

$$\varphi(F, G) = \widetilde{\Psi}_{L_1}\big(\varphi'(F, G)\big) \in \mathrm{Mod}(\Sigma_1) \tag{18}$$

with $\widetilde{\Psi}_{L_1}$ as in Lemma 5.2. Observe that parallel transport with respect to either $\mathcal{H}'_F$ or $\mathcal{H}'_G$ provides diffeomorphisms between $\Sigma_i \setminus (c_i \cup d_i)$ and $\Sigma'_i \setminus \{x_i\}$. (Since the connections agree near the cusps, they give the same diffeomorphisms.) Now we take closed disks $D_i \subset \Sigma'_i$ containing the base points x_i of L_i in their interior. Parallel transport with respect to $\mathcal{H}'_F$ or $\mathcal{H}'_G$ maps them to one holed tori $T_i \subset \Sigma_i$ containing $c_i \cup d_i$. Note that we can assume that the D_i are contained in the part of Σ'_i where F and G evolve according to the cusp merge model (μ_s). Next we observe that for $t_0 > 0$ the Euclidean parallel transport for μ_s maps $(-t_0, x, y, z) \in \mu_s^{-1}(-t_0, 0)$ to $(t_0, x, y, z) \in \mu_s^{-1}(t_0, 0)$ provided that $(t_0, 0)$ is a regular value and the points are not contained in the vanishing sets. From this we can deduce several things. First, it follows that we can assume that D_1 and D_2 are mapped into each other by the parallel transport with respect to both $\mathcal{H}'_F$ and $\mathcal{H}'_G$. Second, with this choice D_1 and T_1 are preserved by $\widetilde{\varphi}'(F, G)$ and $\widetilde{\varphi}(F, G)$, respectively. Third, $\widetilde{\varphi}'(F, G)$ acts on D_1 either as the identity or as a 180 degree rotation (depending on whether the merge coordinates for F and G agree near q). Moreover, in the first case $\widetilde{\varphi}(F, G)$ acts on T_1 as the identity, and in the second as a hyperelliptic involution which has the same effect on c_1 and d_1 as the Δ–twist Δ_{c_1, d_1}. Comparing this with the definition of the map $\widetilde{\Psi}_{L_1}$ we obtain (18). Finally, observe that Theorem B is a consequence of (17), (18), and Proposition 5.3 and the fact that $\varphi(F, G)$ and the isotopy classes of c_i and d_i are independent of the various choices made in the above constructions. $\qquad \square$

5.4. **The case of fold merges.** After our extensive study of cusp merges we now take a brief look at fold merges. As it turns out, the parallel transport problem is much simpler in this case. Let $f\colon X \to B$ be a wrinkled fibration and assume that a square in B intersects the critical image in two parallel fold arcs. We choose a connection $\mathcal{H}$ for f, a reference fiber Σ that maps into the middle region of the square and one fold reference arc for each fold arc, say γ_c and γ_d with corresponding vanishing cycles $c, d \subset \Sigma$. In order to perform a fold merge it is necessary and sufficient that c and d intersect transversely in a single point for some choice of $\mathcal{H}$. We will call this structure a *fold merge configuration* and refer to Σ as the *inner fiber*. In this situation, it is convenient to label the *outer fibers*, that is, the fibers over the other endpoints of γ_c and γ_d, by Σ'_c and Σ'_d, respectively, since the parallel transport provides an identification with the surfaces obtained by surgery on c and d. We choose a regular neighborhood T of $c \cup d$ as in the proof of Lemma 5.2. The parallel transport along γ_c then maps $T \setminus c$ onto a twice punctured disk $D_c \subset \Sigma_c$ in which d appears as an arc d' connecting the two punctures which we can think of as marked points. The analogous statement holds with the roles of c and d interchanged. The crucial observation is that parallel transport yields a diffeomorphism

$$(\Pi^{\mathcal{H}}_{\gamma_c})^{-1} \circ \Pi^{\mathcal{H}}_{\gamma_d}\colon \Sigma_c \setminus D_c \longrightarrow \Sigma_d \setminus D_d$$

between the complements of these disks and up to isotopy there is a unique extension to a diffeomorphism $\Sigma_c \to \Sigma_d$. Moreover, one can show that the parallel transport *after* performing a fold merge will be in this isotopy class. So up to isotopy, there is a unique identification of the outer fibers that can be obtained after merging folds which can already be described *before* actually merging folds.

6. Applications to surface diagrams of 4–manifolds

In this last section shift our attention to simple wrinkled fibrations as defined in Section 2.4. As indicated in the introduction, these maps can be used to obtain a combinatorial descriptions of closed 4–manifolds reminiscent of a Heegaard diagram. After reviewing the structure of these diagram, we will discuss how they change are affected by Williams's basic homotopies that involve cusp merges (Section 6.2), and we will give new examples of surface diagrams of 4–manifolds derived from Lefschetz fibrations and Heegaard diagrams (Section 6.3).

6.1. **Surface diagrams: a review.** Let $w\colon X \to S^2$ be a simple wrinkled fibration. As explained in Section 4.3.3, the set of regular values has two connected components which we refer to as the *higher* and *lower genus regions* depending on the fiber genus which differs by one. We choose one reference point in each region and connect them by fold reference arcs with pairwise disjoint interiors, one passing through each arc of fold values. Note that the orientation of the higher genus region (induced from the standard orientation of S^2) induces a cyclic order of the arcs of fold values. After choosing a connection we obtain simple closed curves $c_1, \ldots, c_l \subset \Sigma$ as the vanishing cycles in the higher genus fiber. Following Williams [Wil11a], the data $\mathfrak{S}_w = (\Sigma; c_1, \ldots, c_l)$ is called the *surface diagram* of w. It follows easily from the results of Section 4.2 that $\mathfrak{S}_w$ is well defined up to isotopy[11] (cf. [Beh13]). Moreover, using Lemma 4.6 we can and will assume that the vanishing cycles are transverse and realize the minimal number of intersections within their isotopy classes. Note that the curves in $\mathfrak{S}_w$ are not arbitrary. They satisfy the *intersection condition* that any two consecutive curves intersect in one

---

[11]For Σ this means that there are canonical isotopy class of diffeomorphisms between any two higher genus reference fibers since the higher genus region is simply connected.

point and the *monodromy condition*

$$\Phi_{c_1}\left(t_{t_{c_l}(c_1)} \circ t_{t_{c_{l-1}(c_l)}} \circ \cdots \circ t_{t_{c_1}(c_2)}\right) = 1. \tag{19}$$

It is shown in [Beh13, Section 4] that if Σ has genus at least three, then any collection of curves satisfying the above properties determines a simple wrinkled fibration up to left-right equivalence. In particular, if the fiber genus is sufficiently high, then w can be recovered from $\mathfrak{S}_w$ as suggested by Williams [Wil10, p.1054].

Remark 6.1. The second author developed an abstract theory of surface diagrams that allows to discuss simple wrinkled fibrations over more general base surface (cf. [Beh13]). In particular, it turns out that intersection condition is enough to build a simple wrinkled fibration over the disk such that the left hand side of (19) describes the monodromy of the boundary fibration. Those familiar with Lefschetz fibrations will probably recognize a familiar pattern here.

A surface diagram depicts a simple wrinkled fibration as seen from the higher genus perspective. Sometimes it is also useful to take a look from the lower genus side. For that purpose Williams introduced *surgered surface diagrams* [Wil11b]. As the name suggests, they are the result of performing surgery on some vanishing cycle in a surface diagram. Instead of giving a formal definition we will only explain their structure. Let $\mathfrak{S} = (\Sigma; c_1, \ldots, c_l)$ be a surface diagram. We parameterize a neighborhood νc_1 by $S^1 \times (-1, 1)$ such that the non-empty intersections $c_i \cap \nu c_1$ correspond to lines of the form $\{\theta\} \times (-1, 1)$ and denote by Σ' the surgered surface obtained by filling the two boundary components of $\Sigma \setminus \nu c_1$ with disks D_1 and D_2 to which we refer as the *surgery disks*. Since Σ can be reconstructed from Σ', the surgery disks and an identification of their boundaries, we can try to capture the whole surface diagram in Σ'. In order to do so we have to understand how the vanishing cycles appear in Σ'. We denote by $c_i' = c_i \setminus \nu c_1$ the remains of c_i after the surgery. Clearly, if c_i is disjoint from c_1, then c_i' is a simple closed curve in Σ' which is disjoint from the surgery disks. However, if c_i intersects c_1, then c_i' is a collection of embedded arcs in Σ' whose endpoints lie on the boundaries of the surgery disks while the interiors are disjoint from the disks. Each surgery disk contains the same number of endpoints and the endpoints on the different disks are matched via the identification of the boundaries of the disks. Finally, the curve c_1 itself appears in Σ' as the boundary of either disk which we also denote by c_1'. The collection $\mathfrak{S}' = (\Sigma'; c_1', \ldots, c_l')$ is then a surgered surface diagram.

Note that this way of picturing surfaces and curves therein is commonly used for Heegaard diagrams. Of course, we could have performed surgery on another vanishing cycle and the resulting diagram might look rather different. If necessary we will use the more precise notation $\mathfrak{S}'_{c_i}$ to indicate on which vanishing cycle we performed surgery, otherwise $\mathfrak{S}'$ will always stand for $\mathfrak{S}'_{c_1}$.

When we draw surgered surface diagrams the identification of the boundaries of the surgery disks is implicitly encoded as follows. Up to isotopy such an identification is determined by specifying one point on each boundary. To obtain such a pair of points we put an adjacent vanishing cycle of the one on which we performed surgery in minimal position so that it will appear as a single arc in the surgered diagram and we take its endpoints. Hence, as soon as we draw such an arc, we know how we have to identify the boundaries of the disks.

6.2. **Homotopy moves for surface diagrams.** In this section we discuss how the surface diagrams of simple wrinkled fibrations change in those of Williams's basic homotopies introduced in [Wil11b] that involve cusp merges. For brevity of notation, given a collection of non-separating simple closed curves $c_1, \ldots, c_k \subset \Sigma$ we write $\mathcal{K}(c_i) = \ker \Phi_{c_i}$ and $\mathcal{K}(c_1, \ldots, c_1) = \cap_i \mathcal{K}(c_i)$. As explained in [Wil11b], a
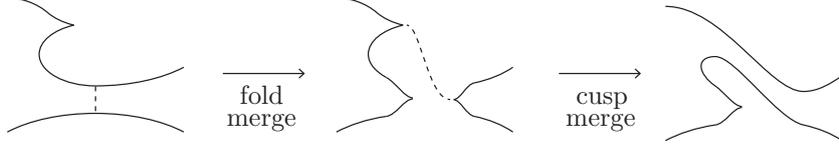
FIGURE 4. The critical values throughout a shift deformation.

multislide deformation is a fold merge, resulting in an additional pair of cusps and a canonical merge arc, followed by a cusp merge along the canonical merge arc. As an immediate consequence of Theorem B we obtain the following.

Proposition 6.2 (Multislides). *Let $w : X \to S^2$ be a simple wrinkled fibration with surface diagram $\mathfrak{S} = (\Sigma; c_1, \ldots, c_l)$ such that for some $1 < k < l$ the curves c_1 and c_k intersect transversely in a single point.*

(i) *If $w' : X \to S^2$ is obtained from w by a multislide deformation involving the two fold arcs with vanishing cycles c_1 and c_k, then there exists an element $\phi \in \mathcal{K}(c_1, c_k)$ such that the surface diagram of w' is given by*

$$\big(\Sigma; c_1, \ldots, c_k, \phi(c_{k+1}), \ldots, \phi(c_l)\big). \tag{20}$$

(ii) *For any $\phi \in \mathcal{K}(c_1, c_k)$, there exists a merge deformation from w to a simple wrinkled fibration whose surface diagram is given by (20).*

A slightly different combination of a fold merge followed by a cusp merge are the so called *(generalized) shift deformations*. The difference is that one of the cusps resulting from an initial fold merge is merged with a cusp adjacent to the other along a canonical merge arc, see Figure 4. Observe that a surface diagram of the resulting fibration is uniquely determined once we specify a framed joining curve mapping to the canonical joining arc. It turns out that there are two distinguished joining curves and we shall reserve the name *shift deformation* for generalized shifts derived from these curves. We will say more about this momentarily after proving the analogue of Proposition 6.2 for generalized shifts.

Proposition 6.3 (Generalized shifts). *Let $w : X \to S^2$ be a simple wrinkled fibration with surface diagram $\mathfrak{S} = (\Sigma; c_1, \ldots, c_l)$ such that for some $1 < k < l$ the curves c_1 and c_k intersect transversely in a single point.*

(i) *If $w' : X \to S^2$ is obtained from w by a generalized shift deformation whose initial fold merge is applied to the fold arcs with vanishing cycles c_k and c_l. Then the surface diagram of w' is given by*

$$\big(\Sigma; c_1, \ldots, c_k, c_l, \chi(c_{k+1}), \ldots, \chi(c_{l-1})\big), \tag{21}$$

where $\chi \in \mathrm{Mod}(\Sigma)$ satisfies $\chi t_{c_l}^{-1} t_{c_k}^{-1} \in \mathcal{K}(c_l)$ and $\chi t_{c_1}^{-1} t_{c_l}^{-1} \in \mathcal{K}(c_1)$.

(ii) *For any $\chi \in \mathrm{Mod}(\Sigma)$ satisfying the conditions in (i) there exists a generalized shift deformation from w to a simple wrinkled fibration whose surface diagram is given by (21).*

Proof. Figure 5 describes a neighborhood of the fold arcs in S^2 corresponding to the vanishing cycles c_1, c_k and c_l decorated with various reference points and oriented reference arcs between them. Let $\mathcal{H}$ be a connection for w and let $\chi \in \mathrm{Mod}(\Sigma)$ be the isotopy class of the diffeomorphism $\Pi_{\gamma_1}^{\mathcal{H}}{}^{-1} \circ \Pi_{\gamma_0}^{\mathcal{H}} : \Sigma \to \Sigma$. Clearly, the surface diagram of w' is given by $\big(\Sigma; c_1, \ldots, c_k, c_l, \chi(c_{k+1}), \ldots, \chi(c_{l-1})\big)$. Observe that the composition $\chi\, t_{c_l}^{-1} t_{c_k}^{-1} t_{c_l}$ preserves c_l up to isotopy and its image under Φ_{c_l} can be considered as the monodromy along the concatenation of $\varepsilon_1, \varepsilon_2$ and ε_3. Since this concatenation is null-homotopic, we have $\chi t_{c_l}^{-1} t_{c_k}^{-1} t_{c_l} \in \mathcal{K}(c_l)$, hence also
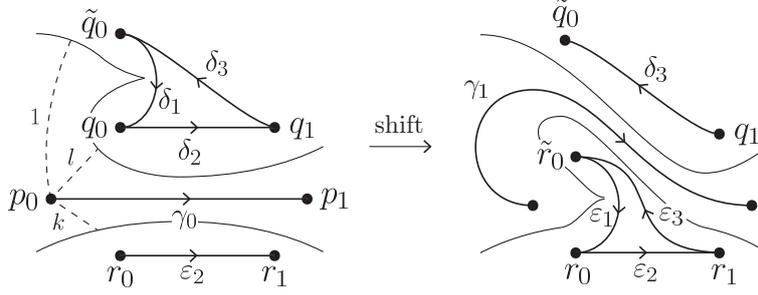
FIGURE 5. The dashed curves describe reference paths which give vanishing cycles c_1, c_k and c_l.

$\chi t_{c_l}^{-1} t_{c_k}^{-1} \in \mathcal{K}(c_l)$. Similar arguments involving the arcs δ_1, δ_2 and δ_3 show that $\chi t_{c_1}^{-1} t_{c_l}^{-1} \in \mathcal{K}(c_1)$ which concludes the proof of (i).

Now suppose that $\chi' \in \mathrm{Mod}(\Sigma)$ also satisfies the conditions in (i). Then we have $\chi \chi'^{-1} \in \mathcal{K}(c_1, c_l)$ and we can use Theorem B to construct a new generalized shift that produced the surface diagram $\big(\Sigma; c_1, \ldots, c_k, c_l, \chi'(c_{k+1}), \ldots, \chi'(c_{l-1}) \big)$. □

Next we explain the actual *shift deformations* alluded to earlier. As before, let $w : X \to S^2$ be a simple wrinkled fibration with surface diagram $(\Sigma; c_1, \ldots, c_l)$ such that c_k and c_l intersect transversely in one point. Let w_0 be the wrinkled fibration obtained from w by a fold merge involving the fold arcs corresponding to c_k and c_l. Let λ_0 be the canonical joining curve for the two cusps created by the fold merge. We consider a based loop γ in S^2 as on the left of Figure 6. Part of γ can be considered as a joining arc for two cusps of $\hat{w}_0$. We claim that there are two distinguished homotopy classes joining curves lifting γ to X. Indeed, each regular value along γ can be connected to the arc of folds to the left by a horizontal line and the vanishing set with respect to some fixed connection that is standard near the cusps is a pair of points in the corresponding fiber. Using Lemmas 4.6 and 4.7 we can show that these points trace out two curves in X connecting the two cusps over γ which we can consider as joining curves, say $\hat{\lambda}_1$ and $\hat{\lambda}_2$. Moreover, arguing as in the proofs of Lemmas 4.6 and 4.7 it follows that the homotopy classes of $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are independent of the connection and the precise choice of horizontal lines. By perturbing the concatenation of λ_0 with the reverse of $\hat{\lambda}_i$ we obtain two joining curves λ_i, $i = 1, 2$, and the cusp merge along either of them with an arbitrary framing completes a shift deformations resulting in simple wrinkled fibrations w_i'. Our goal is to understand the elements $\chi_i \in \mathrm{Mod}(\Sigma)$ that arise from Proposition 6.3(i) applied to w_i.

For that purpose, we consider wrinkled fibrations $\hat{w}_i$ obtained from w_0 by cusp merges along either $\hat{\lambda}_i$ with some framing. The discriminant of $\hat{w}_i$ has three components as shown on the right of Figure 6. Let $\varphi_i \in \mathrm{Mod}(\Sigma)$ be the monodromy along γ with respect to $\hat{w}_i$. Clearly, φ_i preserves c_l and maps c_k to c_1 up to isotopy. Changing the framing of the joining curve if necessary, we can assume that φ_i preserves the orientation of c_l. We choose a reference point $p_0 \in S^2$ in the lower genus region of $\hat{w}_i$ and a fold reference arc from p_0 through the fold arc with vanishing cycle c_l. This path gives a surgered surface diagram in the fiber Σ' over p_0 which we identify with the surgered surface $\Sigma' c_l$. We can collapse the surgery disks to points and obtain a graph in Σ'_{c_l} with two vertices $v_1, v_2 \in \Sigma'_{c_l}$ and edges given by the arcs c_k' and c_1' corresponding to c_k and c_1. These arcs form a loop which goes through both v_1 and v_2. By small perturbations we can obtain two

loops $\eta_1 \subset \Sigma'_{c_l} \setminus \{v_2\}$ and $\eta_2 \subset \Sigma'_{c_l} \setminus \{v_1\}$ where η_i is based at v_i. We orient η_i so that it leaves v_i along c'_k and returns along c'_1.
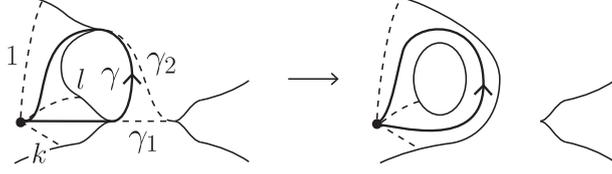

FIGURE 6. Base diagrams of wrinkled fibrations.

Lemma 6.4. *Let $\vec{c}_l \subset T\Sigma$ be a nowhere vanishing section of Tc_l. After changing indices if necessary, there is an integer m such that $\varphi_i \Delta^{2m}_{c_k,c_l} \in \mathrm{Mod}(\Sigma)(\vec{c}_l)$ is sent to the pushing maps along η_i by the map $\Phi^*_{c_l} : \mathrm{Mod}(\Sigma)(\vec{c}_l) \to \mathrm{Mod}(\Sigma'_{c_l}; v_1, v_2)$.*

Proof. The fold circle of $\hat{w}_i$ in the source which is mapped into the inner region of γ bounds a disk. Thus φ_i is contained in the kernel of the composition of $\Phi^*_{c_l}$ with the forgetful map of either v_1 or v_2. Possible after changing the labels we can achieve that $\varphi_i \in \ker(\mathrm{Forget}_{v_i} \circ \Phi^*_{c_l})$. We only give the arguments for φ_2 since φ_1 can be dealt with similarly. The image $\Phi^*_{c_l}(\varphi)$ is a pushing map along some loop $\delta \in \pi_1(\Sigma'_{c_l} \setminus \{v_2\}, v_1)$. Let Π_{v_1,v_2} be the set of homotopy classes of arcs connecting v_1 to v_2. The group $\pi_1(\Sigma'_{c_l} \setminus \{v_2\}, v_1)$ acts on Π_{v_1,v_2} in the obvious way. Since φ maps c_k to c_1, $\delta \cdot c'_k$ is equal to c'_1. Thus, the element $\eta_1^{-1}\delta$ is contained in the stabilizer of c'_k under the action of $\pi_1(\Sigma'_{c_l} \setminus \{v_2\}, v_1)$, which is the infinite cyclic group generated by the loop obtained by connecting v_1 to a sufficiently small circle around v_2 using c'_k. This loop is the image of $\Delta^2_{c_k,c_l}$ under the homomorphism $\Phi^*_{c_l}$. $\square$

Remark 6.5. In general, the two pushing maps along η_1 and η_2 are different. For example, it is easy to see that the two pushing maps give distinct elements of $\mathrm{Mod}(\Sigma'_{c_l}; v_1, v_2)$ if c_1 and c_k are disjoint.

Now let us return to the mapping classes $\chi_i \in \mathrm{Mod}(\Sigma)$ associated via Proposition 6.3 with the shift deformations resulting in w'_i for $i = 1, 2$

Lemma 6.6. *For some integer m we have $\chi_i = \varphi_i \, t_{c_k} t_{c_l} t_{c_k} \Delta^m_{c_k,c_l} \in \mathrm{Mod}(\Sigma)$.*

Proof. We denote by P the subset of $\mathrm{Mod}(\Sigma)$ which maps $c_k \cup c_l$ to $c_l \cup c_1$, and by P' the set of isotopy classes of diffeomorphisms from (Σ'_{c_k,c_l}, v) to (Σ'_{c_1,c_l}, v'). The sets P and P' admit free and transitive actions of $\mathrm{Mod}(\Sigma)(\{c_k, c_l\})$ and $\mathrm{Mod}(\Sigma'_{c_k,c_l}, v)$ defined by compositions as maps, respectively. Using similar definitions as for the surgery homomorphisms used in Section 5.2 we obtain maps

$$\Phi^* : P \to P' \quad \text{and} \quad \Phi^*_{c_k,c_l} : \mathrm{Mod}(\Sigma)(\{c_k, c_l\}) \to \mathrm{Mod}(\Sigma'_{c_k,c_l}; v)$$

which are compatible with the actions on P and P'.[12] Note that both χ_i and φ_i are contained in P. Since the action of $\mathrm{Mod}(\Sigma)(\{c_k, c_l\})$ on P is transitive, we can find some $\xi \in \mathrm{Mod}(\Sigma)(\{c_k, c_l\})$ such that $\xi \cdot \chi_i = \varphi_i$. The image $\Phi^*(\chi_i)$ is the monodromy of $\hat{w}_i$ along γ under a suitable connection, while $\Phi^*(\varphi_i)$ is that along the curve obtained from γ by changing part of γ into the images γ_1 and γ_2 of joining curves (see Figure 6), in particular $\Phi^*(\chi_i) = \Phi^*(\varphi_i)$. Since the action of $\mathrm{Mod}(\Sigma'_{c_k,c_l}; v)$ on P' is free, ξ is contained in the kernel of $\Phi^*_{c_k,c_l}$ which is the infinite cyclic group generated by $t_{c_k} t_{c_l} t_{c_k}$. Since φ_i preserves c_k up to isotopy

---

[12]that is, for $\xi \in P$ and $\varphi \in \mathrm{Mod}(\Sigma)(\{c_k, c_l\})$ we have $\Phi^*(\xi \cdot \varphi) = \Phi^*(\xi) \cdot \Phi^*_{c_k,c_l}(\varphi)$

while χ_i does not, ξ is equal to $(t_{c_k} t_{c_l} t_{c_k})^{2m+1} = t_{c_k} t_{c_l} t_{c_k} \Delta_{c_k,c_l}^m$ for some integer m.

$\square$

Remark 6.7. According to Lemma 6.6, we can obtain χ_i once we determine φ_i. The kernel of Φ_{c_k,c_l}^* in Lemma 6.4 is generated by the Dehn twist along c_l. Moreover, φ_i maps c_k to c_1. Thus, we can easily obtain φ_i from the vanishing cycles of w using Lemma 6.4.

Remark 6.8 (Handleslides and stabilizations). For completeness, we would like to mention that the effect of the two remaining cases of Williams's basic homotopies, namely handleslides and stabilizations, was studied by the second author in [Hay14] Indeed, the effect of stabilizations are completely described by [Hay14, Theorems 6.5 and 6.7]) and the effects of handleslides can easily be understood using [Hay14, Theorem 3.9].

6.3. **Constructions of surface diagrams.** In this section we describe some constructions of simple wrinkled fibrations that involve cusp merges and show how to obtain surface diagrams. In order to do this we have to turn our theoretical understanding of connections in cusp merge into a method which is useful in practice.

Let $f\colon X \to B$ be a wrinkled fibration that contains a cusp merge configuration as depicted in Figure 7 and fix a connection $\mathcal{H}$ that is standard near the cusps. As indicated, we choose reference fibers Σ_i and Σ_i' near the cusps and another one Σ' further away in the lower genus region. We denote the vanishing cycles in Σ_i by c_i and d_i where the c_i correspond to the lower fold arcs.
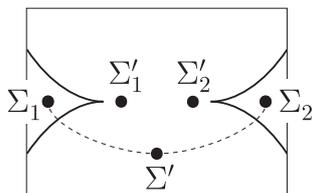

FIGURE 7. Matching surgered surface diagrams.

By parallel transport we can identify Σ' with the surgery on either $c_1 \subset \Sigma_1$ or $c_2 \subset \Sigma_2$ and each possibility gives rise to a pair of surgery disks $D_i^{\pm} \subset \Sigma'$ and an arc d_i' connecting the disks. Recall that with this data we can completely reconstruct the higher genus fiber Σ_i from Σ'. There is now an easy way to identify the higher genus fibers. We simply have to find a *matching isotopy* $\boldsymbol{\phi} = (\phi_t)$ of Σ' that moves the whole configuration $D_2^- \cup d_2' \cup D_2^+$ to $D_1^- \cup d_1' \cup D_1^+$. Clearly, such an isotopy induces a diffeomorphism $\bar{\phi}\colon \Sigma_2 \to \Sigma_1$.

To see that the identifications we obtain in this way are the same as the ones coming from cusp merges, we first observe that the choice of a matching isotopy is essentially equivalent to the choice of a path in $\mathbb{P}(T\Sigma')$. Indeed, it easy to see that from any path $\boldsymbol{\lambda}$ in $\mathbb{P}(T\Sigma')$ that connects a tangent spaces of d_2' and d_1' at interior points, we can construct a matching isotopy which realizes $\mathrm{push}(\boldsymbol{\lambda})$. On the other hand, any matching isotopy gives rise to such a path as the trace of a tangent space to d_2'. For concreteness, let us fix tangent spaces $T_i \in \mathbb{P}(T\Sigma')$ of d_i'. To make the connection to cusp merge homotopies, we denote the tangent lines in the lower genus fibers near the cusps by $L_i \in \mathbb{P}(\Sigma_i')$. We can assume that, after trivializing the lower genus region using $\mathcal{H}$, the lines L_i correspond to the lines T_i. As a consequence, given any path in $\mathbb{P}(T\Sigma')$ from T_2 to T_1, we can construct a framed joining curve covering the obvious joining curve and it is obvious that the
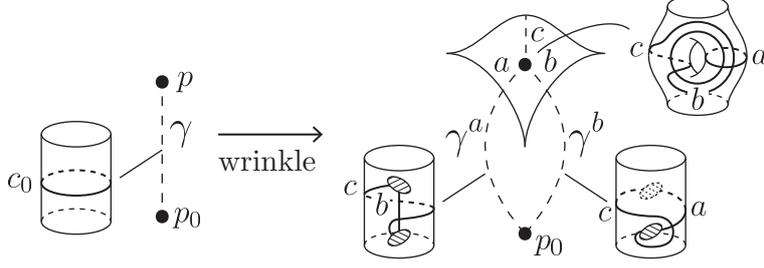
FIGURE 8. Surgered surface diagrams in a wrinkling homotopy.

parallel transport obtained after performing the corresponding cusp merge agrees with the identification obtained from the matching isotopy associated to $\boldsymbol{\lambda}$.

We will discuss examples of this procedure in Sections 6.3.1 and 6.3.2 below.

6.3.1. *Surface diagrams for Lefschetz fibrations.* Let $f \colon X \to B$ be a smooth map from an oriented 4–manifold X to an oriented surface B. A critical point $p \in \mathcal{C}(f)$ is called a *Lefschetz singularity* if there exist orientation preserving complex charts $\varphi \colon U \to \mathbb{C}^2$ and $\phi \colon V \to \mathbb{C}$ around p and $f(p)$, respectively, such that

$$\phi \circ f \circ \varphi^{-1}(z, w) = z^2 + w^2. \tag{22}$$

Furthermore, f is called a *Lefschetz fibration* if all its critical points are Lefschetz singularities. We also consider so called *achiral Lefschetz singularities* which have the same model as Equation (22) but the chart ϕ does not preserve the orientation. A map with only Lefschetz and achiral Lefschetz singularities is called an *achiral Lefschetz fibration.* It is clear from the local models that the critical points of an achiral Lefschetz fibration are isolated. In particular achiral Lefschetz fibrations are not stable. But Theorem 2.1 guarantees that we can perturb any achiral Lefschetz fibration so that it becomes a stable map. An explicit perturbation of the local model was given in [Lek09]:

$$W \colon [0, \infty) \times \mathbb{C}^2 \to \mathbb{C} \colon (s, z, w) \mapsto z^2 + w^2 + s \operatorname{Re}(z).$$

This perturbation is known as *wrinkling* and changes a Lefschetz singularity into a circular singularity which consists of indefinite folds and three indefinite cusps (see Figure 8). Observe that we can obtain a simple wrinkled fibration from a Lefschetz fibration over S^2 by first wrinkling all the Lefschetz singularities and then connecting all the components of singular sets by cusp merges. According to [Lek09], the vanishing cycles of the three folds in a fiber over the inner region of the triangle of critical values of the wrinkled Lefschetz singularity are related to the Lefschetz vanishing cycle as described in Figure 8. Combining this diagram with the method of matching isotopies described in Section 6.1 we can obtain a surface diagram of a Lefschetz fibration once we know its vanishing cycles. We will carry out this procedure for the genus–1 achiral Lefschetz fibration on S^4 due to Matsumoto [Mat82].

Example 6.9. Matsumoto's achiral Lefschetz fibration $f \colon S^4 \to S^2$ is constructed as follows. Let $\pi \colon Y \to D^2$ be a singular fibration with a regular fiber $F \cong T^2$ which has a Lefschetz and an achiral Lefschetz singularities with the same non-separating vanishing cycle $c \subset F$. It is easy to see that the restriction $\pi|_{\partial Y}$ is trivial. We take a bundle isomorphism $\Phi \colon \partial D^2 \times T^2 \to \partial Y$ so that the circle $\Phi(\partial D^2 \times \{*\})$ represents a generator of $H_1(Y; \mathbb{Z})$. We put $X = Y \cup_\Phi (D^2 \times T^2)$. It is not hard to see that X is diffeomorphic to S^4 and admits an achiral Lefschetz fibration $f = \pi \cup \operatorname{pr} \colon X = Y \cup (D^2 \times T^2) \to S^2$, where $\operatorname{pr} \colon D^2 \times T^2 \to D^2$ is

the projection onto the first component. Denote by $p_1 \in S^2$ (resp. $p_2 \in S^2$) the image of the Lefschetz (resp. achiral Lefschetz) singularity of f. After wrinkling the two singularities of f we obtain a wrinkled fibration $f_1 \colon S^4 \to S^2$ whose critical values and vanishing cycles in a regular fiber $f_1^{-1}(p_0)$ are described in the middle of Figure 9. We apply a cusp merge to f_1 along the dotted arrow in the figure. The
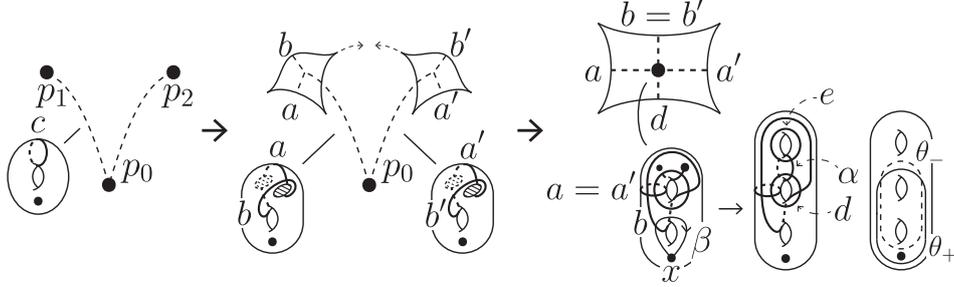

FIGURE 9. The procedure for obtaining a genus–2 simple wrinkled fibration.

resulting fibration $f_2 \colon S^4 \to S^2$ is a simple wrinkled fibration. Vanishing cycles of f_2 in a genus–2 regular fiber Σ are described in the right of Figure 9. We take a small disk $D_0 \subset S^2$ in the higher-genus region. We can deduce from the construction of f in the previous paragraph that an attaching loop of f_2 is $\beta \in \pi_1(\Sigma, x)$ described in the right of Figure 9. We replace a neighborhood of d in Σ with a once punctured torus as shown in Figure 9 and denote the resulting genus–3 surface by $\tilde{\Sigma}$. We take simple closed curves θ_+ and θ_- in Σ_3 as shown in Figure 9. Denote the mapping class $t_{t_a(d)}t_{t_b(a)}t_{t_a(b)}t_{t_d(a)}\mathrm{push}(\beta)t_{\theta_+}t_{\theta_-}^{-1} \in \mathrm{Mod}(\tilde{\Sigma}; x)$ by θ. The mapping class θ is contained in the group $\mathrm{Mod}(\Sigma; x)(d, e)$. Furthermore, we can easily verify that θ satisfies the condition W_3' in [Hay14, §6]. Thus, by [Hay14, Theorem 6.7] a surface diagram of a fibration obtained by stabilizing f_2 is as follows:

$$(\tilde{\Sigma}; a, b, a, d, \tilde{\theta}(\alpha), e, \alpha, d),$$

where $\tilde{\theta}$ denotes the mapping class $t_{t_a(d)}t_{t_b(a)}t_{t_a(b)}t_{t_d(a)}\theta^{-1} = t_{\theta_-}t_{\theta_+}^{-1}\mathrm{push}(\beta)^{-1}$. The simple closed curves in this surface diagram are shown in the upper half of Figure 10.
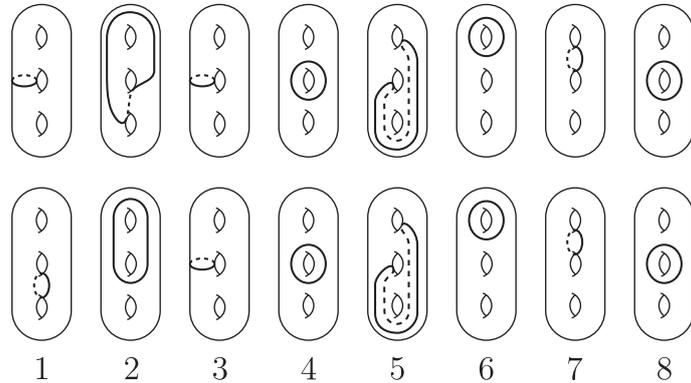

FIGURE 10. Surface diagrams corresponding to the two elements in $\pi_4(S^2) \cong \mathbb{Z}/2\mathbb{Z}$: the upper (resp. lower) one represents the element $1 \in \pi_4(S^2)$ (resp. $0 \in \pi_4(S^2)$).

Remark 6.10. Note that Matsumoto's fibration represents the non-trivial element of $\pi_4(S^2) \cong \mathbb{Z}/2\mathbb{Z}$. Indeed, it can be obtained as a perturbation of the map $h \circ Sh: S^4 \to S^2$, where $h: S^3 \to S^2$ is the Hopf fibration and $Sh: S^4 \to S^3$ is its suspension, which is known to generate $\pi_4(S^2)$ (see [Mat82, §3]). On the other hand, we can also realize the constant homotopy class as follows. One readily checks that the projection $\mathbb{R}^5 \to \mathbb{R}^2$ restricts to a stable map $\pi: S^4 \to S^2$ whose critical locus is a single circle of definite folds mapped injectively into the plane. Using a homotopy described in [Wil10, Example 2] we can trade the definite folds for indefinite ones. Applying stabilization twice to the resulting map, we eventually obtain a genus–3 simple wrinkled fibration. It is not hard to see that the surface diagram of this map is give by lower half of Figure 10. It is interesting to observe that, although both diagrams in Figure 10 describe the 4–manifold S^4, they cannot be related by the moves derived from Williams's basic homotopies since the corresponding simple wrinkled fibrations are not homotopic. Yet the diagrams share a striking resemblance. This indicates that there might be a reasonable diagrammatic way to relate surface diagrams derived from non-homotopic maps which is still an open problem at the time of writing.

6.3.2. *Surface diagrams from Heegaard diagrams.* Let M be a connected, closed, and oriented 3–manifold. In this section we will explain an algorithm to obtain a surface diagram of $S^1 \times M$ from a Heegaard diagram of M. It will follow from the construction that the corresponding simple wrinkled fibration is not surjective and is therefore homotopic to a constant map. Recall that a Heegaard diagram of M is a triple $(\Sigma_g; \boldsymbol{\alpha}, \boldsymbol{\beta})$, where Σ_g is a closed, oriented surface of genus g, and $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_g\}$ and $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_g\}$ are g-tuples of mutually disjoint simple closed curves on Σ_g both of which are linearly independent in $H_1(\Sigma_g)$. Any Heegaard diagram of M can be obtained from a Morse function $h : M \to \mathbb{R}$ with unique minima and maxima mapping to $-g-1$ and $g+1$, respectively, and the Heegaard surface at level 0 such the index 1 critical point with ascending sphere α_i maps to $-g-1+i$ and the index 2 critical point with descending sphere β_i maps to i. Using such a Morse function we obtain a stable map

$$\mathrm{id} \times h : S^1 \times M \to S^1 \times [-g-2, g+2]$$

which we can promote to a stable map $f_0 : S^1 \times M \to S^2$ by collapsing the boundary of the annulus $S^1 \times [-g-2, g+2]$. Note that the critical set of f_0 consists of $2g+2$ circles of folds that are mapped injectively into S^2 as parallel copies of the equator. The two outermost circles are definite while all others are indefinite. Using the deformation explained in [Wil10, Example 2] we can replace the definite folds by indefinite ones (see Figures 11(a) and 11(b)) and we obtain a wrinkled fibration $f_1 : S^1 \times M \to S^2$ (which in particular is surjective, unlike f_0). As a next step we want to move the outermost folds of f_1 into the equatorial region as indicated in Figures 11(b) to 11(d). This is possible by the following result.

Lemma 6.11. *Let D_1, D_2 be two disks in the polar regions of the complements of $\mathcal{D}(f_1)$ and let $\gamma \subset S^2$ be the arc shown in Figure 11(b). There exist diffeomorphisms*

$$\Phi : (S^1 \times M) \setminus \mathrm{Int}\big(f_1^{-1}(D_1) \amalg f_1^{-1}(D_2)\big) \to S^1 \times f_1^{-1}(\gamma) \ and$$
$$\phi : S^2 \setminus \mathrm{Int}(D_1 \amalg D_2) \to S^1 \times I,$$

such that $(\mathrm{id} \times h_1) \circ \Phi = \phi \circ f_1$, where h_1 is the restriction $f_1|_{f_1^{-1}(\gamma)}$.

Proof. Let B_i be a disk in S^2 which contains D_i and either of the outermost images of folds, and is disjoint from the other images of folds of f_1. It is easy to

(a) The map f_0.   (b) The map f_1.   (c)   (d) The map f_2.
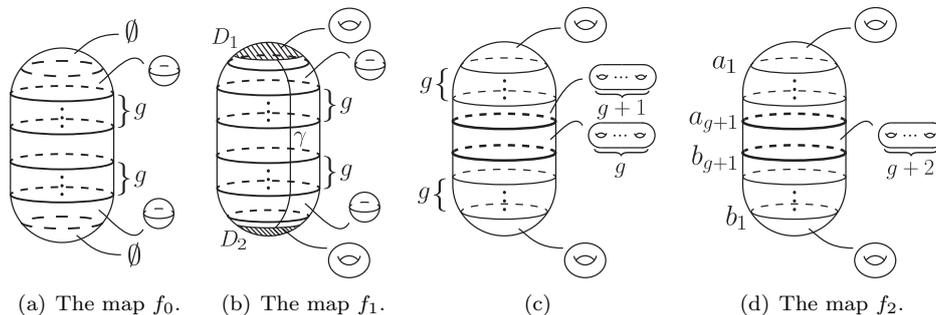
FIGURE 11. The critical values and fibers in various stages of the algorithm.

see by construction of f_1 that there exist desired diffeomorphisms Φ and ϕ defined on $(S^1 \times M) \setminus \mathrm{Int}(f_1^{-1}(B_1) \amalg f_1^{-1}(B_2))$ and $S^2 \setminus \mathrm{Int}(B_1 \amalg B_2)$, respectively. Next, we observe that the manifold $(S^1 \times M) \setminus \mathrm{Int}(f_1^{-1}(D_1) \amalg f_1^{-1}(D_2))$ is obtained from $(S^1 \times M) \setminus \mathrm{Int}(f_1^{-1}(B_1) \amalg f_1^{-1}(B_2))$ by a round 1-handle attachment (for a precise definition of round handles, see [Bay09], for example). This round 1-handle is untwisted and is attached with 0-framing since the monodromy of the higher-genus side of the inner-most fold is trivial. This implies the statement in Lemma 6.11. $\square$
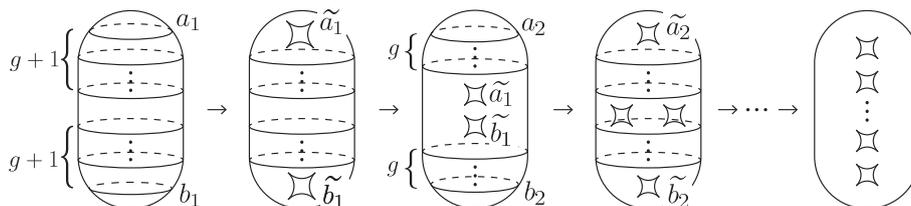
FIGURE 12. Successive application of flip and slips. The far right figure is a base diagram of f_3.

We denote the thus obtained map with critical values as in Figure 11(d) by $f_2 : S^1 \times M \to S^2$ and label the images of the fold circles by $a_1, \ldots, a_{g+1}, b_1, \ldots, b_{g+1}$ as indicated. As shown in Figure 12, we next apply so called *flip and slip* moves to the outermost folds a_1 and b_1 of f_2 to obtain a new map having indefinite circles $\widetilde{a_1}$ and $\widetilde{b_1}$ with four cusps in place of a_1 and a_2. These circles can be moved into the equatorial region and we can repeat this construction until all fold circles of f_2 have been replaced by indefinite circles with four cusps as on the right of Figure 12. We denote the resulting wrinkled fibration by f_3. Finally, we can obtain a simple wrinkled fibration $w_h : S^1 \times M \to S^2$ by applying cusp merges to f_3.

We will now explain how to perform the cusp merges in such a way that we are able to obtain a surface diagram of w_h from the information in the Heegaard diagram for M. First of all, we need to know surgered surface diagrams for all components $\widetilde{a_i}$ and $\widetilde{b_i}$ of the critical value set of f_3. As explained in the previous paragraph, the fibration f_3 is obtained by successive application of flip and slips followed by deformations that move cusped circles to the equatorial region. Using the results of the second author about flip and slips [Hay14, Section 6], it is therefore enough to understand surgered surface diagrams before the flip and slips. The first flip and slips are applied to folds a_1 and b_1, whose lower-genus fibers are tori. According to [Hay14, Theorem 6.7], in order to obtain surgered surface diagrams of $\widetilde{a_1}$ and $\widetilde{b_1}$, we have to look at not only monodromies along the boundaries of the

disks B_i, $i = 1, 2$, which are trivial by Lemma 6.11, but also sections over these disks. According to Lemma 6.11 we can choose an

$$S^1 \times M \cong S^1 \times f_1^{-1}(\gamma) \cup_{\varphi_1 \amalg \varphi_2} (D_1 \times T^2 \amalg D_2 \times T^2)$$

where $\varphi_i : \partial D_i \times T^2 \to S^1 \times \partial f^{-1}(\gamma)$ is a gluing diffeomorphism. By identifying $\partial f_1^{-1}(\gamma)$ with a pair of tori we can regard φ_i as a self-diffeomorphism of $S^1 \times T^2$. Note that we can understand the behavior of the sections over B_1 and B_2 once we know the maps φ_1 and φ_2. Thus, the following lemma enables us to obtain surgered surface diagrams of $\widetilde{a_1}$ and $\widetilde{b_1}$.

Lemma 6.12. *For a suitable identification $\partial f_1^{-1}(\gamma) \cong T^2 \amalg T^2$, there exists an embedding $\theta : S^1 \to T^2$ whose image intersects a vanishing cycle of the fold in one point such that the restriction $\varphi_i|_{S^1 \times \{*\}}$ is isotopic to the map*

$$S^1 \ni z \mapsto (z, \theta(z)) \in S^1 \times T^2.$$

Proof. Let H_1 be a solid torus. We take a small ball B^3 in the interior $\text{Int}(H_1)$. By Lemma 6.11, we can identify $f_1^{-1}(B_i \setminus \text{Int}(D_i))$ with $S^1 \times (H_1 \setminus B^3)$. Thus, we can regard $f_1^{-1}(B_i)$ as the manifold obtained by attaching $D^2 \times T^2$ to $S^1 \times (H_1 \setminus B^3)$. Manifold of this form were extensively studied in the proof of [Hay11, Theorem 4.2]. It is easy to see from the construction of f_1 that the manifold $f_1^{-1}(B_i)$ is diffeomorphic to $S^1 \times D^3$. Thus, by the arguments in [Hay11] the attaching map φ_i satisfies the desired condition. $\square$

The other flip and slips are applied to folds over a_i and b_i ($i \geq 2$) whose lower genus fibers have genus two or higher. Thus, in these cases, we can obtain surgered surface diagrams of $\widetilde{a_i}$ and $\widetilde{b_i}$ once we know the monodromies of the higher-genus sides of a_i and b_i. These monodromies are no longer trivial, yet we can easily obtain them by looking at lifts of surgered diagrams of the previous folds $\widetilde{a_{i-1}}$ and $\widetilde{b_{i-1}}$ (see also the example below).

Example 6.13. Here we consider the genus 0 Heegaard diagram $(S^2, \emptyset, \emptyset)$ of S^3. We denote by $f_i : S^1 \times S^3 \to S^2$, $i = 0, \ldots, 3$ the maps obtained in the various steps of the algorithm explained above. It is easy to see that surgered surface diagrams of f_2 are as described in Figure 13. We choose sections of f_2 over the disks B_1
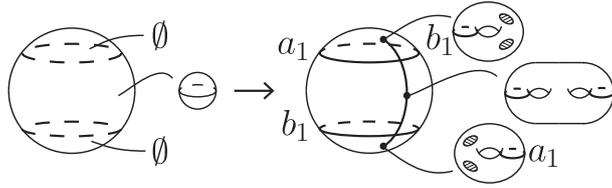

FIGURE 13. Left: a base diagram of f_0. Right: that of f_2.

and B_2 as in the proof of Lemma 6.11. These sections lift the monodromies of the higher-genus sides of the folds a_1 and b_1 to the mapping class group with a marked point x. We can choose a homotopy from f_0 to f_2 and sections so that the monodromy along a_1 (resp. b_1) is equal to the pushing map along θ_1 (resp. θ_2), where θ_i is the oriented loop described in Figure 14(a). Denote by Σ_1 a regular fiber of f_3 which is mapped to the inside of $\widetilde{a_1}$. We consider the simple closed curve $\theta_1^{\pm} \subset \Sigma_1$ as described in Figure 14(c) and put $\varphi_1 = t_{\theta_1^+} t_{\theta_1^-}^{-1}$. By [Hay14, Theorem 6.7] a surface diagram derived from $\widetilde{a_1}$ is $\mathfrak{S}_a = (\Sigma_1; d_1, \varphi_1^{-1}(\delta_1), a_1, \delta_1)$, where δ_1 and d_1 are the simple closed curves in Σ_1 described in Figure 14(b). We

can also verify in a similar manner that a surface diagram derived from the critical set $\widetilde{b_1}$ is $\mathfrak{S}_b = (\Sigma_2; e_1, \epsilon_1, b_1, \varphi_2^{-1}(\epsilon_1))$, where we put $\varphi_2 = t_{\theta_2^+} t_{\theta_2^-}^{-1}$.
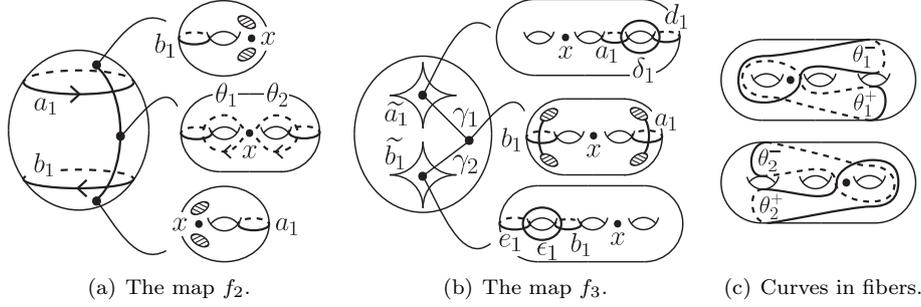

FIGURE 14. Base diagrams and curves.

We consider arcs γ_1 and γ_2 in S^2 as in Figure 14(b). We denote by Σ_0 the fiber on the intersection between γ_1 and γ_2. These paths give surgered surface diagrams which are described in Figure 15(a). We denote these diagrams by $\mathfrak{S}'_a$
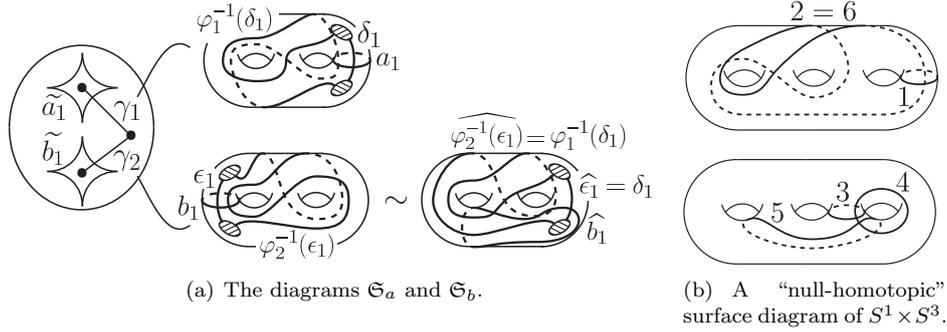

FIGURE 15. (Surgered) surface diagrams.

and $\mathfrak{S}'_b$. To obtain a surface diagram of the fibration after applying a cusp merge we change the diagram $\mathfrak{S}'_b$ by a matching isotopy of Σ_0 (as described in Figure 15(a)) so that the surgery disks and the arc ϵ_1 in $\mathfrak{S}'_b$ coincide with the surgery disks and the arc δ_1 in $\mathfrak{S}'_a$. We denote by $\widehat{b_1}$, $\widehat{\epsilon_1}$ and $\widehat{\varphi_2^{-1}(\epsilon_1)}$ the curves and arcs in Σ_0 obtained after applying the above modification to b_1, ϵ_1 and $\varphi_2^{-1}(\epsilon_1)$, respectively. We can regard $\widetilde{b_1}$ and $\widehat{\varphi_2^{-1}(\epsilon_1)}$ as simple closed curves in Σ_1. Note that the curve $\widehat{\varphi_2^{-1}(\epsilon_1)}$ is equal to $\varphi_1^{-1}(\delta_1)$. We eventually obtain the following surface diagram $(\Sigma_1; d_1, \varphi_1^{-1}(\delta_1), a_1, \delta_1, \widehat{b_1}, \varphi_1^{-1}(\delta_1))$ of $S^1 \times S^3$ which is shown in Figure 15(b).

APPENDIX I. TRUNCATING HOMOTOPIES

In this appendix we will give a way to make a given homotopy compactly supported, that is, we will find another homotopy which is constant outside a compact set and coincides with a given one in some open subset of the support. Although we merely need to find a compactly supported model of a cusp merge in this paper, the recipe we will give here can be applied in much more general situations (see Example I.3).

For a subset $A \subset N \times J$ we denote the intersection $A \cap N \times \{-1\}$ by A_{-1}. Let $\pi : N \times J \to J$ be the projection.

Theorem I.1. *Let $F = (f_t)$ be a homotopy of a map from N to P, $K \subset P$ a closed set diffeomorphic to a p–ball and $R \subset \tilde{F}^{-1}(K \times J)$ a codimension 0 compact submanifold which may have boundaries and corners. Suppose that the following conditions are satisfied:*

- *there exists a diffeomorphism $\nu\partial K \to \partial K \times (-\delta, \delta)$, where $\nu\partial K$ is a tubular neighborhood of ∂K, such that f_t is transverse to $K \times \{s\}$ and the restriction*

$$f_t : f_t^{-1}(\partial K \times \{s\}) \to \partial K \times \{s\}$$

  is locally stable for any t and s (see Section 2.2.1 for the definition of local stability),
- *the boundary ∂R is decomposed into three parts $\partial_c R$, $\partial_s R$ and $\partial_e R$ so that*
  - *the set of corners of R is the union of the intersections $\partial_c R \cap \partial_s R$, $\partial_c R \cap \partial_e R$ and $\partial_s R \cap \partial_e R$,*
  - *$\partial_e R = R \cap (N \times \{-1, 1\})$,*
  - *$\partial_c R$ is contained in $F^{-1}(\partial K)$,*
  - *the restriction $\widetilde{F}|_{\partial_s R} : \partial_s R \to K \times J$ is a surjective submersion.*

Then there exist two neighborhoods $R_1 \subset R_2$ of ∂R and a level preserving diffeomorphism Φ (resp. ϕ) of $N \times J$ (resp. $P \times J$) such that the following conditions hold:

- *the support of Φ is contained in R_2 and that of ϕ is contained in $\nu\partial K$,*
- *$f_{-1} = \phi_t^{-1} \circ f_t \circ \Phi_t$ on R_1, where Φ_t and ϕ_t are the maps satisfying the equations $\Phi(x, t) = (\Phi_t(x), t)$ and $\phi(y, t) = (\phi_t(y), t)$, respectively.*

Remark I.2. Using Theorem I.1 we can obtain a compactly supported homotopy g_t which coincides with f_t in some open subset of the support as follows:

$$g_t(x) = \begin{cases} \phi_t \circ f_{-1} \circ \Phi_t^{-1}(x) & ((x, t) \in R_1 \cup R^c) \\ f_t(x) & ((x, t) \in R) \end{cases}$$

Proof of Theorem I.1. Suppose that $\partial_s R$ is not empty. The restriction $f_{-1} : \partial_s R_{-1} \to K$ is a fiber bundle whose fiber M is a closed manifold. Since K is contractible $\partial_s R_{-1}$ is isomorphic to $M \times K$ as a fiber bundle. The restriction $\pi : \partial_s R \to J$ is also a fiber bundle with a fiber $\partial_s R_{-1}$. Thus $\partial_s R$ is isomorphic to $M \times K \times J$. Let X_0 be an everywhere non-zero vector field on a neighborhood of $\partial_s R$ contained in $\ker(d\tilde{F})$. We can obtain the following bundle isomorphism by integrating X_0:

$$\Psi : \nu\partial_s R \to M \times (-\varepsilon, \varepsilon) \times K \times J,$$

where $\varepsilon > 0$ is sufficiently small and $\nu\partial_s R$ is a neighborhood of $\partial_s R$ in $\tilde{F}^{-1}(K \times J)$. We take the above isomorphism so that $\Psi^{-1}(M \times [0, \varepsilon) \times K \times J)$ is in R. If $\partial_s R$ is empty, then we do not need such an isomorphism. In what follows, we will take vector fields with several conditions assigned to bold symbols. Starred conditions are related with the isomorphism Ψ, which are not needed if $\partial_s R$ is empty.

We denote the union $\partial_c R \cup \Psi^{-1}(M \times (-\varepsilon, \varepsilon) \times \partial K \times J)$ by $\nu_f \partial_c R$, which is a neighborhood of $\partial_c R$ in $F^{-1}(\partial K)$. (Note that $\nu_f \partial_c R = \partial_c R$ if $\partial_s R$ is empty.) By the assumptions the restriction $\pi : \nu_f \partial_c R \to J$ is a fiber bundle. We take a vector field X_1 on $\nu_f \partial_c R$ which satisfies the following conditions:

A the image $d\pi(X_1)$ is equal to $\frac{d}{dt} \in TJ$,

B* the image $d\Psi(X_1)$ is equal to $\frac{d}{dt} \in T(M \times (-\varepsilon, \varepsilon) \times K \times J)$ on $\nu\partial_s R \cap \nu_f \partial_c R$, where t is the coordinate of the interval J.

Integrating X_1 we take the following bundle isomorphism:

$$\Theta : \nu_f \partial_c R \to \nu_f \partial_c R_{-1} \times J.$$

We denote a neighborhood of $\nu_f \partial_c R$ in $F^{-1}(\nu \partial K)$ by $\nu \partial_c R$. Since f_t is transverse to $\partial K \times \{s\}$ for any $s \in (-\delta, \delta)$, the following composition is a submersion:

$$\nu \partial_c R \xrightarrow{F} \nu \partial K \cong \partial K \times (-\delta, \delta) \xrightarrow{\text{proj}} (-\delta, \delta).$$

We denote the above composition by $p : \nu \partial_c R \to (-\delta, \delta)$. We take a vector field X_2 on $\nu \partial_c R$ so that it satisfies the following conditions:

C the image $dp(X_2)$ is equal to the vector field $\frac{d}{ds}$,

D* the image $d\Psi(X_2)$ is $\frac{d}{ds} \in T(M \times (-\varepsilon, \varepsilon) \times K \times J)$, where s is the coordinate for the second component of $\nu \partial K \cong \partial K \times (-\delta, \delta)$.

Using X_2 and Θ we obtain the following bundle isomorphism:

$$\hat{\Theta} : \nu \partial_c R \to \nu_f \partial_c R_{-1} \times (-\delta, \delta) \times J.$$

It is easy to verify that the composition

$$\tilde{F} \circ \hat{\Theta}^{-1} : \nu_f \partial_c R_{-1} \times (-\delta, \delta) \times J \to \partial K \times (-\delta, \delta) \times J$$

preserves the latter two components, that is, there exists a two parameter family $g_{s,t} \in C^\infty(\nu_f \partial_c R_{-1}, \partial K)$ such that $\tilde{F} \circ \hat{\Theta}^{-1}(x, s, t)$ is equal to $(g_{s,t}(x), s, t)$ for any (x, s, t). We denote the composition $\tilde{F} \circ \hat{\Theta}^{-1}$ by $\hat{F}$. By the assumption $g_{s,t}$ is locally infinitesimally stable for any t and s. Thus by the same argument as that in the proofs of Propositions 4.3, 4.5 and Theorem 4.6 in [GG73], we can obtain a vector field ξ on $\nu_f \partial_c R_{-1} \times (-\delta, \delta) \times J$ and a vector field η on $\partial K \times (-\delta, \delta) \times J$ which satisfy the following conditions:

E for any $(x, s, t) \in \nu_f \partial_c R_{-1} \times (-\delta, \delta) \times J$ the vector $\xi(x, s, t)$ is contained in $T\nu_f \partial_c R_{-1}$,

F for any $(y, s, t) \in \partial K \times (-\delta, \delta) \times J$ the vector $\eta(y, s, t)$ is contained in $T\partial K$,

G $t\hat{F}(\xi) + \omega \hat{F}(\eta) = t\hat{F}\left(d\hat{\Theta}(\hat{\Theta}^{-1})^*\left(\frac{\partial}{\partial t}\right)\right) - \hat{F}^*\left(\frac{\partial}{\partial t}\right)$,

H* for any $x \in \hat{\Theta}^{-1}(\nu \partial_s R \cap \nu \partial_c R)$ the image $d(\Psi \circ \hat{\Theta}^{-1})(\xi(x))$ is contained in TK.

Let $\varrho_1 : (-\delta, \delta) \to \mathbb{R}$ be a non-negative smooth function whose value is 1 on $\left[-\frac{\delta}{3}, \frac{\delta}{3}\right]$ and 0 outside $\left[-\frac{2\delta}{3}, \frac{2\delta}{3}\right]$. Since we identify $\nu \partial K$ with $\partial K \times (-\delta, \delta)$ we can regard ϱ_1 as a function defined on $\nu \partial K$, and this function can be extended to that defined on a neighborhood of $P \times J$ so that it vanishes on the complement of $\nu \partial K \times J$. We also denote the extended function by ϱ_1, and let $\tilde{\eta} = \varrho_1 \eta$. Since $\tilde{\eta}$ is compactly supported the vector field $\tilde{\eta} + \frac{\partial}{\partial t}$ on $P \times J$ is complete. We denote the integral curve of it with the initial point $y \in P \times J$ by $c_y(t)$. We define a self-diffeomorphism ϕ of $P \times J$ as follows:

$$\phi(p, t) = c_{(p,-1)}(1 + t).$$

We can deduce from the condition F that the diffeomorphism ϕ preserves $\nu \partial K$.

We denote the set

$$\hat{\Theta}^{-1}\left(\left(\Psi^{-1}\left(M \times \left[-\frac{\varepsilon}{n}, \frac{\varepsilon}{n}\right] \times \partial K \times \{-1\}\right) \cup \nu_f \partial_c R_{-1}\right) \times \left[-\frac{\delta}{n}, \frac{\delta}{n}\right] \times J\right)$$

by $W_{1,n}$, and the set $\Psi^{-1}\left(M \times \left[-\frac{\varepsilon}{n}, \frac{\varepsilon}{n}\right] \times K \times J\right)$ by $W_{2,n}$. We take a vector field $\tilde{\xi}$ on $\nu \partial_c R \cup \nu \partial_s R$ which satisfies the following properties:

I the vector field $\tilde{\xi}$ is equal to $d\hat{\Theta}^{-1}\hat{\Theta}^*(\xi)$ on $W_{1,3}$,

J* for $x = \Psi^{-1}(z, r, q, t) \in W_{2,3}$ the vector $\tilde{\xi}(x)$ is equal to

$$\varrho_1(q)d\hat{\Theta}^{-1}(\xi(\hat{\Theta}(x))) + (1 - \varrho_1(q))d\Psi^{-1}\left(d\tilde{F}_x\left(\frac{\partial}{\partial t}\right)_x + d\phi_{\phi^{-1}(q,t)}\left(\frac{\partial}{\partial t}\right)_{\phi^{-1}(q,t)}\right),$$

where the vectors $d\tilde{F}_x \left(\frac{\partial}{\partial t}\right)_x$ and $d\phi_{\phi^{-1}(q,t)} \left(\frac{\partial}{\partial t}\right)_{\phi^{-1}(q,t)}$ on $K \times J$ are regarded as those on $M \times (-\varepsilon, \varepsilon) \times K \times J$ in the obvious way.

We also take a function $\varrho_2 : N \times J \to \mathbb{R}$ so that $\varrho_2 \equiv 1$ on $W_{1,6} \cup W_{2,6}$ and $\varrho_2 \equiv 0$ outside $W_{1,4} \cup W_{2,4}$. Since the vector field $\varrho_2 \tilde{\xi}$ is compactly supported, the vector field $-\varrho_2 \tilde{\xi} + \frac{\partial}{\partial t}$ on $N \times \mathbb{R}$ is complete. We denote the integral curve of $-\varrho_2 \tilde{\xi} + \frac{\partial}{\partial t}$ with the initial point $x \in \tilde{N}$ by $C_x(t)$. We define a self-diffeomorphism Φ of $N \times J$ as follows:

$$\Phi(x) = \begin{cases} C_{\hat{\Theta}(y_{-1}, s, t)}(1 + t) & (x = \hat{\Theta}^{-1}(y, s, t) \in W_{1,3}) \\ C_{\Psi(z, r, q_{-1}, t)}(1 + t) & (x = \Psi^{-1}(z, r, q, t) \in W_{2,3}) \\ x & \text{(otherwise)}, \end{cases}$$

where $y_{-1} \in \psi^{-1}\left(\partial(\partial_c R) \times \left[-\frac{\varepsilon}{3}, \frac{\varepsilon}{3}\right]\right)$ is $\Theta^{-1}(p_1 \circ \Theta(y), -1)$ and $q_{-1} \in K \times J$ is $(p_1(q), -1)$ (p_1 represents the projection onto the first component). Let R_1 and R_2 denote the unions $W_{1,6} \cup W_{2,6}$ and $W_{1,3} \cup W_{2,3}$, respectively. The support of Φ is contained in R_2. Furthermore it is easy to see that Φ preserves the sets $W_{1,6}$ and $W_{2,6}$. The diffeomorphisms Φ and ϕ preserve the time components. We put $\Phi(x, t) = (\Phi_t(x), t)$ and $\phi(p, t) = (\phi_t(p), t)$.

We denote the composition $\phi^{-1} \circ \tilde{F} \circ \Phi$ by Ω. By the same argument as that in the proof of Theorem 3.3 in [GG73], we can obtain the following equation:

$$d\Omega \left(\frac{\partial}{\partial t}\right) - \Omega^* \left(\frac{\partial}{\partial t}\right) = 0 \text{ on } R_1,$$

We can deduce from this equation that $\phi_t^{-1} \circ f_t \circ \Phi_t$ is equal to f_{-1} on R_1 (see [GG73, Lemma 3.2]). □

Example I.3. Here we consider the cusp merge model $F = (\mu_s)$ defined as follows:

$$\mu_s : \mathbb{R}^4 \to \mathbb{R}^2, \ \mu_s(t, x, y, z) = (t, x^3 - 3(t^2 + s)x + y^2 - z^2).$$

For $\varepsilon > 0$, we denote by K_ε a closed disk neighborhood of the image of the joining curve obtained by smoothing the corners of the rectangle $[-1 - \varepsilon, 1 + \varepsilon] \times [-\varepsilon, \varepsilon]$. Assume that K_ε is sufficiently close to the rectangle so that μ_s is transverse to the boundary ∂K_ε. We take a tubular neighborhood $\nu \partial K_\varepsilon \cong \partial K_\varepsilon \times (-\delta, \delta)$ of ∂K_ε so that the map $\mu^{-1}(\nu \partial K_\varepsilon) \xrightarrow{\mu_s} \nu \partial K_\varepsilon \xrightarrow{p} \partial K_\varepsilon$ has only Morse singularities and thus locally infinitesimally stable for any $s \in [-1, \varepsilon]$, where $p : \nu \partial K_\varepsilon \to \partial K_\varepsilon$ is the projection. Let $(\mathcal{H}_s)$ be a smooth one parameter family of distributions such that $\mathcal{H}_s$ is a connection of μ_s for each s. It is easy to verify that the fiber $\mu_s^{-1}(t, 0)$ is a disk if $t^2 \leq -s$ and a once punctured torus if $t^2 > -s$. We take a two parameter family $(S_{s,t})$ of circles such that $S_{s,t}$ is contained in $\mu_s^{-1}(t, 0)$ and $S_{s,t}$ bounds a compact subsurface containing $(t, 0, 0, 0)$ in $\mu_s^{-1}(t, 0)$. We can obtain a three dimensional submanifold $\partial_s R \subset \tilde{F}^{-1}(K_\epsilon \times [-1, \varepsilon])$ by taking a flow from $(S_{s,t})$ with respect to $\mathcal{H}_s$. The complement of $\partial_s R \cup \tilde{F}^{-1}(\partial K_\varepsilon \times [-1, \varepsilon])$ has a compact component which contains the origin of $\mathbb{R}^4 \times \mathbb{R}$. We denote this component by R and let $\partial_e R = \partial R \cap \mathbb{R}^4 \times \{-1, \varepsilon\}$ and $\partial_c R = \overline{\partial R \setminus (\partial_s R \cup \partial_e R)}$. It is easy to see that K and R satisfy the conditions in Theorem I.1. Thus we can obtain a compactly supported homotopy which coincides with f_s in an open subset of its support.

Note that the above construction of a compactly supported homotopy is also valid in much more general situation. Indeed, the same recipe can be applied to a homotopy g_s if we have a disk K such that the restriction of g_s over ∂K is a Morse function and $\mathcal{C}(\tilde{G})$ is compact. In particular, using Theorem I.1 we can obtain a compactly supported model for each versal unfolding appearing in Theorem 2.6.

APPENDIX II. SYMMETRIES OF FOLDS AND CUSPS

The purpose of this appendix is to explain how to deal with ambiguities that arise when working with local models mentioned. Let $f\colon N \to P$ be a smooth map and suppose that the germ of f at $x \in N$ has a local model $\mathfrak{f}\colon (\mathbb{R}^n, 0) \to (\mathbb{R}^p, 0)$, that is, there are coordinates around x and $f(x)$ in which f is represented by $\mathfrak{f}$. If we choose a second set of coordinates with this property, then the germs of the two coordinate sets differ by germs of local diffeomorphisms ρ of $\mathbb{R}^n$ and λ of $\mathbb{R}^p$, respectively, which preserve the germ of $\mathfrak{f}$ at the origin, that is, we have

$$\mathfrak{f} = \lambda \circ \mathfrak{f} \circ \rho^{-1}.$$

This leads us to study the group of pairs of such local diffeomorphisms.

Symmetries of map-germs. More systematically, we denote by $\mathcal{E}(n, p)$ the set of germs of smooth maps $(\mathbb{R}^n, 0) \to (\mathbb{R}^p, 0)$ and let $\mathcal{R}$ and $\mathcal{L}$ be the groups of local diffeomorphism-germs defined near the origins of $\mathbb{R}^n$ and $\mathbb{R}^p$, respectively; that is,

$$\mathcal{R} = \{\rho \in \mathcal{E}(n, n) \,|\, \det(d\rho_0) \neq 0\} \quad \text{and} \quad \mathcal{L} = \{\lambda \in \mathcal{E}(p, p) \,|\, \det(d\lambda_0) \neq 0\}.$$

Writing $\mathcal{A} = \mathcal{R} \times \mathcal{L}$ for the product we get a split exact sequence

$$1 \longrightarrow \mathcal{R} \overset{i}{\longrightarrow} \mathcal{A} \overset{p}{\longrightarrow} \mathcal{L} \longrightarrow 1 \tag{23}$$

where $i(\rho) = (\rho, \mathrm{id})$ and $p(\rho, \lambda) = \lambda$. The group $\mathcal{A}$ naturally acts on $\mathcal{E}(n, p)$ by the so called *right-left action* $(\rho, \lambda) \cdot \mathfrak{f} = \lambda \circ \mathfrak{f} \circ \rho^{-1}$ which restricts to the *right action* of $\mathcal{R}$ and the *left action* of $\mathcal{L}$, whence the names. The above considerations about local models lead us to study the *right-left symmetries* or $\mathcal{A}$–symmetries of a given germ $\mathfrak{f} \in \mathcal{E}(n, p)$ which are defined as the stabilizer

$$\mathcal{A}(\mathfrak{f}) = \left\{(\rho, \lambda) \in \mathcal{A} \,\middle|\, \lambda \circ \mathfrak{f} \circ \rho^{-1} = \mathfrak{f}\right\}.$$

Of course, there are similar definitions involving $\mathcal{R}$ and $\mathcal{L}$. As a last bit of notation, we denote groups of orientation preserving diffeomorphisms and symmetries by adding a superscript as in $\mathcal{A}^+$, $\mathcal{A}^+(\mathfrak{f})$, etc.

For $n \geq p$ the groups of $\mathcal{A}$–symmetries are typically quite large. But it turns out that for sufficiently well-behaved germs – more precisely for *finitely $\mathcal{A}$–determined* germs which includes the stable ones – they share a convenient property with finite dimensional Lie groups: according to Jänich [Jän78] and Wall [Wal80], they have "maximal compact Lie subgroups" which are unique up to conjugacy and the quotient by such a group is "contractible". However, as the quotation marks indicate, this statement should be taken with a grain of salt. In fact, $\mathcal{A}$ and its subgroups are not topological groups in any natural way.[13] In any case, these delicacies shall not concern us. The interested reader is referred to the exposition of du Plessis and Feragen in [dPF11] which contains the most recent and complete account on these matters. In order to convince the reader that these are in fact the right notions we state the following immediate consequence of the definitions (that we did not give).

Fact II.1. *If $G_{\mathcal{A}}(\mathfrak{f}) \subset \mathcal{A}(\mathfrak{f})$ is a maximal compact subgroup, then any element of $\mathcal{A}(\mathfrak{f})$ can be deformed to an element of $G_{\mathcal{A}}(\mathfrak{f})$ by a smooth family in $\mathcal{A}(\mathfrak{f})$ (where a family of germs $\mathfrak{f}_s \in \mathcal{E}(n, p)$, $s \in I$, is called smooth if there exists a smooth map $F\colon (I \times \mathbb{R}^n, I \times 0) \to (\mathbb{R}^p, 0)$ such that $F(s, \cdot)$ represents $\mathfrak{f}_s$).*

In order to determine maximal compact subgroups of $\mathcal{A}$–symmetries for the fold and cusp singularities, we will use results of du Plessis–Wilson [dPW91] and du Plessis–Feragen [dPF11]. These provide an analogue of the exact sequence (23)

---

[13]This is due to the more general lack of a natural topologies on sets of smooth map-germs.

for maximal compact subgroups and, in addition, completely determine the topology of maximal compact subgroups of $\mathcal{R}(\mathfrak{f})$. We begin with the latter.

Theorem II.2 (du Plessis–Wilson [dPW91])**.** *Let $\mathfrak{f} \in \mathcal{E}(n,p)$ be a stable germ. Then the group $\mathcal{R}(\mathfrak{f})$ of right symmetries of $\mathfrak{f}$ has a maximal compact subgroup $G_{\mathcal{R}}(\mathfrak{f})$ which is unique up to conjugation. Moreover, if $n \geq p$, then*

$$G_{\mathcal{R}}(\mathfrak{f}) \cong \begin{cases} \mathrm{O}(n-p) & \text{if } \mathrm{rk}(d\mathfrak{f}|_0) = p \\ \mathrm{O}(i) \times \mathrm{O}(r-i) & \text{if } \mathrm{rk}(d\mathfrak{f}|_0) = p-1 \\ \{1\} & \text{if } \mathrm{rk}(d\mathfrak{f}|_0) \leq p-2 \end{cases} \tag{24}$$

where in the second case r and i denote the rank and index of the intrinsic second derivative $\delta^2|_0\mathfrak{f}\colon \ker(d\mathfrak{f}|_0) \times \ker(d\mathfrak{f}|_0) \to \mathrm{coker}(d\mathfrak{f}|_0)$.

Before we can state the exact sequence involving maximal compact subgroups of $\mathcal{R}(\mathfrak{f})$ and $\mathcal{A}(\mathfrak{f})$, we have to understand the image of $\mathcal{A}(\mathfrak{f})$ in $\mathcal{L}$ for stable $\mathfrak{f}$. It is easy to see that if $(\rho, \lambda) \in \mathcal{A}(\mathfrak{f})$, then $\lambda \in \mathcal{L}$ satisfies the following two conditions.

- *The discriminant condition:* λ preserves the (germ of) the discriminant $\mathcal{D}(\mathfrak{f}) = f(\mathcal{C}(f))$.
- *The cokernel condition:* If $\mathrm{rk}\, d\mathfrak{f}_0 = p-1$ and $r \neq 2i$, then $d\lambda_0$ preserves the orientation of $\mathrm{coker}(d\mathfrak{f}_0)$.

These conditions define a subgroup of $\mathcal{L}$ which is denoted by $\mathrm{Inv}^{\sim}(\mathcal{D}(\mathfrak{f}))$ in [dPW91]. According to [dPW91, Lemma 4.1], $\mathcal{A}(\mathfrak{f})$ is mapped onto $\mathrm{Inv}^{\sim}(\mathcal{D}(\mathfrak{f}))$ if $\mathfrak{f}$ is stable.

Theorem II.3 (du Plessis–Feragen [dPF11])**.** *Let $\mathfrak{f} \in \mathcal{E}(n,p)$ be a stable germ such that any linear compact subgroup of $\mathrm{Inv}^{\sim}(\mathcal{D}(\mathfrak{f}))$ is contained in the image of a compact subgroup of $\mathcal{A}(\mathfrak{f})$. Then there are maximal compact subgroups $G_{\mathcal{R}}(\mathfrak{f}) \subset \mathcal{R}(\mathfrak{f})$, $G_{\mathcal{A}}(\mathfrak{f}) \subset \mathcal{A}(\mathfrak{f})$, and $G_{\mathcal{D}}(\mathfrak{f}) \subset \mathrm{Inv}^{\sim}(\mathcal{D}(\mathfrak{f})) \cap \mathrm{GL}(p)$ such that (23) restricts to a split exact sequence*

$$1 \longrightarrow G_{\mathcal{R}}(\mathfrak{f}) \overset{i}{\longrightarrow} G_{\mathcal{A}}(\mathfrak{f}) \overset{p}{\longrightarrow} G_{\mathcal{D}}(\mathfrak{f}) \longrightarrow 1. \tag{25}$$

Remark II.4. A stronger form of this theorem was originally stated by du Plessis and Wilson in [dPW91, (1.5)] but an error in their proof was pointed out in [dPF11, Ch. 4.2].

Symmetries of folds and cusps. Now we turn to the computation of maximal compact subgroups of $\mathcal{A}$–symmetries of the fold and cusp singularities for $n = 4$ and $p = 2$. Recall that these were defined in terms of the local models

$$\mathfrak{F}_{\pm}(t,x,y,z) = (t, x^2 + y^2 \pm z^2)$$
$$\mathfrak{C}_{\pm}(t,x,y,z) = (t, x^3 + 3tx + y^2 \pm z^2).$$

From the formulas some immediate linear symmetries are apparent. In fact, the linear $\mathcal{R}$–symmetries are easily determined as

$$\mathcal{R}(\mathfrak{F}_+) \cap \mathrm{GL}(4) \cong \mathrm{O}(3) \qquad \mathcal{R}(\mathfrak{F}_-) \cap \mathrm{GL}(4) \cong \mathrm{O}(2,1)$$
$$\mathcal{R}(\mathfrak{C}_+) \cap \mathrm{GL}(4) \cong \mathrm{O}(2) \qquad \mathcal{R}(\mathfrak{C}_-) \cap \mathrm{GL}(4) \cong \mathrm{O}(1,1)$$

where $\mathrm{O}(3)$ and $\mathrm{O}(2,1)$ act on the (x,y,z)–coordinates while $\mathrm{O}(2)$ and $\mathrm{O}(1,1)$ act on the (y,z)–plane. (Note that $\mathrm{O}(r,s)$ is not compact for $r,s \geq 1$, a maximal compact subgroup is given by $\mathrm{O}(r) \times \mathrm{O}(s) = \mathrm{O}(r,s) \cap \mathrm{O}(r+s)$.) Moreover, there are some additional linear $\mathcal{A}$–symmetries, one for each $\mathfrak{F}_{\pm}$, given by

$$(t,x,y,z) \mapsto (-t,x,y,z) \quad \text{and} \quad (u,v) \mapsto (-u,v), \tag{26}$$

and another one for $\mathcal{C}_-$, namely

$$(t,x,y,z) \mapsto (t,-x,z,y) \quad \text{and} \quad (u,v) \mapsto (u,-v). \tag{27}$$

Obviously, these have order two and therefore contribute copies of $\mathbb{Z}_2$ in $\mathcal{A}(\mathfrak{F}_\pm)$ and $\mathcal{A}(\mathfrak{C}_-)$. We will show that the maximal compact subgroups of linear symmetries are also maximal in the full groups of $\mathcal{A}$–symmetries.

Proposition II.5. *The $\mathcal{A}$–symmetries of the fold and cusp singularities have maximal compact subgroups of the form*

$$G_\mathcal{A}(\mathfrak{F}_+) \cong \mathrm{O}(3) \ltimes \mathbb{Z}_2 \qquad G_\mathcal{A}(\mathfrak{F}_-) \cong \big(\mathrm{O}(2) \times \mathrm{O}(1)\big) \ltimes \mathbb{Z}_2$$

$$G_\mathcal{A}(\mathfrak{C}_+) \cong \mathrm{O}(2) \qquad G_\mathcal{A}(\mathfrak{C}_-) \cong \big(\mathrm{O}(1) \times \mathrm{O}(1)\big) \ltimes \mathbb{Z}_2$$

generated by the linear symmetries exhibited above. In particular, for orientation preserving $\mathcal{A}$–symmetries we have maximal compact subgroups

$$G_\mathcal{A}^+(\mathfrak{F}_+) \cong \mathrm{SO}(3) \qquad G_\mathcal{A}^+(\mathfrak{F}_-) \cong \mathrm{O}(2)$$

$$G_\mathcal{A}^+(\mathfrak{C}_+) \cong \mathrm{SO}(2) \qquad G_\mathcal{A}^+(\mathfrak{C}_-) \cong \mathbb{Z}_2$$

where $\mathrm{O}(2)$ is embedded into $\mathrm{O}(2) \times \mathrm{O}(1)$ via $A \mapsto (A, \det A)$ and the $\mathbb{Z}_2$ inside $\mathrm{O}(1) \times \mathrm{O}(1)$ is generated by $(t, x, y, z) \mapsto (t, x, -y, -z)$.

Proof. Let $\mathfrak{f}$ be either $\mathfrak{F}_\pm$ or $\mathfrak{C}_\pm$. One readily checks that

$$\ker(d\mathfrak{f}|_0) = \langle \partial_x, \partial_y, \partial_z \rangle \quad \text{and} \quad \mathrm{im}(d\mathfrak{f}|_0) = \langle \partial_u = d\mathfrak{f}|_0(\partial_t) \rangle$$

so that $\mathrm{coker}(d\mathfrak{f}|_0)$ can be identified with the span of ∂_v. Furthermore, it is simply a matter of writing out the definitions to see that in these identifications the intrinsic second derivatives $\delta_0^2 \mathfrak{F}_\pm$ and $\delta_0^2 \mathfrak{C}_\pm$ correspond to the Hessians of the functions $x^2 + y^2 \pm z^2$ and $x^3 + y^2 \pm z^2$, respectively. So according to Theorem II.2, the maximal compact subgroups of $\mathcal{R}$–symmetries satisfy

$$G_\mathcal{R}(\mathfrak{F}_+) \cong \mathrm{O}(3) \qquad G_\mathcal{R}(\mathfrak{F}_-) \cong \mathrm{O}(2) \times \mathrm{O}(1)$$

$$G_\mathcal{R}(\mathfrak{C}_+) \cong \mathrm{O}(2) \qquad G_\mathcal{R}(\mathfrak{C}_-) \cong \mathrm{O}(1) \times \mathrm{O}(1)$$

and these groups are realized by the linear symmetries discussed above.

In order to apply Theorem II.3 we have to study the compact subgroups of $\mathrm{Inv}^\sim(\mathcal{D}(\mathfrak{f})) \cap \mathrm{GL}(2)$ which can be determined completely in the cases at hand. Indeed, it is easy to see that the only non-trivial linear diffeomorphism that preserves the discriminants of the cusps

$$\mathcal{D}(\mathfrak{C}_\pm) = \big\{ (-\tau^2, -2\tau^3) \,\big|\, \tau \in \mathbb{R} \big\}$$

is given by $(u, v) \mapsto (u, -v)$ which violates to cokernel condition for $\mathfrak{C}_+$ whereas the cokernel is empty for $\mathfrak{C}_-$. Therefore, we have

$$\mathrm{Inv}^\sim(\mathcal{D}(\mathfrak{C}_+)) \cap \mathrm{GL}(2) = 1 \quad \text{and} \quad \mathrm{Inv}^\sim(\mathcal{D}(\mathfrak{C}_-)) \cap \mathrm{GL}(2) = \mathbb{Z}_2.$$

and similar reasoning yields

$$\mathrm{Inv}^\sim(\mathcal{D}(\mathfrak{F}_\pm)) \cap \mathrm{GL}(2) = \left\{ \left(\begin{smallmatrix} a & b \\ 0 & c \end{smallmatrix} \right) \,\middle|\, ac \neq 0, \ c > 0 \right\}$$

which has a unique compact subgroup, a copy of $\mathbb{Z}_2$ generated by $(u, v) \mapsto (-u, v)$.

Finally, according to (27) and (26) any compact subgroup of $\mathrm{Inv}^\sim(\mathcal{D}(\mathfrak{f})) \cap \mathrm{GL}(2)$ can be lifted to $\mathcal{A}(\mathfrak{f})$ and the claims about the structure of $\mathcal{G}_\mathcal{A}(\mathfrak{f})$ now follow from Theorem II.3. $\square$

References

[ADK05] D. Auroux, S. K. Donaldson, and L. Katzarkov, *Singular Lefschetz pencils*, Geom. Topol. **9** (2005), 1043–1114.

[Bay09] R. İ. Baykur, *Topology of broken Lefschetz fibrations and near-symplectic four-manifolds*, Pacific J. Math. **240** (2009), no. 2, 201–230.

[Beh13] S. Behrens, *On 4-manifolds, folds and cusps*, Pacific J. Math. **264** (2013), no. 2, 257–306.

[Cer70] J. Cerf, *La stratification naturelle des espaces de fonctions différentiables réelles et le théorème de la pseudo-isotopie*, Inst. Hautes Études Sci. Publ. Math. **39** (1970), 5–173 (French).

[Chi77] E. A. E. Chincaro, *Bifurcations of Whitney maps*, Notas da Universidade Federal de Minas Gerais, Instituto de Ciencias Exatas, Brasil (1977).

[CMWA02] T. Cooper, D. Mond, and R. Wik Atique, *Vanishing topology of codimension 1 multi-germs over $\mathbb{R}$ and $\mathbb{C}$*, Compositio Math. **131** (2002), no. 2, 121–160.

[DS03] S. Donaldson and I. Smith, *Lefschetz pencils and the canonical class for symplectic four-manifolds*, Topology **42** (2003), no. 4, 743–785.

[dPF11] A. du Plessis and A. Feragen, *The structure of groups of multigerm equivalences* (2011), preprint, available at `http://arxiv.org/abs/1110.1981v1`.

[dPV01] A. du Plessis and H. Vosegaard, *Characterisation of strong smooth stability*, Math. Scand. **88** (2001), no. 2, 193–228.

[dPW95] A. du Plessis and C. T. C. Wall, *The geometry of topological stability*, London Mathematical Society Monographs. New Series, vol. 9, The Clarendon Press, Oxford University Press, New York, 1995.

[dPW91] A. du Plessis and L. C. Wilson, *Right-symmetry of mappings*, Singularity theory and its applications, Part I (Coventry, 1988/1989), Lecture Notes in Math., vol. 1462, Springer, Berlin, 1991, pp. 258–275.

[EE69] C. J. Earle and J. Eells, *A fibre bundle description of Teichmüller theory*, J. Differential Geometry **3** (1969), 19–43.

[FM11] B. Farb and D. Margalit, *A Primer on Mapping Class Groups*, Princeton Mathematical Series, vol. 49, Princeton University Press, Providence, RI, 2011.

[GK11a] D. T. Gay and R. C. Kirby, *Indefinite Morse 2-functions; broken fibrations and generalizations* (2011), preprint, available at `http://arxiv.org/abs/1102.0750`.

[GK11b] ———, *Fiber-connected, indefinite Morse 2-functions on connected n-manifolds*, Proc. Natl. Acad. Sci. USA **108** (2011), no. 20, 8122–8125.

[GK12] ———, *Trisecting 4-manifolds* (2012), preprint, available at `http://arxiv.org/abs/1205.1565v3`.

[Gib79] C. G. Gibson, *Singular points of smooth mappings*, Research Notes in Mathematics, vol. 25, Pitman (Advanced Publishing Program), Boston, Mass.-London, 1979.

[GG73] M. Golubitsky and V. Guillemin, *Stable mappings and their singularities*, Springer-Verlag, New York, 1973. Graduate Texts in Mathematics, Vol. 14.

[HW73] A. Hatcher and J. Wagoner, *Pseudo-isotopies of compact manifolds*, Société Mathématique de France, Paris, 1973. Astérisque, No. 6.

[Hay11] K. Hayano, *On genus-1 simplified broken Lefschetz fibrations*, Algebr. Geom. Topol. **11** (2011), no. 3, 1267–1322.

[Hay14] ———, *Modification rule of monodromies in an R2–move*, Algebr. Geom. Topol. **14** (2014), no. 4, 301–342.

[Iva02] N. V. Ivanov, *Mapping class groups*, Handbook of geometric topology, North-Holland, Amsterdam, 2002, pp. 523–633.

[Jän78] K. Jänich, *Symmetry properties of singularities of C^∞-functions*, Math. Ann. **238** (1978), no. 2, 147–156, DOI 10.1007/BF01424772. MR512820 (80k:58021)

[Lek09] Y. Lekili, *Wrinkled fibrations on near-symplectic manifolds*, Geom. Topol. **13** (2009), no. 1, 277–318. Appendix B by R. İnanç Baykur.

[Lev65] H. I. Levine, *Elimination of cusps*, Topology **3** (1965), no. suppl. 2, 263–296.

[Mar82] J. Martinet, *Singularities of smooth functions and maps*, London Mathematical Society Lecture Note Series, vol. 58, Cambridge University Press, Cambridge-New York, 1982.

[ML89] L. E. Mata-Lorenzo, *A note on the codimension of a map*, J. Math. Anal. Appl. **137** (1989), no. 1, 37–45, DOI 10.1016/0022-247X(89)90271-0. MR981921 (90e:58009)

[Mat70] J. N. Mather, *Stability of C^∞ mappings. V. Transversality*, Advances in Math. **4** (1970), 301–336 (1970).

[Mat71] J. N. Mather, *Stability of C^∞ mappings. VI: The nice dimensions*, Proceedings of Liverpool Singularities-Symposium, I (1969/70), Springer, Berlin, 1971, pp. 207–253. Lecture Notes in Math., Vol. 192.

[Mat82] Y. Matsumoto, *On 4-manifolds fibered by tori*, Proc. Japan Acad. Ser. A Math. Sci. **58** (1982), no. 7, 298–301.

[Mor65] B. Morin, *Formes canoniques des singularités d'une application différentiable*, C. R. Acad. Sci. Paris **260** (1965), 6503–6506.

[Nic11] L. Nicolaescu, *An invitation to Morse theory*, 2nd ed., Universitext, Springer, New York, 2011.

[Per07] T. Perutz, *Lagrangian matching invariants for fibred four-manifolds. I*, Geom. Topol. **11** (2007), 759–828.

[Per08] ———, *Lagrangian matching invariants for fibred four-manifolds. II*, Geom. Topol. **12** (2008), no. 3, 1461–1542.

[Rie87] J. H. Rieger, *Families of maps from the plane to the plane*, J. London Math. Soc. (2) **36** (1987), no. 2, 351–369.

[RR91] J. H. Rieger and M. A. S. Ruas, *Classification of $\mathcal{A}$-simple germs from k^n to k^2*, Compositio Math. **79** (1991), no. 1, 99–108.

[Sae06] O. Saeki, *Elimination of definite fold*, Kyushu J. Math. **60** (2006), no. 2, 363–382.

[Tak71] F. Takens, *A solution to a problem of Thom*, Manifolds - Amsterdam 1970, Lecture Notes in Math., vol. 197, Springer, Berlin, 1971, pp. 231.

[Wal80] C. T. C. Wall, *A second note on symmetry of singularities*, Bull. London. Math. Soc. **12** (1980), no. 1, 347–354.

[Wal81] ———, *Finite determinacy of smooth map-germs*, Bull. London Math. Soc. **13** (1981), no. 6, 481–539.

[Wal09] ———, *Transversality in families of mappings*, Proc. Lond. Math. Soc. (3) **99** (2009), no. 1, 67–99.

[Was75] G. Wassermann, *Stability of unfoldings in space and time*, Acta Math. **135** (1975), no. x, 57–128.

[Wil10] J. D. Williams, *The h-principle for broken Lefschetz fibrations*, Geom. Topol. **14** (2010), no. 2, 1015–1061.

[Wil11a] ———, *Topology of surface diagrams of smooth 4-manifolds*, Proc. Natl. Acad. Sci. USA **108** (2011), no. 20, 8126–8130.

[Wil11b] ———, *Uniqueness of surface diagrams of smooth 4-manifolds* (2011), available at `http://arxiv.org/abs/1103.6263`.

[Wil14] ———, *Existence of 2-parameter crossings* (2014), available at `http://arxiv.org/abs/1411.1742`.

ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS, HUNGARIAN ACADEMY OF SCIENCES, REÁLTANODA UTCA 13–15, H–1053 BUDAPEST, HUNGARY
  E-mail address: `sbehrens@renyi.hu`

DEPARTMENT OF MATHEMATICS, GRADUATE SCHOOL OF SCIENCE, HOKKAIDO UNIVERSITY, KITA 10, NISHI 8, KITA-KU, SAPPORO, HOKKAIDO 060-0810, JAPAN
  E-mail address: `k-hayano@math.sci.hokudai.ac.jp`