Smooth 4–Manifolds and Surface Diagrams

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch–Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich–Wilhelms–Universität Bonn

vorgelegt von

Stefan Behrens

aus

Hannover

Bonn, November 2013

Angefertigt mit Genehmigung der Mathematisch–Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich–Wilhelms–Universität Bonn

1. Gutachter: Prof. Dr. Peter Teichner
2. Gutachter: Prof. Dr. Matthias Kreck

Tag der Promotion: 20. Mai 2014
Erscheinungsjahr: 2014

Contents

Chapter 1

Introduction

1.1 The Mysteries of Dimension Four

Dimension four is different. It marks the border between low dimensional and high dimensional topology and, while its two sides are well explored, the border itself is still largely uncharted territory. Since this thesis is ultimately concerned with the topology of smooth 4–manifolds, we begin with a brief review of the state of affairs. In this discussion, all manifolds are assumed to be closed, connected and oriented, and all homeomorphisms preserve orientations.

For a modern topologist there are two main flavors of manifolds: topological and smooth. It is also common to speak of the smooth and topological *categories* in this context. In all dimensions but four, both categories can be studied by the similar or even the same means. In dimensions up to three, there is no difference between the smooth and the topological categories and manifolds can be effectively studied by a mixture of hands-on techniques such as handlebody theory and more sophisticated geometric methods. In dimensions five and higher, essentially everything is governed by the powerful machinery of surgery theory and the s–cobordism theorem which allow to translate classification problems into homotopy theory and algebra. This includes the difference between the smooth and the topological category which turns out to be finite (in the sense that a topological manifold admits at most finitely many non-diffeomorphic smooth structures).

In contrast, in dimension four, topological and smooth manifolds are studied by drastically different means. By the groundbreaking work of Freedman [25], topological 4–manifolds are to some extent accessible to the high dimensional techniques. The key to these methods is the so called Whitney trick which is the main tool to match geometry and algebra. This turns out to be possible if one restricts ones attention to so called "good" fundamental groups, and the main open problem is the question whether all groups are good. For smooth 4–manifolds, the situation is much worse. On the one hand, the methods to show that two smooth 4–manifolds are diffeomorphic are very limited. Either one is lucky enough to be able to write down a concrete diffeomorphism, or one has to resort to handlebody theory in the guise of Kirby calculus. Although far from useless, both methods are rather primitive and usually not very effective. On the other hand, in order to tell two smooth 4–manifolds apart, a set of intricate invariants have been devised using ideas from gauge theory and Floer homology. Using these invariants and many ingenious con-

structions, it has been shown that – unlike in all other dimensions – topological 4–manifolds that admit one smooth structure tend to have infinitely many different ones. In other words, smooth 4–manifolds can have infinitely many *exotic copies*, that is, smooth 4–manifolds which are homeomorphic but not diffeomorphic to the given model. However, as powerful as these invariants are, they cannot distinguish all 4–manifolds. In fact, for technical reasons, they are only defined for 4–manifolds which contain a surface with positive self-intersection (this condition is usually phrased as $b_2^+ \geq 1$). In particular, they cannot be used directly to tackle the smooth 4–dimensional Poincaré conjecture, which states that any smooth 4–manifold which is homotopy equivalent to S^4 should be diffeomorphic to S^4. Furthermore, all these invariants are conjectured to contain the same information and there are ongoing programs to establish relations between them. The bottom line is, although the past 30 years have brought many insights, the topology of smooth 4–manifolds is still full of mysteries: it is not known whether all smooth 4–manifolds have infinitely many exotic copies, there is not a single (smoothable) topological 4–manifold whose smooth structures have been classified, and there is no structure theory in sight.

This thesis will not change the situation dramatically. However, one thing to take away from the above discussion is that the theory of smooth 4–manifolds has been stagnant and new idea should be pursued. And this is what we will do. Loosely speaking, we will cut 4 into $2 + 2$ by looking at certain maps from 4–manifolds to surfaces which fail to be fiber bundles in a controlled way. Such a map exhibits a 4–manifold as a singular family of surfaces parametrized by the target surface, providing a link from smooth 4–manifolds to surface topology. This idea has been around for a long time but has spiked in popularity in the recent years. We will give a short review in the next section and start a thorough discussion in Chapter 3.

But before, we would like to mention some positive results about 4–manifolds. First and foremost, there is Freedman's celebrated classification of simply connected topological 4–manifolds.

Theorem (Freedman [25])**.** *Let X be a simply connected topological 4–manifold. Then X is determined up to homeomorphism by its intersection form Q_X and Kirby–Siebenmann invariant* $\mathrm{ks}(X) \in \mathbb{Z}_2$.

Recall that Q_X is the unimodular, symmetric bilinear form defined on $H_2(X)$ as the Poincaré dual to the cup product form on $H^2(X)$, and $\mathrm{ks}(X)$ is zero if and only if $X \times \mathbb{R}$ is smoothable. These two invariants are not completely independent. In fact, if Q_X is even, then $\mathrm{ks}(X) \equiv \sigma(X)/8 \mod 2$. Here, Q_X is called *even* if all "squares" $Q_X(x, x)$ are even numbers (and *odd* otherwise), and $\sigma(X)$ is the *signature* of X defined as the difference of the numbers of positive and negative eigenvalues of $Q_X \otimes \mathbb{Q}$. Moreover, a complimentary theorem of Freedman states that any possible pair of a unimodular symmetric bilinear form Q (over $\mathbb{Z}$) and $k \in \mathbb{Z}_2$ can be realized by a simply connected topological 4–manifold. From an algebraic perspective, unimodular symmetric bilinear forms over $\mathbb{Z}$ fall into two categories: definite and indefinite. The definite ones are the great unknown, while the definite ones are easily classified.

Theorem (Serre [59])**.** *Let $Q\colon V \times V \to \mathbb{Z}$ be a symmetric bilinear form defined on a free Abelian group V. Then Q is determined up to isometry by its rank, signature and type (even or odd).*

In this light, the following theorem of Donaldson is all the more miraculous.

Theorem (Donaldson [15]). *Let X be a smooth 4–manifold. If Q_X is definite, then it is diagonalizable over $\mathbb{Z}$.*

Not only does this theorem exclude many topological 4–manifolds from being smoothable, it also shows that no complicated definite intersection forms appear for smooth 4–manifolds! So there is some way in which smooth 4–manifolds are well-behaved, after all. One can even write down a complete list of possible intersection forms of smooth 4–manifolds, these are $k(1) \oplus l(-1)$ and $mE_8 \oplus n\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, where $k, l, m, n \in \mathbb{Z}$, $k, l, n \geq 0$, and E_8 is the unique even, definite form of rank 8. The only remaining question here is which combinations of m and n can be realized and the famous $\frac{11}{8}$*–conjecture* states that $n \geq 6|m|$ (which is equivalent to $b_2(X) \geq \frac{11}{8}|\sigma(X)|$). The upshot of this discussion is the following beautiful "cross-category" result which follows from the three mentioned theorems.

Corollary (Freedman, Donaldson, Serre). *Smooth, simply connected 4–manifolds are classified up to homeomorphism by their Euler characteristic, signature and type.*

1.2 Singular Fibrations on 4–Manifolds

At this point we have said everything we had to say about topological 4–manifolds and we add the property "smooth" to our list of standing assumptions on manifolds. As mentioned above, we will later consider 4–manifolds as singular families of surfaces parametrized by another surface. What follows is a brief overview of the history of this philosophy.

Lefschetz fibrations and symplectic 4–manifolds. The idea of sweeping out a geometric object by smaller sub-objects is quite common in algebraic geometry. It was used extensively by Lefschetz in his study of complex projective varieties (see Lamotke's beautiful survey [42]). Since we are interested in 4–manifolds, we will focus on projective surfaces and follow [32, Chapter 8.1]. Given such a surface $V \subset \mathbb{C}\mathrm{P}^N$, we choose a generic linear subspace $A \subset \mathbb{C}\mathrm{P}^N$ that is transverse to V and has complementary (complex) dimension $N-2$. If we write A as the common zero locus of two linear homogeneous polynomials p_0 and p_1, then we obtain a family of hyperplanes $H_t \subset \mathbb{C}\mathrm{P}^N$ containing A parametrized by $t = [t_0 : t_1] \in \mathbb{C}\mathrm{P}^1$ where H_t is the zero locus of $t_0 p_0 + t_1 p_1$. The intersection $P = A \cap V \subset V$ is a finite number of points and each $\Sigma_t = H_t \cap V$ is a (possibly singular) complex curve containing P. Moreover, for each $x \in V \setminus P$ there is a unique $t \in \mathbb{C}\mathrm{P}^1$ such that $x \in \Sigma_t$. This observation gives rise to a holomorphic map $p\colon X \setminus P \to \mathbb{C}\mathrm{P}^1$ which is known as a *Lefschetz pencil* with *axis* A and *base locus* P. It turns out that p has only finitely many critical points which locally look like the quadratic polynomial $z^2 + w^2$ defined on $\mathbb{C}^2$ – so called *Lefschetz singularities* – and around its base points p looks like the projection $\mathbb{C}^2 \setminus \{0\} \to \mathbb{C}\mathrm{P}^1$ which is easily seen from the construction. In particular, after blowing up the base points one obtains a *Lefschetz fibration* $\tilde{p}\colon V\#|P|\overline{\mathbb{C}\mathrm{P}^2} \to \mathbb{C}\mathrm{P}^1$.

The notion of Lefschetz pencils and fibrations is easily generalized to arbitrary smooth 4–manifolds.[1] A *Lefschetz fibration* is simply a surjective smooth map $X \to B$ onto a (real) surface B – the so called *base surface* – with only finitely many critical points which are all of Lefschetz type. A *Lefschetz pencil* is a Lefschetz fibration on $X \setminus P$ over $S^2 \cong \mathbb{C}\mathrm{P}^1$ where the base locus $P \subset X$ is a (non-empty) finite set where the map is modeled on the projection $\mathbb{C}^2 \setminus \{0\} \to \mathbb{C}\mathrm{P}^1$. The repeated use of the word "base" is a little unfortunate, but it does not lead to any confusion because, by definition, Lefschetz fibrations have empty base loci and the base surface of Lefschetz pencils is always $\mathbb{C}\mathrm{P}^1$.

As opposed to the case of projective surfaces, Lefschetz pencils and fibrations are significantly harder to construct on more general 4–manifolds. First of all, it is not always possible. According to a theorem of Gompf, total spaces of Lefschetz pencils admit symplectic structures and the same holds for Lefschetz fibrations with only few exceptions, see [32, p.401ff.]. In particular, Lefschetz pencils can only exist on symplectic 4–manifolds and it is a deep results due to Donaldson [16] that this is always the case. Both results combined can be considered as a topological characterization of symplectic 4–manifolds.

Theorem (Donaldson [16], Gompf [32])**.** *A 4–manifold is symplectic if and only if it admits a Lefschetz pencil.*

In particular, topological properties of symplectic 4–manifolds should also be visible through the eyes of Lefschetz pencils and fibrations. One of the most important features of symplectic 4–manifolds is that they have non-trivial Seiberg–Witten invariants and generally interact very well with Seiberg–Witten theory. Among other things, Taubes [61] had shown that the Seiberg–Witten invariants of symplectic 4–manifolds can be expressed as a certain count of pseudoholomorphic curves which he called the *Gromov invariant*, thus providing a geometric interpretation. In search for a counterpart on the Lefschetz side, Donaldson and Smith introduced their *standard surface count* [17] which, roughly, counts certain pseudoholomorphic multisections of a Lefschetz pencil or fibration. It was later shown by Usher [66] that the standard surface count agrees with the Gromov invariant.

Near-symplectic 4–manifolds and broken Lefschetz fibrations. Obviously, not all 4–manifolds are symplectic. But from the point of view of Seiberg–Witten theory, the relevant one come quite close. We already mentioned that the Seiberg–Witten invariants are only defined for 4–manifolds with $b_2^+ \geq 1$. It turns out that the latter condition implies the existence of a closed 2–form which is non-degenerate outside of a closed 1–dimensional submanifold, that is, a disjoint union of finitely many embedded circles. Such a 2–form is called a *near-symplectic structure*. In search for a general geometric interpretation of the Seiberg–Witten invariants, Taubes set out a program to extend his methods to the near-symplectic setting [62, 63]. This turned out to be a challenging task and, as far as the author knows, Taubes's program is still ongoing. But the interest in near-symplectic structures raised the question for an analogue of the Donaldson–Gompf correspondence, which was eventually discovered by Auroux, Donaldson and Katzarkov [3].

[1] According to Matsumoto [48], non-holomorphic Lefschetz fibrations were first studied by Moishezon [51, p.162].

Theorem (Auroux–Donaldson–Katzarkov [3]). *A 4–manifold is near-symplectic if and only if it admits a broken Lefschetz pencil.*

The trade-off for the degeneracy of the 2–form is the appearance of an additional type of singularities in broken Lefschetz pencils, so called *indefinite folds* which we will discuss in great detail later. Just as the degeneracy loci of near-symplectic forms, indefinite folds appear in 1–dimensional families. In fact, the fold locus is the direct counterpart of the degeneracy locus in the above correspondence; circles of indefinite folds have also become known as *round singularities* in this context. So instead of working with near-symplectic forms, one can also work with broken Lefschetz pencils and fibrations. In particular, instead of trying to generalize the Gromov invariant to the near-symplectic setting, one can also attempt to extend the standard surface count to the broken setting. Such an effort was made by Perutz who introduced his *Lagrangian matching invariants* [57, 58]. However, this approach also presents severe technical difficulties and it is still neither known if these invariants agree with the Seiberg–Witten invariants, nor if they actually are invariants.

The passage to all 4–manifolds. Although Auroux, Donaldson and Katzarkov had mainly focused on near-symplectic 4–manifolds, they exhibited a broken Lefschetz fibration on S^4 over S^2 (see Example 1 in [3, Section 8.2]). This simple example turned out to be surprisingly influential. Since S^4 is not near-symplectic, it led to the question whether all 4–manifolds admit broken Lefschetz fibrations. The first advance in this direction was made by Gay and Kirby [27] who used handle-body techniques to prove the existence of so called *achiral* broken Lefschetz fibrations. Since Gay and Kirby could not avoid achiral Lefschetz singularities (which are modeled on $z^2 + \overline{w}^2$) using their methods, they speculated that this might indeed be impossible. But shortly after, Lekili and Baykur [45] gave arguments to remove achiral singularities, thus proving the general existence of broken Lefschetz fibrations. Two further, independent existence proofs were given by Baykur [4] and Akbulut–Karakurt [1] almost at the same time. In fact, the statement is stronger.

Theorem (Gay–Kirby, Lekili, Baykur, Akbulut–Karakurt). *Let X be a 4–manifold. Then any map $X \to S^2$ is homotopic to a broken Lefschetz fibration.*

Lekili [45] and Baykur [4] both used methods from the singularity theory of smooth maps in their proofs. Lekili realized that it was possible to trade (regular and achiral) Lefschetz singularities for so called *indefinite cusp* singularities and vice versa. Maps with only folds and cusps have a long history in singularity theory and Lekili's observation opened the door to study broken Lefschetz fibrations as well as families thereof in this context. This led to a change of focus away from Lefschetz singularities to maps with folds and cusps. In particular, Williams [67] introduced a class of maps called *simple wrinkled fibrations*[2] which have a particularly simple critical point structure: the critical locus is a single circle consisting of indefinite folds and finitely many cusps, which is mapped injectively into S^2. Simple wrinkled fibrations are closely related to broken Lefschetz fibrations. In fact, Lekili's perturbations can be used to turn a simple wrinkled fibration into a broken Lefschetz fibration with a single circle of folds (as studied by Baykur in [5]). Williams also proved the existence of these maps.

[2] Williams uses the term "simplified purely wrinkled fibrations".

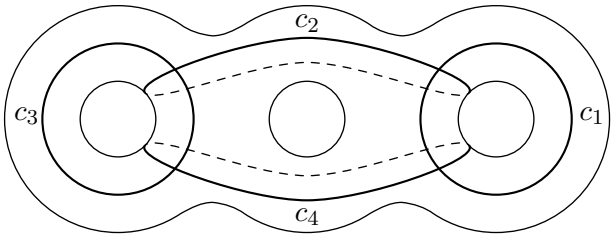


Figure 1: A surface diagram of $S^1 \times S^3 \# S^1 \times S^3$ due to Hayano [35].

Theorem (Williams [67])**.** *Any map from a 4–manifold to S^2 is homotopic to a simple wrinkled fibration.*

Meanwhile, Gay and Kirby initiated an in-depth study of maps from 4–manifolds to surfaces [28–30] which have only fold and cusp singularities. This ends our historical outline as far as the various singular fibration structures are concerned. In our presentation we will essentially reverse the historical order. In Section 2.2 we give a brief review of maps from 4–manifolds to surfaces from the perspective of singularity theory and from Chapter 3 onward we will mainly be interested in what we call *wrinkled fibrations* (these are maps which only have indefinite folds and cusps). Broken Lefschetz fibrations and Lefschetz fibrations will merely be used occasionally for motivational purposes or as convenient tools.

Since their inception, broken Lefschetz fibrations and related structures have received considerable attention in the research literature (see [5–8, 33–35, 40, 69], for example). We will point out several other developments throughout the text. Although the original motivation from Seiberg–Witten theory seems to have been lost somewhere along the way, it has only been put aside in order to obtain a better understanding of the newly discovered structures. As this understanding is improving, so are the chances of finding a connection.

Surface diagrams. Williams's existence theorem for simple wrinkled fibrations together with another observation of Williams is the starting point of our work. An important feature of Lefschetz fibrations over S^2 is the classical observation that they are accessible via handlebody theory and can be described more or less combinatorially in terms of configurations of simple closed curves on surfaces [32, 39]. Given a Lefschetz fibration, one can associate to each critical point a *Lefschetz vanishing cycle* which is a simple closed curve in a fixed regular fiber. Moreover, from the vanishing cycles one can recover the fibration up to a suitable notion of equivalence using a handlebody construction (see Section 4.3 for more details). This was extended to the broken setting by Baykur [5]. The fold singularities contribute additional *fold vanishing cycles* which behave slightly differently than Lefschetz vanishing cycles. In the setting of simple wrinkled fibrations there are only fold singularities and Williams suggested recording the fold vanishing cycles in a regular fiber. Abstractly, this leads to a closed, oriented surface Σ decorated with a cyclically indexed set of simple closed curves $c_1, \ldots, c_l \subset \Sigma$, such that c_i and c_{i+1} (where $l+1=1$) have geometric intersection number one. Williams called this structure a *surface diagram* [68], an example is shown in Figure 1. He also noticed that the surface diagram of a simple wrinkled fibration contains enough information to reconstruct

the total space [67, Corollary 2] which easily follows from results of Lekili [45] and Baykur [5] via a detour over broken Lefschetz fibrations[3]. Combined with the existence theorem for simple wrinkled fibration over S^2 this leads to the following intriguing consequence.

Corollary (Williams [67])**.** *All 4–manifolds can be described by surface diagrams.*

However, there is a caveat with this simple statement. Remember that we restricted our attention to closed 4–manifolds in the present discussion. Unfortunately, it turns out that an arbitrary surface diagram as described above does *not* describe a closed 4–manifold unless a certain *trivial monodromy* condition is satisfied. This adds a lot of subtlety to the theory and we refer to Chapter 4 for further details. The study of simple wrinkled fibrations, surface diagrams, and how they relate to the topology of 4–manifolds is the central focus of our work.

1.3 Summary of Results

We now describe the contents of this thesis and state our main results. The main body of the text is divided into three parts which each have a slightly different focus.

Part I is mostly of preliminary nature. After reviewing some background material in Chapter 2, we turn to wrinkled fibrations and broken Lefschetz fibrations in Chapter 3. We collect the central definitions and summarize important results surrounding these classes of maps. In addition, in Section 3.2 we develop a theory of parallel transport in the context of wrinkled fibrations which provides a solid framework for the discussion of vanishing cycles.

In Part II we develop a self-contained theory of simple wrinkled fibrations and surface diagrams. In the process we extend and clarify various aspects of the work of Williams. Recall that Williams showed how to extract surface diagrams from simple wrinkled fibrations over S^2 and how their total spaces can be recovered from surface diagrams assuming that the genus is at least three. We go further and show that the map itself can be recovered as well. Unfortunately, the statement is a little convoluted since we include the low genus cases.

Theorem 1.1 (Correspondence over the sphere)**.** *Let $\mathcal{SWF}_g(S^2)$ and $\mathcal{SD}^0_g$ be the sets of equivalence classes of genus g simple wrinkled fibrations over S^2 and surface diagrams with trivial monodromy, respectively. There is a surjective map*

$$\mathcal{SWF}_g(S^2) \longrightarrow \mathcal{SD}^0_g$$

whose point preimages have a transitive action of the group $\pi_1\big(\operatorname{Diff}(\Sigma_{g-1}), \mathrm{id}\big)$. In particular, the map is bijective for $g \geq 3$.

We refer to Chapter 4 for the relevant definitions. In particular, we would like to highlight the trivial monodromy condition which is hidden in Williams's approach. Without this condition we obtain a different correspondence result for certain simple wrinkled fibrations over the disk which is much cleaner.

[3] See the discussion following Corollary 4.19 on page 52.

Theorem 1.2 (Correspondence over the disk)**.** *Equivalence classes of surface diagrams correspond bijectively to equivalence classes of descending simple wrinkled fibrations over the disk.*

The main ideas used to prove Theorems 1.1 and 1.2 are outlined in Section 4.4. Our discussion closely parallels the relation of Lefschetz fibrations and their vanishing cycles which is reviewed in Section 4.3. However, the actual proofs are carried out in a more general setting. In fact, we define simple wrinkled fibrations over arbitrary base surfaces and explain how their study reduces to that of certain simple wrinkled fibrations over the annulus which we call *annular*. We then give a purely combinatorial definition of surface diagrams, whereas Williams always discusses surface diagrams in relation to simple wrinkled fibrations. We also introduce *generalized surface diagrams* as a combinatorial counterpart for annular simple wrinkled fibrations as well as a natural notion of equivalence of (generalized) surface diagrams. The following is the central result of Part II.

Theorem 1.3 (Annular correspondence)**.** *There is a bijective correspondence between annular simple wrinkled fibrations and generalized surface diagrams, both considered up to equivalence.*

The proof is rather lengthy and occupies most of Chapter 5. We finally deduce Theorems 1.1 and 1.2 in Section 5.3. In the course of the proof we will obtain a good understanding how (generalized) surface diagrams encode the topology of the total spaces of the corresponding simple wrinkled fibrations. In particular, we will see that simple wrinkled fibrations are directly accessible to handlebody theory without the previously customary detour over broken Lefschetz fibrations. This aspect is revisited and further elucidated in Chapter 6 in the context of simple wrinkled fibrations over the disk and the sphere. We described the structure of the handle decompositions induced by these maps, explain how to draw Kirby diagrams, and discuss some examples.

In Part III we leave simple wrinkled fibrations aside and focus on the interplay between the combinatorics of surface diagrams and the topology of the 4–manifolds they describe. In Chapter 7 we show how certain cut-and-paste operations on 4–manifolds can be realized in terms of surface diagrams, including connected sums with $S^2 \times S^2$ and $\mathbb{C}\mathrm{P}^2$ with either orientation. These techniques turn out to be a key ingredient in the proof of our next main result, namely the classification of closed 4–manifolds that can be described by surface diagrams of the lowest genus, which is the subject of Chapter 8. Our result should be compared to the attempts of Hayano [33] and Baykur–Kamada [6] to classify 4–manifolds with genus one broken Lefschetz fibrations.

Theorem 1.4. *A closed 4–manifold admits a surface diagram of genus one if and only if it is diffeomorphic to $kS^2 \times S^2$ or $m\mathbb{C}\mathrm{P}^2 \# n\overline{\mathbb{C}\mathrm{P}^2}$ where $k, m, n \geq 1$.*

In Chapter 9 we take on the task of understanding how surface diagrams encode basic homotopy information of 4–manifolds and give some applications. We first discuss the fundamental and homology groups as well as intersection forms in Section 9.1, for which we obtain descriptions in terms of surface diagrams. Using these we derive an easily verifiable obstruction for surface diagrams to have trivial

monodromy, see Section 9.2. We also discuss spin and spin^c structures in Section 9.3. As a final application, we elaborate on the following theorem in Section 9.4.

Theorem 1.5. *Let $w\colon X \to S^2$ be a simple wrinkled fibration with surface diagram $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$. If X is simply connected and $[\Sigma] = 0 \in H_2(X)$, then the homeomorphism type of X is determined by the homology classes $[c_i] \in H_1(\Sigma)$.*

An interesting observation is that the *diffeomorphism type* of X a priori depends on the *isotopy classes* of the curves $c_i \subset \Sigma$. Furthermore, the difference between isotopy and homology for curves on a surface is measured in terms of the *Torelli group* of the surface, which is the non-linear and mysterious part of the mapping class group. This suggests the possibility of a relation between exotic smooth structures on 4–manifolds and Torelli groups.

Finally, we include three appendices which provide detailed proofs of some results that qualify as "mathematical folklore" which are used in the main body of the text. In Appendix A we show how the intersection form of a closed 4–manifold can be computed from a Kirby diagram. This is common knowledge in case that the underlying handle decomposition is free of 1– and 3–handles but we are not aware of a treatment of the general situation in the literature. The purpose of Appendix B is to explain how the presence of a singular fibration leads to a geometric interpretation of the set of spin^c structures of a given 4–manifold. This is mainly an elaboration on ideas of Taubes [64] and Perutz [58]. Appendix C deals with a problem in 3–dimensional Morse theory which is relevant for the proof of Theorem 1.3. Using methods of Cerf [13] and basic mapping class group theory we show that a canceling pair of critical points of a Morse function defined on an orientable 3–manifold can be canceled in an essentially unique way.

Several results in this thesis, including Theorems 1.1 to 1.4, have already appeared in the author's article [7]. However, our exposition here is much more detailed and several proofs have been completely rewritten in order to improve clarity. In particular, we make consistent use of the theory of parallel transports which allows to make many arguments more precise.

Acknowledgments

First and foremost, I want to thank my advisor Peter Teichner for all his support, encouragement, guidance, and for giving me the opportunity to work on 4–manifolds and low dimensional topology although my background was rather different. Next, I want to express my gratitude to my co-advisor Matthias Kreck, not only for his interest in my work, but also for teaching me a lot about the topology of manifolds and topology in general. His inspiring lectures at that I attended in my undergraduate days at the University of Bonn shaped my view on these subjects. Also, İnanç Baykur was sort of my unofficial third advisor during his time in Bonn and I am grateful for all his help. I had many interesting conversations, either in person or via email, with many other mathematicians. I would like to highlight Kenta Hayano, Jonathan Williams, and David Gay who shared their insights with me and have helped me understand many things better.

I am also indebted to the Max Planck Society and especially to the Max Planck Institute for Mathematics in Bonn where I spent the past years as an IMPRS scholar. I have thoroughly enjoyed the experience and I will always keep this time of my life in good memory. Besides the terrific mathematical atmosphere, the non-scientific staff deserves a special mention. They made all the necessary bureaucracy and other annoyances as easy as they should be but usually never are.

Outside of mathematics, words are not enough to express what I owe to Cathalin Recko. She is my safe haven in the real world. Besides proofreading parts of my thesis, she has bravely put up with my moods in the various "final" stages of the writing process and kept me focused. Last but not least, I want to thank my parents who have always been there for me and encouraged me to follow my dreams. Without their loving support I would not have been able to get this far.

Part I

Preliminaries

Chapter 2

Background Material

The main theme of our work is to study smooth 4–manifolds in terms of curve configurations in surfaces which are derived from certain maps to the 2–sphere. In doing so, we will primarily rely on notions and methods from such varied fields as

- singularity theory of smooth maps,
- mapping class groups of surfaces, and
- handlebody theory.

As a service to the reader, as well as to set up some terminology, we include short reviews of the necessary background from each of these three subjects. We deliberately risk being overly detailed and suggest that the reader only skim this chapter in order to become acquainted with our notation and terminology, and come back to it when he or she feels that more information is needed.

General assumptions. We will work exclusively in the smooth category, meaning that all manifolds and maps that appear are assumed to be smooth. In addition, we make the standing assumptions that all manifolds are compact, connected, and oriented. Deviations will be explicitly indicated. Occasionally, we will also restate some of these assumptions for emphasis. We will freely use basic results from differential topology such as the tubular neighborhood theorem, the isotopy extension theorem, and transversality theory as covered in [11], for example. We often use notation such as νA to indicate an open neighborhood of a subset A of some manifold. If A is a submanifold, then we implicitly assume that νA is a tubular neighborhood.

Orientation conventions. Since we are working with oriented manifolds, we have to settle on some conventions for induced orientations. If M is a manifold with boundary, then we orient ∂M by the *outward normal first* convention. This is the convention that gives the unit circle $S^1 \subset \mathbb{R}^2$ the counterclockwise orientation when thought of as the boundary of the unit disk D^2. Furthermore, if $p\colon E \to B$ is an oriented fiber bundle with fiber F, then we require that the orientations on E, B, and the fibers $E_b = p^{-1}(b)$ are related by the *fiber first convention*. In the case of the trivial bundle $E = F \times B$ this means that we consider the fiber as the first factor and require that the product orientation of the right hand side agrees with the orientation on E. This generalizes to arbitrary bundles by the choice of a bundle atlas with fiber wise orientation preserving transition maps.

Framings. Let M be an n–dimensional manifold and let $K \subset M$ be a submanifold of codimension k. Recall that a *normal framing* or simply a *framing* for K is a trivialization of its normal bundle. Equivalently, a framing consists of k point wise linearly independent normal vector fields along K (that is, sections of $TM|_K$ that are nowhere tangent to K). We use the latter interpretation and the obvious notion for *homotopies* of framings. We will allow ourselves the common inaccuracy to blur the distinction between framings and their homotopy classes. This is justified by the tubular neighborhood theorem, which establishes a bijection between framings up to homotopy and extensions of the inclusion $K \subset M$ to embeddings $K \times D^k \hookrightarrow M$ up to ambient isotopy, and the latter structures are what one is actually interested in most of the time. If both M and K are oriented, then a framing is already determined up to homotopy by the choice of $(k-1)$ normal vector fields, since the orientations specify the last vector field up to homotopy. In particular, for $k = 1$ the orientations determine a canonical homotopy class of framings and for $k = 2$ it is enough to specify a single normal vector field up to homotopy. We will sometimes be sloppy and use this method even when K is only orientable but not oriented. In those situations it is to be implicitly understood that the framing is only determined after choosing an orientation. Usually this issue will arise in constructions which require the choice of an orientation but whose result turns out to be independent of this choice.

(Co-)Homology. Be default, (co-)homology groups are taken with integer coefficients, that is, $H_k(X)$ always means $H_k(X;\mathbb{Z})$ and other coefficient groups will be indicated explicitly. We will freely appeal to various forms of Poincaré duality (as found in Bredon's book [10, p.348ff.], for example) and denote the corresponding isomorphisms by PD. Finally, as customary in low dimensional topology, we usually think of homology classes as represented by submanifolds. For an oriented submanifold $S \subset M$ we denote its homology class by $[S] \in H_k(M)$ where $k = \dim S$. Furthermore, we equip the homology groups with the *intersection product* $H_k(M) \times H_l(M) \to H_{k+l-n}(M)$ defined as $[S] \cdot [T] = [S \pitchfork T]$ where the symbol $\pitchfork$ indicates a transverse intersection (possibly after an implicit perturbation of either S or T).

2.1 4–Manifolds and Kirby Calculus

We begin by reviewing some basic facts about 4–manifolds and their handle decompositions. The latter will be our main tool for relating properties of maps onto surfaces as well as combinatorial structures in their fibers to the topology of the source manifold.

Handle decompositions. Roughly speaking, handle decompositions are a manifold version of cell decompositions and they are a central tool in manifold topology. We briefly recall the 4–dimensional situation. For a detailed account we refer to Gompf and Stipsicz [32, Ch. 4]. A (4–dimensional) *k–handle* is a copy of $h^k = D^k \times D^{4-k}$ and can be thought of as a thickened k–cell. A k–handle is attached to a 4–manifold X via an embedding $\varphi\colon S^1 \times D^2 \hookrightarrow \partial X$ resulting in a new 4–manifold $X \cup_\varphi h^k$. The subset $S^k \times D^{4-k}$ of ∂h^k is called the *attaching region*

and $S^k \times \{0\}$ is the *attaching sphere*; both are usually implicitly identified with their images in ∂X. Similarly, $D^k \times S^{3-k}$ and $\{0\} \times S^{3-k}$ are called the *belt region* and *belt sphere* and so are their counterparts in $\partial(X \cup_\varphi h^k)$. There are two further important subsets of h^k, namely the *core* $D^k \times \{0\}$ and the *cocore* $\{0\} \times D^{4-k}$ which are bounded by the attaching sphere and belt sphere, respectively.

A manifold that is obtained from the empty set by a sequence of handle attachments is called a *handlebody*. Since only 0–handles can be attached to the empty set, every handlebody must have a zero handle. A *handle decomposition* of a 4–manifold X is a diffeomorphism from X to a handlebody. It is well known that every 4–manifold admits a handle decomposition. Moreover, one can always arrange the following extra properties (see [32, Ch. 4.2]) which we shall henceforth assume:

- the handles are attached in order of increasing index,
- there is a unique 0–handle, and
- there is at most one 4–handle which is needed if and only if X is closed.

Given such a handle decomposition of X we denote by $X^{\leq k}$ the union of all handles of index at most k and call this the *k–skeleton* of X. An important observation is that for closed X, equipped with a handle decomposition as above, the 2–skeleton $X^{\leq 2}$ already determines X up to diffeomorphism. This follows from results of Laudenbach and Poénaru [43], who implicitly show that any orientation preserving diffeomorphism of $\#^k(S^1 \times S^2)$ extends across $\natural^k(S^1 \times D^3)$, and the observation that the union of the 3– and 4–handles of X are diffeomorphic to $\natural^k(S^1 \times D^3)$ with k the number of 3–handles.[1] In other words, if the boundary of a 2–handlebody diffeomorphic to $\#^k(S^1 \times S^2)$, then up to diffeomorphism there is a unique way to attach 3– and 4–handles to obtain a closed 4–manifold. This is very convenient, since 4–dimensional 2–handlebodies can actually be visualized by 3–dimensional pictures, as discussed below.

The last general fact about handle decompositions we will need is that they can be used to compute homology groups in very much the same spirit as cell decompositions. To a handle decomposition of a 4–manifold X one can associate the *handle complex*

$$C_4(X) \xrightarrow{\partial_4} C_3(X) \xrightarrow{\partial_3} C_2(X) \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X)$$

where $C_k(X)$ is the free Abelian group generated by the k–handles and the differential ∂_k counts the intersections between the attaching spheres of the k–handles with the belt spheres of the $(k-1)$–handles in $\partial X^{\leq k}$. As shown in [32, p.111], this is a chain complex which computes the homology of X. Note that the assumption about 0– and 4–handles force ∂_4 and ∂_1 to vanish, so that all interesting information is concentrated in ∂_2 and ∂_3. Moreover, for closed X the fact that X is determined by $X^{\leq 2}$ shows that the whole homological information about X is encoded only in the map $\partial_2 \colon C_2(X) \to C_1(X)$.

Kirby diagrams. As mentioned before, the structure of a 2–handlebody can be described by a 3–dimensional picture known as a *Kirby diagram*. There are two

[1] A more explicit account can be found in an article of Montesinos [52, Theorems 1&2].

different ways to deal with 1–handles and we will employ the "dotted circle notation" so that a Kirby diagram consists of the following data:

- An unlink $\mathcal{U} = U_1 \cup \cdots \cup U_{h_1}$ in S^3 where each component is decorated with a dot – these are the "dotted circles".
- A framed link $\mathcal{L} = L_1 \cup \cdots \cup L_{h_2}$ in the complement $S^3 \setminus \mathcal{U}$ where the framing of L_i is specified by its *framing coefficient* $f_i \in \mathbb{Z}$ which measures the difference to the framing induced from a Seifert surface for L_i in S^3, also known as the *0–framing*.

A Kirby diagram $(\mathcal{U}, \mathcal{L})$ encodes the 2–handlebody part of a 4–manifold X as follows. The ambient S^3 is thought of as the boundary of the unique 0–handle of X, which is implicitly identified with D^4. Each dotted circle represents a 1–handle, albeit in a slightly subtle way: we choose pairwise disjoint spanning disks for all components of $\mathcal{U}$, push their interiors into the interior of D^4, and carve out open tubular neighborhoods of these disks from D^4. Up to diffeomorphism, this process turns out to have the same effect as attaching 1–handles to D^4 (see [32, Ch. 5.4]) so that the unlink $\mathcal{U}$ represents $X^{\leq 1}$. Note that this description of $X^{\leq 1}$ naturally identifies $\partial X^{\leq 1}$ with the 0–surgery on $\mathcal{U}$ so that the complement of a neighborhood $\nu\mathcal{U} \subset S^3$ can be considered as part of $\partial X^{\leq 1}$. With this understood, the framed components of $\mathcal{L}$ simply specify the attaching regions of the 2–handles so that the description of $X^{\leq 2}$ is complete. As a side note, there does not seem to be a convenient way to include information about 3–handles in a Kirby diagram so that an arbitrary 4–manifold with boundary cannot be described. However, for studying closed 4–manifolds (up to diffeomorphism) this is irrelevant by the mentioned results of Laudenbach and Poénaru [43]. This makes Kirby diagrams a powerful tool in the context of closed 4–manifolds.

Handle moves and Kirby calculus. Of course, handle decompositions and thus Kirby diagrams of 4–manifolds are not unique. However, it is known that any two handle decompositions of a given 4–manifold are related by isotopies of the attaching maps, including the so called *handle slides*, and the *creation/cancellation* of pairs of handles of adjacent index (see [32, Theorem 4.2.12]). Translated into the language of Kirby diagrams, these handle moves are commonly known as *Kirby calculus* and provide a visually accessible method for proving that two 4–manifolds are diffeomorphic. Very roughly, given a Kirby diagram $(\mathcal{U}, \mathcal{L})$ the isotopies for the 1–handles appear as isotopies and band sums among the components of $\mathcal{U}$ (dragging $\mathcal{L}$ along), while the 2–handle isotopies affect $\mathcal{L}$ by isotopies and band sums with components of both $\mathcal{L}$ and $\mathcal{U}$. The creation/cancellation of handle pairs takes the form of insertion/deletion of an isolated Hopf link with one dotted and one 0–framed component (for pairs of index 1 and 2) or an isolated 0–framed unknot (for index 2 and 3). For more detailed descriptions we refer to [32, Chs. 5.1&5.4]. In some arguments in Section 6.3 and Chapter 7 we will use Kirby calculus so that some familiarity with the subject is useful, although not strictly necessary because the manipulations are rather simple.

Intersection forms. For a 4–manifold X the intersection product and the canonical isomorphism $H_0(X) \cong \mathbb{Z}$ give rise to a symmetric bilinear map

$$H_2(X) \times H_2(X) \longrightarrow \mathbb{Z}$$

called the *intersection pairing* and denoted by $a \cdot b$. The intersection pairing vanishes on torsion classes and thus descends to a symmetric bilinear form Q_X on the free Abelian group $H_2(X)/\text{torsion}$, the so called *intersection form* of X. After passing to rational coefficients, Q_X can be diagonalized and the numbers of positive and negative eigenvalues are homotopy invariants of X denoted by $b_2^+(X)$ and $b_2^-(X)$, respectively. The difference $\sigma(X) = b_2^+(X) - b_2^-(X)$ is called the *signature* of X.

2.2 Singularities of Smooth Maps

Starting with Chapter 3 we will study certain maps from 4–manifolds to surfaces which are characterized by their critical point structure. In order to put the central definitions into a proper context, it is useful to know some general facts about smooth maps and their singularities. However, we want to emphasize that the main purpose of this section is simply to convince the reader that if one studies maps from 4–manifolds to surfaces, then it is natural to consider maps with only folds and cusps. Of course, we will also explain what folds and cusps are. The more delicate parts of the discussion will actually not be used later on. General references for the singularity theory of smooth maps are the textbooks of Golubitsky and Guillemin [31] and Arnol'd et al. [2]. We will mostly follow [31].

For the moment, we consider two smooth manifolds M and N of arbitrary dimensions m and n, respectively. For simplicity we assume that M is closed and that N has empty boundary. However, both may be non-orientable and/or have several components. Given a smooth map $f\colon M \to N$ we denote its differential by $df\colon TM \to TN$. Recall that $p \in M$ is called a *critical point* of f (or a *singularity*) if the rank of df_p is not maximal, and the that image of a critical point is called a *critical value*. We refer to the sets $\mathcal{C}_f$ of critical points and $f(\mathcal{C}_f)$ of critical values as the *critical locus* and the *critical image* of f. Sometimes we will also denote the critical locus by $\mathcal{C}(f)$ for aesthetic reasons.

Since arbitrary smooth maps can be very complicated, one of the goals of singularity theory is to find reasonably large sets of maps with nice properties that rule out as much pathological behavior as possible. In order fill the word large with meaning, it is necessary to equip $C^\infty(M,N)$ with a topology. For compact M there is a natural choice, namely the C^∞ *topology* of uniform convergence of all partial derivatives.[2] An obvious interpretation of large subsets of $C^\infty(M,N)$ is to require that they are open and dense. While this interpretation is not perfect, it serves for our purposes and we will stick with it.

As a warm up, let us take a look at the situation of real valued functions. Recall that for $f \in C^\infty(M,\mathbb{R})$ and a critical point $p \in \mathcal{C}_f$ there is a well defined notion of

[2] Of course, in order to define this topology one has to make some choices, such as Riemannian metrics or atlases, but for compact M all choices give rise to the same topology. In the non-compact case one has to resort to a finer topology with better properties known as the *Whitney topology* or *strong C^∞ topology* (see [31, p.42ff.]). But these delicacies shall not concern us.

second derivative at p. This takes the form of a symmetric bilinear form on T_pM and is called the *Hessian* of f at p. Then p is called *non-degenerate* if its Hessian is non-degenerate and the *index* of p is defined as the number of negative eigenvalues of the Hessian. Functions with only non-degenerate critical points are known as *Morse functions*. Following Cerf [13] we call a Morse function that is injective on its critical points *excellent*. The following is well known.

Theorem 2.1. *(Excellent Morse functions) Let M be a smooth n–manifold. The set of all smooth functions $f\colon M \to \mathbb{R}$ such that*

(a) f is a Morse function, that is, all its critical points are non-degenerate, and

(b) f is injective on its critical points

is open and dense in $C^\infty(X, \mathbb{R})$. Moreover, near each non-degenerate critical point f has a local model of the form

$$(x_1, \dots, x_n) \mapsto -x_1^2 - \cdots - x_k^2 + x_{k+1}^2 + \cdots + x_n^2 \tag{2.1}$$

for some $k \in \{1, \dots, n\}$.

Since we will frequently work with local models for smooth maps, we want to make absolutely clear what we mean by this.

Definition 2.2. Let $f\colon M \to N$ be a smooth map. We say that f has a *local model* $F\colon \mathbb{R}^m \to \mathbb{R}^n$ at $p \in M$ if there are local coordinates centered at $p \in M$ and $f(p) \in N$ such that the coordinate representation of f agrees with F. If either M or N are oriented the coordinates are required to respect orientations.

Many different proofs for Theorem 2.1 are available in the literature. One that is very conceptual and anticipates a generalization to maps between arbitrary manifolds can be found in [31, Ch. II.6][3]. The key ideas are the notion of transversality and the language of jet spaces, culminating in the so called *Multijet Transversality Theorem* [31, Theorem II.4.13]. The latter is an extremely powerful tool whose importance in differential topology can hardly be overstated. Unfortunately, we can neither formulate nor explain this result without going too far astray. But just to give a rough sketch of the proof of Theorem 2.1: the openness and denseness of conditions (a) and (b) follows from two applications of the transversality theorem, and the construction of the local models is the content of the classical *Morse Lemma* [31, Theorem II.6.9].

Remark 2.3. The whole discussion above, including Theorem 2.1, extends verbatim to the case when the target $\mathbb{R}$ is replaced by an oriented, 1–dimensional manifold. It is convenient to speak of (excellent) Morse functions in this context as well.

As indicated, Theorem 2.1 admits a generalization to maps between arbitrary manifolds and we will state the version for maps from 4–manifolds to surfaces in Theorem 2.9 below. However, to give the proper context we embark on a small digression about the general situation. Note that condition (a) in Theorem 2.1 is concerned with the nature of critical points in the source, while condition (b)

[3] As stated, Theorem 2.1 follows from Theorem II.6.2, Propositions II.6.6 and II.6.13, and Theorem II.6.9 in [31].

addresses how critical points are mapped into the target. Both conditions have natural generalizations to arbitrary dimension.

The basic idea for finding an analogue of condition (a) is to partition the critical points of a smooth map $f\colon M^m \to N^n$ according to their level of degeneracy. More precisely, for a non-negative integer $r \in \mathbb{Z}_{\geq 0}$ we consider the sets

$$\mathcal{S}_r(f) = \{p \in M \,|\, \mathrm{rk}(df_p) = \min\{m,n\} - r\} \subset M$$

of points where f *drops rank by* r. Note that $\mathcal{S}_0(f)$ consists of the regular points while each critical point lies in exactly one $\mathcal{S}_r(f)$ with $r > 0$. Of course, $\mathcal{S}_r(f)$ is empty if r exceeds the dimension of either M or N. As explained in [31, p.143], for sufficiently nice f the sets $\mathcal{S}_r(f)$ are submanifolds of M of codimensions

$$\operatorname{codim} \mathcal{S}_r(f) = r^2 + r|m-n|. \tag{2.2}$$

This observation suggests the following inductive scheme which is attributed to Thom [65]. Given a sequence of integers $r_1, \ldots, r_{k+1} \in \mathbb{Z}_{\geq 0}$ and assuming that $\mathcal{S}_{r_1,\ldots,r_k}(f)$ is a submanifold of M (and also of $\mathcal{S}_{r_1,\ldots,r_{k-1}}(f)$ if $k > 1$) we define

$$\mathcal{S}_{r_1,\ldots,r_{k+1}}(f) = \mathcal{S}_{r_{k+1}}(f|_{\mathcal{S}_{r_1,\ldots,r_k}(f)}).$$

Note that if $\mathcal{S}_{r_1,\ldots,r_{k+1}}(f)$ again turns out to be a submanifold of $\mathcal{S}_{r_1,\ldots,r_k}(f)$, then according to equation (2.2) it has positive codimension there unless $r_{k+1} = 0$. In particular, the above process becomes stagnant after finitely many steps, in the sense that f has no critical points when restricted to $\mathcal{S}_{r_1,\ldots,r_k}(f)$ so that $\mathcal{S}_{r_1,\ldots,r_{k+1}}(f)$ is empty for $r_{k+1} > 0$. So eventually, if f is nice enough for everything to work out, the source M is partitioned into submanifolds of the form $\mathcal{S}_{r_1,\ldots,r_k,0}(f)$. The general analogue of condition (a) in Theorem 2.1 is that the above process can be carried out. The fact that this is possible for a dense set of maps in $C^\infty(M,N)$ was proved by Boardman [9]; we follow [31, p.157] and call such maps *Boardman maps*.

The partition of M is commonly known as the *Thom-Boardman stratification* (or *TB stratification*, for short). We will usually ignore the top stratum $\mathcal{S}_0(f)$ of regular points and only focus on the stratification of the critical locus. Again, a proper exposition of the Thom-Boardman stratification requires the notion of jet spaces and our discussion should be taken with a grain of salt. The interested reader is referred to [31, Ch. VI.5] and [9] for more details.

Remark 2.4. Surprisingly, the set of Boardman maps fails to be open in general. According to Wilson [70], its openness depends on the dimensions of M and N. More precisely, the Boardman maps form an open set if and only if either $n < 4$ or $2n > 3m - 4$.

We now shift our attention to maps $f\colon X \to B$ from a closed 4–manifold X to a surface B. Before stating the analogue of condition (b) in Theorem 2.1 we want to discuss two important examples of Boardman maps.

Example 2.5 (Fold models)**.** Consider the maps $\mathrm{F}_\pm\colon \mathbb{R}^4 \to \mathbb{R}^2$ given by

$$\mathrm{F}_\pm(t,x,y,z) = \big(t, x^2 + y^2 \pm z^2\big). \tag{2.3}$$

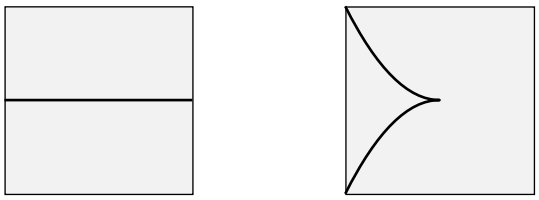


Figure 2: The critical images of the fold (left) and cusp models (right). In both pictures the t–parameter increases from left to right and the origin of $\mathbb{R}^2$ is at the center.

We will refer to these maps as the *fold models*, which come in two flavors called *definite* (F_+) and *indefinite* (F_-). A direct calculation shows that

$$d\mathrm{F}_\pm(t,x,y,z) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2x & 2y & \pm 2z \end{pmatrix}$$

from which we immediately see that $\mathcal{C}(\mathrm{F}_\pm) = \{(\tau,0,0,0) \,|\, \tau \in \mathbb{R}\}$ is a 1–dimensional submanifold of $\mathbb{R}^4$ and that $\mathrm{F}_\pm$ drops rank by 1 at each critical point. In particular, we find $\mathcal{C}(\mathrm{F}_\pm) = \mathcal{S}_1(\mathrm{F}_\pm)$. Moreover, the restriction of $\mathrm{F}_\pm$ to its critical locus obviously has full rank so that $\mathrm{F}_\pm$ is a Boardman map and its TB stratification has only one non-empty singular stratum, namely $\mathcal{C}(\mathrm{F}_\pm) = \mathcal{S}_{1,0}(\mathrm{F}_\pm)$.

Example 2.6 (Cusp models)**.** Next we take a look at the (*indefinite* and *definite*) *cusp models* $\mathrm{C}_\pm \colon \mathbb{R}^4 \to \mathbb{R}^2$ given by

$$\mathrm{C}_\pm(t,x,y,z) = \big(t, x^3 + 3tx + y^2 \pm z^2\big). \tag{2.4}$$

Again, we compute the differential and obtain

$$d\mathrm{C}_\pm(t,x,y,z) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 3t & 3(x^2+t) & 2y & \pm 2z \end{pmatrix}.$$

The critical locus is a parabola cut out by the equations $x^2 + t = y = z = 0$. In particular, it is again a 1–dimensional submanifold. Also, just as for the fold models, the rank drops by 1 at all critical points, so that $\mathcal{C}(\mathrm{C}_\pm) = \mathcal{S}_1(\mathrm{C}_\pm)$. If we parametrize the critical locus by the curve $(-\tau^2, \tau, 0, 0)$, then the restriction of $\mathrm{C}_\pm$ corresponds to the map

$$\tau \mapsto \mathrm{C}_\pm(-\tau^2,\tau,0,0) = (-\tau^2, -\tau^3)$$

whose differential is non-zero for $\tau \neq 0$ but vanishes for $\tau = 0$. It follows that the origin of $\mathbb{R}^4$ is contained in $\mathcal{S}_{1,1,0}$. So $\mathrm{C}_\pm$ is also a Boardman map and its TB stratification takes the more complicated form

$$\begin{array}{rcccc} \mathcal{C}(\mathrm{C}_\pm) & = & \mathcal{S}_{1,0}(\mathrm{C}_\pm) & \amalg & \mathcal{S}_{1,1,0}(\mathrm{C}_\pm) \\ & = & \big\{(-\tau^2,\tau,0,0) \,\big|\, \tau \neq 0\big\} & \amalg & \{(0,0,0,0)\}. \end{array}$$

The critical images of the fold and cusp models are shown in Figure 2; the cuspidal shape is the reason for the name cusp.

It turns out that in some sense these examples capture the full local complexity of singularities of maps from 4–manifolds to surfaces. Indeed, if $f \colon X \to B$ is

a Boardman map, then the only candidates for non-empty TB strata are $\mathcal{S}_{1,0}(f)$ and $\mathcal{S}_{1,1,0}(f)$. According to equation (2.2) the codimension of $\mathcal{S}_r(f)$ is r^2+2r which exceeds the dimension of X as soon as $r \geq 2$. So $\mathcal{C}_f = \mathcal{S}_1(f)$ has codimension 3 and is thus a 1–dimensional submanifold of X. Similarly, one argues that $\mathcal{S}_{1,s}$ has to be empty for $s \geq 2$ while $\mathcal{S}_{1,1}(f)$ has codimension 1 in $\mathcal{S}_1(f)$, making it a finite set of points. These two types of critical points have special names.

Definition 2.7 (Folds and cusps)**.** Let $f\colon X \to B$ be a smooth map from a 4–manifold to a surface. A critical point $p \in \mathcal{C}_f$ is called a *fold point* if f drops rank by 1 at p, $\mathcal{C}_f$ is a 1–dimensional submanifold near p, and $f|_{\mathcal{C}_f}$ has non-zero derivative at p. Similarly, $p \in \mathcal{C}_f$ is called a *cusp point* if f drops rank by 1 at p, $\mathcal{C}_f$ is a 1–dimensional submanifold near p, and the derivative of $f|_{\mathcal{C}_f}$ vanishes at p. We denote the sets of fold and cusp points of f by $\mathcal{C}_f^{\mathrm{fo}}$ and $\mathcal{C}_f^{\mathrm{cu}}$, respectively.

The following is clear from the definitions.

Lemma 2.8. *A map $f\colon X \to B$ from a 4–manifold to a surface is a Boardman map if and only if all its critical points are folds and cusps. In that case we have $\mathcal{S}_{1,0}(f) = \mathcal{C}_f^{\mathrm{fo}}$ and $\mathcal{S}_{1,1,0}(f) = \mathcal{C}_f^{\mathrm{cu}}$.*

Now, remember that we are still looking for an analogue of condition (b) in Theorem 2.1, which should be a condition on how $\mathcal{C}_f$ is mapped into B. Loosely speaking, the restriction of f to $\mathcal{C}_f$ and all its TB strata should be as regular as possible. More precisely, since the cusp stratum $\mathcal{S}_{1,1,0}(f)$ is a finite set of points, it is natural to require that it is mapped injectively into B. Similarly, in the light of the Whitney's immersion theorem [31, Thm. II.5.7], f should restrict to an immersion with normal crossings on the 1–dimensional fold stratum $\mathcal{S}_{1,0}(f)$. Lastly, the images of $\mathcal{S}_{1,0}(f)$ and $\mathcal{S}_{1,1,0}(f)$ should be disjoint, because a point and a line in the plane are generically disjoint. We can now state the analogue of Theorem 2.1 for maps from 4–manifolds to surfaces.

Theorem 2.9 (Maps from 4–manifolds to surfaces)**.** *Let X be a closed 4–manifold and B a surface. The set of smooth maps $f\colon X \to B$ such that*

(a) f is a Boardman map, that is, all its critical points are folds or cusps, and
(b) $f|_{\mathcal{C}_f^{\mathrm{fo}}}$ is an immersion with normal crossings and f is injective on $f^{-1}(f(\mathcal{C}_f^{\mathrm{cu}}))$

is open and dense in $C^\infty(X, B)$. Moreover, each fold or cusp point is locally modeled on the fold or cusp models $\mathrm{F}_\pm$ and $\mathrm{C}_\pm$ from Examples 2.5 and 2.6.

Proof. The denseness of the conditions (a) and (b) is a special case of [31, Theorem VI.5.2, p.157]. (Note that condition (b) is equivalent to "Condition NC" stated on p.157 of [31].) Moreover, condition (a) is open by Remark 2.4 (since $\dim B < 4$), and the openness of condition (b) follows from the multijet transversality theorem as in the proof of [31, Proposition VI.5.6, p.158] with the additional input that M is compact. Finally, the construction of local models for folds and cusps is due to Morin [53] and Levine [46, p.154f.]. □

Remark 2.10 (Definite or indefinite?). Given a map $f\colon X \to B$ with a fold or cusp point $p \in \mathcal{C}_f$, there are two candidates for a local model, the definite or the indefinite

one. As in the case of Morse functions, one can decide which one fits by studying a notion of second derivative. As explained in [31, p. 152] there is a well defined symmetric bilinear map

$$\delta_p^2 f\colon \ker df_p \times \ker df_p \longrightarrow \operatorname{coker} df_p = T_{f(p)}B/\operatorname{im} df_p$$

which naturally generalizes the Hessian (see also [2, p.60ff.]). If p is a fold or cusp point, then $\operatorname{coker} df_p$ is 1–dimensional and the choice of a non-zero vector provides an identification with the real line. Using such an identification, $\delta_p^2 f$ becomes a symmetric bilinear form on $\ker df_p$. (For example, in the fold and cusp models, $\delta_p^2 f$ can be identified with the Hessian of either $x^2+y^2-z^2$ or $x^3+y^2-z^2$.) If all non-zero eigenvalues of $\delta_p^2 f$ have the same sign, then the singularity is definite. If there are eigenvalues of both sign, then we have an indefinite singularity. Note that these conditions are independent of the choice of identification $\operatorname{coker} df_p \cong \mathbb{R}$.

To summarize, for a map $f\colon X \to B$ from a 4–manifold to a surface that satisfies the conditions in Theorem 2.9 the critical locus $\mathcal{C}_f \subset X$ is a 1–dimensional submanifold which decomposes into finitely many open arcs, the connected components of the fold locus $\mathcal{C}_f^{\text{fo}} = \mathcal{S}_{1,0}(f)$, whose ends limit to the cusp locus $\mathcal{C}_f^{\text{cu}} = \mathcal{S}_{1,1,0}(f)$ which is a finite set. The critical points are locally mapped into B according to the models discussed in Examples 2.5 and 2.6 (see also Figure 2) and the only multiple points that occur in the critical image are transverse intersections between the images of arcs of folds. For brevity we will sometimes refer to both, the arcs of folds in X and their images in B, as *fold arcs*. This should not cause any confusion since it will usually be clear from the context where the arcs in question live.

Remark 2.11 (Stability). Another important concept in singularity theory that we have not discussed so far is the notion of stability. A smooth map $f\colon M \to N$ is called *stable* if every $g \in C^\infty(M,N)$ sufficiently close to f is *equivalent* to f, that is, there are diffeomorphisms ϕ of M and ψ of N such that $g = \psi \circ f \circ \phi^{-1}$. For 1–dimensional N it is a classical fact that f is stable if and only if it is an excellent Morse function (see [31, Ch. II.6]). Moreover, it turns out that the conditions stated in Theorem 2.9 characterize the stable maps from 4–manifolds to surfaces. This is well known folklore in singularity theory and can be proved along the lines of [31, IV.6.3, Theorem 6.3] where the details are worked out for the more complicated case of maps between 4–manifolds. Curiously enough, while Boardman maps fail to be open in certain dimensions, stable maps are not always dense (see [31, p.160ff.]). However, in the context that is most relevant to us the two notions agree and we can enjoy the best of both worlds.

2.3 Surfaces and Their Mapping Class Groups

The regular fibers of a smooth map $f\colon X \to B$, with both X and B oriented, are compact, oriented surfaces and these fibers will play an important role later on, and we are naturally led into the theory of surfaces and their mapping class groups. This itself is a vast subject and, in the author's experience, many different conventions are in use, sometimes making it difficult to decide whether a statement in some reference actually applies to a situation at hand. For this reason we will give very precise definitions. As a general reference we use the book of Farb and Margalit [24].

Surfaces and Simple Closed Curves. By a *surface* we mean a compact, orientable, 2–dimensional manifold Σ, possibly with non-empty boundary, and possibly equipped with a finite set of *marked points* $P \subset \Sigma \setminus \partial\Sigma$ which is usually not mentioned explicitly. A *simple closed curve* in Σ is a closed, connected, 1–dimensional submanifold $a \subset \Sigma$ which neither meets the boundary nor the marked points. In other words, it is the image of an embedding $S^1 \hookrightarrow \Sigma \setminus (\partial\Sigma \cup P)$. Similarly, a *simple arc* $r \subset \Sigma$ is the image of a proper embedding $([0,1],\{0,1\}) \hookrightarrow (\Sigma, \partial\Sigma \cup P)$. Note however, that the embedding is not part of the data. Also, simple closed curves and simple arcs are unoriented objects according to our definition, but at times it will be convenient to choose orientations in order to speak of (integral) homology classes or fundamental group elements. In those situations we will sometimes use the notation $\vec{a}$ for an oriented simple closed curve $a \subset \Sigma$. Simple closed curves and other objects related to Σ are usually considered up to ambient isotopy in Σ via isotopies which leave the boundary as well as the marked points fixed. We use the notation $a \sim b$ to indicate that two given objects are isotopic.

Intersection numbers. Let $a, b \subset \Sigma$ be a pair of simple closed curves. There are several ways to count intersections between a and b. The crudest way is to simply count the number of points in $a \cap b$ (sometimes called the *numerical intersection number*). But this count might not be finite and it is certainly not invariant under isotopies. With respect to these properties, a better approach is the *geometric intersection number*

$$i(a,b) = \min\{\#(\alpha \cap \beta) | \alpha \sim a,\ \beta \sim b,\ \alpha \pitchfork \beta\} \in \mathbb{N} \tag{2.5}$$

where the symbol $\pitchfork$ indicates a transverse intersection. Obviously, $i(a,b)$ is finite, isotopy invariant, and also symmetric in a and b. A third way of counting requires that a, b, and Σ are oriented. In this situation the *algebraic intersection number* is defined as

$$\langle a,b\rangle = \langle a,b\rangle_\Sigma = \langle [a],[b]\rangle_{H_1(\Sigma)} \in \mathbb{Z}$$

where bracket on the right hand side denotes the intersection form on $H_1(\Sigma)$. In contrast to the geometric intersection number, $\langle a,b\rangle$ is skew symmetric and only depends on the homology classes of a and b. However, both $i(a,b)$ and $\langle a,b\rangle$ always have the same parity (even or odd) and satisfy the inequality

$$|\langle a,b\rangle| \leq i(a,b). \tag{2.6}$$

Note that the left hand side is actually independent of the chosen orientations.

As far as computability is concerned, assuming that a and b intersect transversely, it is obviously easiest to determine $\#(a \cap b)$ by simply counting points, followed by $\langle a,b\rangle$ where one has to count with signs, and then there is $i(a,b)$ which is harder to come by. In order to compute $i(a,b)$ one has to bring a and b in *minimal position*, that is, one has to find isotopic curves that are transverse and minimize the number of intersections. This seems difficult from the outset but it can be done in finitely many steps due to the so called *bigon criterion*. Following [24, Section 1.2.4] we say that a and b form a *bigon*, if there is an embedded disk $\Delta \subset \Sigma$ whose interior is disjoint from a and b, and whose boundary is a union of an arc of a and an arc of b intersecting in exactly two points (see Figure 3). Given such a bigon, one can push b

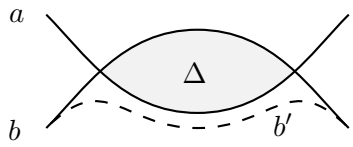


Figure 3: Two curves $a, b \subset \Sigma$ forming a bigon Δ, and a curve b' obtained from b by a Whitney move across Δ.

across Δ by a 2–dimensional Whitney move resulting in a curve b' with two fewer intersections with a. An immediate consequence of the following result is that one can bring a and b in minimal position by finitely many Whitney moves.

Proposition 2.12 (Bigon criterion)**.** *Let $a, b \subset \Sigma$ be simple closed curves intersecting transversely. If a and b are not in minimal position, then they form a bigon.*

Proof. See [24, Proposition 1.7]. □

Later on we will frequently encounter pairs of simple closed curves with geometric or algebraic intersection number one and we find it convenient to introduce the following terminology.

Definition 2.13 (Duality for curves)**.** Two simple closed curves $a, b \subset \Sigma$ are called

(a) *algebraically dual* if $\langle a, b\rangle = 1$ (for some choice of orientations),

(b) *weakly dual* if $i(a, b) = 1$, and

(c) *strongly dual* if a and b intersect transversely in a single point.

Remark 2.14. In [7] we used the term "geometrically dual" instead of "weakly dual". However, it has been brought to our attention that this was misleading since "geometrically dual" is commonly used in the sense of "strongly dual", especially in higher dimensional contexts. In order to prevent this potential confusion, we decided to change our terminology.

Diffeomorphisms and Mapping Class Groups of Surfaces. We turn to diffeomorphisms of surfaces. Let $\mathrm{Diff}^+_\partial(\Sigma, P)$ be the group of orientation preserving diffeomorphisms that restrict to the identity on $\partial\Sigma$ and preserve the set P of marked points (possibly permuting the points). The *mapping class group* of Σ is defined as

$$\mathrm{Mod}(\Sigma) = \pi_0\big(\mathrm{Diff}^+_\partial(\Sigma, P))\big).$$

Simple closed curves play an important role in the theory of mapping class groups. On the one hand, a simple closed curve $a \subset \Sigma$ gives rise to an element $\tau_a \in \mathrm{Mod}(\Sigma)$ called the (right-handed) *Dehn twist* about a, which depends only on the isotopy class of a. A diffeomorphism representing τ_a is obtained by taking out an annulus neighborhood of a, applying a full right-handed twist to it, and gluing it back in. Similarly, a simple arc $r \subset \Sigma$ connecting two distinct marked points gives rise to a *half-twist* $\bar{\tau}_r \in \mathrm{Mod}(\Sigma)$, represented by a diffeomorphism supported in a disk neighborhood D of r, ambiently isotoping r onto itself by a clockwise 180 degree rotation, thus permuting the two marked points, while keeping ∂D fixed. (Note that $\bar{\tau}_r$ is a square root of $\tau_{\partial D}$ in $\mathrm{Mod}(\Sigma)$.) We will frequently use the following

elementary facts about Dehn twists. First of all, if $a \subset \Sigma$ is a simple closed curve and $f \in \mathrm{Diff}(\Sigma)$, then the Dehn twist about $f(a)$ is given by

$$\tau_{f(a)} = f\tau_a f^{-1} \in \mathrm{Mod}(\Sigma) \tag{2.7}$$

where the right hand side is to be understood as the isotopy class of fT_af^{-1} where $T_a \in \mathrm{Diff}^+_\partial(\Sigma)$ is some representative of τ_a (see [24, p.73f.]). Furthermore, if $a, b \subset \Sigma$ is a pair of weakly dual curves (that is, $i(a,b) = 1$), then we have

$$\tau_a\tau_b(a) \sim b \tag{2.8}$$

and the so called *braid relation*

$$\tau_a\tau_b\tau_a = \tau_b\tau_a\tau_b \in \mathrm{Mod}(\Sigma) \tag{2.9}$$

holds (see [24, p.78f.]).

It is well known that $\mathrm{Mod}(\Sigma)$ is generated by Dehn twists and half-twists (see [24, Chapter 4], for example). On the other hand, mapping classes can be effectively studied by their action on (isotopy classes of) simple closed curves and simple arcs. In particular, it is desirable to understand the effect of Dehn twists on simple closed curves. While this can be quite complicated for high geometric intersection numbers, the situation becomes accessible on the level of homology.

Proposition 2.15 (Picard–Lefschetz formula)**.** *Let Σ be a surface, $a \subset \Sigma$ a simple closed curve and let $x \in H_1(\Sigma)$. Then for any orientation on a we have*

$$(\tau_a^k)_* x = x + k\langle [a], x\rangle [a]. \tag{2.10}$$

In particular, if b is an oriented simple closed curve, then

$$[\tau_a^k(b)] = [b] + k\langle [a], [b]\rangle [a]. \tag{2.11}$$

Proof. See [24, Proposition 6.3]. □

Remark 2.16. The Picard–Lefschetz formula is particularly useful when $\Sigma = T^2$. In that case, mapping classes are completely determined by their action on homology (see [24, Theorem 2.5], for example).

Another useful tool is the so called *change of coordinates principle* which roughly states that any two configurations of simple closed curves on a surface with the same intersection pattern can be mapped onto each other by a diffeomorphism. We will only use the following special cases. For details we refer to [24, Chapter 1.3].

Proposition 2.17 (Change of coordinates principle)**.** *If $a, b \subset \Sigma$ is a pair of non-separating simple closed curves, then there exists some $\phi \in \mathrm{Diff}^+_\partial(\Sigma, P)$ such that $\phi(a) = b$. Furthermore, if a, b and a', b' are two pairs of strongly dual curves, then there is some $\phi \in \mathrm{Diff}^+_\partial(\Sigma, P)$ such that $\phi(a) = a'$ and $\phi(b) = b'$.*

It is worthwhile mentioning that the higher homotopy groups of the diffeomorphism groups of compact surfaces are well understood. In fact, the homotopy type of the identity components (and thus of all components) was determined by Earle, Eells, and Schatz.

Theorem 2.18 (Earle–Eells, Earle–Schatz [19,20])**.** *Let Σ be a compact, connected, orientable surface and let $\mathrm{Diff}^0_\partial(\Sigma)$ be the identity component of $\mathrm{Diff}_\partial(\Sigma)$.*

(i) If Σ is closed of genus $g \geq 2$, then $\mathrm{Diff}^0_\partial(\Sigma) = \mathrm{Diff}^0(\Sigma)$ is contractible. Furthermore, the canonical inclusions $\mathrm{SO}(3) \hookrightarrow \mathrm{Diff}^0(S^2)$ and $T^2 \hookrightarrow \mathrm{Diff}^0(T^2)$ are homotopy equivalences.

(ii) If $\partial\Sigma \neq \emptyset$, then $\mathrm{Diff}^0_\partial(\Sigma)$ is contractible.

Proof. See [19, p.21, Corollary] for (i) and [20, p.170, 2nd Theorem] for (ii). □

This has many important consequences. We would like to point out three of them. The first concerns spaces of simple closed curves. We refer to [37, Chs. 2.6&2.7] for a general discussion of spaces of embeddings and submanifolds as well as a proof of the following result.

Corollary 2.19. *Let Σ be a closed, orientable surface of genus at least two. Then the space of non-separating simple closed curves has weakly contractible components (that is, all homotopy groups of each component are trivial).*

Proof. See [37, Theorem 2.7.H]. □

Another consequence concerns what we call the *automorphisms* of $\Sigma \times S^1$, that is, the orientation and fiber preserving diffeomorphism of $\Sigma \times S^1$ considered as the trivial Σ–bundle over S^1. (More generally, we refer to orientation and fiber preserving diffeomorphisms of total spaces of oriented fiber bundles as automorphisms.) These form a group, denoted by $\mathrm{Aut}(\Sigma \times S^1)$, whose elements can be identified with maps from S^1 into $\mathrm{Diff}^+(\Sigma)$, or in other words, loops in $\mathrm{Diff}^+(\Sigma)$. By fixing a fiber and identifying it with Σ, we obtain a short exact sequence of groups

$$1 \to \pi_1\big(\mathrm{Diff}(\Sigma), \mathrm{id}\big) \to \pi_0\big(\mathrm{Aut}(\Sigma \times S^1)\big) \to \mathrm{Mod}(\Sigma) \to 1 \tag{2.12}$$

where the map $\pi_0\big(\mathrm{Aut}(\Sigma \times S^1)\big) \to \mathrm{Mod}(\Sigma)$ induced by restricting an automorphism to a fixed fiber which is canonically identified with Σ. Note that this sequence is split by sending $\phi \in \mathrm{Diff}^+(\Sigma)$ to the *constant* automorphism $\phi \times \mathrm{id}_{S^1}$. Combining the exact sequence above with Theorem 2.18 we immediately see that if the genus of Σ is at least two, then every automorphism of $\Sigma \times S^1$ is isotopic to a constant one. The following easy consequence will be important for our purposes.

Corollary 2.20. *Let Σ be a closed, oriented surface. Then $\phi \in \mathrm{Aut}(\Sigma \times S^1)$ extends to an automorphism of $\Sigma \times S^1$ if and only if it is isotopic to a constant one. In particular, if Σ has genus at least two, then all automorphism extend.*

Finally, we would like to mention the classification of surface bundles over the 2–sphere.

Corollary 2.21. *Let $X \to S^2$ be a surface bundle with closed fibers of genus g.*

(i) If $g = 0$, then X is diffeomorphic to $S^2 \times S^2$ or $\mathbb{C}\mathrm{P}^2 \# \overline{\mathbb{C}\mathrm{P}^2}$.

(ii) If $g = 1$, then X is diffeomorphic to $T^2 \times S^2$, $S^1 \times S^3$ or $S^1 \times L(n,1)$.

(iii) If $g \geq 2$, then X is diffeomorphic to $\Sigma_g \times S^2$.

Proof. Let Σ_g be a closed surface of genus g. It follows from Corollary 2.20 that Σ_g–bundles over S^2 are classified by elements of $\pi_1\big(\mathrm{Diff}(\Sigma_g);\mathrm{id}\big)$. For $g \neq 1$ the above classification is then easily deduced from Theorem 2.18. The genus one case is slightly more complicated and is covered in [6, Lemma 10]. □

Chapter 3

Wrinkled Fibrations and Related Structures

3.1 Wrinkled Fibrations and Broken Lefschetz Fibrations

After all these preliminaries we narrow in closer toward the core of our work. We start with one of the central definitions. Let M be a 4–manifold and B a surface, both satisfying our standing assumptions[1].

Definition 3.1 (Wrinkled fibrations)**.** A *wrinkled fibration* is a map $f\colon M \to B$ satisfying $\partial M = f^{-1}(\partial B)$ and the following conditions:

(a) All critical points of f are indefinite folds and cusps.

(b1) The cusp locus $\mathcal{C}_f^{\mathrm{cu}}$ does not meet ∂M and f is injective on $f^{-1}\big(f(\mathcal{C}_f^{\mathrm{cu}})\big)$.

(b2) The fold locus $\mathcal{C}_f^{\mathrm{fo}}$ is transverse to ∂M and f restricts to an immersion with normal crossings on $\mathcal{C}_f^{\mathrm{fo}}$.

Two wrinkled fibrations $f\colon M \to B$ and $f'\colon M' \to B'$ are called *equivalent* if there are orientation preserving diffeomorphisms $\hat{\phi}\colon M \to M'$ and $\check{\phi}\colon B \to B'$ such that $f' = \check{\phi}\, f\, \hat{\phi}^{-1}$.

Our general philosophy is to consider wrinkled fibrations as generalized fiber bundles and we will borrow some terminology from this context. We will usually refer to the source and target as *total space* and *base*, respectively, and we call the preimages of points *fibers*.

A closely related class of maps are the so called *broken Lefschetz fibrations*. Their definition is almost the same as Definition 3.1 except that cusps are replaced by *Lefschetz singularities*. The latter are defined in terms of the local model

$$\mathrm{L}\colon \mathbb{C}^2 \to \mathbb{C}, \quad (z,w) \mapsto z^2 + w^2$$

using complex local coordinates. More precisely, broken Lefschetz fibrations have only indefinite folds and Lefschetz singularities and satisfy the obvious analogues

[1] As a reminder, these are smooth, compact, connected and oriented.

of conditions (b1) and (b2) in Definition 3.1. Broken Lefschetz fibrations without indefinite folds are simply called *Lefschetz fibrations*. In fact, we have introduced the maps in reverse chronological order. As described in Section 1.2, the correct order is: first Lefschetz, then broken Lefschetz, and finally wrinkled fibrations.

Remark 3.2. It is important that Lefschetz singularities are modeled using orientation preserving charts. The use of orientation reversing charts for the Lefschetz model leads to so called *achiral* Lefschetz singularities. These can also be modeled in orientation preserving charts by $(z, w) \mapsto z^2 + \overline{w}^2$. The point is that in orientation preserving complex charts Lefschetz fibrations are locally holomorphic while achiral singularities disrupt this property. For folds and cusps this does not really matter because both models have an orientation reversing symmetry.

Remark 3.3. Before moving on we would like to point out the unfortunate diversity of terminology used in the field. The term "wrinkled fibration" was originally introduced by Lekili [45] as a mixture of broken Lefschetz fibrations and the"wrinkled maps" studied by Eliashberg and Mishachev [21] (these are certain maps with only folds and cusps). However, Lekili allowed his "wrinkled fibrations" to have Lefschetz singularities and used "purely wrinkled fibration" for maps with only indefinite folds and cusps; he also did not require (b1) or (b2). Williams [67] adopted Lekili's terminology of "purely wrinkled fibrations" but implicitly added the conditions ($b1$) and ($b2$). Meanwhile, what we call wrinkled fibrations is called "indefinite generic map" by Baykur [4] whereas Gay and Kirby [28] use the name "indefinite Morse 2–function". Unfortunately, it is not foreseeable which terminology will eventually catch on.

In order to make the connection with Section 2.2, we note that if M is closed, then a wrinkled fibration $f\colon M \to B$ is just a map as in Theorem 2.9 with the additional assumption that all critical points are indefinite. So in some sense, wrinkled fibrations can be considered as analogues of Morse functions without local extrema. In the case of non-empty boundary the conditions ($b1$) and ($b2$)[2] imply that wrinkled fibrations restrict to excellent Morse functions[3] over their boundary components. As explained in Section 1.2, the questions of existence and uniqueness of wrinkled fibrations have a convoluted history that is largely intertwined with the analogous questions for broken Lefschetz fibrations. Since all the due credits were already given in Section 1.2, we limit ourselves to stating the most general existence result due to Gay and Kirby [28]. The uniqueness will be discussed in Section 3.4.

Theorem 3.4 ([28, Theorem 1.1])**.** *Let $f\colon M \to B$ be a map from a 4–manifold to a surface such that $\partial M = f^{-1}(\partial B)$ and $f\colon \partial M \to \partial B$ is an excellent Morse function. Then f is homotopic relative to ∂M to a wrinkled fibration if and only if $f_*\pi_1(M)$ has finite index in $\pi_1(B)$.*

In particular, for $B = S^2$ the finite index condition is always satisfied so that all maps $M \to S^2$ are homotopic to wrinkled fibrations. In this case one can even do better and obtain *simple* wrinkled fibrations as observed by Williams [67], see Definition 4.1 and Theorem 4.2 below.

[2]The conditions (b1) and (b2) are most likely open and dense for maps $f\colon M \to B$ with $\partial M = f^{-1}(\partial B)$ but we are not aware of any reference. However, we will actually not need this.

[3]See Remark 2.3 on page 17 for the definition.

Let us discuss some immediate consequences of the definition of wrinkled fibrations. It follows from the discussion in Section 2.2 that $\mathcal{C}_f \subset M$ is a properly embedded, 1–dimensional submanifold – in other words, $\mathcal{C}_f$ is the disjoint union of a finite number of embedded circles in the interior and properly embedded arcs. Moreover, the critical image $f(\mathcal{C}_f) \subset B$ is immersed except for the finitely many cusps, and its complement $B \setminus f(\mathcal{C}_f)$ has finitely many connected components to which we refer as *regions*. The condition that $\partial M = f^{-1}(\partial B)$ implies that all regular fibers are closed, orientable surfaces, which are oriented by the fiber first convention and the orientations of M and B. As mentioned before, we will usually think of wrinkled fibrations as singular families of surfaces parametrized by the base. Since f restricts to a proper submersion over each region, such a restriction is a fiber bundle. In particular, if two fibers are mapped to the same region, then they must be diffeomorphic. In Section 3.3 we will study how the topology of the fibers changes near indefinite folds and cusps. In particular, we will see that wrinkled fibrations are automatically surjective (see the discussion after Definition 3.12). But first we have to introduce some tools to relate different fibers in a wrinkled fibration.

3.2 Parallel Transport in Wrinkled Fibrations

As mentioned above, we not only want to understand the fibers of a wrinkled fibration individually, but also how they fit together. For that purpose we will generalize the concept of parallel transport in fiber bundles along arcs in the base (see [18, p.225ff.], for example). The main technical tool is a consistent choice of complements for the tangent spaces of the fibers.

Definition 3.5. Let $f\colon M \to B$ be a wrinkled fibration. The *vertical distribution* of f is the kernel of its differential $\mathcal{V}^f = \ker(df) \subset TM$. A *horizontal distribution* for f, denoted by $\mathcal{H} \subset TM$, is the orthogonal complement of $\mathcal{V}^f$ with respect to some Riemannian metric on M.

Note that f restricts to a submersion outside of $\mathcal{C}_f$ so that the vertical distribution $\mathcal{V}^f$ has constant rank 2 on $M \setminus \mathcal{C}_f$ and consists of the tangent spaces to the fibers. However, at critical points the rank of $\mathcal{V}^f$ jumps up to 3 so that neither $\mathcal{V}^f$ nor any horizontal distribution $\mathcal{H}$ are vector bundles unless f is a submersion. Nevertheless, it is possible to speak of vector fields on M with values in $\mathcal{V}^f$ or $\mathcal{H}$.

Remark 3.6. The set of all horizontal distributions for a given wrinkled fibration is by definition a quotient of the space of Riemannian metrics[4] and we give it the quotient topology. Since the space of metrics is connected, it follows that the space of horizontal distributions for a given wrinkled fibration is connected as well.

Now let us fix a wrinkled fibration $f\colon M \to B$ and a horizontal distribution $\mathcal{H}$ for f. A smooth curve $\tilde{r}\colon [0,1] \to M$ is called *$\mathcal{H}$–horizontal* if its velocity vectors, denoted by $\dot{\tilde{r}}(\tau)$, are contained in $\mathcal{H}$. The main idea for the definition of parallel transport is to find $\mathcal{H}$–horizontal lifts for curves in the base.

[4]The space of metrics is topologized as a subspace of the sections of the second symmetric power of the cotangent bundle with its standard topology.

Proposition 3.7. *Let $f\colon M \to B$ be a wrinkled fibration, $\mathcal{H}$ a horizontal distribution for f, and $r\colon [0,1] \to B$ an embedded arc such that $r(0)$ and $r(1)$ are regular values. For convenience we write $\Sigma_t = f^{-1}\big(r(t)\big)$.*

(i) For any $t \in [0,1]$ and any regular point $p \in \Sigma_t$ there is a unique $\mathcal{H}$–horizontal lift $\tilde{r}_{t,p}\colon I_{t,p} \to M$ of r, defined on a maximal open interval $I_{t,p} \subset [0,1]$ containing t, such that $\tilde{r}_{t,p}(t) = p$. Furthermore, the curves $\tilde{r}_{t,p}$ have left and right limits in either Σ_0, Σ_1 or $\mathcal{C}_f$.

(ii) If $\mathcal{H}$ depends smoothly on some auxiliary parameters, then so do lifts $\tilde{r}_{t,p}$.

Proof. Let R be the image of r and let $Y = f^{-1}(R)\setminus\mathcal{C}_f$. Since f restricted to $M\setminus\mathcal{C}_f$ is a submersion, Y is a non-compact smooth 3–manifold with boundary $\Sigma_+^R \amalg \Sigma_-^R$. Let Γ_0 be the unique $\mathcal{H}$–horizontal lift of the velocity vector field $\dot{r}$. Then Γ_0 is necessarily tangent to Y, thus providing a vector field on Y whose flow generates the desired $\mathcal{H}$–horizontal lifts $\tilde{r}_{t,p}$. Moreover, if $\mathcal{H}$ depends smoothly on some parameters, then so do Γ_0 and its integral curves.

It remains to study the limiting behavior of these lifts. For that purpose we let $\tilde{r}$ be some lift of r defined over an open interval $(a,b) \subset [0,1]$. We will show that $\tilde{r}$ can be extended to the closed interval $[a,b]$. Since the situation is symmetric, we only give the arguments for b. For convenience, we choose a Riemannian metric on M which induces $\mathcal{H}$ and a metric on B such that

$$|\Gamma_0|_M = |df(\Gamma_0)|_B = |\dot{r}|_B. \tag{3.1}$$

Let (t_n) be a sequence in (a,b) converging to b. By the compactness of M the sequence $\tilde{r}(t_n)$ must have an accumulation point $\tilde{b} \in M$ and the continuity of r implies that $f(\tilde{b}) = r(b)$. We have to distinguish two cases: either $\tilde{b}$ is regular point of f or it is a critical point. If $\tilde{b}$ is a regular point, then (a,b) must intersect the interval $I_{b,\tilde{b}}$ so that $\tilde{r}$ is just the restriction of the maximal lift $\tilde{r}_{b,\tilde{b}}$. In particular, $r(t)$ converges to $\tilde{b}$ as t approaches b. On the other hand, if $\tilde{b}$ is a critical point, then a priori the sequence $\tilde{r}(t_n)$ could have several accumulation points, or there could be another sequence with a different accumulation point. In any case, it is enough to treat the situation that t_n has two different accumulation points. By the above arguments, these have to be critical points in the fiber $f^{-1}(b)$. But each fiber contains at most finitely many critical points so that the curve $\tilde{r}(t)$ must oscillate rapidly near b. In particular, the absolute value of $\dot{\tilde{r}} = \Gamma_0$ cannot be bounded near b. But this is impossible since $\dot{r}$ is bounded by compactness and we have $|\Gamma_0|_M = |\dot{r}|_B$ by equation (3.1). □

Using Proposition 3.7 we can make the following definition which is similar to the notions of ascending and descending manifolds in Morse theory. We use the same notation as in Proposition 3.7.

Definition 3.8. Given a pair (t,p) with $p \in \Sigma_t$ a regular point, we say that $\tilde{r}_{t,p}$ *runs into* $\mathcal{C}_f$ (resp. *emerges from* $\mathcal{C}_f$) if its left (resp. right) limit lies in $\mathcal{C}_f$. We define the *vanishing sets* of r as follows:

$$V_0^{\mathcal{H}}(r) = \big\{p \in \Sigma_0 \big|\ \tilde{r}_{0,p} \text{ runs into } \mathcal{C}_f\big\}$$
$$V_1^{\mathcal{H}}(r) = \big\{q \in \Sigma_1 \big|\ \tilde{r}_{1,q} \text{ emerges from } \mathcal{C}_f\big\}$$

Moreover, we define the *parallel transport* along r with respect to $\mathcal{H}$

$$\mathrm{PT}_r^{\mathcal{H}} \colon \Sigma_0 \setminus V_0^{\mathcal{H}}(r) \to \Sigma_1 \setminus V_1^{\mathcal{H}}(r)$$

by sending $x \in \Sigma_0$ to $\tilde{r}_{0,x}(1) \in \Sigma_1$.

Standard results on the smooth dependence of solutions of ordinary differential equations on their initial conditions imply that $\mathrm{PT}_r^{\mathcal{H}}$ is a diffeomorphism. Note that if r is an arc of regular values, then its vanishing sets are empty and we recover the notion of parallel transport in bundles. We will discuss more interesting situations in Section 3.3 below. Next we want to investigate how the vanishing sets and parallel transport depend on $\mathcal{H}$.

Corollary 3.9. *Let f and r be as in Proposition 3.7 and let $\mathcal{H}$ and $\mathcal{H}'$ be two horizontal distributions for f. Then the vanishing sets of r with respect to $\mathcal{H}$ and $\mathcal{H}'$ are ambiently isotopic in the reference fibers.*

Proof. This follows from part (ii) of Proposition 3.7 and the fact that the space of horizontal distributions for f is connected (see Remark 3.6). □

Conversely, we now show that all ambient isotopies of the vanishing sets can be realized by changing the horizontal distribution. Recall that $\mathrm{Diff}^0(\Sigma)$ is our notation for the identity component of $\mathrm{Diff}(\Sigma)$.

Lemma 3.10. *Let $f \colon M \to B$ be wrinkled fibration, and $r \colon [0,1] \to B$ an embedded arc of regular values. As before, we let $\Sigma_t = f^{-1}(r(t))$.*

(i) If $\mathcal{H}$ and $\mathcal{H}'$ are horizontal distributions for f, the $\mathrm{PT}_r^{\mathcal{H}}$ and $\mathrm{PT}_r^{\mathcal{H}'}$ are isotopic.

(ii) Conversely, if $\mathcal{H}$ is a horizontal distribution, then for any $\phi \in \mathrm{Diff}^0(\Sigma_0)$ and $\psi \in \mathrm{Diff}^0(\Sigma_1)$ there exists a horizontal distribution $\mathcal{H}'$ which agrees with $\mathcal{H}$ outside of an arbitrarily small neighborhood of $R = r([0,1])$ and satisfies $\mathrm{PT}_r^{\mathcal{H}} = \psi \circ \mathrm{PT}_r^{\mathcal{H}} \circ \phi^{-1}$.

Proof. The first claim again follows from part (ii) of Proposition 3.7 and the connectivity of the space of horizontal distributions. To prove the second claim, we can assume that $B = \mathbb{R}^2$, equipped with coordinates (τ, σ), and that $r(\tau) = (\tau, 0)$ by restricting to a neighborhood of $R = r([0,1])$. Let H_τ and H_σ be the unique $\mathcal{H}$–horizontal lifts of the coordinate vector fields ∂_τ and ∂_σ. Then $\mathcal{H}$ is spanned by H_τ and H_σ. Next observe that $f^{-1}(R)$ is a 3–dimensional cobordism from Σ_0 to Σ_1 and that

$$r^{-1} \circ f \colon f^{-1}(R) \longrightarrow [0,1]$$

is a Morse function without critical points. Moreover, H_τ restricts to a gradient-like vector field for $r^{-1} \circ f$ whose flow induces to the parallel transport $\mathrm{PT}_r^{\mathcal{H}}$. According to [54, Lemma 2.28], we can find another gradient-like vector field H' on $f^{-1}(R)$ whose flow induces $\psi \circ \mathrm{PT}_r^{\mathcal{H}} \circ \phi^{-1}$. Using standard arguments we can extend H' to a vector field on M with the following properties:

- H' agrees with H_τ outside of $f^{-1}(U)$ where $U \subset B$ is a small open neighborhood of R.
- $df(H') = \rho \partial_\tau$ for some positive function $\rho \in C^\infty(M)$.

Then H' and H_σ span a new horizontal distribution $\mathcal{H}'$ which and $\mathrm{PT}_r^{\mathcal{H}'}$ agrees with the gradient flow of H' by construction. □

Corollary 3.11. *Consider f, $\mathcal{H}$ and r as in Proposition 3.7 and let $\mathcal{H}$ and $\mathcal{H}'$ be two horizontal distributions for f. For all $\phi \in \mathrm{Diff}^0(\Sigma_0)$ and $\psi \in \mathrm{Diff}^0(\Sigma_1)$ there exists a horizontal distribution $\mathcal{H}'$ such that*

$$V_0^{\mathcal{H}'}(r) = \phi\big(V_0^{\mathcal{H}}(r)\big), \quad V_1^{\mathcal{H}'}(r) = \psi\big(V_1^{\mathcal{H}}(r)\big), \quad \textit{and} \quad \mathrm{PT}_r^{\mathcal{H}'} = \psi \circ \mathrm{PT}_r^{\mathcal{H}} \circ \phi^{-1}.$$

Proof. Apply Lemma 3.10 near the endpoints of r. □

It turns out that the vanishing sets as well as the parallel transport diffeomorphism of an embedded arc $r\colon [0,1] \to B$ depend only on the oriented image of r, which we denote by $R = r([0,1])$. Indeed, any orientation preserving reparametrization of r only changes the flow used to define the lifts in the proof of Proposition 3.7 by rescaling with a bounded function, which only affects the speed of the integral curves but not the flow lines. So we will only use R from now on unless we specifically need a parametrization.

Lastly, although we have only discussed parallel transport and vanishing sets for embedded arcs, there are obvious generalization to more general setting such as immersed arcs, piecewise smooth arcs, and closed curves. Moreover, we will usually refer to the parallel transport along a closed curve of regular values as the *monodromy* along the curve.

3.3 Folds, Cusps, and Vanishing Cycles

We now try to compare the fibers over the different regions of a wrinkled fibration $f\colon M \to B$. For that purpose, we focus on two neighboring regions in $B \setminus f(\mathcal{C}_f)$, that is, two regions which are separated by a fold arc. We consider an embedded arc $R \subset B$ with regular endpoints that intersects the critical image of f transversely in one fold point, say $R \pitchfork f(\mathcal{C}_f) = \{f(p)\}$. Then the fold model implies that $f|_{f^{-1}(R)}\colon f^{-1}(R) \to R$ is a Morse function with a single critical point at p which has index 1 or 2, depending on the choice of an orientation of R.[5] The following definition is central to all subsequent developments.

Definition 3.12 (Reference arcs)**.** Let $f\colon M \to B$ be a wrinkled fibration. A *reference arc* for f is an embedded arc $R \subset B$ with the following properties:

(a) the endpoints of R are regular values,

(b) R intersects the critical image $f(\mathcal{C}_f)$ transversely in one fold point, and

(c) R is oriented such that $f|_{f^{-1}(R)}\colon f^{-1}(R) \to R$ has an index 2 critical point.

The endpoints of R are also called *reference points*. The fibers over the initial and final reference point are denoted by Σ_R and Σ'_R, respectively, and are called the *reference fibers* of R.

[5] Indeed, we can choose coordinates around p and $f(p)$ such that f appears as the fold model $(t,x,y,z) \mapsto (t, x^2+y^2-z^2)$ and R corresponds to $\{0\} \times \mathbb{R}$ in the target of the model. Then $f|_{f^{-1}(R)}$ is locally given by $\pm(x^2+y^2-z^2)$.

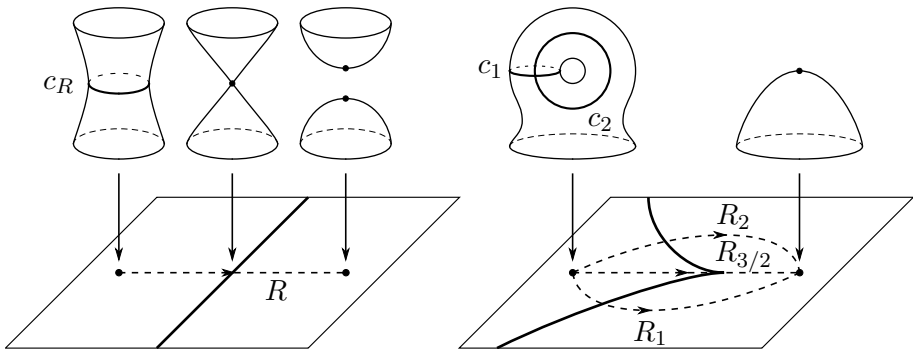


Figure 4: The fibers and vanishing sets in the models for indefinite folds and cusps

Given a reference arc R for f, basic Morse theory already gives qualitative information about the relation of the reference fibers Σ_R and Σ'_R. Indeed, since $f|_{f^{-1}(R)}$ has a single index 2 critical point, Σ'_R is related to Σ_R by a surgery on a simple closed curve. It follows that the Euler characteristic of the fibers increases by 2 along R, that is, $\chi(\Sigma'_R) = \chi(\Sigma_R) + 2$. If both reference fibers happen to be connected, then this is equivalent to saying that the genus decreases by one along R. In this case we will usually call Σ_R and Σ'_R the *higher* and *lower genus fiber* of R, respectively. Note in particular that this discussion also shows that wrinkled fibration cannot have empty fibers. In other words, they are automatically surjective. Using the machinery of parallel transport developed in the previous section we can also quantify the relation between the fibers over neighboring regions.

Lemma 3.13. *Let $f\colon M \to B$ be a wrinkled fibration equipped with a horizontal distribution $\mathcal{H}$. If $R \subset B$ is a reference arc, then the vanishing sets of R consist of a simple closed curve (that is, an embedded S^1)*

$$c_R = V_0^{\mathcal{H}}(R) \subset \Sigma_R$$

and two distinct points (that is, an embedded S^0)

$$\{w_R, z_R\} = V_1^{\mathcal{H}}(R) \subset \Sigma'_R.$$

Furthermore, the parallel transport $\mathrm{PT}_R^{\mathcal{H}}\colon \Sigma_R \setminus c_R \to \Sigma'_R \setminus \{w_R, z_R\}$ provides an identification of Σ'_R with the surgery on $c_R \subset \Sigma_R$.

Definition 3.14 (Vanishing cycles)**.** The simple closed curve $c_R \subset \Sigma_R$ appearing in Lemma 3.13 is called the *(fold) vanishing cycle* of R.

The situation is illustrated on the left side of Figure 4 (which is in fact an accurate reflection of the fibers in the indefinite fold model).

Remark 3.15. Recall that the surgery on $c_R \subset \Sigma_R$ is usually defined by removing a tubular neighborhood of Σ and filling in the resulting boundary components with disks. However, there is a different interpretation of this process which is more convenient for our purposes. We simply consider the complement $\Sigma_R \setminus c_R$ and take

its endpoint compactification[6]. One can show that the smooth structure on $\Sigma_R \setminus c_R$ can be extended across the endpoints and the resulting manifold diffeomorphic to the usual interpretation of surgery. In fact, this amounts to the same arguments that show that the usual surgery results in a well-defined smooth manifold.

Proof of Lemma 3.13. We fix a parametrization $r\colon [0,1] \to R \subset B$ and consider the Morse function

$$f_R = r^{-1} \circ f|_{f^{-1}(R)}\colon f^{-1}(R) \to [0,1]$$

with its unique critical point $p \in f^{-1}(R)$. Let g be a Riemannian metric on M that induces the horizontal distribution $\mathcal{H}$, and let Γ be the gradient of f_R with respect to $g|_{f^{-1}(R)}$. Then by definition Γ takes values in $\mathcal{H}$ and we can assume that the unit length vector field $\Gamma_0 = \Gamma/|\Gamma|$ is the horizontal lift of $\dot{r}$ (possibly after rescaling the metric). But this means that, up to reparametrization, the $\mathcal{H}$–horizontal lifts of r constructed in Proposition 3.7 are exactly the non-constant gradient flow lines of f_R. In particular, the vanishing sets of R agree with the descending and ascending spheres of p (considered as a critical point of f_R) in Σ_R and Σ'_R. Since p has index 2, the descending sphere has dimension 1 while the ascending sphere has dimension 0. In other words, the vanishing sets are a simple closed curve $c_R \subset \Sigma_R$ and two distinct points $\{w_r, z_r\} \in \Sigma'_R$, as claimed.

Finally, it is clear that the parallel transport along R identifies the endpoint compactifications of $\Sigma_R \setminus c_R$ and $\Sigma'_R \setminus \{w_r, z_r\}$. The latter is canonically identified with Σ'_R while the former can be considered as the surgery on $c_R \subset \Sigma_R$ as explained in Remark 3.15. □

Remark 3.16. One can think of wrinkled fibrations as "surface valued Morse functions". In this analogy, horizontal distributions correspond to gradient-like vector fields and the vanishing cycles can be considered as analogues of the descending spheres of critical points. Recall that an excellent Morse function can be recovered up to equivalence[7] from its set of critical values, the topology of intermediate regular level sets, and the descending spheres of the critical points above the level sets (see [50], for example). Put slightly differently, an excellent Morse function decomposes into (necessarily trivial) fiber bundles over the intervals of regular values, and the descending spheres control how these are glued together. Similarly, a wrinkled fibration $f\colon M \to B$ decomposes into a disjoint union of surface bundles over $B \setminus f(\mathcal{C}_f)$ and the vanishing cycles contain information how these are glued together. However, the situation is slightly more complicated for the following reasons. First of all, there can be non-simply connected regions over which f is a non-trivial bundle. So the knowledge of the topology of the fibers alone is not sufficient. Second, two regions can meet along several different fold arcs. In this case the corresponding vanishing cycles can be different, but they must be compatible in some sense and the compatibility conditions are not always obvious. Third, there

[6] Recall that there is an abstract notion of *ends* of a topological space. The *endpoint compactification* is obtained by adding one point for each end and declaring a neighborhood basis for each endpoint. We will only consider endpoint compactifications of complements of curves or finite sets in surfaces. In these cases it is intuitively clear what the ends and their neighborhoods are and we will not discuss the details. For a formal definition see [26, p.60], for example.

[7] that is, up to composition with diffeomorphisms of source and target

are additional gluing ambiguities cause by fibers whose diffeomorphism groups fail to be simply connected. However, if all regions are simply connected and all fibers are connected of genus at least two, then the knowledge of the fibers and vanishing cycles is enough to recover f. We will discuss this phenomenon extensively in the case of *simple* wrinkled fibrations (see Definition 4.1) in Part II. More general cases have been studied by Gay and Kirby [30].

Of course, the vanishing cycle c_R associated to a reference arc R depends on the choice of a horizontal distribution, but according to Corollary 3.9 its isotopy class does not. Next we want to understand the dependence of c_R on R.

Definition 3.17. Two reference arcs $R_1, R_2 \subset B$ are called *isotopic* if there is an ambient isotopy that moves R into R' through reference arcs. Furthermore, they are called *strictly isotopic* if they have the same reference points and are isotopic relative to the reference points.

Lemma 3.18. *Let $f\colon M \to B$ be a wrinkled fibration and let $\mathcal{H}$ be a horizontal distribution for f. If $R_1, R_2 \subset B$ are strictly isotopic reference arcs, then their vanishing sets are isotopic.*

Proof. Let R_t, $t \in [1,2]$, be a 1–parameter family of reference arcs obtained from a strict isotopy from R_1 to R_2. We first assume that all R_t agree in a neighborhood of the critical image. In this situation the claim follows from the fact that the parallel transport in fiber bundles has the property that two homotopic paths (relative to their endpoints) induce isotopic parallel transport diffeomorphisms (see [18, p.226ff.], for example). If R_t is also allowed to move near the critical image, then the situation can be reduced to the fold model as follows. Assume that R_0 meets the critical image in $f(p)$. We choose model coordinates for f around p and let $U \subset B$ be the support of the coordinates around $f(p)$. For small t the intersection point of R_t with $f(\mathcal{C}_f)$ is contained in U and we can assume that the image of R_t in $\mathbb{R}^2$ contains the arc $\{\rho(t)\} \times [-1,1]$. Now the fold model provides canonical identifications of the vanishing sets of $\{\rho(t)\} \times [-1,1]$ and, combined with the property of parallel transport in bundles, this argument shows that the vanishing sets of R_t depend smoothly on t. □

The above result is closely related to the fact that for a 1–parameter family of Morse functions $f_t\colon Y \to [0,1]$ on a fixed 3–manifold Y the ascending and descending spheres move by isotopies. In fact, the fold model can be considered as a model for a 1–parameter family of 3–dimensional Morse functions near a critical point. Similarly, the cusp model is related to 3–dimensional Morse theory in that it models the cancellation of a pair of critical points of index 1 and 2. This observation will allow us to understand the vanishing cycles of the two fold arcs near a cusp. Recall that it is necessary and sufficient for a pair of critical points of a Morse function to cancel each other that the index 1 ascending sphere and the index 2 descending sphere (with respect to some metric) intersect transversely in one point in an intermediate level.

Definition 3.19. Let $f\colon M \to B$ be a wrinkled fibration. Two reference arcs R_1 and R_2 for f with common reference points are called *adjacent* if their union $R_1 \cup R_2$ bounds a disk in B that contains exactly one cusp.

A visual account of this definition and the following lemma can be found in the right side of Figure 4.

Lemma 3.20. *Let $f\colon M \to B$ be a wrinkled fibration and let $\mathcal{H}$ be a horizontal distribution for f. If R_1 and R_2 are adjacent reference arcs with initial reference fiber Σ, then the corresponding vanishing cycles $c_1, c_2 \subset \Sigma$ are weakly dual, that is, they have geometric intersection number one*[8]*.*

Proof. For convenience, we assume that the union $R_1 \cup R_2$ is smooth. This can always be achieved by a perturbation near the common reference points and, according to Lemma 3.18, such a perturbation preserves the isotopy class of the vanishing cycles. We orient the circle $R_1 \cup R_2$ be taking the reverse orientation on R_1 and observe that f restricts to a circle valued Morse function over $R_1 \cup R_2$. By removing a neighborhood of the lower genus reference point from $R_1 \cup R_2$, we obtain an honest Morse function with a pair critical points of index 1 and 2. The important observation is that c_2 appears as the descending sphere of the index 2 point, while c_1 is the ascending sphere of the index 1 point (since we reversed the orientation of R_1). Now we can use a similar localization idea as in the proof of Lemma 3.18. We choose model coordinates around the cusps and isotope the union $R_1 \cup R_2$ into the support of the model coordinates in such a way that R_1 and R_2 stay reference arcs throughout the isotopy. This reduces the problem to a study of the cusp model, since the vanishing cycles c_1 and c_2 can be recovered up to isotopy from those of the isotoped reference arcs (by parallel transport along the path traces out be the reference points and using Lemma 3.18). But in the cusp model the critical points of the Morse function cancel, meaning that the vanishing cycles corresponding to c_1 and c_2 intersect transversely in one point. □

Remark 3.21. Another interesting case is to consider parallel transport along the intermediate arc $R_{3/2}$ shown in Figure 4 which passes directly through a cusp. It can be shown that vanishing set in the higher genus fiber is an embedded figure eight which appears as the union $c_1 \cup c_2$ of two simple closed curves intersecting transversely in one point, while in the lower genus fiber the vanishing set is a single point x. Moreover, these vanishing sets can be related to the vanishing sets of the reference arcs R_1 and R_2 by considering families of reference arcs starting with R_i and converging to $R_{3/2}$. What happens is that the vanishing sets of R_i will converge to x on the lower genus side and to the simple closed curve c_i on the higher genus end. However, the proofs of these claims require lengthy and rather painful computations in the cusp model. Since we will not use these results, we will not prove them here.

Remark 3.22 (Lefschetz vanishing cycles). The theory of parallel transport and vanishing sets immediately generalizes to include Lefschetz singularities and thus can be applied to (broken) Lefschetz fibrations. In that setting one usually considers arcs running into Lefschetz points, originating from a regular value. For such an arc R it is classically known that one also obtains a simple closed curve c_R in the initial regular fiber Σ. These curves are the famous *Lefschetz vanishing cycles* (see [32, Ch. 8.2], for example). As seen in Figure 5, the local picture is similar to the case of a fold arc, which explains the name fold vanishing cycles.

[8] See page 22 for the definitions.

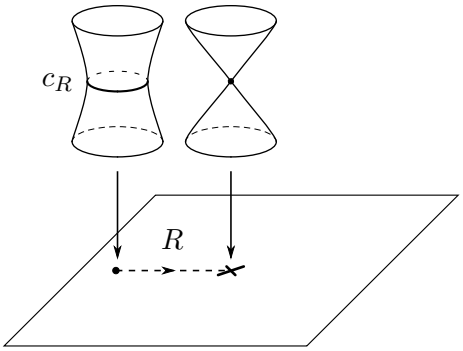


Figure 5: The vanishing cycle of a Lefschetz singularity

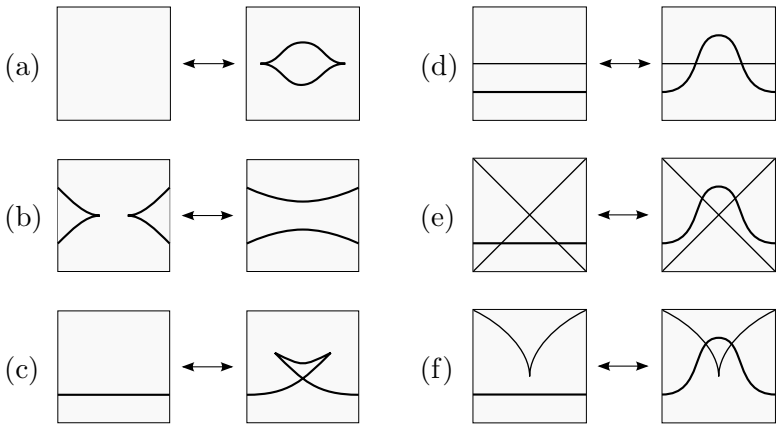


Figure 6: Elementary homotopies of wrinkled fibrations as seen in the base: (a) birth/death, (b) merge, (c) flip, (d) R_2, (e) R_3 and (f) fold-cusp crossing.

3.4 Moves for Wrinkled Fibrations

We briefly summarize what is known about the uniqueness of wrinkled fibrations up to homotopy. In [45] Lekili gave a list of local modifications for wrinkled fibrations which take the form of homotopies supported in balls in the total space; we will refer to these as *elementary deformations*. Lekili also used the term *moves* because he used them to manipulate critical images much like link diagrams are manipulated by the Reidemeister moves (see Figure 6). The elementary deformations can be subdivided into two families. The members of the first family change the structure of the critical locus, while the members of the category only affect the way the critical locus is mapped into the base. Since we will not make essential use of these moves we will only describe them briefly and refer to [45] for more details (see also [28,67]). We begin by describing the first family.

Birth. The *birth deformation* is described by the local model

$$B_s\big(t,x,y,z\big) = \big(t, x^3 - 3(s-t^2)x + y^2 - z^2\big).$$

For $s < 0$ there are no critical points and for $s > 0$ the critical locus is an indefinite circle with two cusps. In the latter case, the critical image appears eye shaped as in Figure 6(a). One can show that the fibers outside and inside the eye are open

disks and once punctured tori, respectively. The reverse deformation of a birth is naturally called a *death*.

Merge. Another model which is looks quite similar to that of the birth deformation describes the *merge deformation*

$$M_s\big(t,x,y,z\big) = \big(t, x^3 + 3(s-t^2)x + y^2 - z^2\big).$$

Its effect on the critical image is shown in Figure 6(b). Throughout the deformation two cusps approach each other and eventually merge, while in the reverse deformation two fold points approach each other. It is thus intuitive to speak of *cusp merge* and *fold merge* instead of merge and inverse merge. The fibers for $s < 0$ are again disks outside the cusps and once punctured tori inside. For $s > 0$ the two outer regions have open disks as fiber while the middle strip has once punctured tori.

Flip. The first family is completed by the *flip deformations* modeled on

$$FL_s\big(t,x,y,z\big) = \big(t, x^4 - sx^2 + tx + y^2 - z^2\big).$$

As shown in Figure 6(c) it begins with a single fold arc for $s < 0$ and introduces two cusps and a double point when passing to $s > 0$. Considered as a map $\mathbb{R}^5 \to \mathbb{R}^3$ the flip deformation is known as swallowtail or dovetail singularity. The fibers for $s < 0$ are exactly as in the fold model and once can show that inside the swallowtail shaped region the fibers are twice punctured tori.

The elements of the second family all have a local model of the form

$$P_s\big(t,x,y,z\big) = \big(t, x^2 + y^2 - z^2 \pm s\rho(t)\big), \quad s \in [0,1]$$

where $\rho\colon \mathbb{R} \to [0,1]$ is a smooth function satisfying $\rho(0) = 1$ and $\rho(t) = 0$ for $|t| \gg 0$. Note that the initial map P_0 is just the standard indefinite fold model and the deformation leaves the critical locus unchanged while pushing the critical image from the straight line $\mathbb{R} \times \{0\}$ to the graph of ρ. Such a deformation is not interesting by itself, but only in relation to other singularities present in a wrinkled fibration. Again, there are three basic cases shown in Figure 6(d), (e), and (f).

R_2 moves. As shown in Figure 6(d) an R_2 move starts with two parallel fold arcs in the base and pushes one across the other, resemblant of the type two Reidemeister move for link diagrams.

R_3 moves. Similarly, an R_3 move shown in Figure 6(e) mimics a Reidemeister move of type three where one arc is pushed across a crossing of two other arcs.

Fold-cusp crossings. The final deformation is very similar to an R_2 move in that a fold arc moves across an arc of critical values that stays stationary, only this time the stationary arc contains a cusp. This deformation is shown in Figure 6(f) and carries the suggestive name *fold-cusp crossing*.

The following result was originally conjectured by Lekili [45, p.309] and first proved by Williams [67, Theorem 1] for closed base surfaces. The general case is due to Gay and Kirby [28, Theorem 1.2].

Theorem 3.23 (h–principle for wrinkled fibrations)**.** *If two wrinkled fibrations are homotopic (rel. boundary), then they can be connected by a finite sequence of elementary deformations and homotopies through wrinkled fibrations.*

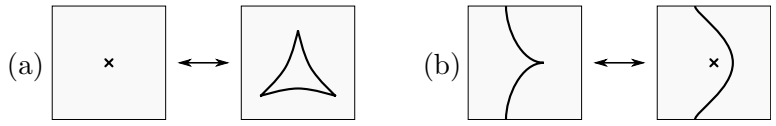


Figure 7: (a) Wrinkling and (b) unsinking a Lefschetz singularity.

Remark 3.24. Figure 6 shows how the critical image of a given wrinkled fibration can evolve in homotopies. Conversely, it is natural to ask to what extent these modifications can be prescribed. More precisely, given a wrinkled fibration whose critical image contains some configuration shown in Figure 6, is it possible to find an elementary deformation realizing the corresponding modification in Figure 6? It turns out that the answer is usually no. The only known exceptions are the passages from left to right in (a) and (c), that is, prescribing the effect of birth and flip deformations is always possible. In all other cases extra conditions are necessary. For example, in order to realize the passage from left to right in (b) by a cusp merge it is sufficient that the fiber along an arc connecting the cusps are connected (see [8, Corollary 3.12] for a necessary and sufficient condition). Conversely, going from right to left in (b) via a fold merge is possible if and only if the vanishing cycles of the fold arcs, measured in the interior region with respect to a suitable horizontal distribution and reference arcs, intersect transversely in one point (see [8, Section 3.4]). In the remaining cases there are similar obstructions coming from vanishing cycle configurations but precise conditions have yet to be worked out. Further details about vanishing cycles in merge deformations and R_2 moves can be found in [8] and [35].

Trading cusps and Lefschetz singularities. Lekili [45] also showed how to go back and forth between broken Lefschetz fibrations and wrinkled fibrations via homotopies. One direction follows from the behavior of Lefschetz singularities under small perturbations which was well known. In fact, a small perturbation of the local model converts it into a map with only indefinite folds and cusps. For example, the critical locus of the perturbation

$$\mathrm{L}_s(z,w) = z^2 + w^2 + 2s\bar{z}, \quad s \geq 0$$

is the circle $\{|z| = s, w = 0\}$ and contains three cusps for $s > 0$, one for each solutions of $z^3 = s^3$. The critical image of L_ϵ, $\epsilon > 0$, appears as on the right side of Figure 7(a). Furthermore, one can show that all fibers are connected which implies that all critical points must be indefinite. (The fibers outside the triangle are cylinders and inside they are twice punctured tori.) Lekili called the process of passing from $\mathrm{L} = \mathrm{L}_0$ to L_ϵ for some $\epsilon > 0$ *wrinkling* a Lefschetz singularity [45]. Conversely, one can also trade a cusp for a Lefschetz singularity and an arc of indefinite folds as in Figure 7(b). This process is called *unsinking* a Lefschetz singularity and we refer to [45, Section*2].

Moreover, Lekili observed that wrinkling and unsinking work equally well with achiral Lefschetz singularities, which can thus be converted to Lefschetz singularities and folds. This, together with the results of [27], proves the existence of broken Lefschetz fibrations on closed 4–manifolds.

Part II

Simple Wrinkled Fibrations and Surface Diagrams

Chapter 4

Definitions and The Correspondence

We specialize from general wrinkled fibrations to a class of maps which we call *simple* wrinkled fibrations. As the name suggests, these maps have a particularly simple critical point structure. They were introduced by Williams [67] who proved that all maps from closed 4–manifolds to S^2 are homotopic to simple wrinkled fibrations. Williams also extracted combinatorial gadgets from simple wrinkled fibrations over S^2 which have become known as *surface diagrams*, and showed how the total space of a simple wrinkled fibrations can be recovered up to diffeomorphism from a surface diagram. However, the arguments involved a detour over broken Lefschetz fibrations. Here we develop a more natural approach to studying the interplay between simple wrinkled fibrations and surface diagrams which allows us to prove precise correspondence results. In particular, our methods not only recover the total spaces of simple wrinkled fibrations from surface diagrams but also the maps.

4.1 Simple Wrinkled Fibrations

We start with the central definition which is due to Williams [67, p.1052].

Definition 4.1 (Simple wrinkled fibrations)**.** A wrinkled fibration $w\colon M \to B$ is called *simple* if the following conditions are satisfied:

(a) $\mathcal{C}_f$ is connected and contained in the interior of M.

(b) f is injective on $\mathcal{C}_f$.

(c) All fibers of f are connected.

(d) $\mathcal{C}_f$ contains a cusp.

Two simple wrinkled fibrations are called *equivalent* if they are equivalent as wrinkled fibrations (see Definition 3.1). Lastly, the *genus of* w is the maximal genus among all fibers.

Williams's original definition[1] was slightly different in that he only considered the case $B = S^2$ and did not require the presence of cusps. While the latter assumption

[1] Williams actually used the name "simplified purely wrinkled fibrations".

might seem peculiar at this point, it ultimately leads to a more uniform theory (see Remark 5.10). Wrinkled fibrations that satisfy the conditions (a), (b), and (c) but fail to have cusps must be treated separately. These are either surface bundles or maps with a single circle of indefinite folds; in the latter case we will speak of *fake* simple wrinkled fibrations. Simple wrinkled fibrations are interesting from the perspective of 4–manifold topology because of the following existence result, which is also due to Williams [67, Corollary 1, p.1052].

Theorem 4.2 (Williams [67])**.** *Let X be a closed, oriented 4–manifold. Then any map $X \to S^2$ is homotopic to a simple wrinkled fibration of arbitrarily high genus.*

This makes simple wrinkled fibrations over S^2 a potential tool for studying closed, oriented 4–manifolds. In fact, this is the central theme the present work. Although we will only use Theorem 4.2 as a black box, we include a brief sketch of the proof in order to give an idea of the involved ideas.

Proof of Theorem 4.2 (sketch). The proof relies on a different existence theorem of Gay and Kirby [27] which asserts the following. Let X be a closed, oriented 4–manifold and let $\Sigma \subset X$ be a closed, oriented surface with trivial self-intersection. Then there exists a broken achiral Lefschetz fibration[2] $X \to S^2$ with Σ as a fiber [27, Theorem 1.1]. Using Lekili's wrinkling modification, these maps can be transformed into wrinkled fibrations (see Section 3.4). As Williams points out, the Gay–Kirby construction starts by identifying a neighborhood of Σ in X with $\Sigma \times D^2$, which amounts to choosing a framing for the normal bundle of Σ, and the projection $\Sigma \times D^2 \to D^2$ is then extended to X in a controlled way. So according to the Pontrjagin–Thom theorem, all homotopy classes of maps to S^2 can be realized by wrinkled fibrations. Williams also notes that the Gay–Kirby construction produces maps which are injective on their critical loci and that the wrinkling modification preserves this property. Finally, he uses the moves discussed in Section 3.4 to devise an algorithm to turn a wrinkled fibration which is injective on its critical locus into a simple wrinkled fibration by a homotopy. Moreover, the genus of a simple wrinkled fibration over S^2 can be increased by performing a so called flip-and-slip move discussed in Remark 4.4 below. □

Remark 4.3. The Gay–Kirby construction [27] relies on deep results in 3–dimensional contact topology.[3] This somewhat unnatural dependence for Theorem 4.2 could be removed by refining the singularity theory based approach of Baykur [4] to the existence of wrinkled fibrations to produce maps which are injective on their critical locus.

Remark 4.4 (Flip-and-slip moves)*.* For simple wrinkled fibrations over S^2 there is an important homotopy known as a *flip-and-slip move*, which was originally introduced by Baykur [4] in the context of broken Lefschetz fibrations. A flip-and-slip move increases the genus of a simple wrinkled fibration by one and introduces four additional cusps. The effect on the critical image is shown in Figure 8. The move begins with two flips preformed on the same fold arc, introducing four new cusps and two

[2]That is, a broken Lefschetz fibration which is also allowed to have achiral Lefschetz singularities.

[3]More precisely, these are Eliashberg's classification of overtwisted contact structures and the Giroux correspondence between contact structures and open book decompositions.

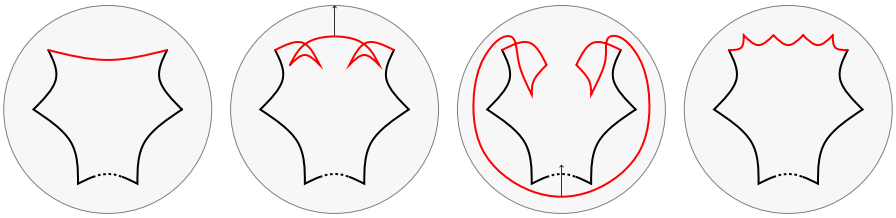


Figure 8: A "flip-and-slip move" performed on a simple wrinkled fibration over S^2. (The pictures show the complement of a disk in the lower genus region of the initial fibration.)

new regions with fiber genus $g+1$ (where g is the genus of the original simple wrinkled fibration). Then follows a sequence of fold-cusp crossings and R_2 moves that "slips" the indicated fold arc across the lower genus region, eventually eliminating it completely, and connecting the two regions with fiber genus $g+1$. For more details see [35, 67].

Although we are mainly interested in simple wrinkled fibrations over S^2, we will mostly work with fibrations over disks or annuli. These naturally arise by removing neighborhoods of regular fibers from fibrations over S^2. More generally, the study of simple wrinkled fibrations over arbitrary base surfaces can be reduced to the study of certain fibrations over the annulus. We will say more about this below. We first fix some ground rules for notation.

Notation. From now on we reserve the letters X, Z, and W (possibly with further decorations) for the following types of objects:

- X always stands for a closed 4–manifold.
- Z is used for compact 4–manifolds with connected boundary.
- W indicates a compact 4–manifold with two boundary components.

Moreover, whenever X, Z, or W are used it is implicitly understood that the manifold in question carries a winkled (or broken Lefschetz) fibrations over a sphere, disk, or annulus, respectively. So the letter used to denote a 4–manifold determines the designated target of fibration structures on the manifold. For example, whenever some simple wrinkled fibration is mentioned in the context of a 4–manifold Z, then the base has to be a disk. With this understood, we will sometimes abbreviate $f\colon Z \to D^2$ as (Z, f).

Now let us take a closer look at the structure of simple wrinkled fibrations. Given a genus g simple wrinkled fibration $w\colon M \to B$, the discussion in Section 2.2 shows that $\mathcal{C}_w \subset M$ is a smoothly embedded circle and that w restricts to a topological embedding of $\mathcal{C}_w$ into B. In fact, $w|_{\mathcal{C}_w}$ is a smooth homeomorphism whose inverse fails to be smooth only at the cusps. A neighborhood of $w(\mathcal{C}_w)$ appears as in Figure 9. It follows that $B \setminus w(\mathcal{C}_w)$ has at most two connected components, at least one of which must have fibers of genus g. Since the singular locus is non-empty, it must contain indefinite fold points and by Lemma 3.13 there must be a second region with fibers obtained by a surgery on a simple closed curve in a genus g fiber. But all fibers are connected, so the second region must have fibers of genus $g-1$. We will refer to these two regions as the *higher genus region* and the *lower genus region*.

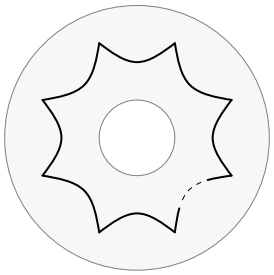


Figure 9: A neighborhood of the critical image of a simple wrinkled fibration.

Building on these observations we divide the base surface into three pieces

$$B = B_+ \cup A \cup B_-$$

where A is a annular neighborhood of the critical image as in Figure 9, and $B_\pm$ are the closures of the complement of A; the subscript in $B_\pm$ indicates whether the surface is contained in the higher or lower genus region of w. Since A contains all the critical values, w restricts to surface bundles over $B_\pm$ and, although complicated in their own right, surface bundles over surfaces are well studied. To conclude, the unexplored part of w is the restriction to $w^{-1}(A)$, which is a simple wrinkled fibration over the annulus A with the additional property that the critical image is boundary parallel. These kinds of fibrations can thus be considered as the cores of general simple wrinkled fibrations.

Definition 4.5. A simple wrinkled fibration $w\colon W \to A$ is called *annular* if the base surface A is an annulus, and the critical image $w(\mathcal{C}_w) \subset A$ is boundary parallel.

The upshot of the above discussion is that a general simple wrinkled fibration decomposes into an annular simple wrinkled fibration and a pair of surface bundles. As a consequence, the study of general simple wrinkled fibrations can be divided into studying annular simple wrinkled fibrations, surface bundles, and how they can be glued together. The latter two can be translated into the language of mapping class groups of surfaces. Indeed, surface bundles correspond to representations of the fundamental group of the base surface in the mapping class group of the fiber, and the study of fiber preserving diffeomorphisms of surface bundles over S^1 (up to isotopy) boils down to studying the centralizer of the monodromy in the mapping class group of the fiber. This leaves the problem of understanding annular simple wrinkled fibrations for which we provide a combinatorial counterpart in Theorem 1.3.

We end this section with some elementary constructions of simple wrinkled fibrations. These provide useful examples and also illustrate how one should think of simple wrinkled fibrations.

Example 4.6 (Surface bundles)**.** Let $\pi\colon P \to B$ be a surface bundle with closed fibers of genus g. Then performing a birth deformation on π gives rise to a simple wrinkled fibration of genus $g+1$ with two cusps.

Example 4.7 (Lefschetz fibrations)**.** Let $f\colon M \to B$ be a Lefschetz fibration (possibly achiral) with closed fibers of genus g. By wrinkling all the Lefschetz points we obtain a wrinkled fibration with a number of "triangles" of critical values as in

Figure 7(a), one for each Lefschetz point. By suitably merging cusps we can turn this configuration into a single cusped circle resulting in a simple wrinkled fibration of genus $g+1$.

In this light, it is natural to think of simple wrinkled fibrations as a common generalization of surface bundles and Lefschetz fibrations. The following example illustrates the classification approach outlined after Definition 4.5.

Example 4.8 (Fake simple wrinkled fibrations over S^2)**.** Consider a 3–dimensional cobordism Ω from a genus g surface Σ_g to a surface Σ_{g-1} of one genus lower, equipped with a Morse function $\mu\colon \Omega \to I$ with exactly one critical point of index 2. By taking the product with S^1 we obtain a map $\mu \times \mathrm{id}\colon \Omega \times S^1 \to I \times S^1$ whose critical locus is a circle of indefinite folds, which is embedded into $I \times S^1$ by $\mu \times \mathrm{id}$, and all fibers of $\mu \times \mathrm{id}$ are connected. In other words, $\mu \times \mathrm{id}$ is a fake simple wrinkled fibration of genus g. In fact, one can show that all fake annular simple wrinkled fibrations with non-empty critical locus are of this form, see [3, Section 8.1]. By definition, the boundary components of $\Omega \times S^1$ are diffeomorphic to $\Sigma_g \times S^1$ and $\Sigma_{g-1} \times S^1$ so that we can close off to a fake simple wrinkled fibration over S^2 by gluing in copies of $\Sigma_g \times D^2$ and $\Sigma_{g-1} \times D^2$ and extending $\mu \times \mathrm{id}$ by projecting onto the D^2 factors. If $g \geq 3$ (so that $g-1 \geq 2$), then it follows from Corollary 2.20 that the gluing on both sides of $\Omega \times S^1$ is unique up to equivalence. Moreover, it is known that the closed 4–manifold obtained by this procedure is diffeomorphic to $\Sigma_{g-1} \times S^2 \# S^1 \times S^3$ (see [5, Example 3.1]). For $g \leq 2$ gluing ambiguities arise and the closed 4–manifold is not uniquely determined. For example, if we use the trivial gluing (that is, the identity map) on the higher genus side, then the arguments in [5, Example 3.1] show that

$$(\Sigma_g \times D^2) \cup_{\mathrm{id}} (\Omega \times S^1) \cong \Sigma_{g-1} \times D^2 \# S^1 \times S^3.$$

As a consequence, filling in the remaining boundary component with $\Sigma_{g-1} \times D^2$ results manifolds of the form $P \# S^1 \times S^3$ where P is a Σ_{g-1}–bundle over S^2 (see Corollary 2.21). Finally, a non-trivial gluing on the higher genus side is only possible for $g = 1$ which results in a family of closed 4–manifolds with finite cyclic fundamental groups (see [33, Main Theorem B]), including the famous wrinkled (or broken Lefschetz) fibration on S^4 discovered by Auroux, Donaldson and Katzarkov [3].

4.2 Surface Diagrams

We now introduce the combinatorial counterparts of simple wrinkled fibrations which also appeared in the work of Williams, first implicitly in [67] and later explicitly in [68]. Again, our definitions differ slightly from those of Williams and we will explain the differences in Remark 4.20 at the end of this chapter. At this point we recommend that the reader (re-)familiarize himself with the notions of intersection numbers and duality for simple closed curves discussed on page 22.

Definition 4.9 (W–chains and –cycles)**.** Let Σ be a surface and let $\Gamma = (c_1, \ldots, c_l)$ be an ordered collection of simple closed curves $c_i \subset \Sigma$.

(1) Γ is called a *W–chain* if $i(c_i, c_{i+1}) = 1$ for all $i < l$.

(2) A W–chain G is called a *W–cycle* if it is *closed* in the sense that $i(c_l, c_1) = 1$.

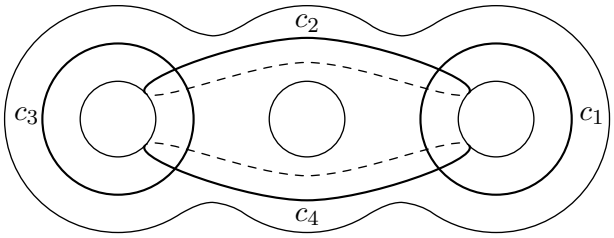


Figure 10: An example of a surface diagram.

(3) If Γ is a W–chain and satisfies $i(c_l, \psi(c_1)) = 1$ for some $\psi \in \mathrm{Diff}(\Sigma)$, then it is called a *$\psi$–twisted W–cycle*.

In other words, a *W*–chain is an ordered collection of curves $c_1, \ldots, c_l \subset \Sigma$ such that curves with adjacent index are weakly dual. To avoid misunderstandings, we want to emphasize that we do *not* assume that adjacent curves are strongly dual, that is, they do not necessarily intersect transversely in a single point. In fact, we do not even assume that the intersections are transverse. However, in many situations only the isotopy classes of the vanishing cycles are important and we are free to assume these properties whenever they make things simpler. Also, there are no restrictions on the intersections between non-adjacent curves in W–chains, as opposed to the more common "chains" of simple closed curves used in mapping class group theory. The latter satisfy $i(c_i, c_j) = 0$ for $|i - j| > 1$ (see [24, p.107ff.], for example). Altogether, W–chain can look rather complicated.

Remark 4.10. In [7] W–chains and W–cycles were called "circuits" and "closed circuits", respectively. However, in the meantime we realized that various other mathematicians in the field were using variations of the chain/cycle terminology in conversations and we decided to adopt this. We added the prefix W to avoid confusion with the above mentioned chains of curves. We leave it to the reader's imagination to choose his favorite interpretation of the letter, some suggestions are Williams, wrinkled, or wild.

Definition 4.11 (Surface diagrams)**.** A *surface diagram* is a pair $\mathfrak{S} = (\Sigma, \Gamma)$ consisting of a closed, oriented surface Σ and a W–cycle $\Gamma = (c_1, \ldots, c_l)$ in Σ. The curves $c_i \subset \Sigma$ are called the *vanishing cycles* of $\mathfrak{S}$.

An example of a surface diagram is shown in Figure 10. Note that a W–cycle must contain at least two curves because a single curve is not weakly dual to itself. In particular, surface diagrams always contain at least two curves.

Remark 4.12. For convenience we will only draw surface diagrams whose vanishing cycles are in *general position* such that all intersections are transverse. Obviously, general position can always be achieved by isotopies of the vanishing cycles. One can even go further and bring the diagram intro *minimal position*, meaning that the vanishing cycles are in general position and for $i \neq j$ we have $i(c_i, c_j) = \#(c_i \pitchfork c_j)$. In fact, an arbitrary collection of simple closed curves $a_1, \ldots, a_k \subset \Sigma$ can be brought into minimal position, for example by isotoping the curves to geodesics with respect to some constant curvature metric and taking parallel push offs in the case that the collection contained isotopic curves (see [24, p.29ff.]).

As indicated above, we want to consider two surface diagrams as equivalent if their vanishing cycles are isotopic. But it turns out that we ultimately need a more flexible notion of equivalence. We will give the complete definition of equivalence in the context of the following generalization of surface diagrams.

Definition 4.13 (Generalized surface diagrams)**.** A *generalized surface diagram* is a triple $\mathfrak{G} = (\Sigma, \Gamma, \psi)$ where Σ is a closed, oriented surface, $\psi\colon \Sigma \to \Sigma$ is an orientation preserving diffeomorphism, and Γ is a ψ–twisted W–cycle. The curves in Γ are also called the *vanishing cycles* and ψ is called the *twist*.

So surface diagrams are exactly those generalized surface diagrams (Σ, Γ, ψ) that are *untwisted* in the sense that the twist ψ is isotopic to the identity. Note that generalized surface diagrams are allowed to have only one curve, but then the twist has to be non-trivial.

Definition 4.14 (Equivalence)**.** Let $\mathfrak{G} = (\Sigma, \Gamma, \psi)$ and $\mathfrak{G}' = (\Sigma', \Gamma', \psi')$ be two generalized surface diagrams of the same length with vanishing cycles c_i and c'_j. We say that $\mathfrak{G}$ and $\mathfrak{G}'$ are *equivalent* if they are related by a finite sequence of the following operations:

- *Isotopy:* $\Sigma = \Sigma'$, $c_i \sim c'_i$, and $\psi \sim \psi'$ (possibly with different isotopies)
- *Diffeomorphism:* $c'_i = \phi(c_i)$ and $\psi' = \phi^{-1}\psi\phi$ for some orientation preserving diffeomorphism $\phi\colon \Sigma \to \Sigma'$
- *Twisted cyclic permutation:* $\Sigma = \Sigma'$, $\psi = \psi'$, and $\Gamma' = \big(c_2, \dots, c_l, \psi(c_1)\big)$

Twisted cyclic permutation may seem strange at first sight and its origin will only become clear in Section 5.1.1. But observe that for a surface diagram it simply amounts to a cyclic permutation of the vanishing cycles.

To complete our terminology surrounding surface diagrams we give another definition which might seem rather unmotivated at this point. Let $\mathfrak{G} = (\Sigma, \Gamma, \psi)$ be a generalized surface diagram with $\Gamma = (c_1, \dots, c_l)$. For the moment it will be convenient to blur the distinction between isotopy classes and their representatives, both for diffeomorphisms and simple closed curves. Since c_1 and c_{i+1} are weakly dual for $i < l$, equation (2.8) shows that the product of Dehn twists $\tau_{c_i}\tau_{c_{i+1}}$ maps c_i to c_{i+1}. By iteration, we see that the product $\tau_{c_{l-1}}\tau_{c_l} \dots \tau_{c_1}\tau_{c_2}$ maps c_1 to c_l, and the condition that c_l and $\psi(c_1)$ are weakly dual implies that $\tau_{c_l}\tau_{\psi(c_1)}$ maps c_l to $\psi(c_1)$. Altogether, it follows that the mapping class

$$\psi^{-1}\tau_{c_l}\tau_{\psi(c_1)}\tau_{c_{l-1}}\tau_{c_l} \dots \tau_{c_1}\tau_{c_2} \in \mathrm{Mod}(\Sigma)$$

preserves the isotopy class of c_1. We denote the subgroup of $\mathrm{Mod}(\Sigma)$ of mapping classes that fix c_1 up to isotopy by $\mathrm{Mod}(\Sigma)(c_1)$. Denoting by Σ_{c_1} the surface obtained by surgery along c_1, there is a homomorphism

$$\Phi_{c_1}\colon \mathrm{Mod}(\Sigma)(c_1) \longrightarrow \mathrm{Mod}(\Sigma_{c_1})$$

obtained by choosing a representative that fixes c_1 as a set, thus giving a diffeomorphism $\Sigma \setminus c_1$, and considering Σ_{c_1} as the endpoint compactification of Σ_{c_1}. We will explain this construction in more detail Section 5.2.1.

Definition 4.15 (Monodromy)**.** Let $\mathfrak{G} = (\Sigma, \Gamma, \psi)$ be as above. The mapping class

$$\mu(\mathfrak{G}) = \Phi_{c_1}\left(\psi^{-1}\,\tau_{c_l}\tau_{\psi(c_1)}\,\tau_{c_{l-1}}\tau_{c_l}\ldots\tau_{c_1}\tau_{c_2}\right) \in \mathrm{Mod}(\Sigma_{c_1})$$

is called the *monodromy* of $\mathfrak{G}$.

Now that we have introduced our main objects of interest, namely simple wrinkled fibrations and (generalized) surface diagrams, we want to understand how they are related. As a motivation we briefly recall the relation between Lefschetz fibrations and their vanishing cycles.[4]

4.3 Interlude: Lefschetz Fibrations and Their Vanishing Cycles

For simplicity we restrict our attention to Lefschetz fibrations over D^2 or S^2. General references are [32, Chs. 8.1&8.2] and [41, Ch. 3] (see also [39, 48]). Since this material is well known and only serves as a motivation, we will not give more precise references. We first discuss Lefschetz fibrations over the disk. Let $f\colon Z \to D^2$ be such a map. By an appropriate choice of reference arcs[5] for the critical points one can record all the corresponding Lefschetz vanishing cycles in one regular fiber Σ. The result is an ordered collection of simple closed curves $\lambda_1, \ldots, \lambda_k \subset \Sigma$ and it is easy to understand how these curves depend on the choice of reference arcs. Moreover, one can show that the monodromy of the boundary fibration $f|_{\partial Z}\colon \partial Z \to S^1$ (which is only well-defined up to isotopy) is given by the word in Dehn twists

$$\tau_{\lambda_k}\cdots\tau_{\lambda_1} \in \mathrm{Mod}(\Sigma). \tag{4.1}$$

Conversely, given an ordered collection of simple closed curves $\lambda_1, \ldots, \lambda_k \subset \Sigma$ in a closed, oriented surface – or equivalently, a word in positive Dehn twists as in equation (4.1) – one can construct a Lefschetz fibration by the following procedure.

(1) Take $\Sigma \times D^2$ and choose $\theta_1, \ldots, \theta_k \in S^1$ ordered according to the orientation.

(2) Build a 4–manifold Z by attaching 2–handles to $\Sigma \times D^2$ along $\lambda_i \times \{\theta_i\} \subset \Sigma \times S^1$ with respect to the framing in $\Sigma \times S^1$ induced by the inclusion $\lambda_i \times \{\theta_i\} \subset \Sigma \times \{\theta_i\}$ corrected by -1.

(3) Then the projection $\Sigma \times D^2 \to D^2$ extends to a Lefschetz fibration on Z over a slightly larger disk in a natural such that there is one critical point for each 2–handle.

We pause to give a definition for later reference.

Definition 4.16. (Fiber framing) Let Y be a 3–manifold, $\Sigma \subset Y$ a regular fiber of a smooth map $Y \to S^1$, and $c \subset \Sigma$ a simple closed curve. Then the framing of c in Y induced by the inclusion $c \subset \Sigma$ is called the *fiber framing*.

[4] See pages 27 and 36 for the definitions.
[5] See Figure 5.

In the light of the above we will refer to 2–handles attached along simple closed curves in fibers with "fiber framing -1" as *Lefschetz handles*. It is well known that the constructions give rise to a bijective correspondence between Lefschetz fibrations over the disk up to equivalence and ordered sequences of simple closed curves up to certain moves.

Remark 4.17. The above discussion easily generalizes to achiral Lefschetz fibrations[6]. These correspond to words in positive and negative Dehn twists since the achiral singularities contribute negative Dehn twists about their vanishing cycles to the analogue of the formula (4.1) for the boundary monodromy. The construction of achiral Lefschetz fibrations with prescribed words is almost the same except that for each negative twist one has to attach a 2–handle along the corresponding curve with "fiber framing $+1$" instead of -1.

Now let us discuss Lefschetz fibrations over S^2. Given such a map $f\colon X \to S^2$ we obtain a Lefschetz fibration over the disk by removing a fibered neighborhood of a regular fiber. For concreteness, let $\nu\Sigma = f^{-1}(\text{int}D)$ where $D \subset S^2$ is disk over regular values and let $Z = X \setminus \nu\Sigma$. The monodromy of $f|_{\partial Z}\colon \partial Z \to \partial D$ is necessarily trivial so that for any collection of curves $\{\lambda_i\}$ extracted from $f|_Z$ we must have

$$\tau_{\lambda_k} \cdots \tau_{\lambda_1} = 1 \in \text{Mod}(\Sigma).$$

Conversely, if an ordered sequence of simple closed curves $\lambda_1, \ldots, \lambda_k \subset \Sigma$ satisfies this condition, then the corresponding Lefschetz fibration $f\colon Z \to D^2$ over the disk has trivial boundary monodromy. One can then close off to a fibration over S^2 using a fiber preserving diffeomorphism $\varphi\colon \Sigma \times S^1 \to \partial Z$ and extending f to a map

$$X = Z \cup_\varphi (\Sigma \times D^2) \longrightarrow D^2 \cup D^2 = S^2$$

the projection $\Sigma \times D^2 \to D^2$. Moreover, Corollary 2.20 shows that the different choices of φ give equivalent extensions if Σ has genus at least two, while ambiguities can appear for genus at most one.[7]

The above discussion can be summarized by saying that Lefschetz fibrations over the disk correspond to arbitrary words in Dehn twists, while Lefschetz fibrations over the sphere correspond to factorizations of the identity into positive Dehn twists. Lefschetz fibrations thus provides a bridge between 4–manifold topology and mapping class groups of surfaces which has proved to be very useful (see [41] for a survey of results in this area).

4.4 An Outline of The Correspondence

We now want to paint a similar picture for simple wrinkled fibrations and surface diagrams. Here we focus on simple wrinkled fibrations over the disk and the sphere, although most of the work will actually be done in the context of annular wrinkled fibrations. Before stating the results, we observe that there are two substantially different types of simple wrinkled fibrations over the disk. The critical image of a simple wrinkled fibration $w\colon Z \to D^2$ divides the target into an open sub-disk and

[6] See Remark 3.2 on page 28 for the definition.

[7] It turns out that there is no ambiguity for genus one but this not obvious, see [32, p.311].

a half-open annulus containing $S^1 = \partial D^2$. Either of these can be the higher genus region. If the open disk is the higher genus region, then the fiber genus drops while moving toward the boundary. We will thus speak of *descending* simple wrinkled fibrations in this situation, and of *ascending* simple wrinkled fibrations in the other. From the perspective of surface diagrams it turns out to be more natural to work with descending simple wrinkled fibrations and, in fact, we will never encounter any ascending ones. With this understood, we will eventually prove the following correspondence results which were already stated on page 7.

Theorem 1.2 (Correspondence over the disk)**.** *Equivalence classes of surface diagrams correspond bijectively to equivalence classes of descending simple wrinkled fibrations over the disk.*

Theorem 1.1 (Correspondence over the sphere)**.** *Let $\mathcal{SWF}_g(S^2)$ and $\mathcal{SD}^0_g$ be the sets of equivalence classes of genus g simple wrinkled fibrations over S^2 and surface diagrams with trivial monodromy, respectively. There is a surjective map*

$$\mathcal{SWF}_g(S^2) \longrightarrow \mathcal{SD}^0_g$$

whose point preimages have a transitive action of the group $\pi_1\big(\operatorname{Diff}(\Sigma_{g-1}), \mathrm{id}\big)$. In particular, the map is bijective for $g \geq 3$.

Note that these results are instances where the philosophy of Remark 3.16 is successful. In fact, they can by now be considered as a special case of a more general result of Gay and Kirby [30, Theorem 1] which appeared while the author was writing [7]. As mentioned before, we will deduce Theorems 1.1 and 1.2 from a correspondence result for annular simple wrinkled fibrations and generalized surface diagrams.

Theorem 1.3 (Annular correspondence)**.** *There is a bijective correspondence between annular simple wrinkled fibrations and generalized surface diagrams, both considered up to equivalence.*

The details are deferred to Chapter 5. We now give a non-technical outline of the constructions involved in the correspondences over the disk and the sphere which closely resembles the discussion of Lefschetz fibrations in the previous section.

Let $w\colon Z \to D^2$ be a descending simple wrinkled fibrations over the disk. To obtain a surface diagram we fix a point in the higher genus region and consider its fiber Σ. We number the fold arcs in the critical image from 1 to l according to the boundary orientation induced from the higher genus region and choose one reference arc for each fold arc. The resulting collection of fold vanishing cycles $c_1, \dots, c_l \subset \Sigma$ turns out to have the structure of a W–cycle so that $\mathfrak{S}_w = (\Sigma; c_1, \dots, c_l)$ is a surface diagram. This process is explained in detail in the discussion surrounding Lemma 5.3. For the boundary fibration $w|_{\partial Z}\colon \partial Z \to S^1$, we find a surface bundle over S^1 whose fiber is not Σ but a surface of lower genus. In Proposition 5.12 we explain how to identify the boundary fibers with the surface Σ_{c_1} (obtained from Σ by surgery on c_1) and show that the monodromy of $w|_{\partial Z}$ corresponds to the mapping class

$$\mu(\mathfrak{S}_w) = \Phi_{c_1}\big(\tau_{c_l}\tau_{c_1}\,\tau_{c_{l-1}}\tau_{c_l}\dots\tau_{c_1}\tau_{c_2}\big) \in \operatorname{Mod}(\Sigma_{c_1})$$

from Definition 4.15. This also explains why we call $\mu(\mathfrak{S}_w)$ the monodromy of $\mathfrak{S}_w$.

The other way around, given a surface diagram $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$ we can construct a descending simple wrinkled fibration by the following recipe.

(1) Start with $\Sigma \times D^2$.

(2) Build a 4–manifold $Z_{\mathfrak{S}}$ by attaching fiber framed 2–handles to $\Sigma \times D^2$ along $c_i \times \{\theta_i\} \subset \Sigma \times S^1$ for some $\theta_1, \dots, \theta_l \in S^1$ ordered according to the orientation of S^1.

(3) Extend the projection $\Sigma \times D^2 \to D^2$ to a descending simple wrinkled fibration $w_{\mathfrak{S}} \colon Z_{\mathfrak{S}} \to D^2$ (over a larger disk, strictly speaking) in a natural way.

The difficult part is the last step and we refer to Lemma 5.8 and its proof for the details. Quite notably, the only difference to the construction of Lefschetz fibrations lies in the framings of the 2–handles. We simply have to use the fiber framing instead of correcting it by -1. The following terminology will be useful.

Definition 4.18. (Fold handles) A 2–handle attached along a simple closed curve in a fiber with respect to the fiber framing is called a *fold handle*.

The passage to fibrations over S^2 is exactly the same as in the Lefschetz case. Given a simple wrinkled fibration $w \colon X \to S^2$ we remove a fibered neighborhood of a lower genus fiber to obtain a descending simple wrinkled fibration over a disk[8] with surface diagram $\mathfrak{S}_w$. The boundary monodromy is necessarily trivial so that

$$\mu(\mathfrak{S}_w) = 1 \in \mathrm{Mod}(\Sigma_{c_1}),$$

again by Proposition 5.12. Conversely, if a surface diagram $\mathfrak{S}$ has trivial monodromy, then $w_{\mathfrak{S}} \colon Z_{\mathfrak{S}} \to D^2$ has trivial boundary monodromy and can be closed off to a simple wrinkled fibration over S^2. More specifically, we can choose a fiber preserving diffeomorphism $\varphi \colon \Sigma_{g-1} \times S^1 \to \partial Z_{\mathfrak{S}}$ (where g is the genus of Σ) and extend $w_{\mathfrak{S}}$ to a simple wrinkled fibration

$$w_{\mathfrak{S}}^{\varphi} \colon X_{\mathfrak{S}}^{\varphi} = Z_{\mathfrak{S}} \cup_{\varphi} (\Sigma_{g-1} \times D^2) \longrightarrow D^2 \cup D^2 \cong S^2$$

using the projection $\Sigma_{g-1} \times D^2 \to D^2$. The ambiguity for the choice of φ is controlled by the group $\pi_1\big(\mathrm{Diff}(\Sigma_{g-1}), \mathrm{id}\big)$ which vanishes for $g \geq 3$ by Theorem 2.18.

One thing that should be remembered from this discussion is that simple wrinkled fibrations are directly accessible via handlebody theory. In fact, a surface diagram of a descending simple wrinkled fibration over the disk (Z, w) can be considered as a construction manual for a handle decomposition of Z. For simple wrinkled fibrations over the sphere the process of closing off makes the handlebody approach a little more cumbersome. This issue will be discussed in more detail in Section 5.3. Nevertheless, we obtain a rather direct link between surface diagrams and the topology of the 4–manifolds they describe. This observation will be further exploited in Part III.

Combining Theorems 1.1 and 4.2, we arrive at the following intriguing consequence which was already observed by Williams [67, p.1054, Corollary 2].

Corollary 4.19 (Williams [67])**.** *All closed, oriented 4–manifolds can be specified up to diffeomorphism by surface diagrams of genus $g \geq 3$ with trivial monodromy.*

[8] Note that if we removed a neighborhood of a higher genus fiber, then we would obtain an ascending simple wrinkled fibration over a disk.

Curiously, Williams actually did not explain how to construct a 4–manifold from a surface diagram, although it was clear at the time that this can be done via a detour over broken Lefschetz fibrations using results of Lekili [45] and Baykur [5]. More precisely, Lekili's wrinkling modification discussed in Section 3.4 shows how to turn a simple wrinkled fibration into a broken Lefschetz fibration and the vanishing cycles of the latter are determined by the surface diagram. Baykur's handlebody approach to broken Lefschetz fibrations [5] then allows to reconstruct the total space. Finding a more intrinsic proof for Corollary 4.19 was our main motivation for proving Theorem 1.1.

Remark 4.20. We now comment on the difference between our definition of surface diagrams and Williams's original one [68, Definition 2]. Williams defines a surface diagram as the collection of vanishing cycles associated to a simple wrinkled fibration over S^2 of genus at least three. In our language this corresponds to surface diagrams of genus $g \geq 3$ with trivial monodromy. We found that Williams's definition two drawbacks. First, it makes it difficult to consider surface diagrams as abstract combinatorial gadgets, mostly because the relation between the implicit trivial monodromy condition and the vanishing cycles does not become clear. Second, it excludes low genus simple wrinkled fibrations from the discussion which, in spite of the gluing ambiguities, seem to be more accessible than fibrations of genus three and higher. Our goal was to provide an abstract definition of surface diagrams which can be studied without the reference to simple wrinkled fibrations, just as Heegaard diagrams of 3–manifolds can be discussed without reference to Morse functions or Heegaard splittings.

Chapter 5

The Correspondence: Proofs

The present chapter is the technical core of Part II. We first prove the annular correspondence theorem which we restate again.

Theorem 1.3 (Annular correspondence)**.** *There is a bijective correspondence between annular simple wrinkled fibrations and generalized surface diagrams, both considered up to equivalence.*

The proof is contained in Section 5.1. In Section 5.2 we give a geometric interpretation of the monodromy of generalized surface diagrams as defined in Definition 4.15. Finally, in Section 5.3 we deduce Theorems 1.1 and 1.2.

5.1 The Annular Correspondence

For convenience, we denote by $a\mathcal{SWF}$ the set of equivalence classes of annular simple wrinkled fibrations and by $g\mathcal{SD}$ the set of equivalence classes of generalized surface diagrams. In Sections 5.1.1 and 5.1.2 we will define maps

$$a\mathcal{SWF} \longrightarrow g\mathcal{SD} \qquad \text{and} \qquad g\mathcal{SD} \longrightarrow a\mathcal{SWF}$$

and show that they are mutually inverse, thus proving Theorem 1.3.

5.1.1 From Annular Simple Wrinkled Fibrations to Generalized Surface Diagrams

The first step is to assign generalized surface diagrams to annular simple wrinkled fibrations. As indicated in Section 4.4, we need a method to record all (fold) vanishing cycles in one fiber. The key ingredients are the notions of reference arcs and their adjacency discussed in Section 3.3. The following concept makes sense for arbitrary base surfaces.

Definition 5.1 (Reference systems)**.** Let $w\colon M \to B$ be a simple wrinkled fibration with l cusps (and thus l fold arcs). A *reference system* for w is a collection of pairwise disjoint reference arcs $\mathcal{R} = \{R_1, \dots, R_l\}$ with common reference points such that

(a) R_i and R_{i+1} are adjacent for $i < l$,

(b) each fold arc of w meets exactly one reference arc, and

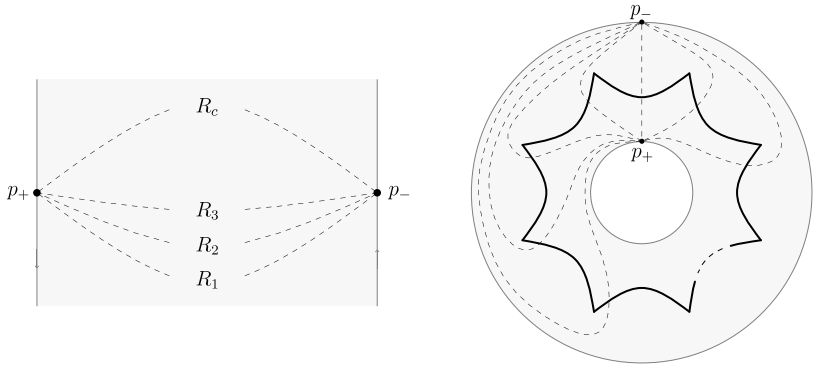


Figure 11: A reference system for an annular simple wrinkled fibration. On the left hand side, the annulus is drawn as a rectangle where the top and bottom are identified.

(c) the reference arcs leave the initial reference point in order of increasing index.

Generalizing the concept of isotopy for reference arcs, we say that reference systems are *isotopic* if they are ambiently isotopic in B through reference systems, and *strictly isotopic* if the ambient isotopy fixes the reference points.

In the annular case we will always assume that the reference points lie on the boundary of the base annulus as shown in Figure 11.

Remark 5.2. Condition (c) might need some further explanation. Suppose that we have a collection of embedded arcs in an oriented surface B, all emanating from a point $b \in B$, and having pairwise disjoint interiors. Depending on whether b lies in the interior or the boundary of B we consider a small disk or half disk D around b whose boundary intersects all arcs transversely once. The boundary orientation on ∂D then induces a cyclic order on the arcs which is easily seen to be independent of the choice of D. By declaring one arc as the first, the cyclic order becomes an absolute order. Moreover, if $b \in \partial B$ and the arcs are properly embedded, then there is a natural choice for the first arc, namely the one that intersects ∂D after the part $\partial D \cap \partial B$.

Now let $w\colon W \to A$ be an annular simple wrinkled fibration. We denote the boundary components of A by $\partial_\pm A$, where $\partial_+ A$ lies in the higher genus region, and we let $\partial_\pm W = w^{-1}(\partial_\pm A)$. We fix a reference system $\mathcal{R} = (R_1, \dots, R_l)$ and a horizontal distribution $\mathcal{H}$ for w and denote the higher and lower genus reference fibers of $\mathcal{R}$ by Σ and Σ', respectively. According to Lemma 3.13, each R_i has a vanishing cycle $c_i \subset \Sigma$ and we let $\Gamma = (c_1, \dots, c_l)$. Finally, we denote the parallel transport with respect to $\mathcal{H}$ around $\partial_\pm A$ by

$$\psi = \mathrm{PT}^{\mathcal{H}}_{\partial_+ A}\colon \Sigma \longrightarrow \Sigma \quad \text{and} \quad \mu = \mathrm{PT}^{\mathcal{H}}_{\partial_- A}\colon \Sigma' \longrightarrow \Sigma'$$

which we will refer to as the *higher and lower genus monodromy* of w.

Before continuing we would like to point out that the part of $\mathcal{R}$ that is contained in the lower genus region will be largely irrelevant in the subsequent discussion. The reason is that it does not affect the vanishing cycles. In fact, it will not be used until Section 5.2 where we relate μ, which lives on the lower genus side, to the vanishing

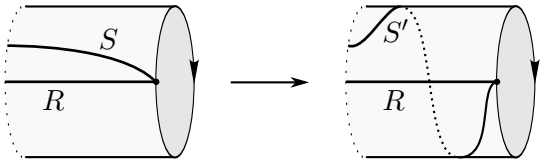


Figure 12: Swinging an arc around a boundary component.

cycles. However, we will keep track of the full reference systems in the proofs below and one should not get distracted by the fact that we seemingly ignore half of the information.

Lemma 5.3. *With the notation introduced above, the triple* $\mathfrak{G}_w^{\mathcal{R},\mathcal{H}} = (\Sigma, \Gamma, \psi)$ *is a generalized surface diagram.*

In the proof we will need the following construction. Let B be an oriented surface and let $R, S \subset B$ be two properly embedded arcs with disjoint interiors. Suppose that they both meet a boundary component $\partial_i B \subset \partial B$ transversely in the same point such that S enters $\partial_i B$ after R (as explained in Remark 5.2). Then we can modify S by moving its endpoint along $\partial_i B$ as indicated in Figure 12, resulting in a new arc S' which enters $\partial_i B$ before R and whose interior is still disjoint from R. We say that S' is obtained from S by *swinging once around* $\partial_i B$.[1]

Proof of Lemma 5.3. Since the reference arcs R_i and R_{i+1} are adjacent for $i < l$, it follows from Lemma 3.20 that c_i and c_{i+1} are weakly dual. In order to see that $\psi(c_1)$ and c_l are weakly dual, we swing R_1 once around $\partial_+ A$ and once negatively around $\partial_- A$. The result is a new reference arc R_1' which is by construction adjacent to R_l so that its vanishing cycle c_1' is weakly dual to c_l. In addition, R_1' is homotopic to the concatenation $(\partial_+ A)^{-1} * R_1 * (\partial_- A)^{-1}$ which shows that the parallel transport around $\partial_+ A$ maps c_1 to c_1', in other words $c_1' = \psi(c_1)$. Since c_l' and c_1 are weakly dual, so are their images under ψ, showing that c_l and $\psi(c_1)$ are weakly dual. □

Of course, the generalized surface diagram $\mathfrak{G}_w^{\mathcal{R},\mathcal{H}}$ not only depends on w but also on the choices of $\mathcal{R}$, $\mathcal{H}$. However, it turns out that its equivalence class does not.

Lemma 5.4. *The equivalence class of* $\mathfrak{G}_w^{\mathcal{R},\mathcal{H}}$ *is independent of* $\mathcal{R}$ *and* $\mathcal{H}$.

Proof. It follows from Corollary 3.9 and part (i) of Lemma 3.10 that a different choice of $\mathcal{H}$ only changes ψ and the vanishing cycles by isotopies. Of course, Σ does not change at all so that $\mathfrak{G}_w^{\mathcal{R},\mathcal{H}}$ is independent of $\mathcal{H}$ up to isotopy.

To investigate the $\mathcal{R}$–dependence, we note that $\mathcal{R}$ is determined up to (strict) isotopy by the (strict) isotopy class of the first reference arc R_1. This follows from the observation that R_1 cuts the base annulus A into a disk and there is a unique reference arc adjacent to R_1. Proceeding by induction, we can reconstruct $\mathcal{R}$ up to isotopy. Now suppose that $\mathcal{S} = (S_1, \ldots, S_l)$ is another reference system for w and let d_i be the vanishing cycle of S_i.

[1]Note that swinging around $\partial_i B$ is not the same as performing a boundary parallel Dehn twist. Dehn twists are supported in the interior of B and thus fix a collar of the boundary. In particular, they cannot change the order of arcs at the boundary and, moreover, in Figure 12 a boundary parallel Dehn twist applied to S would produce an arc that intersects R in its interior.

We first assume that $\mathcal{S}$ and $\mathcal{R}$ have the same reference points and that S_1 is strictly isotopic to R_k for some k. Since Σ and ψ depend solely on the reference points, we only have to discuss the relation of the vanishing cycles. If $k = 1$, then the arguments above show $\mathcal{R}$ and $\mathcal{S}$ are strictly isotopic and, according to Lemma 3.18, c_i' is isotopic to c_i for all i. For $k = 2$ we argue as follows. As in the proof of Lemma 5.3 we swing R_1 around $\partial_\pm A$ to obtain a reference arc R_1' adjacent to R_l whose vanishing cycle is isotopic to $\psi(c_1)$. Since S_1 and R_2 are strictly isotopic, it follows as in the case $k = 1$ that the reference system $\mathcal{R}' = (R_2, \dots, R_l, R_1')$ is strictly isotopic to $\mathcal{S}$ and we deduce that $(d_1, \dots, d_l) \sim (c_2, \dots, c_l, \psi(c_1))$. But this shows that $\mathfrak{G}_w^{\mathcal{R},\mathcal{H}}$ and $\mathfrak{G}_w^{\mathcal{S},\mathcal{H}}$ differ by isotopies and a twisted cyclic permutation and are therefore equivalent. For arbitrary k we repeat the above argument.

Finally, for general $\mathcal{R}$ and $\mathcal{S}$ we can assume by swinging that S_1 and R_1 hit the same fold arc. Obviously, S_1 and R_1 cannot be strictly isotopic if their reference points differ, however, they are isotopic since we allow ambient isotopies of A that move the boundary. And since $\mathcal{R}$ and $\mathcal{S}$ are determined by R_1 and S_1 they are isotopic as well. After a strict isotopy of either $\mathcal{R}$ or $\mathcal{S}$ we can assume that they agree in a neighborhood of the critical image. Moreover, we can find an isotopy from $\mathcal{S}$ to $\mathcal{R}$ that fixes this neighborhood so that the movement of $\mathcal{R}$ and $\mathcal{S}$ is confined to regular values of w. Throughout such an isotopy the higher genus reference point of $\mathcal{S}$ traces out a curve δ in $\partial_+ A$, and each S_i is homotopic to the concatenation of σ and R_i. But now we can appeal to the fact that in fiber bundles, homotopic curves give rise to isotopic parallel transport diffeomorphisms, which shows that the parallel transport along δ provides a diffeomorphism from $\mathfrak{G}_w^{\mathcal{S},\mathcal{H}}$ to a generalized surface diagram isotopic to $\mathfrak{G}_w^{\mathcal{R},\mathcal{H}}$. This finishes the proof. □

Remark 5.5. As an addendum to the last step of the proof, note that if we connect $\mathcal{R}$ and $\mathcal{S}$ by a different isotopy, then we obtain a different curve δ' in $\partial_+ A$ connecting the reference points and the concatenation $\delta^{-1}\delta'$ is a closed curve in $\partial_+ A$. In particular, the parallel transports along δ and δ' differ by some power of ψ so that the diffeomorphisms that relate $\mathfrak{G}_w^{\mathcal{S},\mathcal{H}}$ and $\mathfrak{G}_w^{\mathcal{R},\mathcal{H}}$ are in fact unique up to isotopy and powers of ψ.

As a consequence, to an annular simple wrinkled fibration $w\colon W \to A$ we can associate a well-defined equivalence class of generalized surface diagrams – simply denoted by $\mathfrak{G}_w$ from now on – and if the higher genus monodromy ψ is isotopic to the identity, then $\mathfrak{G}_w$ consists of surface diagrams. We thus define the desired map as

$$a\mathcal{SWF} \to g\mathcal{SD}, \quad w \mapsto \mathfrak{G}_w$$

and the next lemma shows that this is well-defined.

Lemma 5.6. *Equivalent annular simple wrinkled fibrations have equivalent generalized surface diagrams.*

Proof. Let $w\colon W \to A$ and $w'\colon W' \to A'$ be annular simple wrinkled fibrations and assume that they are equivalent via a commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{\hat{\phi}} & X' \\ {\scriptstyle w}\downarrow & & \downarrow{\scriptstyle w'} \\ A & \xrightarrow{\check{\phi}} & A' \end{array}$$

where $\hat{\phi}$ and $\check{\phi}$ are orientation preserving diffeomorphisms. Moreover, let $\mathfrak{G}$ be a generalized surface diagram for w, induced by a reference system $\mathcal{R} = (R_1, \dots, R_l)$ and a horizontal distribution $\mathcal{H}$. It suffices to exhibit a single generalized surface diagram for w' which is equivalent to $\mathfrak{G}$ since all surface diagrams of w' are equivalent by Lemma 5.4. To construct such a diagram we use $\hat{\phi}$ and $\check{\phi}$ to transfer $\mathcal{R}$ and $\mathcal{H}$ over to w'. More precisely, we take the reference system $\check{\phi}(\mathcal{R}) = \big(\check{\phi}(R_1), \dots, \check{\phi}(R_l)\big)$ for w' and the horizontal distribution $\hat{\phi}_*\mathcal{H}$ obtained by pushing $\mathcal{H}$ forward. By construction the corresponding surface diagram for w' is then diffeomorphic to $\mathfrak{G}$ via a diffeomorphism between the reference fibers induced by $\hat{\phi}$. □

5.1.2 From Generalized Surface Diagrams to Annular Simple Wrinkled Fibrations

We now describe a procedure to build annular simple wrinkled fibrations with prescribed generalized surface diagrams. The construction generalizes the one outlined in Section 4.4. We begin by setting up some notation. Given a surface diffeomorphism $\psi\colon \Sigma \to \Sigma$, we consider its *mapping torus*

$$\Sigma(\psi) = \big(\Sigma \times [0,1]\big)/(x,1) \sim (\psi(x),0).$$

As a temporary notation, we denote the image in $\Sigma(\psi)$ of a subset $E \subset \Sigma \times [0,1]$ under the quotient map by $[E]_\psi \subset \Sigma(\psi)$. The mapping torus $\Sigma(\psi)$ is a 3–manifold and it comes a canonical map

$$p_\psi\colon \Sigma(\psi) \to S^1, \quad p_\psi(x,\theta) = e^{2\pi i\theta}$$

which is easily seen to be a submersion. In what follows we will identify S^1 with the quotient $[0,1]/\{0,1\}$ via the map $\theta \mapsto e^{2\pi i\theta}$. Moreover, when writing $\theta \in S^1$ we will implicitly think of the unique representative in $[0,1)$ and expressions of the form $\theta_1 < \theta_2$ are to be understood in this sense. For brevity we also write the unit interval as $I = [0,1]$.

In the subsequent arguments we will make essential use of the following construction of Morse functions that appears in Milnor's classic [50].

Remark 5.7. Let Σ be a surface and let $c \subset \Sigma$ be a simple closed curve. In [50, Theorem 3.12] Milnor constructs a cobordism $\omega(\Sigma, c)$ from Σ to Σ_c together with a Morse function

$$f_c\colon \omega(\Sigma,c) \longrightarrow D^1 = [-1,1]$$

with a single critical point of index 2 whose descending sphere is isotopic to c. We briefly recall the construction. For the omitted details we refer to [50, p.30ff.]. Consider the subset $H \subset \mathbb{R}^3$ consisting of all points (x,y,z) which satisfy the inequalities

$$-1 \le -x^2 - y^2 + z^2 \le 1 \quad \text{and} \quad (x^2+y^2)z^2 \le \sinh^2(1)\cosh^2(1).$$

Milnor defines the cobordism $\omega(\Sigma, c)$ as a quotient

$$\omega(\Sigma,c) = \big((\Sigma \setminus c \times D^1) \amalg H\big)/\sim$$

where the identification depends on the choice of an embedding $\varphi\colon S^1 \times D^1 \hookrightarrow \Sigma$ with $\varphi(S^1 \times \{0\}) = c$. However, this choice is essentially unique since simple closed

curves in surface have canonical framings. The Morse function on $\omega(\Sigma, c)$ is then defined by

$$\begin{aligned} \Sigma \setminus c \times D^1 \ni (p, s) \quad &\rightsquigarrow \quad f_c(p, s) = s \\ H \ni (x, y, z) \quad &\rightsquigarrow \quad f_c(x, y, z) = -x^2 - y^2 + z^2. \end{aligned}$$

This process can be interpreted as an inherently smooth version of a 3–dimensional 2–handle attachment to $\Sigma \times I$ along $c \times \{1\}$, together with an extension of the projection $\Sigma \times I \to I$ to a Morse function across the handle. Milnor also shows in [50, Theorem 3.13] that any other pair (W, f), where W is a cobordism from Σ to Σ_c and $f\colon W \to [a, b]$ is a Morse function with a single index 2 critical point whose descending sphere is isotopic c, is equivalent[2] to $(\omega(\Sigma, c), f_c)$. It is clear from the construction that a diffeomorphism $\phi\colon \Sigma \to \Sigma'$ gives rise to a commutative triangle

$$\begin{array}{ccc} \omega(\Sigma, c) & \xrightarrow{\bar{\phi}} & \omega(\Sigma', \phi(c)) \\ & {\scriptstyle f_c} \searrow \quad \swarrow {\scriptstyle f_{\phi(c)}} & \\ & [-1, 1] & \end{array} \tag{5.1}$$

where $\bar{\phi}$ is a diffeomorphism that restricts to $\phi \times \mathrm{id}$ on $\Sigma \setminus c \times D^1$.

Lemma 5.8. *Let $\mathfrak{G} = (\Sigma, \Gamma, \psi)$ be a generalized surface diagram. There exists an annular simple wrinkled fibration*

$$w_{\mathfrak{G}}\colon W_{\mathfrak{G}} \longrightarrow S^1 \times I,$$

well-defined up to equivalence, with surface diagram $\mathfrak{G}$. Furthermore, the total space $W_{\mathfrak{G}}$ has a relative handle decomposition on $\Sigma(\psi)$ with fold handles[3] attached along $[c_i \times \theta_i]_\psi$ for some $0 < \theta_1 < \cdots < \theta_l < 1$.

Proof. The construction is divided into three steps.

Step 1: We consider the product $W_1 = \Sigma(\psi) \times \left[0, \frac{1}{3}\right]$ and define a map

$$w_1 = p_\psi \times \mathrm{id}\colon W_1 \longrightarrow S^1 \times \left[0, \tfrac{1}{3}\right].$$

Note that we can identify each fiber

$$w_1^{-1}(\theta, s) = [\Sigma \times \{\theta\}]_\psi \times \{s\}, \quad \theta \in [0, 1)$$

with Σ using the injectivity of the quotient map $\Sigma \times I \to \Sigma(\psi)$ on $\Sigma \times [0, 1)$.

Step 2: This is the core of the construction. As in Remark 5.7, we construct a 3–dimensional cobordism $\omega(\Sigma, c_i)$ from Σ to Σ_{c_i} together with a Morse function

$$f_i\colon \omega(\Sigma, c_i) \longrightarrow \left[\tfrac{1}{3}, \tfrac{2}{3}\right]$$

with a single critical point p_i of index 2 (with $f(p_i) = \frac{1}{2}$, say). Next we choose pairwise disjoint, closed intervals $\Theta_i \subset (0, 1) \subset S^1$ such that $\Theta_i < \Theta_{i+1}$ for $i < l$ and define

$$\Omega_i = \Theta_i \times \omega(\Sigma, c_i), \quad B_i = \Theta_i \times \left[\tfrac{1}{3}, \tfrac{2}{3}\right] \subset S^1 \times I, \quad \text{and}$$

[2] in the sense that $f = \psi \circ f_c \circ \phi^{-1}$ for diffeomorphisms ϕ and ψ between the sources and targets

[3] That is, fiber framed 2–handles (see Definition 4.18 on page 51.)

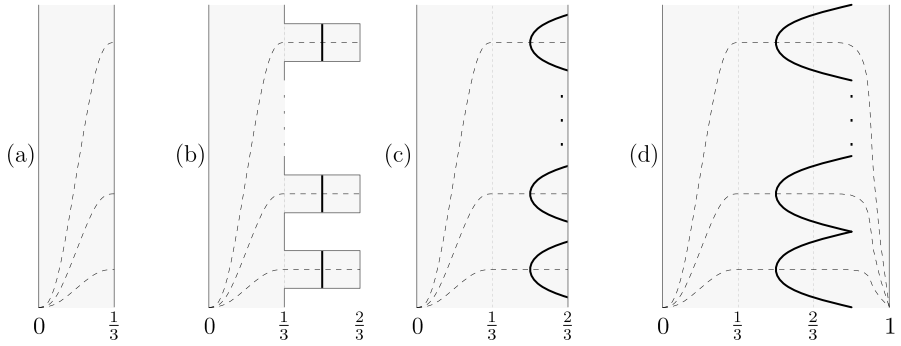


Figure 13: Building annular simple wrinkled fibrations from generalized surface diagrams. Each picture shows an annulus (top and bottom are understood to be identified). Bold lines are critical values while dashed lines are reference arcs.

$$F_i = \mathrm{id}_{\Theta_i} \times f_i \colon \Omega_i \longrightarrow B_i.$$

By construction the critical locus of F_i is $\Theta_i \times \{p_i\}$ which is an arc of indefinite folds mapping to $\Theta_i \times \{\frac{1}{2}\} \subset B_i$. In particular, F_i is a wrinkled fibration over the square B_i. Note that we have canonical identifications of both $F_i^{-1}(\Theta_i \times \{\frac{1}{3}\})$ and $w_1^{-1}(\Theta_i \times \{\frac{1}{3}\})$ with $\Sigma \times \Theta_i$ so that w_1 and the F_i can be glued together to a wrinkled fibration

$$w_1 \cup \big(\amalg_i F_i\big) \colon W_1 \cup \big(\amalg_i \Omega_i\big) \longrightarrow \big(S^1 \times \big[0, \tfrac{1}{3}\big]\big) \cup \big(\cup_i B_i\big) \subset S^1 \times I$$

onto a "bumpy annulus" inside $S^1 \times I$ with one arc of folds on each bump, see Figure 13(b). Moreover, the obvious corners in this construction can be smoothed using standard methods.

Observe that gluing $\Omega_i = \Theta_i \times \omega(\Sigma, c_i)$ to W_1 can be considered as attaching a fold handle along the copy of c_i in $w_1^{-1}(\theta_i, \frac{1}{3}) \cong \Sigma$ where $\theta_i \in \Theta_i$ is some interior point. Indeed, recall that $\omega(\Sigma, c_i)$ can be considered as a 3–dimensional 2–handle attachment along c_i, and that the product of a 3–dimensional 2–handle with an interval is a 4–dimensional 2–handle. Moreover, the attachment is fiber framed since the framing is obtained by taking the product of the 3–dimensional framing with an interval, and 3–dimensional 2–handles are automatically attached with the fiber framing. Conversely, the above construction can be reinterpreted as follows. Whenever a fold handle is attached to the total space of a wrinkled fibration, then the fibration extends uniquely up to equivalence across the handle with an extra arc of folds (over a bump added to the base surface) – whence the terminology.

As a consequence, the 4–manifold $W_2 = W_1 \cup \big(\amalg_i \Omega_i\big)$ is obtained by a sequence of fold handle attachments to $W_1 = \Sigma(\psi) \times \big[0, \frac{1}{3}\big]$ and therefore has the desired relative handle decomposition on $\Sigma(\psi)$. For cosmetic reasons we flatten the bumps of the target annulus of $w_1 \cup \big(\amalg_i F_i\big)$ (using an isotopy of $S^1 \times I$ with support in a small neighborhood of $S^1 \times \big[\frac{1}{3}, \frac{2}{3}\big]$) and denote the resulting wrinkled fibration by

$$w_2 \colon W_2 \longrightarrow S^1 \times \big[0, \tfrac{2}{3}\big].$$

Step 3: By construction $w_2 \colon W_2 \to S^1 \times \big[0, \frac{2}{3}\big]$ restricts to an excellent Morse

function on $\partial_2 W_2 = w_2^{-1}\big(S^1 \times \{\frac{2}{3}\}\big)$ and we denote the restriction by

$$w_{2,\partial}\colon \partial_2 W_2 \longrightarrow S^1 \times \{\tfrac{2}{3}\} \cong S^1.$$

The crucial observation is that all critical points of $w_{2,\partial}$ are arranged in canceling pairs, which follows from the condition that $\Gamma = (c_1, \dots, c_l)$ is a ψ–twisted W–cycle. Indeed, the fold handle attachment along c_i results in two consecutive critical values and, going around $S^1 \times \{\frac{2}{3}\}$ according to the boundary orientation, the first critical value has index 2 while the second has index 1. Furthermore, for $i < l$ the ascending sphere of the index 1 point of the ith fold handle and the descending sphere of the index 2 point of the $(i+1)$st fold handle can be identified with c_i and c_{i+1}, respectively. But these two curves are weakly dual so that the critical points can be canceled. Similarly, the ascending and descending spheres of the lth and the 1st fold handles correspond to c_l and $\psi(c_1)$ because of the twisted gluing in the construction of $\Sigma(\psi)$. As explained in Appendix C, we can choose a *path of death* for $w_{2,\partial}$ (see Definition C.2 on page 132), that is, a homotopy emanating from $w_{2,\partial}$ to a submersion which realizes the cancellation of critical points in a minimal way. Let $\mathfrak{w}_t$ be such a path of death, conveniently parametrized by $t \in [\frac{2}{3}, 1]$ such that $\mathfrak{w}_{2/3} = w_{2,\partial}$. To finish the construction we let

$$W_{\mathfrak{G}} = W_2 \cup_{\partial_2 W_2} \big(\partial_2 W_2 \times [\tfrac{2}{3}, 1]\big)$$

and note that $W_{\mathfrak{G}}$ is diffeomorphic to W_2 from which it inherits the desired handle decomposition. Finally, we extend w_2 to across $\partial_2 W_2 \times [\frac{2}{3}, 1]$ as the *trace* of the homotopy $(\mathfrak{w}_t)$ – that is, the map $(x,t) \mapsto (\mathfrak{w}_t(x), t)$ – and thus obtain a map

$$w_{\mathfrak{G}}\colon W_{\mathfrak{G}} \to S^1 \times [0,1].$$

As explained in Appendix C on page 136, while tracing out the path of death the fold arcs of w_2 eventually end in a cusp and the critical image of $w_{\mathfrak{G}}$ appears as in Figure 13(d) which shows that $w_{\mathfrak{G}}$ is an annular simple wrinkled fibration. Moreover, the reference system for $w_{\mathfrak{G}}$ shown in Figure 13 recovers $\mathfrak{G}$.

It remains to show that $w_{\mathfrak{G}}$ is well-defined up to equivalence. Observe that the only essential choices made in the construction were those of $f_i\colon \omega(\Sigma, c_i) \to [\frac{1}{3}, \frac{2}{3}]$ and the path of death $(\mathfrak{w}_t)$. We already mentioned in Remark 5.7 that each f_i is unique up to equivalence so that the same holds for $F_i = \mathrm{id}_{\Theta_i} \times f_i$. Finally, it follows from Corollary C.5 that the trace of $(\mathfrak{w}_t)$ used to extend w_2 is unique up to equivalence as well. This finishes the proof. □

Our next result shows that the construction in Lemma 5.8 exhaust all equivalence classes of annular simple wrinkled fibrations and that annular simple wrinkled fibrations can be recovered up to equivalence from any of their generalized surface diagrams.

Lemma 5.9. *Let $w\colon W \to A$ be an annular simple wrinkled fibration with generalized surface diagram $\mathfrak{G}$. Then w is equivalent to $w_{\mathfrak{G}}\colon W_{\mathfrak{G}} \to S^1 \times I$ as constructed in Lemma 5.8.*

Proof. The proof of Lemma 5.8 can essentially be reversed. But this requires some preliminary considerations. For simplicity, we assume that $A = S^1 \times I$. As before

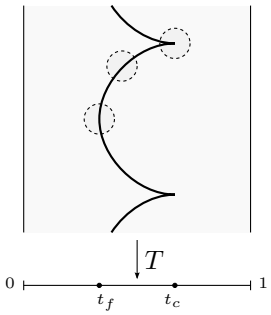


Figure 14: The interaction of T and critical values of w.

we fix an identification $S^1 \cong [0,1]/\{0,1\}$ and denote the coordinates on $S^1 \times I$ by θ and t. For convenience we let

$$S^1_t = S^1 \times \{t\} \quad \text{and} \quad I_\theta = \{\theta\} \times I.$$

Intuitively, we want the critical image of w to appear as on the right side of Figure 11. More formally, we will assume the following conditions which can all be achieved by a suitable reparametrization of the base annulus.

(a) Each I_θ contains exactly one critical value of w, and if I_θ contains a fold point, then the intersection $I_\theta \cap w(\mathcal{C}_w)$ is transverse.

(b) The critical image of w is contained in $S^1 \times [t_f, t_c]$ for some $0 < t_f < t_c < 1$.

(c) All folds are mapped to $S^1 \times [t_f, t_c)$ while the cusps are mapped to $S^1_{t_c}$.

(d) The image of the fold arcs meets S^1_t transversely for $t \in (t_f, t_c)$ and has quadratic tangencies with $S^1_{t_f}$.

Note that the last condition implies that the projection $T\colon S^1 \times I \to I$ restricts to a Morse function on the fold part of $w(\mathcal{C}_w)$. Assuming the above properties, we claim that the map

$$f = T \circ w\colon W \longrightarrow I$$

is a Morse function on W with only critical points of index 2. This can be seen by a direct investigation of f around its critical points. Since T is a submersion, we have $\mathcal{C}_f \subset \mathcal{C}_w$ and we are thus led to study f near critical points of w. According to the properties (c) and (d), we have to distinguish three cases: $p \in \mathcal{C}_w$ is either a cusp point, a fold point such that $T|_{w(\mathcal{C}_w)}$ is regular at $w(p)$, or p is a fold point and $w(p)$ is a local minimum of $T|_{w(\mathcal{C}_w)}$; see Figure 14 for an illustration. It turns out that only the minima of $T|_{w(\mathcal{C}_w)}$ contribute critical points of f. Indeed, if p is either a cusp point or a fold point where $T|_{w(\mathcal{C}_w)}$ is regular, then we can find model coordinates around p such that (up to a sign) the map T corresponds to the *vertical projection* in the fold or cusp model, that is, the projection of $\mathbb{R}^2$ onto the first factor. As a consequence, f is modeled around p by one of the compositions

$$(t, x, y, z) \mapsto (t, x^3 + 3tx + y^2 - z^2) \mapsto t$$
$$(t, x, y, z) \mapsto (t, x^2 + y^2 - z^2) \mapsto \pm t$$

which shows that p is a regular point of f. In the remaining case, when p is a fold point and $T|_{w(\mathcal{C}_w)}$ is minimal at $w(p)$, Figure 14 suggests that T should be

modeled on the *horizontal projection* in the fold model. Of course, this is not quite possible because the horizontal projection is not a Morse function on the critical image of the fold model. But this obstacle is easily overcome by a change of coordinates. Composing the fold model with either of the diffeomorphisms of $\mathbb{R}^2$ given by $(t, s) \mapsto (t, s \pm t^2)$, we obtain two new models

$$(t, x, y, z) \mapsto (t, x^2 + y^2 - z^2 \pm t^2) \tag{5.2}$$

which we call the $\pm$*–modifications* of the fold model. In the $\pm$–modification the critical image is parametrized by $t \mapsto (t, \pm t^2)$ so that the horizontal projection restricts to the Morse function $t \mapsto \pm t^2$. What is more, note that the composition of the modified models with the horizontal projection is the standard model for a 4–dimensional Morse critical point of index 2 or 3. Returning to the function $f\colon W \to I$ and the fold point $p \in \mathcal{C}_w$ with $w(p)$ a minimum of $T|_{w(\mathcal{C}_w)}$, we can find modified model coordinates around p such that, up to a sign, T corresponds to the horizontal projection. In order to find out which modified model fits, we observe that the Euler characteristic of the fibers of w increases as T increases, while in either of the modified models the Euler characteristic decreases in the positive horizontal directions. So T must be modeled on the negative horizontal projection, and since $T|_{w(\mathcal{C}_w)}$ has a minimum at p, we have to use the $(-)$–modification. In particular, we see that p is a non-degenerate critical point of f and has index 2.

With these remarks in place, we return to the study of w. We choose a horizontal distribution $\mathcal{H}$ and a reference system $\mathcal{R} = (R_1, \dots, R_l)$ for w which induce the generalized surface diagram $\mathfrak{G} = (\Sigma, \Gamma, \psi)$. It will be convenient to use the notation

$$W_{[s,t]} = f^{-1}([s, t]) \quad \text{for } s, t \in I,\ s < t.$$

Note that $f^{-1}([s, t]) = w^{-1}\big(S^1 \times [s, t]\big)$ and that w maps $f^{-1}(t) = w^{-1}(S^1_t)$ onto S^1_t.

In order to prove the equivalence of w and $w_{\mathfrak{G}}$, we choose $\sigma, \tau \in I$ close to t_f such that $\sigma < t_f < \tau < t_c$, and try to relate the pieces $W_{[0,\sigma]}$, $W_{[\sigma,\tau]}$, and $W_{[\tau,1]}$ to the three steps in the in construction of $w_{\mathfrak{G}}$ in the proof of Lemma 5.8.

Step 1: Using the parallel transport with respect to $\mathcal{H}$ it is easy to construct a commutative diagram

$$\begin{array}{ccc} W_{[0,\sigma]} & \xrightarrow{\cong} & \Sigma(\psi) \times [0, \sigma] \\ & {\scriptstyle w}\searrow \quad \swarrow{\scriptstyle p_\psi \times \mathrm{id}} & \\ & S^1 \times [0, \sigma] & \end{array}$$

where $p_\psi\colon \Sigma(\psi) \to S^1$ is the canonical fibration of the mapping torus. For example, for $\theta < 1$ and $t \le \sigma$ we can obtain a diffeomorphism

$$\Sigma = w^{-1}(0, 0) \xrightarrow{\cong} w^{-1}(\theta, \tau)$$

by parallel transport along the arc in $S^1 \times [0, \sigma]$ given by $s \mapsto (s\theta, st)$, $s \in [0, 1]$. This directly identifies the restriction of w to $W_{[0,\sigma]}$ with the result of the first step in the construction of $w_{\mathfrak{G}}$.

Step 2: Note that the restriction of w to $W_{[0,\tau]}$ closely resembles the map $w_2 \colon W_2 \to \left[0, \frac{2}{3}\right]$ obtained in the second step of the construction of $w_{\mathfrak{G}}$. It is a wrinkled fibration with l properly embedded fold arcs with boundary on $S^1_\tau = S^1 \times \{\tau\}$, each bounding a half disk in $S^1 \times I$, and over S^1_τ we find a circle valued Morse function with l canceling pairs of critical points. Moreover, the Morse function f tells us that $W_{[0,\tau]}$ is diffeomorphic to $W_{[0,\sigma]}$ with l 2–handles attached. We now take a closer look a these 2–handles.

Let $p_1, \dots, p_l \in W$ be the critical points of f labeled such that $w(p_i)$ lies on the fold arc corresponding to R_i. For convenience, we assume that

$$R_i \cap \big(S^1 \times [\sigma, \tau]\big) = \{\theta_i\} \times [\sigma, \tau] \qquad \text{and} \qquad R_i \cap w(\mathcal{C}_w) = w(p_i) = (\theta_i, t_f),$$

for some $0 < \theta_1 < \dots < \theta_l < 1$. In other words, R_i is a straight line near the critical image and the intersection agrees with the minimum of $T|_{w(\mathcal{C}_w)}$ on the corresponding fold arc. This can always be achieved by an isotopy of $\mathcal{R}$ which only changes the vanishing cycles by isotopies (which can then be compensated by modifying $\mathcal{H}$ according to Lemma 3.10). Next we choose a gradient-like vector field ξ for f as follows. As in the proof of Lemma 3.13, the restriction of f to $w^{-1}(R_i)$ is a Morse function and has a gradient-like vector field ξ_i which takes values in $\mathcal{H}$. Moreover, the descending sphere of the pair $(f|_{w^{-1}(R_i)}, \xi_i)$ is exactly the vanishing cycle $c_i \subset \Sigma$. We define ξ to be ξ_i on $w^{-1}(R_i) \cap W_{[\sigma,\tau]}$ and extend it arbitrarily to a gradient-like vector field for f. Now everything is set up so that the descending manifold of (f, ξ) for $p_i \in \mathcal{C}_f$ intersects $f^{-1}(\sigma) = w^{-1}(S^1_\sigma)$ in the copy of c_i in the fiber $w^{-1}(\theta_i, \sigma)$. It follows that the 2–handle corresponding to p_i is attached along this copy of c_i and the framing must be the fiber framing since it restricts to the framing of the 3–dimensional 2–handle corresponding to the critical point of $(f|_{w^{-1}(R_i)}, \xi_i)$.

To sum up, $W_{[0,\tau]}$ is obtained from $W_{[0,\sigma]}$ by fold handle attachments in the same way as W_2 is obtained from W_1. It now follows from the uniqueness of fold handle attachments mentioned in the proof of Lemma 5.8 that $w|_{W_{[0,\tau]}}$ is equivalent to w_2.

Step 3: Since f has no critical values in $[\tau, 1]$ we can use the flow of ξ to obtain a diffeomorphism $f^{-1}(\tau) \times [\tau, 1] \xrightarrow{\cong} W_{[\tau,1]}$. Moreover, we can assume that ξ maps to ∂_t under dw. Since $f^{-1}(\tau) = w^{-1}(S^1_\tau)$, we obtain a commutative diagram

$$\begin{array}{ccc} w^{-1}(S^1_\tau) \times [\tau, 1] & \xrightarrow{\cong} & W_{[\tau,1]} \\ \downarrow{\scriptstyle \mathrm{pr}_2} & \searrow & \downarrow{\scriptstyle w} \\ [\tau, 1] & \xleftarrow{T} & S^1 \times [\tau, 1]. \end{array}$$

But this shows that the restriction of w to $W_{[\tau,1]}$ can be considered as the trace of a path of death for the circle valued Morse function $w \colon w^{-1}\big(S^1_\tau\big) \to S^1_\tau$ which matches the third step in the construction of $w_{\mathfrak{G}}$. □

Remark 5.10. We can now explain why we require simple wrinkled fibrations to have cusps. In the above proof it was crucial that we could arrange the restriction of T to the fold part of the critical image to be a Morse function with only local minima. Each minimum then contributed an index 2 critical points of f. However, if there were no cusps but only a circle of indefinite folds, then there would also have to be local maxima which would contribute index 3 critical points of f. (T would have

to be modeled on the negative horizontal projection in the (+)–modification of the fold model.) So the presence of cusps guarantees that the total space of an annular simple wrinkled fibration has a relative handle decomposition on the higher genus boundary component with only 2–handles. For fake simple wrinkled fibrations 3–handles become necessary. Note that this is in accordance with Baykur's work on handle decompositions of broken Lefschetz fibrations where each circle of indefinite folds contributes a 2–handle and a 3–handle, see [5, Ch. 2]. Finally, we would like to mention that the observation that certain projections of wrinkled fibrations are Morse functions was made independently by Gay and Kirby in their work on Morse 2–functions, see [28, p.43ff.].

Returning to the proof of Theorem 1.3, we want to define a map

$$g\mathcal{SD} \longrightarrow a\mathcal{SWF}$$

by sending a generalized surface diagram $\mathfrak{G}$ to the annular simple wrinkled fibration $w_{\mathfrak{G}}\colon W_{\mathfrak{G}} \to S^1 \times I$ constructed in Lemma 5.8. Assuming that this map is well-defined, it follows from Lemmas 5.8 and 5.9 that it is an inverse for the map $a\mathcal{SWF} \to g\mathfrak{S}$ constructed in Section 5.1.1. The proof of Theorem 1.3 is completed with the following lemma.

Lemma 5.11. *If $\mathfrak{G}$ and $\mathfrak{G}'$ are equivalent, then $w_{\mathfrak{G}}$ and $w_{\mathfrak{G}'}$ are equivalent.*

Proof. First, it is clear from the proof of Lemma 5.8 that if $\mathfrak{G}$ and $\mathfrak{G}'$ are isotopic, then $w_{\mathfrak{G}}$ and $w_{\mathfrak{G}'}$ are equivalent. Second, if $\mathfrak{G}$ and $\mathfrak{G}'$ differ by a twisted cyclic permutation, then we claim that $w_{\mathfrak{G}}$ has a reference system whose diagram is isotopic to $\mathfrak{G}'$ and Lemma 5.9 shows that $w_{\mathfrak{G}}$ is equivalent to $w_{\mathfrak{G}'}$. Indeed, as we saw in the proof of Lemma 5.4, twisted cyclic permutations can be realized by suitably swinging reference arcs. So we simply have to take the reference system for $w_{\mathfrak{G}}$ shown in Figure 13 and swing the first or last reference arc to obtain a reference system that induces $\mathfrak{G}'$. Lastly, suppose that $\mathfrak{G} = (\Sigma, \Gamma, \psi)$ and $\mathfrak{G}' = (\Sigma', \Gamma', \psi')$ are diffeomorphic via a diffeomorphism $\phi\colon \Sigma \to \Sigma'$. We claim that ϕ induces a diffeomorphism $\Phi\colon W_{\mathfrak{G}} \to W_{\mathfrak{G}'}$ such that $w_{\mathfrak{G}} = w_{\mathfrak{G}'} \circ \Phi$. To construct Φ we go through the steps of the proof of Lemma 5.8. Let W_i and W_i', $i = 1, 2$, be the manifolds resulting from the first two steps for $\mathfrak{G}$ and $\mathfrak{G}'$, respectively. Since we are assuming that $\psi' = \phi\psi\phi^{-1}$, we obtain a diffeomorphism $\Sigma(\psi) \to \Sigma'(\psi')$ induced by $\phi \times \mathrm{id}\colon \Sigma \times I \to \Sigma' \times I$. In turn, this provides a diffeomorphism $\Phi_1\colon W_1 \to W_1'$ which maps the attaching regions of the fold handles in $W_{\mathfrak{G}}$ to those of $W_{\mathfrak{G}'}$. It then follows from the discussion surrounding equation (5.1) in Remark 5.7 that Φ_1 extends across the fold handles to $\Phi_2\colon W_2 \to W_2'$. Finally, the extension to the rest $W_{\mathfrak{S}}$ follows from Corollary C.5. □

5.2 Identifying the Monodromy

Let $w\colon W \to A$ be an annular simple wrinkled fibration with generalized surface diagram $\mathfrak{G} = (\Sigma, \Gamma, \psi)$. As on page 54, we fix a reference system $\mathcal{R}$ and a horizontal distribution $\mathcal{H}$ for w that induce $\mathfrak{G}$ and denote that higher and lower genus monodromies of w by

$$\psi = \mathrm{PT}^{\mathcal{H}}_{\partial_+ A}\colon \Sigma \longrightarrow \Sigma \quad \text{and} \quad \mu = \mathrm{PT}^{\mathcal{H}}_{\partial_- A}\colon \Sigma' \longrightarrow \Sigma'.$$

By definition, ψ is part of the data of $\mathfrak{G}$ and we will now give an interpretation of μ in terms of $\mathfrak{G}$. Recall from Lemma 3.13 that the parallel transport with respect to $\mathcal{H}$ along the first reference arc of $\mathcal{R}$ induces an identification of Σ' with the surface Σ_{c_1} obtained by surgery on $c_1 \subset \Sigma$.

Proposition 5.12. *Let $w\colon W \to A$ be an annular simple wrinkled fibration with generalized surface diagram $\mathfrak{G} = (\Sigma, \Gamma, \psi)$. Under the identification $\Sigma' \cong \Sigma_{c_1}$ explained above, the monodromy diffeomorphism μ is represents the monodromy of Σ*

$$\mu(\mathfrak{G}) = \Phi_{c_1}\big(\psi^{-1}\ \tau_{c_l}\tau_{c_1}\ \tau_{c_{l-1}\tau_{c_l}} \cdots \tau_{c_1\tau_{c_2}}\big) \in \mathrm{Mod}(\Sigma_{c_1}).$$

Moreover, all representatives of $\mu(\mathfrak{G})$ can be realized in this way.

Before going into the proof, we embark on a small digression and describe the map Φ_{c_1} appearing in the definition of $\mu(\mathfrak{G})$ in more detail. These so called *surgery homomorphisms* have also appeared in the work of Baykur [5, p.214] on broken Lefschetz fibrations and their importance for the theory of (simple) wrinkled fibrations is highlighted by results of Hayano [35] and Hayano and the author [8]. A different approach to relating the higher and lower genus monodromy from the persepective of the lower genus side can be found in [3, Section 8.1].

5.2.1 Mapping Class Groups and Surgery

Let Σ be a closed, orientable surface of genus g and let $c \subset \Sigma$ be a non-separating simple closed curve. As before, we denote by $\mathrm{Mod}(\Sigma)(c)$ the subgroup of $\mathrm{Mod}(\Sigma)$ consisting of all elements that fix the isotopy class of c, and by Σ_c the surface obtained by surgery on $c \subset \Sigma$. Recall that we consider Σ_c as the endpoint compactification of $\Sigma \setminus c$. We define the *surgery homomorphism*

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

as follows. By the isotopy extension theorem, every $\theta \in \mathrm{Mod}(\Sigma)(c)$ has a representative $T \in \mathrm{Diff}^+(\Sigma)$ which satisfies $T(c) = c$. Such a T restricts to a diffeomorphism of $\Sigma \setminus c$ which, in turn, extends uniquely to a diffeomorphism $\widehat{T}$ of the endpoint compactification Σ_c. We define $\Phi_c(\theta)$ as the mapping class represented by $\widehat{T}$.

In order to see that Φ_c is well-defined, we first observe that if Σ has genus one, then Φ_c takes values in a trivial group and there is nothing to check. For genus two and higher we have to show that two representatives of θ which preserve c as a set are isotopic through diffeomorphisms with this property. But as explained in [36, Ch.7.5], this follows from the fact that the space of simple closed curves isotopic to c is simply connected (see Corollary 2.19).

It is sometimes useful to think of Φ_c as a composition. Observe that Σ_c contains two distinguished points, namely the endpoints, and the construction used to define Φ_c produces diffeomorphisms which preserve these points. We can thus consider Σ_c as a surface with marked points, for which we shall use the notation Σ_c^*, and we obtain a homomorphism $\Phi_c^*\colon \mathrm{Mod}(S)(c) \to \mathrm{Mod}(\Sigma_c^*)$. Composing with the *forgetful map* $\mathcal{F}\colon \mathrm{Mod}(\Sigma_c^*) \to \mathrm{Mod}(\Sigma_c)$, which simply forgets the marked points, we recover Φ_c. The advantage of this interpretation is that both Φ_c^* and $\mathcal{F}$ are known to be surjective and their kernels are known. The kernel of Φ_c^* is generated by the Dehn twist about c (see [24, Proposition 3.20]), and that of $\mathcal{F}$ is isomorphic to the

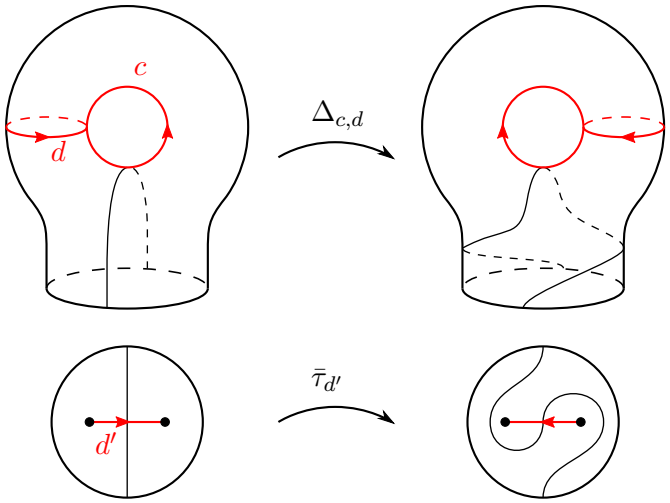


Figure 15: The effects of Δ–twists and half-twists.

braid group of Σ_c on two strands (see [24, Theorem 9.1]). This shows that Φ_c is surjective and gives some idea about its kernel.

Let us discuss some special elements of $\mathrm{Mod}(\Sigma)(c)$ and their images under Φ_c.

Example 5.13. Let $b \subset \Sigma$ be a simple closed curve with $i(b,c) = 0$. Then b can be made disjoint from c by an isotopy and the Dehn twist τ_b is obviously contained in $\mathrm{Mod}(\Sigma)(c)$. Moreover, a Dehn twist about a curve isotopic to b and disjoint from c gives a representative for $\Phi_c(\tau_b)$.

Example 5.14 (Δ–twists)**.** More interesting elements can be obtained from simple closed curves $d \subset \Sigma$ with $i(c,d) = 1$. We define the Δ–*twist* about c and d as

$$\Delta_{c,d} = (\tau_c\tau_d)^3 = (\tau_c\tau_d\tau_c)^2 = (\tau_d\tau_c\tau_d)^2 = (\tau_d\tau_c)^3 = \Delta_{d,c}$$

where the equalities follow from the braid relation (2.9). A repeated application of equation (2.8) shows that $\Delta_{c,d}$ fixes c (and also d) up to isotopy. In order to obtain a more geometric interpretation of $\Delta_{c,d}$ we assume that c and d intersect transversely in one point. Then a neighborhood $\nu(c \cup d)$ is a one-holed torus and a representative for $\Delta_{c,d}$ is obtained by a half-rotation of $\nu(c \cup d)$ relative its boundary as shown in Figure 15. It is clear from the picture that such a half rotation preserves the curve c (while reversing its orientation) and thus gives rise to a representative of $\Phi_c\big(\Delta_{c,d}\big)$. Observe that in the surgered surface Σ_c the curve d appears as an arc d' connecting the two endpoints while the one-holed torus $\nu(c \cup d)$ becomes a disk containing d'. Moreover, the half rotation in $\nu(c \cup d)$ descends to the half-twist[4] about d' as shown in Figure 15. Put differently, we have $\Phi_c^*\big(\Delta_{c,d}\big) = \bar{\tau}_{d'} \in \mathrm{Mod}(\Sigma_c^*)$. But since half-twists are realized by isotopies of Σ_c, they are annihilated by the forgetful map and we have

$$\Phi_c\big(\Delta_{c,d}\big) = \mathcal{F}(\bar{\tau}_{d'}) = 1 \in \mathrm{Mod}(\Sigma_c).$$

It turns out that the elements discussed in Examples 5.13 and 5.14 are enough to generate the group $\mathrm{Mod}(\Sigma)(c)$.

[4]See page 23 for the definition.

Lemma 5.15. *Let $c \subset \Sigma$ be a non-separating simple closed curve. Then* $\mathrm{Mod}(\Sigma)(c)$ *is generated by elements of the form τ_b where $i(b,c) = 0$ and $\Delta_{c,d}$ where $i(c,d) = 1$. Moreover, the elements $\Delta_{c,d}$ are contained in the kernel of Φ_c and so is τ_c.*

Proof. As mentioned above, there is a short exact sequence

$$1 \longrightarrow \langle \tau_c \rangle \longrightarrow \mathrm{Mod}(\Sigma)(c) \xrightarrow{\Phi_c^*} \mathrm{Mod}(\Sigma_c^*) \longrightarrow 1$$

and it is well known that the group $\mathrm{Mod}(\Sigma_c^*)$ is generated by Dehn twists about curves that are disjoint from the marked points together with half-twists about arcs connecting the marked points (see [24, Corollary 4.15]). The Dehn twists obviously lift to $\mathrm{Mod}(\Sigma)(c)$ since they are supported in $\Sigma \setminus c \subset \Sigma_c$. Moreover, the arguments in Example 5.14 show that half-twists in $\mathrm{Mod}(\Sigma_c^*)$ lift to Δ–twists in $\mathrm{Mod}(\Sigma)(c)$. The claim about generators for $\mathrm{Mod}(\Sigma)(c)$ thus follows from the exact sequence. The fact that τ_c and all Δ–twists are annihilated by Φ_c was observed before. □

So in principle, one can compute $\Phi_c(\theta)$ by expressing θ in terms of Dehn twists disjoint from c and Δ–twists involving c, which we call the *standard generators* of $\mathrm{Mod}(\Sigma)(c)$. However, this can be difficult in practice. For example, if θ is given as an word Dehn twists that might involve twist about curves that intersect c, then it is at all not clear how to rewrite θ in terms of the standard generators of $\mathrm{Mod}(\Sigma)(c)$. It would be desirable to have an algorithm for this, but no such algorithm is known. On the subject of unknown things, it is conceivable that the kernel of Φ_c is actually generated by τ_c and the Δ–twists involving c but we have not been able to prove this.

5.2.2 The Proof of Proposition 5.12

We return to the problem of identifying the lower genus monodromy of an annular simple wrinkled fibration $w\colon W \to A$ in terms of a generalized surface diagram $\mathfrak{G} = (\Sigma, \Gamma, \psi)$. We first discuss this in the context of for a fake annular simple wrinkled fibration, that is, for a map $w\colon W \to S^1 \times I$ whose critical locus is a single circle of indefinite folds which is mapped injectively onto a boundary parallel circle in $S^1 \times I$. Such maps are also known as *round cobordisms* (see [5, 33]).

Lemma 5.16. *Let $w\colon W \to A$ be a round cobordism and let $R \subset A$ be a reference arc from $\partial_+ A$ to $\partial_- A$ with higher and lower genus reference fiber Σ and Σ'. Let $c \subset \Sigma$ be the vanishing cycle of R with respect to a horizontal distribution $\mathcal{H}$ and let $\phi\colon \Sigma' \to \Sigma_c$ be the diffeomorphism induced by parallel transport along R. Then the higher and lower genus monodromies ψ and μ satisfy*

$$\Phi_c(\psi^{-1}) = \phi\, \mu\, \phi^{-1} \in \mathrm{Mod}(\Sigma_c)$$

on the level of mapping class groups.

Proof. Again, the statement is trivial if Σ has genus one. We can assume that the genus of Σ is at least two. We consider a 1–parameter family of arcs $R_t \subset \Sigma$ such that $R_0 = R$, $\partial R_t = \partial R$ for all t, R_t is a reference arc for $t < 1$, and R_1 agrees with the concatenation $\partial_+ A^{-1} * R * \partial_- A$, see Figure 16. According to

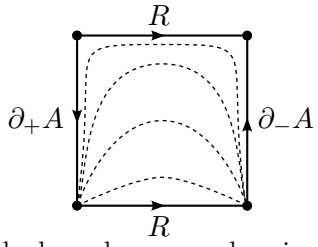


Figure 16: Relating the boundary monodromies of a round cobordism.

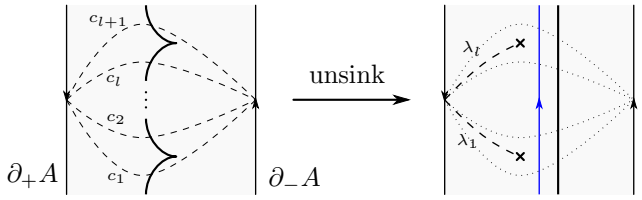


Figure 17: Turning an annular simple wrinkled fibration into a broken Lefschetz fibrations.

Lemma 3.13, we obtain 1–parameter families of vanishing cycles $c_t \subset \Sigma$ and diffeomorphisms $\phi_t \colon \Sigma' \to \Sigma_{c_t}$ with $c_0 = c$ and $\phi_0 = \phi$. But all the vanishing cycles are isotopic by Lemma 3.18 and, since the component of c in the space of simple closed curve is simply connected (see Corollary 2.19), we can use Lemma 3.10 to modify $\mathcal{H}$ in the complement of R such that $c_t = c$ for all t. In particular, we now have a 1–parameter family of diffeomorphism $\phi_t \colon \Sigma' \to \Sigma_c$. Furthermore, the higher genus monodromy ψ maps c to itself so that the canonical extension of $\psi|_{\Sigma \setminus c}$ to Σ_c, denoted by $\hat{\psi}$, represents $\Phi_c(\psi)$. Now, it follows from the definition of R_1 that

$$\phi_1 = \hat{\psi} \circ \phi \circ \mu \colon \Sigma' \to \Sigma_c$$

which finishes the proof since $\phi_1 \phi^{-1} \in \mathrm{Diff}(\Sigma_c)$ is isotopic to the identity. □

Proof ofProposition 5.12. As in the beginning of this section, let $w \colon W \to A$ be an annular simple wrinkled fibration together with a horizontal distribution $\mathcal{H}$ and a reference system $\mathcal{R} = (R_1, \dots, R_l)$. As shown in Figure 17, we turn w into a broken Lefschetz fibration β_w using Lekili's unsinking modification[5] to trade all cusps for Lefschetz singularities in the higher genus region. Let λ_i be the Lefschetz vanishing cycle of the Lefschetz singularity created by unsinking the cusp between the ith and $(i+1)$st fold arc with respect to the indicated reference arc. It is well known that the monodromy around a loop that encircles the ith (Lefschetz) reference arc in a counterclockwise fashion is the Dehn twist about λ_i. Since the blue circle in Figure 17 is homotopic to the successive concatenation of such loops and a loop that travels once around $\partial_+ A$ in the negative direction, the monodromy around this loop satisfies

$$\tilde{\mu} = \psi^{-1} \tau_{\lambda_l} \dots \tau_{\lambda_1} \in \mathrm{Mod}(\Sigma).$$

Moreover, the restriction of β_w to the region bounded by the blue circle and $\partial_- A$ is a round cobordism. Since β_w agrees with w near $\partial_- A$, it follows from Lemma 5.16 that the lower genus monodromy μ of w represents $\Phi_{c_1}(\tilde{\mu})$.

[5]This was explained in Section 3.4.

It remains to show that $\Phi_{c_1}(\tilde{\mu}) = \mu(\mathfrak{G})$. According to Lekili [45, p.285], we have $\lambda_i = \tau_{c_i}(c_{i+1})$ where $c_{l+1} = \psi(c_1)$ so that by equation (2.7)

$$\tilde{\mu} = \psi^{-1}\ \tau_{c_l}\tau_{c_{l+1}}\tau_{c_l}^{-1}\ \dots\ \tau_{c_1}\tau_{c_2}\tau_{c_1}^{-1}.$$

Moreover, we can move all the inverses to the left by a repeated application of the relation $\tau_{c_i}\tau_{c_{i+1}}\tau_{c_i}^{-1} = \tau_{c_{i+1}}^{-1}\tau_{c_i}\tau_{c_{i+1}}$ (which follows from the braid relation (2.9) for c_i and c_{i+1}) and we obtain

$$\tilde{\mu} = \psi^{-1}\ \tau_{c_{l+1}}^{-l}\ \tau_{c_l}\tau_{c_{l+1}}\ \cdots\ \tau_{c_1}\tau_{c_2}.$$

Finally, another application of equation (2.7) gives

$$\psi^{-1}\tau_{c_{l+1}} = \psi^{-1}\tau_{\psi(c_1)} = \psi^{-1}(\psi\tau_{c_1}\psi^{-1}) = \tau_{c_1}\psi^{-1}$$

and we end up with

$$\tilde{\mu} = \tau_{c_1}^{-l}\ \psi^{-1}\ \tau_{c_l}\tau_{c_{l+1}}\ \cdots\ \tau_{c_1}\tau_{c_2}$$

and the fact that $\tau_{c_1} \in \ker \Phi_{c_1}$ finishes the proof. □

Remark 5.17. Of course, there are many different lifts of $\mu_{\mathfrak{G}}$ to $\mathrm{Mod}(\Sigma)$. For example, using the braid relation as in the proof above we can also write

$$\begin{aligned}\tilde{\mu}(\mathfrak{G}) &= \Phi_{c_1}\big(\psi^{-1}\tau_{c_l}\tau_{c_1}\ \tau_{c_{l-1}}\tau_{c_l}\ \cdots\ \tau_{c_1}\tau_{c_2}\big)\\ &= \Phi_{c_1}\big(\tau_{c_1}^{l}\psi^{-1}\tau_{c_l}\tau_{c_1}\ \tau_{c_{l-1}}\tau_{c_l}\ \cdots\ \tau_{c_1}\tau_{c_2}\big)\\ &= \Phi_{c_1}\big(\psi^{-1}(\tau_{c_l}\tau_{c_1}\tau_{c_l})(\tau_{c_{l-1}}\tau_{c_l}\tau_{c_{l-1}})(\tau_{c_1}\tau_{c_2}\tau_{c_1})\big).\end{aligned}$$

The latter expression can also be useful.

5.3 The Correspondence over the Disk and the Sphere

It remains to prove the correspondence theorem for surface diagrams and simple wrinkled fibrations over the disk and the sphere. The following terminology will be useful.

Definition 5.18 (Fibered neighborhoods)**.** Let $f\colon M \to B$ be wrinkled fibration and let $b \in B$ be a regular value. An open (or closed) *fibered neighborhood* of $\Sigma = f^{-1}(b)$ with *base* $D \subset B$ is a neighborhood of the form $f^{-1}(D) \subset M$ where $D \subset B$ is an open (or closed) disk of regular values containing b as an interior point.

For convenience we recall the statements before their proofs.

Theorem 1.2 (Correspondence over the disk)**.** *Equivalence classes of surface diagrams correspond bijectively to equivalence classes of descending simple wrinkled fibrations over the disk.*

Proof. Let $w\colon Z \to D^2$ be a descending simple wrinkled fibration and let N be a closed fibered neighborhood of a higher genus fiber Σ with base $D \subset D^2$. Then w restricts to an annular simple wrinkled fibration on $Z \setminus \mathrm{int}N$ and, since ∂N is diffeomorphic to $\Sigma \times S^1$, any generalized surface diagram (Σ, Γ, ψ) for $w|_W$ must be untwisted, that is, ψ is isotopic to the identity. In other words, $\mathfrak{S}_w = (\Sigma, \Gamma)$ is an honest

surface diagram in the sense of Definition 4.11. Moreover, the map $(Z, w) \mapsto \mathfrak{S}_w$ is well-defined on equivalence classes by one half of Theorem 1.3 and the obvious observation that equivalent descending simple wrinkled fibrations restrict to equivalent annular simple wrinkled fibrations.

Now let $\mathfrak{S} = (\Sigma, \Gamma)$ be a surface diagram. By construction, the higher genus boundary component of $W_{\mathfrak{S}}$ is the trivial mapping torus $\Sigma(\mathrm{id})$ which is canonically identified with $\Sigma \times S^1$. We can thus close off the annular simple wrinkled fibration $(W_{\mathfrak{S}}, w_{\mathfrak{S}})$ to a descending simple wrinkled fibration

$$w_{\mathfrak{S}} \colon Z_{\mathfrak{S}} = \Sigma \times D^2 \cup_{\mathrm{id}} W_{\mathfrak{S}} \longrightarrow D^2 \cup (S^1 \times I) \cong D^2$$

which we also denote by $w_{\mathfrak{S}}$. We claim that the map $\mathfrak{S} \mapsto (Z_{\mathfrak{S}}, w_{\mathfrak{S}})$ is also well-defined on equivalence classes. Indeed, if $\mathfrak{S} = (\Sigma, \Gamma)$ is equivalent to $\mathfrak{S}' = (\Sigma', \Gamma')$, then the diffeomorphism $\Phi \colon W_{\mathfrak{S}} \to W_{\mathfrak{S}'}$ constructed in the proof of Lemma 5.11 restricts to the higher genus boundary components as $\phi \times \mathrm{id} \colon \Sigma \times S^1 \to \Sigma \times S^1$ for some diffeomorphism $\phi \colon \Sigma \to \Sigma'$. Using the canonical extension $\phi \times \mathrm{id}\, \Sigma \times D^2 \to \Sigma' \times D^2$ we can therefore extend Φ to a diffeomorphism $Z_{\mathfrak{S}} \to Z_{\mathfrak{S}'}$ which establishes the equivalence of $(Z_{\mathfrak{S}}, w_{\mathfrak{S}})$.

It remains to show that both constructions are mutually inverse on the level of equivalence classes. Obviously, $(Z_{\mathfrak{S}}, w_{\mathfrak{S}})$ is mapped back to surface diagram $\mathfrak{S}$. Now suppose that another simple wrinkled fibration (Z, w) is sent to $\mathfrak{S}$. We have to show that (Z, w) is equivalent to $(Z_{\mathfrak{S}}, w_{\mathfrak{S}})$. As above, we take a fibered neighborhood $N \subset Z$ of a higher genus fiber of w. Then by Theorem 1.3 the restriction of w to $W = Z \setminus \mathrm{int} N$ is equivalent to $(W_{\mathfrak{S}}, w_{\mathfrak{S}})$. Moreover, the restriction of w to N is clearly equivalent to the projection $\Sigma \times D^2 \to D^2$. But the problem is that these equivalences might not match along ∂N. In fact, we can only conclude that (Z, w) is equivalent to a simple wrinkled fibration of the form

$$(\Sigma \times D^2) \cup_{\varphi} W_{\mathfrak{S}} \to D^2 \cup (S^1 \times I)$$

where φ is a fiber preserving diffeomorphism of $\Sigma \times S^1$; or in the language introduced on page 25, $\varphi \in \mathrm{Aut}(\Sigma \times S^1)$ is an automorphism of the trivial bundle $\Sigma \times S^1$. If we can show that φ extends across either $\Sigma \times D^2$ or $W_{\mathfrak{S}}$, then the proof is finished.

If Σ has genus at least two, then according to Corollary 2.20 all such automorphisms extend to $\Sigma \times D^2$. So problems can only arise when Σ is a torus and for simplicity we assume that $\Sigma = T^2$. In this case, Corollary 2.20 allows us to restrict our attention to (isotopy classes of) automorphisms corresponding to non-trivial elements of $\pi_1(\mathrm{Diff}(T^2), \mathrm{id})$. Such automorphisms cannot be extended across $T^2 \times D^2$, so our only chance is to extend them across $W_{\mathfrak{S}}$. Fortunately, this is possible. We can assume that the first vanishing cycles in $\mathfrak{S}$ are the curves $a = S^1 \times \{1\}$ and $b = \{1\} \times S^1$ in $T^2 = S^1 \times S^1 \subset \mathbb{C}^2$. Since a and b generate $\pi_1(T^2)$ (whose base point $(1, 1)$ we omit from the notation), it follows from Theorem 2.18 that $\pi_1(\mathrm{Diff}(T^2), \mathrm{id})$ is generated by loops of the form

$$h^a_\theta(z, w) = \left(e^{2\pi i \rho(\theta)} z, w\right) \quad \text{and} \quad h^b_\theta(z, w) = \left(z, e^{2\pi i \rho(\theta)} w\right) \qquad (\theta \in [0, 1])$$

where $\rho \colon [0, 1] \to [0, 1]$ is a smooth function such that for a sub-interval $[\sigma, \tau] \subset [0, 1]$ we have $\rho|_{[0,\sigma]} \equiv 0$ and $\rho|_{[\tau,1]} \equiv 1$. It is thus enough to extend the corresponding automorphisms across $W_{\mathfrak{S}}$; we denote these by $h^a, h^b \in \mathrm{Aut}(T^2 \times S^1)$. We explain

how this for h^a, the argument for h^b is completely analogous. Recall that $W_{\mathfrak{S}}$ was built by attaching fold handles to $T^2 \times S^1 \times I$ in the boundary $T^2 \times S^1 \times \{1\}$ which we identify with $T^2 \times S^1$ for brevity of notation. The attachment of the fold handle along a was realized by gluing on a copy of $\Theta_a \times \omega(T^2, a)$ where $\Theta_a \subset S^1$ is an interval and $\omega(T^2, a)$ was obtained from Milnor's construction in Remark 5.7. By an appropriate choice of ρ we can assume that h^a restricts to the identity outside of $T^2 \times \Theta_a$. The crucial observation is that h^a_θ maps a to itself for all $\theta \in \Theta_a$ (in fact, for all $\theta \in S^1$). As explained in Remark 5.7, this implies that h^a_θ extends to a self-diffeomorphism of $\omega(T^2, a)$ and all these extensions piece together to an extension over $\Theta_a \times \omega(T^2, a)$. Since h^a is the identity outside of $T^2 \times \Theta_a$, we obtain the desired extension to all $W_{\mathfrak{S}}$. □

Theorem 1.1 (Correspondence over the sphere)**.** *Let $\mathcal{SWF}_g(S^2)$ and $\mathcal{SD}^0_g$ be the sets of equivalence classes of genus g simple wrinkled fibrations over S^2 and surface diagrams with trivial monodromy, respectively. There is a surjective map*

$$\mathcal{SWF}_g(S^2) \longrightarrow \mathcal{SD}^0_g$$

whose point preimages have a transitive action of the group $\pi_1\big(\mathrm{Diff}(\Sigma_{g-1}), \mathrm{id}\big)$. In particular, the map is bijective for $g \geq 3$.

Proof. Let $w\colon X \to S^2$ be a simple wrinkled fibration. We argue exactly as in the proof of Theorem 1.2. By removing a fibered neighborhood of a lower genus fiber we obtain a descending simple wrinkled fibration over a disk whose boundary fibration has trivial monodromy. According to Proposition 5.12, the corresponding surface diagram $\mathfrak{S}_w$ must has trivial monodromy. Then sending the equivalence class of (X, w) to that of $\mathfrak{S}_w$ is the desired map.

Now let $\mathfrak{S}$ be a genus g surface diagram with trivial monodromy. Then the corresponding descending simple wrinkled fibration $w_{\mathfrak{S}}\colon Z_{\mathfrak{S}} \to D^2$ restricts to a trivial Σ_{g-1}–bundle on $Z_{\mathfrak{S}}$, again by Proposition 5.12. The choice of a fiber preserving diffeomorphism $\varphi\colon \Sigma_{g-1} \times S^1 \to \partial Z_{\mathfrak{S}}$ then determines a simple wrinkled fibration

$$w^\varphi_{\mathfrak{S}}\colon X^\varphi_{\mathfrak{S}} \to S^2$$

whose surface diagram is $\mathfrak{S}$. It follows that the map $(X, w) \to \mathfrak{S}_w$ is surjective. Note that this time there is no canonical choice for φ and different choices can result in inequivalent simple wrinkled fibrations over S^2 (see Section 6.3.1 for an example). However, by Corollary 2.20 the ambiguity is controlled by $\pi_1\big(\mathrm{Diff}(\Sigma), \mathrm{id}\big)$. This finishes the proof. □

Remark 5.19. The correspondence result for simple wrinkled fibrations over S^2 emphasizes the importance of the monodromy of surface diagrams. One might wonder whether there are a priori restrictions on the monodromy of surface diagrams. However, this is not the case as the following easy argument shows. Since the mapping class group of a closed surface is generated by Dehn twists, it follows from the discussion in Section 4.3 that every mapping class can be realized as the boundary monodromy of an achiral Lefschetz fibration over D^2. But wrinkling and merging as in Example 4.7 turns such an achiral Lefschetz fibrations into a descending simple wrinkled fibration while preserving the boundary fibration. Hence, every

mapping class appears as the boundary monodromy of a descending simple wrinkled fibration over D^2 and therefore also as the monodromy of a surface diagram by Proposition 5.12.

Chapter 6

Simple Wrinkled Fibrations over the Disk and the Sphere

We continue where we left off in Section 4.4 and return to simple wrinkled fibrations over D^2 and S^2. By Theorems 1.1 and 1.2 these correspond to honest surface diagrams as defined in Definition 4.11 and, in fact, we will not encounter any generalized surface diagrams anymore. Throughout Chapter 5, we have gained some valuable insights into the interactions of simple wrinkled fibrations, surface diagrams, and the topology of 4–manifolds. Here we summarize these in the context of simple wrinkled fibrations over D^2 and S^2.

6.1 Handle Decompositions

Let $w\colon Z \to D^2$ be a descending simple wrinkled fibration with surface diagram $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$. According to Theorem 1.2, we can identify (Z, w) with the fibration $(Z_{\mathfrak{S}}, w_{\mathfrak{S}})$ constructed in Lemma 5.8. Recall that $Z_{\mathfrak{S}}$ was built by attaching fold handles to $\Sigma \times D^2$ along copies of the c_i in boundary fibers cyclically ordered around S^1. If we choose a handle decomposition of Σ (as usual, with unique 0– and 2–handles), then $Z_{\mathfrak{S}}$ becomes a handlebody, and an identification of (Z, w) with $(Z_{\mathfrak{S}}, w_{\mathfrak{S}})$ gives rise to a handle decomposition of Z. These handle decompositions consist of:

- A 0–handle and $2g$ 1–handles coming from (the handle decomposition of) Σ where g is the genus of Σ
- A 2–handle coming from Σ which we will call the *fiber 2–handle*
- l further 2–handles, namely the *fold handles*

In particular, the total spaces of descending simple wrinkled fibrations have handle decompositions with handles of index at most two and can thus can be described faithfully in terms of Kirby diagrams. This observation will be very useful later on.

Now let $w\colon X \to S^2$ be a simple wrinkled fibration over S^2 with higher and lower genus fibers Σ and Σ', respectively, and let $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$ be a surface diagram for w. By removing a fibered neighborhood $\nu\Sigma'$ we obtain a descending simple wrinkled fibration on $Z = X \setminus \nu\Sigma'$, and an identification with $Z_{\mathfrak{S}}$ provides a handle decomposition of Z. We would like to complete this to a handle

decomposition of X. For that purpose, we choose a (fiber preserving) diffeomorphism $\varphi\colon \Sigma' \times D^2 \to \nu\Sigma'$ and a handle decomposition of Σ' with a unique 2–handle given by a disk $D \subset \Sigma'$. Then $D \times D^2$ can be considered as a 2–handle of a relative handle decomposition of $\Sigma' \times D^2$ on its boundary $\Sigma' \times S^1$ and the 0– and 1–handles of Σ' contribute 3–handles and a 4–handle. In this way, we obtain a handle decomposition of $X \cong Z \cup_\varphi (\Sigma' \times D^2)$. Note that, whereas the 3– and 4–handles are for the most part negligible when working with closed 4–manifolds, the additional 2–handle appearing in the passage from Z to X is important. For obvious reasons we will call it the *last 2–handle*. The last 2–handle is the more mysterious part in the handle decomposition of X as it turns out to be rather intractable in practice. Recall that if the genus of w is at least three, then the pair (X, w) is determined up to equivalence by $\mathfrak{S}$. In particular, the attaching information of the last 2–handle should be encoded in $\mathfrak{S}$. This is indeed the case but with the currently available techniques this information is hard to extract in practice. We will say more about this problem later.

For now, let us take a closer look at the last 2–handle. For simplicity, we assume that $Z = Z_{\mathfrak{S}}$ so that Z is obtained from $\Sigma \times D^2$ by attaching fold handles. Using the notation above, the attaching curve of the last 2–handle is given by

$$\kappa' = \varphi(\{p\} \times S^1) \subset \partial Z$$

where $p \in D$ is an interior point and its framing is determined by the choice of a tangent vector $v \in T_p\Sigma'$. Since the order of 2–handle attachments is interchangeable we could also attach the last 2–handle directly to $\Sigma \times D^2$. In order to see the attaching curve in $\Sigma \times S^1$, we note that κ' is a section of $w|_{\partial Z}$ and appeal to the following lemma.

Lemma 6.1. *Let $w\colon W \to A$ be an annular simple wrinkled fibration. Then any section of w defined over one of the boundary components of A, say $\gamma_\pm\colon \partial_\pm A \to \partial_\pm W$, extends to a section $\Gamma_\pm\colon A \to W$. Moreover, any normal framing of $\gamma_\pm$ in $\partial_\pm W$ extends to a normal framing of $\Gamma_\pm$ in W.*

The idea for the proof is to push the section through the cobordism using parallel transport. We postpone the details for a moment and continue with our discussion of the last 2–handle. As a consequence of Lemma 6.1, we can extend κ' to a section of the annular simple wrinkled fibration given by the fold handle attachments to $\Sigma \times D^2$ and thus get a section $\tilde{\kappa}\colon S^1 \to \Sigma \times S^1$ which in turn can be considered as a map

$$\kappa\colon S^1 \longrightarrow \Sigma.$$

Moreover, if κ' was framed by a tangent vector $v \in T_p\Sigma'$, then the construction produces a lift $\boldsymbol{\kappa}\colon S^1 \to T\Sigma$. However, observe that we are not claiming that $\tilde{\kappa}$ or κ are uniquely determined by κ'.

Definition 6.2 (Closing curves)**.** Let $w\colon X \to S^2$ be a simple wrinkled fibration with surface diagram $\mathfrak{S}$. Any curve $\kappa\colon S^1 \to \Sigma$ (or $\boldsymbol{\kappa}\colon S^1 \to T\Sigma$) obtained by the above procedure is called a *(framed) closing curve* for $\mathfrak{S}$.

To summarize, the 2–skeleton of a handle decomposition of X is completely determined by $\mathfrak{S}$ and a framed closing curve. This statement is true regardless

of the genus of $\mathfrak{S}$. So one way to understand the last 2–handles would be to find framed closing curves for surface diagrams with trivial monodromy. As explained by Hayano [35, Remark 6.6], the problem of finding closing curves can be attacked by mapping class group techniques and his methods can be adapted to obtain framings as well. However, this process is usually hard to carry out in practice and we will not make use of it. In Remark 6.6 we explain a different approach for finding the last 2–handle based on Kirby calculus.

We now give the pending proof of Lemma 6.1.

Proof of Lemma 6.1. Let $\gamma_\pm\colon \partial_\pm A \to \partial_\pm W$ be sections of w. Let $\mathcal{H}$ be a horizontal distribution for w. We choose a parametrization $\alpha\colon S^1 \times I \to A$ and consider the arcs $I_\theta = \alpha(\{\theta\} \times I)$. We denote by $V_\pm^\theta \subset \partial_\pm W$ the vanishing sets of I_θ and we let $V_\pm = \cup_\theta V_\pm^\theta$. Observe that if the image of $\gamma_\pm$ is disjoint from $V_\pm$, then the parallel transport with respect to $\mathcal{H}$ immediately provides the desired extensions of $\gamma_\pm$ to a section of w over all of A. It is therefore enough to show that one can always find $\mathcal{H}$ and α such that $V_\pm$ does not meet $\gamma_\pm$.

For that purpose we first choose $\mathcal{H}$ arbitrarily and for α we require the same properties as in the proof of Lemma 5.9. Then each I_θ contains a unique critical value and is either transverse to a fold arc or passes through a cusp. In the former case, the structure of the vanishing sets is determined in Lemma 3.13: V_+^θ is a simple closed curve (the fold vanishing cycle of I_θ) and V_-^θ consists of two points. If I_θ passes through a cusp, one can argue in two ways. Either one determines the vanishing sets in the cusp model as indicated in Remark 3.21 or one uses Lekili's unsinking move to reduce the problem to the known cases of folds and Lefschetz singularities. We choose the latter approach.

Let $\beta_w\colon W \to A$ be the broken Lefschetz fibration obtained from unsinking all the cusps. Since the unsinking deformations can be chosen with supports in arbitrary small neighborhoods of the cusps, it suffices to extend $\gamma_\pm$ to sections of β_w. Now, with respect to β_w any I_θ has one transverse intersection with the fold locus while for finitely many θ it will also pass through a Lefschetz singularity. In the latter situation the structure of the vanishing sets is as follows. We consider an intermediate fiber along I_θ whose image lies between the Lefschetz and the fold point and let V_0^θ be the union of the vanishing sets of the critical points. As before, the fold point contributes a fold vanishing cycle in V_0^θ and two points in V_-^θ. According to Remark 3.22, the Lefschetz point gives rise to a Lefschetz vanishing cycle in both V_+^θ and V_0^θ. Moreover, we can assume that the fold and Lefschetz vanishing cycles in V_0^θ are transverse (possibly after a small perturbation of $\mathcal{H}$ and appealing to Lemma 3.10). Depending on whether the fold and vanishing cycles intersect in V_0^θ, the parallel transport along I_θ takes the fold vanishing cycle either to a simple closed curve in V_+^θ disjoint from the Lefschetz vanishing cycle or a collection of open arcs that limit to the Lefschetz vanishing cycle. Similarly, the Lefschetz vanishing cycle of V_0^θ appears in V_-^θ as a simple closed curve or a collection of arcs connecting the two points coming from the folds. However, we do not need to know the precise nature of the vanishing sets. What is important is that they are certainly small enough to be made disjoint from $\gamma_\pm$ (which meets each fiber in a single point) by a fiber preserving isotopy which can be realized by modifying $\mathcal{H}$. Finally, it follows from Lemma 3.10 that any fiber preserving ambient isotopy of $V_\pm$ in $\partial_\pm W$ can be realized by a suitable change of $\mathcal{H}$. □

6.2 Drawing Kirby Diagrams

Our next goal is to describe convenient methods to draw Kirby diagrams for the handle decompositions discussed in the previous section. Since the handle decompositions have the same structure as those coming from Lefschetz fibrations (except for the framings), the same holds Kirby diagrams. A discussion in the Lefschetz setting can be found in [32, Ch. 8.2] (see also [5]). As mentioned in Section 2.1, we use the "dotted circle notation" for 1–handles in order to have well-defined framing coefficients. By a slight abuse of notation we will sometimes blur the distinction between handles and their representations in Kirby diagrams.

Simple Wrinkled Fibrations over the disk

Let $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$ be a surface diagram of genus g and let $(Z_{\mathfrak{S}}, w_{\mathfrak{S}})$ be the associated descending simple wrinkled fibration over the disk. Since $Z_{\mathfrak{S}}$ is obtained by attaching fold handles to $\Sigma \times D^2$, in order to draw a Kirby diagram for $Z_{\mathfrak{S}}$, we need a one for $\Sigma \times D^2$ in which the boundary fibers are as clearly visible as possible. We will use two different diagrams for $\Sigma \times D^2$, one is more useful for abstract reasoning while the other is better suited for Kirby calculus – we call these the *tactical* and the *practical* diagrams , see Figures 18 and 19. Both are obtained from the following recipe:

- Choose a handle decomposition of Σ with unique 0– and 2–handles.
- Embed the 1–skeleton Σ° (that is, the union of 0– and 1–handles) into $\mathbb{R}^3$.
- Put a dotted circle around each 1–handle.
- Decorate the boundary of Σ° with a 0.

Then each dotted circle represents a (4–dimensional) 1–handle in the product handle decomposition of $\Sigma \times D^2$ and the fiber 2–handle is attached along $\partial\Sigma^\circ$ with the 0–framing. In such a diagram a boundary fiber $\Sigma \times \{\theta\}$, $\theta \in S^1$, can be visualized as Σ° capped off with the core of the fiber 2–handle. Moreover, by thickening Σ° in $\mathbb{R}^3$ we obtain an interval worth of boundary fibers, say $\Sigma \times \Theta$ where $\Theta \subset S^1$ is an interval that contains the interval $[\theta_1, \theta_l]$ needed for the fold handle attachments.[1]

The tactical approach. The first diagram for $\Sigma \times D^2$, shown in Figure 18, is induced from the obvious embedded 1–skeleton Σ° of Σ. Figure 18 also shows a symplectic basis $a_1, b_1, \dots, a_g, b_g$ for $H_1(\Sigma)$, oriented such that $\langle a_i, b_i \rangle_\Sigma = 1$. As for orientations, we require that the orientation of the fiber agrees with the standard orientation of the plane so that, according to the fiber first convention, the positive S^1–direction points out of the paper toward the reader.

It is now easy to locate the attaching curves of the fold handles in Figure 18. Assuming that the vanishing cycles $c_i \subset \Sigma$ are disjoint from the 2–handle of Σ, which can always be achieved, we draw c_1 in one surface layer in Figure 18, then go to a higher layer to draw c_2, and so on. To complete the Kirby diagram for $Z_{\mathfrak{S}}$ it remains to determine the framing coefficients of the fold handles. At this point it

[1] It does not seem to be possible to see the full circle of fibers in Figure 18. However, this is not necessary for our purposes.

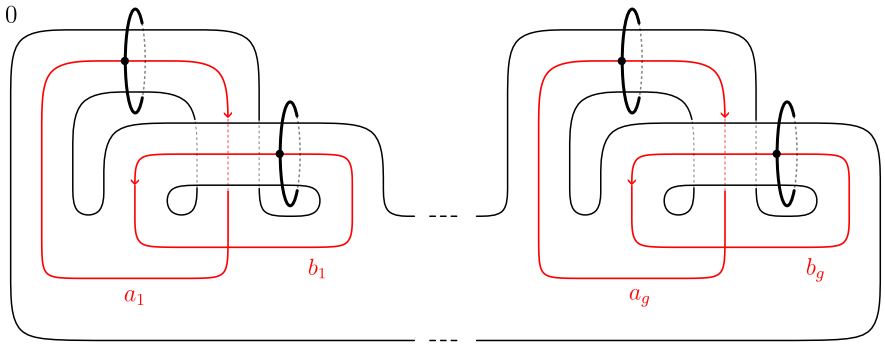


Figure 18: The "tactical diagram" $\Sigma \times D^2$ where fiber and blackboard framing agree. The red curves show a basis for $H_1(\Sigma)$.

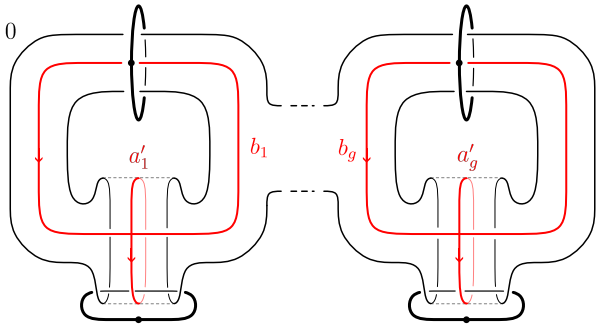


Figure 19: The "practical diagram" for $\Sigma \times D^2$.

comes in handy that the fiber framing equals the blackboard framing in Figure 18. The framing coefficient of latter can be obtained by a count of crossings, which turns out to be expressible in terms of intersection data in Σ. In a similar fashion we can also describe the linking numbers for two curves in different fibers. To state the result we denote the cohomology class of an oriented simple closed curve $c \subset \Sigma$ expressed in terms of the basis $\{a_i, b_i\}$ by

$$[c] = \sum_{i=1}^{g} \left(n_{a_i}(c)\, a_i + n_{b_i}(c)\, b_i \right) \in H_1(\Sigma).$$

Remark 6.3. There are two other useful interpretations of $n_{a_i}(c)$ and $n_{b_i}(c)$.

(i) Since the basis $\{a_i, b_i\}$ of $H_1(\Sigma)$ satisfies $\langle a_i, b_j \rangle_\Sigma = \delta_{ij}$ we have $n_{a_i}(c) = \langle c, b_i \rangle_\Sigma$ and $n_{b_i}(c) = \langle a_i, c \rangle_\Sigma$.

(ii) Let A_i and B_i be the dotted circles in Figure 18 that link with a_i and b_i, respectively. If we orient A_i such that $\mathrm{lk}(A_i, a_i) = 1$ and draw $c \subset \Sigma$ in Figure 18, then we have $n_{a_i}(c) = \mathrm{lk}(c, A_i)$ and similar arguments apply to $n_{b_i}(c)$.

Lemma 6.4. *Let $c \subset \Sigma$ be a simple closed curve and let $\theta \in \Theta \subset S^1$. Then the framing coefficient for the fiber framing of $c_\theta = c \times \{\theta\} \subset \Sigma_g \times S^1$ in Figure 18 is given by*

$$\mathrm{ff}(c_\theta) = \sum_{i=1}^{g} n_{a_i}(c)\, n_{b_i}(c). \tag{6.1}$$

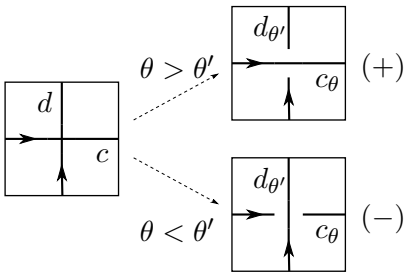


Figure 20: A positive intersection of c and d in Σ turns into a positive or negative crossing of c_θ and $d_{\theta'}$ depending on θ and θ'.

Furthermore, if $c, d \subset \Sigma$ are oriented simple closed curves and $\theta' \neq \theta \in \Theta$, then the linking number of c_θ and $d_{\theta'}$ in Figure 18 is

$$\mathrm{lk}(c_\theta, d_{\theta'}) = \frac{1}{2}\left(\epsilon_{\theta\theta'}\langle c, d\rangle_\Sigma + \sum_{i=1}^{g}\left[n_{a_i}(c)\, n_{b_i}(d) + n_{a_i}(d)\, n_{b_i}(c)\right]\right) \tag{6.2}$$

where $\epsilon_{\theta\theta'}$ denotes the sign of $\theta - \theta'$.

Note that the right hand side of equation (6.1) requires the choice of an orientation of c but both orientations result in the same number.

Proof. In order to compute the framing coefficient $\mathrm{ff}(c_\theta)$ we first push $c \subset \Sigma$ off the 2–handle of Σ if necessary so that c_θ is completely visible in Figure 18. Since the fiber framing and blackboard framing of c_θ agree, the framing coefficient is given by the *writhe* of c_θ in the diagram, that is, the signed count[2] of crossings with some chosen orientation (see [32, Proposition 4.5.8]). From the way the diagram is drawn it is clear that each crossing of c_θ is caused by c running over both a_i and b_i for some i so that the writhe of c_θ is given by the right hand side of equation (6.1).

The statement about linking numbers follows from a similar count of crossings. Again, if necessary we first push c and d off the 2–handle of Σ and we also make them transverse. Recall that the linking number of two oriented knots in S^3 can be computed from any link diagram as half of the signed number of crossings (see [12, p.63], for example). The second term on the right hand side of equation (6.2) arises just as above. The first term can be understood as follows. Each intersection point of c and d in Σ_g contributes a crossing in the diagram. Now, the sign of the crossing depends on two things: the sign of the intersection point in Σ and the information which strand is on top in the diagram. From Figure 20 we see that the contribution of each crossing is as in equation (6.1). □

Example 6.5. Consider the curves $c = a_1$ and $d = b_1$. Then equation (6.1) shows that $\mathrm{ff}(c) = \mathrm{ff}(d) = 0$. Of course, this is also clear from Figure 18. More interestingly, let us look at the curves $\tau_c^k(d)$ for $k \in \mathbb{Z}$ whose framing coefficients in Figure 18 are not that obvious. By the Picard–Lefschetz formula (Proposition 2.15) we have $[\tau_c^k(d)] = [d] + k[c] = k[a_1] + [b_1]$ and equation (6.1) shows that $\mathrm{ff}(\tau_c^k(d)) = k$. This example will be useful later on since we will frequently

[2]Our sign convention for crossings is illustrated in Figure 20.

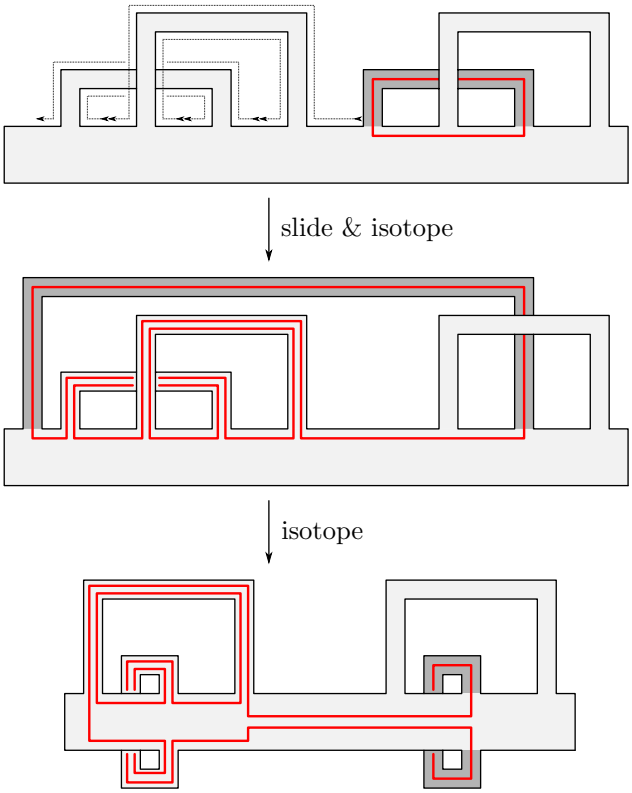


Figure 21: Understanding the relation between Figures 18 and 19.

normalize arbitrary pairs of simple closed curves $c, d \subset \Sigma$ that intersect transversely in one point to $(c, d) = (a_1, b_1)$ using Proposition 2.17.

Returning to the surface diagram $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$ and the corresponding 4–manifold $Z_{\mathfrak{S}}$, the upshot of Lemma 6.4 and Remark 6.3 is that Figure 19 provides an interface between linking information in a Kirby diagram for $Z_{\mathfrak{S}}$ and intersection numbers in Σ. This will be further exploited in Section 9.1.2 where we describe the intersection form of $Z_{\mathfrak{S}}$ purely in terms of $\mathfrak{S}$.

The practical approach. Although the tactical Kirby diagrams of simple wrinkled fibrations based on Figure 18 are useful for abstract reasoning, they tend to get quite confusing when one actually tries to draw them. This makes them not very well suited as a starting point for Kirby calculus. For practical purposes, it is better to start with a cleaner diagram for $\Sigma \times D^2$ such as the one shown in Figure 19. As indicated, this diagram is obtained from a different embedded 1–skeleton $\Sigma^\bullet$ of Σ. A symplectic basis $\{a_i', b_i\}$ for $H_1(\Sigma)$ appears again as meridians for the dotted circles. One can also think of $\Sigma^\bullet$ as the obvious spanning disk of the unknot representing the fiber 2–handle, surgered to avoid the punctures by the dotted circle on the bottom such that the tubes resulting from the surgeries wrap around the upper halves of the dotted circles. This is interpretation might be helpful since we typically draw the diagrams as in Figure 23 where the fiber 2–handle has been slightly isotoped and the fiber is a little harder to spot.

Locating the attaching curves of the fold handles works exactly as in the tac-

tical approach. However, in Figure 19 the fiber framing does not agree with the blackboard framing. In order to understand the framing coefficients, we observe that Figure 19 can be obtained from Figure 18 by a series of 1–handles slides. The reason is that the underlying embedded 1–skeleta Σ° and $\Sigma^\bullet$ of Σ can be related by a sequence of embedded 1–handle slides as shown in Figure 21 for the genus 2 case. There it is also shown how the curve a_2 is dragged along throughout the slides. This picture easily generalizes to arbitrary genus and the translation to the 4–dimensional context is standard. The important observation is that the 1–handles slides do not affect b_i while a_i appears in Figure 19 as a curve freely homotopic to

$$a_i' * [b_{i-1}, a_{i-1}'] * \cdots * [b_1, a_1'] \in \pi_1(\Sigma)$$

where $[x, y] = xyx^{-1}y^{-1}$. In particular, a_i is homologous to a_i' in Σ. Now we can compute the framing coefficients in Figure 19 as follows. Suppose that we have located in curve c_θ in Figure 18 and we compute its fiber framing using equation (6.1). We can then locate c_θ in Figure 19 by dragging it around throughout the 1–handle slides. The latter do not affect the framing coefficient so that equation (6.1) continues to hold. But since a_i and a_i' are homologous in Σ the framing coefficient of c_θ in Figure 19 can be computed by the same formula as in equation (6.1) with a_i replaced by a_i', that is,

$$\mathrm{ff}(c_\theta) = \sum_{i=1}^{g} n_{a_i}(c)\, n_{b_i}(c) = \sum_{i=1}^{g} n_{a_i'}(c)\, n_{b_i}(c).$$

Moreover, the numbers $n_{a_i'}$ and n_{b_i} can be interpreted as linking numbers as in Remark 6.3.

Locating the Last 2–Handle

Now suppose that $\mathfrak{S} = (\Sigma; c_1, \ldots, c_l)$ has trivial monodromy. Suppose that we are also given a framed closing curve $\boldsymbol{\kappa}\colon S^1 \to T\Sigma$ and let $\kappa\colon S^1 \to \Sigma$ be its projection to Σ. Recall that the last 2–handle is attached along the section of $\Sigma \times S^1$ given by $\tilde{\kappa}(\theta) = (\kappa(\theta), \theta)$ and the framing is determined by the normal vector field induced by $\boldsymbol{\kappa}$. So in order to locate the attaching curve of the last 2–handle in Figures 18 and 19, we have to understand how sections of $\Sigma \times S^1$ appear. The method explained below works for both pictures.

Note that there is a canonical bijection between sections of $\Sigma \times S^1$ and maps from S^1 into Σ induced by the projection onto Σ. A distinguished homotopy class of sections are those of the form $\{*\} \times S^1$ corresponding to the constant loop based at $* \in \Sigma$. One such constant section is given by the belt circle of the fiber 2–handle in $\Sigma \times D^2$ which is isotopic to a meridian of the attaching curve. In other words, a meridian of the attaching curve of the fiber 2–handle in the Kirby diagram represents a constant section. We fix such a meridian $\mathfrak{m}$ and the corresponding constant section $\{*\} \times S^1$. Now, for an arbitrary section such as $\tilde{\kappa}$, which projects to κ, we can assume that $\kappa(\theta) = *$ for some $\theta \in S^1$. We start at the point corresponding to $(*, \theta) \in \Sigma \times S^1$ and travel once around the meridian and then follow the copy of κ in $\Sigma \times \{\theta\}$. The resulting curve in $\Sigma \times S^1$ is homotopic to $\tilde{\kappa}$ and we can recover $\tilde{\kappa}$ up to isotopy by resolving the self-intersections as follows. The intersection of the

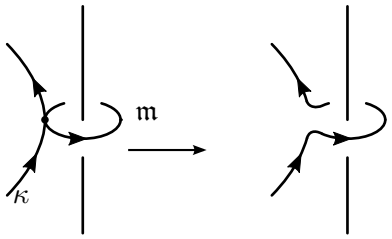


Figure 22: Resolving a crossing of a closing curve and a meridian for the fiber 2–handle.

meridian and κ is resolved according to the orientations as shown in Figure 22, and at each double point of κ we push the strand that is traversed first to a lower θ–level in the diagram. Finally, the framing coefficient of the last 2–handle can be computed by hand in concrete examples. One simply has to push the attaching curve off itself in the direction of the normal vector field determined by $\boldsymbol{\kappa}\colon S^1 \to T\Sigma$ and compute the linking number.

However, instead of starting with a framed closing curve, one can also try to find possible attaching curves for the last 2–handle directly using Kirby calculus. The following method is explained in the context of Lefschetz fibrations in [32, p.299f.] and we refer there for further details.

Remark 6.6 (The last 2–handle via Kirby calculus). Suppose that we have a Kirby diagram for $Z_{\mathfrak{S}}$. By trading all dotted circles for 0–framed unknots we obtain a surgery diagram for $\partial Z_{\mathfrak{S}}$. We then modify this diagram using 3–dimensional Kirby moves until we arrive at a known diagram for $\Sigma' \times S^1$ in which we can recognize the circle $\{*\} \times S^1$ and its product framing. For example, we know that in Figure 18 or 19 this framed circle appears as a 0–framed meridian of the fiber 2–handle. Finally, we pull the framed circle back to the original diagram by dragging it along while we undo the Kirby moves again. Since we have the a priori information that $\partial Z_{\mathfrak{S}}$ and $\Sigma' \times S^1$ are diffeomorphic, by Kirby's theorem[3] this strategy always works. However, one should not expect that a suitable sequence of Kirby moves will be easy to find in general.

6.3 Some Examples

We consider some simple examples of surface diagrams with trivial monodromy, mainly in order to illustrate our methods for drawing Kirby diagrams.

6.3.1 Surface Diagrams with Two Curves

We begin by studying the simplest possible surface diagrams, namely those that contain only two curves. Up to equivalence there is a unique such diagram for each genus $g \geq 1$; this follows from the change of coordinates principle (Proposition 2.17). We can thus fix our favorite model surface Σ_g and our favorite pair of curves $a, b \subset \Sigma_g$ that intersect transversely in one point and restrict our attention to the surface diagram $\mathfrak{S} = (\Sigma_g; a, b)$ shown in Figure 23. To demonstrate the methods developed

[3] See Theorem 5.3.6 in [32], for example.

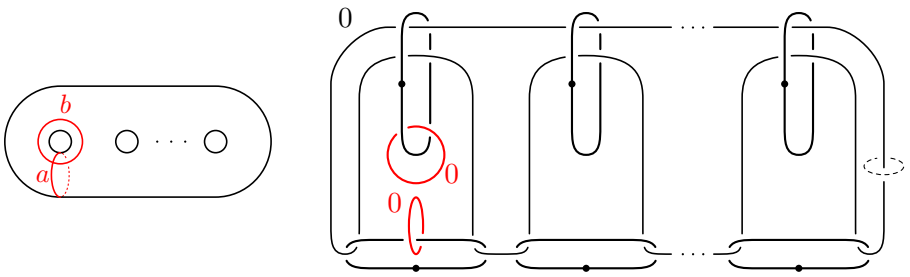


Figure 23: The surface diagram $\mathfrak{S} = (\Sigma_g; a, b)$ and a Kirby diagram for $Z_{\mathfrak{S}}$. The dashed meridian for the fiber 2–handle is an attaching curve for the last 2–handle.

in Section 5.2, let us compute the monodromy. We have

$$\mu(\Sigma; a, b) = \Phi_a\big(\tau_b\tau_a\tau_a\tau_b\big) = \Phi_a\big(\tau_a^{-1}\tau_a\tau_b\tau_a\tau_a\tau_b\tau_a\tau_a^{-1}\big) = \Phi_a\big(\tau_a^{-1}\Delta_{a,b}\tau_a^{-1}\big) = 1$$

where the last equality uses Lemma 5.15. It follows that $\mathfrak{S}$ has trivial monodromy and thus arises as the surface diagram of a simple wrinkled fibration over S^2. However, it turns out that the above calculation was not really necessary since the Kirby diagram in Figure 23 immediately shows that $Z_{\mathfrak{S}}$ is diffeomorphic to $\Sigma_{g-1} \times D^2$. Indeed, we can simply unlink the fiber 2–handle from the two left most 1–handles by sliding over the fold handles and then cancel the 1–handles against the fold handles. What remains is the genus $g-1$ version of Figure 19. In particular, we have also succeeded in carrying out half of the procedure for finding the last 2–handle explained in Remark 6.6. In the diagram for $\Sigma_{g-1} \times D^2$ the last 2–handle can be attached along a 0–framed meridian of the fiber 2–handle, resulting in the double of $\Sigma_{g-1} \times D^2$, that is, $\Sigma_{g-1} \times S^2$. Moreover, we can take such a meridian in a region that was not affected by the Kirby moves so that it pulls back to a 0–framed meridian for the fiber 2–handle in the original diagram for $Z_{\mathfrak{S}}$.

Altogether, we see that $(\Sigma_g; a, b)$ with our choice of last 2–handle attachment describes the closed 4–manifold $\Sigma_{g-1} \times S^2$, that is, the trivial Σ_{g-1}–bundle over S^2. In fact, we should have seen this coming. Recall from Example 4.6 that Σ_{g-1} bundle over S^2 can be turned into a genus g simple wrinkled fibration with two cusps by performing a birth homotopy. The corresponding surface diagram has two curves and is thus equivalent to $(\Sigma_g; a, b)$. In particular, the non-trivial S^2– and T^2–bundles over S^2 have the same surface diagrams which shows that the ambiguity of closing off really matters in low genus situations.

6.3.2 Doubles

Let Σ be an oriented surface and let $\Gamma = (c_1, \dots, c_l)$ be a W–chain in Σ (not necessarily a W–cycle). Then we can form a W–cycle

$$D\Gamma = (c_1, \dots, c_{l-1}, c_l, c_{l-1}, \dots, c_2)$$

which we call the *double* of Γ.

Lemma 6.7. *The surface diagram $(\Sigma; D\Gamma)$ has trivial monodromy.*

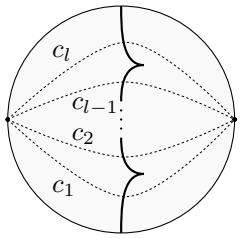


Figure 24: The critical image of $w_\Gamma\colon P_\Gamma \to D^2$

Proof. For convenience write $\tau_i = \tau_{c_i}$. As explained in Remark 5.17, the monodromy of $(\Sigma, D\Gamma)$ can be lifted from $\mathrm{Mod}(\Sigma_{c_1})$ to $\mathrm{Mod}(\Sigma)$ as

$$\begin{aligned}\tilde{\mu} &= (\tau_2\tau_1\tau_2)\dots(\tau_{l-2}\tau_{l-1}\tau_{l-2})(\tau_{l-1}\tau_l\tau_{l-1})(\tau_l\tau_{l-1}\tau_l)(\tau_{l-1}\tau_{l-2}\tau_{l-1})\dots(\tau_1\tau_2\tau_1)\\ &= (\tau_2\tau_1\tau_2)\dots(\tau_{l-2}\tau_{l-1}\tau_{l-2})\Delta_{c_{l-1},c_l}(\tau_{l-1}\tau_{l-2}\tau_{l-1})\dots(\tau_1\tau_2\tau_1)\end{aligned}$$

that is, we have $\Phi_{c_1}(\tilde{\mu}) = \mu(\Sigma; D\Gamma)$. Our goal is to factor the above expression into a sequence of Δ–twists involving c_1 which are annihilated by Φ_{c_1} according to Lemma 5.15. The key observation is that

$$\begin{aligned}&(\tau_{l-2}\tau_{l-1}\tau_{l-2})\ \Delta_{c_{l-1},c_l}\ (\tau_{l-1}\tau_{l-2}\tau_{l-1})\\ =&(\tau_{l-2}\tau_{l-1}\tau_{l-2})\ \Delta_{c_{l-1},c_l}\ (\tau_{l-2}\tau_{l-1}\tau_{l-2})\\ =&(\tau_{l-2}\tau_{l-1}\tau_{l-2})\ \Delta_{c_{l-1},c_l}\ (\tau_{l-2}\tau_{l-1}\tau_{l-2})^{-1}\ \Delta_{c_{l-2},c_{l-1}}\\ =&\Delta_{\tau_{l-2}\tau_{l-1}\tau_{l-2}(c_{l-1}),\tau_{l-2}\tau_{l-1}\tau_{l-2}(c_l)}\ \Delta_{c_{l-2},c_{l-1}}\\ =&\Delta_{c_{l-2},\tau_{l-2}\tau_{l-1}\tau_{l-2}(c_l)}\ \Delta_{c_{l-2},c_{l-1}}.\end{aligned}$$

Applying this repeatedly, we eventually obtain

$$\tilde{\mu} = \Delta_{c_1,\delta_l}\Delta_{c_1,\delta_{l-1}}\dots\Delta_{c_1,\delta_2}$$

where $\delta_k = \tau_1\tau_2\tau_1\dots\tau_{k-2}\tau_{k-1}\tau_{k-2}(c_k)$. □

As a consequence, $(\Sigma; D\Gamma)$ describes a closed 4–manifold together with a simple wrinkled fibration over S^2. However, this can also be seen directly. In fact, we now show that $(\Sigma, D\Gamma)$ arises as the surface diagram of a simple wrinkled fibration over S^2 defined on the double of a compact 4–manifold P_Γ associated to Γ.

The construction of P_Γ follows a familiar pattern. We take $\Sigma \times D^2$, attach fold handles to $c_i \times \{\theta_i\}$ with $\theta_i \in S^1$ ordered according to the orientation, and extend the projection $\Sigma \times D^2 \to D^2$ across the fold handles as in the second step of the proof of Lemma 5.8. The result is an oriented 4–manifold P_Γ together with a wrinkled fibration over disk which restricts to a circle valued Morse function on ∂P_Γ. Since Γ is a W–chain, we can proceed as in third step of the proof of Lemma 5.8 and trade critical points on the boundary for cusps in the interior. However, since we do not assume that Γ is a W–cycle, there is one pair of critical points that cannot be canceled, namely the index 1 point of the lth fold handle and the index 2 point of the first fold handle. Altogether we obtain a wrinkled fibration $w_\Gamma\colon P_\Gamma \to D^2$ with critical image as in Figure 24. Now we simply double this map, that is, we define

$$\widehat{w}_\Gamma = w_\Gamma \cup w_\Gamma\colon DP_\Gamma = P_\Gamma \cup_{\mathrm{id}} \overline{P_\Gamma} \longrightarrow D^2 \cup_{\mathrm{id}} \overline{D^2} \cong S^2$$

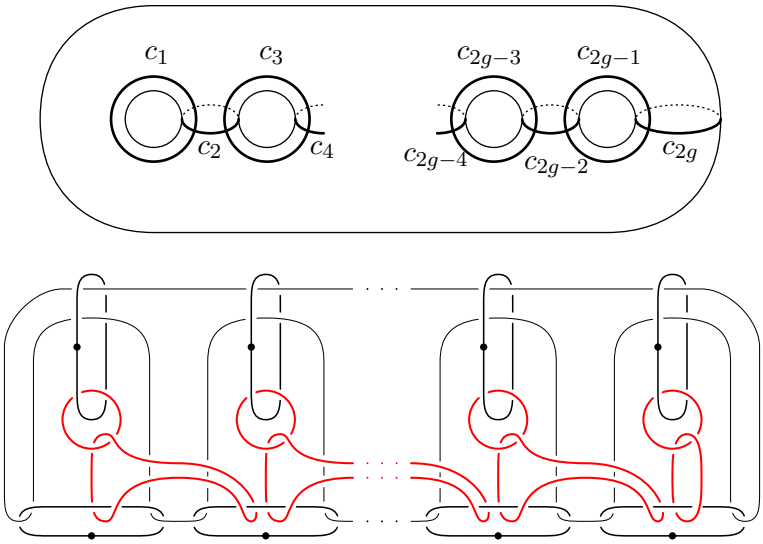


Figure 25: The non-separating chain $\mathrm{Ch}_{g,2g}$ and the 4–manifold $P_{\mathrm{Ch}_{g,2g}}$. (All 2–handles are 0–framed and the fold handles are drawn in red.)

and it is clear from the construction that this is a simple wrinkled fibration. Indeed, the singular arcs of the two copies of w_Γ piece together to one singular circle in the double which is mapped injectively into S^2. Moreover, we can choose a reference system for $\widehat{w}_\Gamma$ by taking the reference arcs in D^2 indicated in Figure 24 and their copies in $\overline{D^2}$. However, we can neglect the outermost reference arcs in $\overline{D^2}$ since the corresponding fold arcs of $\widehat{w}_\Gamma$ are already covered by the reference arcs in D^2. Of course, the vanishing cycles of the reference arcs in $\overline{D^2}$ are the same as their counterparts from D^2 but because of the reversed orientation they are recorded in reverse order. It follows that the surface diagram of $\widehat{w}_\Gamma$ is

$$(\Sigma; c_1, \dots, c_{l-1}, c_l, c_{l-1}, \dots, c_2) = (\Sigma, D\Gamma).$$

We have thus proved:

Lemma 6.8. *Let Γ be a W–chain in Σ. Then the surface diagram $(\Sigma, D\Gamma)$ obtained from doubling describes the double DP_Γ of the 4–manifold P_Γ described above.*

Using the doubling construction it is easy to write down many surface diagrams with trivial monodromy. We end this chapter by working out a concrete series of examples.

Example 6.9 (Doubles of non-separating chains)**.** Let Σ_g be an oriented surface of genus g. We consider the non-separating chains of curves in Σ_g

$$\mathrm{Ch}_{g,l} = \big(c_1, \dots, c_l\big), \quad l \leq 2g$$

with $c_i \subset \Sigma_g$ as in Figure 25, and the doubled surface diagrams

$$\mathfrak{Ch}_{g,l} = \big(\Sigma_g; D\mathrm{Ch}_{g,l}\big) = \big(\Sigma_g; c_1, \dots, c_l, c_{l-1}, \dots, c_2\big).$$

For convenience, we assume $g \geq 3$ so that $\mathfrak{Ch}_{g,l}$ unambiguously describes the double of the 4–manifold $P_{\mathrm{Ch}_{g,l}}$ constructed above. We claim that

$$P_{\mathrm{Ch}_{g,l}} \cong \begin{cases} \Sigma_{g-k} \times D^2 & \text{if } l = 2k \\ \Sigma_{g-k} \times D^2 \natural S^1 \times D^3 & \text{if } l = 2k-1 \end{cases} \tag{6.3}$$

which can be seen using Kirby calculus. Figure 25 also shows a Kirby diagram of $P_{\mathrm{Ch}_{g,l}}$ for the maximal chain length $l = 2g$; in the cases where $l \leq 2g$ one simply has to erase the rightmost fold handles. Note that the upper left fold handle is a 0–framed meridian for the upper left 1–handle. As indicated in Figure 26, it can be used to unlink the fiber 2–handle from the 1–handle so that the 1–handle can be canceled against the fold handle. This maneuver has the additional effect that the fiber 2–handle also becomes unlinked from the lower left 1–handle. Moreover, if $l \geq 2$ then the lower left 1–handle is linked geometrically once by the lower left fold handle and the pair can be canceled; if not, then and isolated 1–handle remains. By iterating this procedure we can eventually cancel all fold handles against 1–handles and what is left is a Kirby diagram of $\Sigma_{g-k} \times D^2$ (as in Figure 19) with k as in equation (6.3) and an isolated 1–handle for odd l which represents a boundary sum with $S^1 \times D^3$.

We now pass to the doubles. After some minor relabeling, we conclude that for $h \geq 2$ and $k \geq 1$ the surface diagrams $\mathfrak{Ch}_{h+k,2k}$ and $\mathfrak{Ch}_{h+k,2k-1}$ describe

$$\Sigma_h \times S^2 \quad \text{and} \quad \Sigma_h \times S^2 \# S^1 \times S^3,$$

respectively. In particular, we have found genus g surface diagrams for these manifolds for all $g \geq h+1$. Moreover, in equation (9.1) we will obtain a formula for the Euler characteristic which excludes surface diagrams of genus $g \leq h$ in the case at hand so that we have surface diagrams for every possible genus.

Remark 6.10. The surface diagrams for $\Sigma_h \times S^2$ and $\Sigma_h \times S^2 \# S^1 \times S^3$ obtained in Example 6.9 were discovered independently by Hayano [35, Example 6.9 & Remark 6.9] using different methods. Hayano's arguments are more complicated than ours but they also give more information. In fact, Hayano shows that simple wrinkled fibrations $\Sigma_h \times S^2 \to S^2$ corresponding to $\mathfrak{Ch}_{h+k,2k}$ is homotopic to the projection onto S^2 via k flip-and-slip moves. Similarly, he shows that $\mathfrak{Ch}_{h+k,2k-1}$ represents a simple wrinkled fibration $\Sigma_h \times S^2 \# S^1 \times S^3 \to S^2$ homotopic to one of the fake simple wrinkled fibrations discussed in Example 4.8. In contrast, our method are limited to identifying the total spaces of these fibrations.

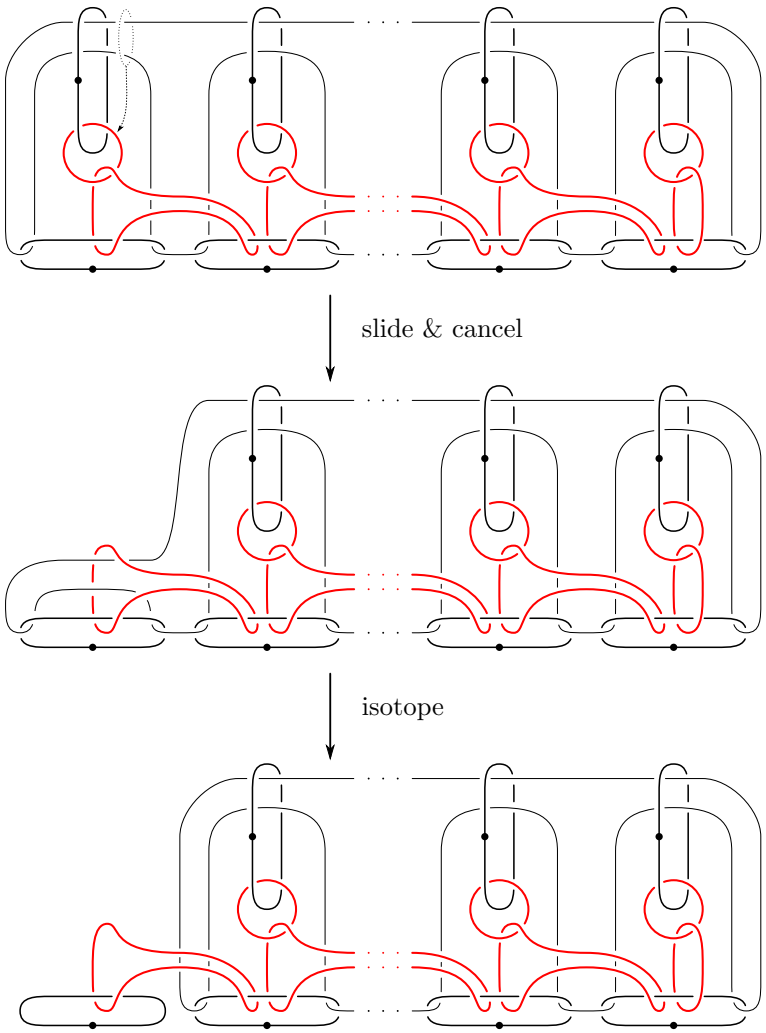


Figure 26: Identifying $P_{\mathrm{Ch}_{g,l}}$ as in equation (6.3).

Part III

The Topology of Surface Diagrams

According to Corollary 4.19, all closed 4–manifolds can be described by surface diagrams with trivial monodromy. Moreover, we saw in Chapter 6 that surface diagrams naturally encode handle decompositions of 4–manifolds. Building on this observation, we investigate further how the topology of 4–manifolds is reflected in the combinatorics of surface diagrams. Since isotopic surface diagrams give handle decompositions with isotopic attaching maps, it will be convenient to consider simple closed curves up to isotopy. Once more for emphasis:

In the remaining chapters we will blur the distinction between simple closed curves and their isotopy classes.

In Chapter 7 we study how certain cut-and-paste operations on 4–manifolds can be realized by modifying their surface diagrams. We discuss connected sums with $S^2 \times S^2$, $\mathbb{C}\mathrm{P}^2$ and $\overline{\mathbb{C}\mathrm{P}^2}$, as well as surgeries on curves and 2–spheres. Our main application of these techniques is given in Chapter 8 where we obtain a complete classification of closed 4–manifolds that admit surface diagrams of genus one, the lowest possible genus.

In the final Chapter 9 we address the question how some basic information about the homotopy type of a 4–manifold can be extracted from a given surface diagram. We give surface diagram descriptions of the fundamental group, homology, and the intersection form, see Section 9.1. As a first application, we obtain an easily accessible obstruction for a surface diagram to have trivial monodromy in Section 9.2. We then go on to discuss spin and spin^c structures in Section 9.3.

Chapter 7

Substitutions in Surface Diagrams

Let $\mathfrak{S} = (\Sigma, \Gamma)$ be a surface diagram with $\Gamma = (c_1, \dots, c_l)$ and let $(Z_{\mathfrak{S}}, w_{\mathfrak{S}})$ be the corresponding simple wrinkled fibration over the disk. Furthermore, let Λ' be an arbitrary W–chain in Σ that starts with c_r and ends with c_s for some $1 \leq r \leq s \leq 1$. If we remove $\Lambda = (c_r, \dots, c_s)$ from Γ and replace it with Λ', then the result is a new W–cycle Γ' and thus a new surface diagram $\mathfrak{S}' = (\Sigma, \Gamma')$.

Definition 7.1 (Substitutions)**.** Let $\mathfrak{S}$ and $\mathfrak{S}'$ be as above. We say that $\mathfrak{S}'$ is obtained from $\mathfrak{S}$ by a *substitution* $\Lambda \rightsquigarrow \Lambda'$.

On the 4–manifold side, a substitution can naturally be interpreted as a cut-and-past operation. Roughly, $Z_{\mathfrak{S}}$ contains a submanifold of the form P_Λ (as discussed on page 83) and $Z_{\mathfrak{S}'}$ is obtained by removing P_Λ and replacing it with $P_{\Lambda'}$. However, this observation is not very useful unless we have a good understanding of P_Λ and $P_{\Lambda'}$.

Definition 7.2. Let $\mathcal{O}$ be some operation on 4–manifolds. We say that the substitution $\Lambda \rightsquigarrow \Lambda'$ *corresponds to* $\mathcal{O}$ if $Z_{\mathfrak{S}'}$ is diffeomorphic to $\mathcal{O}(Z_{\mathfrak{S}})$.

Another problem is to keep track of the monodromy. Indeed, if $\mathfrak{S}$ has trivial monodromy and thus describes a closed 4–manifold, then we would like $\mathfrak{S}'$ to have trivial monodromy as well. Of course, this is wrong in general. For example, $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$ is obtained from the diagram $(\Sigma; c_1, c_l)$ by a substitution, and while the latter has trivial monodromy (see Section 6.3.1), we know from Remark 5.19 that the monodromy of $\mathfrak{S}$ can be any element of $\mathrm{Mod}(\Sigma_{c_1})$.

Definition 7.3. Let $\mathfrak{S}'$ be obtained from $\mathfrak{S}$ by a substitution $\Lambda \rightsquigarrow \Lambda'$ as above. We say that the substitution is *monodromy preserving* if $\mu(\mathfrak{S}') = \mu(\mathfrak{S})$.

In order to verify this property for a given substitution one can argue in two ways. On the one hand, one can simply compute the monodromies before and after the substitution and show directly that they agree. On the other hand, one can also consider the corresponding 4–manifolds and try to show that the effect of the substitution does not change the boundary. One can then deduce a posteriori that the substitution must have trivial monodromy.

In the following we will discuss some simple substitutions which correspond to well known operations on 4–manifolds. These substitutions happen to be monodromy preserving and are therefore applicable in the context of closed 4–manifolds and simple wrinkled fibrations over S^2. The central tool are the handle decompositions of simple wrinkled fibrations and their Kirby diagrams discussed in Sections 6.1 and 6.2.

Remark 7.4. Similar substitution techniques have been studied for Lefschetz fibrations by Endo et al. in [22, 23] where certain rational blowdowns are realized by the corresponding notion of substitutions. Our goals are more modest in the sense that the we discuss much more basic cut-and-paste operations. However, we will gain some first practical insights into how surface diagrams encode geometric information and it is quite conceivable that many other interesting operations can be rephrased as substitutions in surface diagrams.

7.1 Blow-Ups and Stabilizations

In the context of 4–manifolds the operations of taking connected sums with $\overline{\mathbb{CP}^2}$ and $S^2 \times S^2$ are commonly known as *blow-up* and *stabilization*. We will also use this terminology for connected sums with $\mathbb{CP}^2$ and the non-trivial S^2–bundle over S^2, which is diffeomorphic to $\mathbb{CP}^2 \# \overline{\mathbb{CP}^2}$. For convenience, we let

$$\mathcal{S}_k = \begin{cases} S^2 \times S^2, & k \text{ even} \\ \mathbb{CP}^2 \# \overline{\mathbb{CP}^2}, & k \text{ odd} \end{cases}$$

and recall that a Kirby diagram for $\mathcal{S}_k$ is given by the $(0, k)$–framed Hopf link.

Lemma 7.5 (Blow-up and stabilization)**.** *Let $\mathfrak{S} = (\Sigma; \Gamma)$ be a surface diagram and let $a, b \subset \Sigma$ be two consecutive vanishing cycles in $\mathfrak{S}$.*

(i) A substitution of the form

$$\big(a, b\big) \rightsquigarrow \big(a, \tau_a^{\pm 1}(b), b\big) \tag{7.1}$$

corresponds to a blow-up with $\pm\mathbb{CP}^2$.

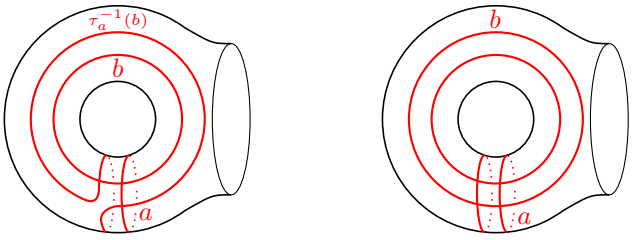
(ii) A substitution of the form

$$\big(a, b\big) \rightsquigarrow \big(a, b, \tau_b^k(a), b\big), \quad k \in \mathbb{Z} \tag{7.2}$$

corresponds to a stabilization with $\mathcal{S}_k$.

In particular, the substitutions (7.1) *and* (7.2) *are monodromy preserving.*

For obvious reasons, we will refer to the substitutions described in equations (7.1) and (7.2) as *blow-up* and *stabilization substitutions*. Of course, the substitutions are reversible and whenever a surface diagram contains a configuration as in the right hand sides of the equations, the associated 4–manifold must be a blow-up or stabilization of some other 4–manifold. These configurations are shown in Figure 27 and we call them *blow-up* and *stabilization configurations*. The following proof of Lemma 7.5 is a simplification of that given by the author in [7, Lemma 5.1] using a localization idea inspired by Hayano's proof of Lemma 7.8 below (see [35, Lemma 6.13]).



$(a, \tau_a^{-1}(b), b)$ (a, b, a, b)

Figure 27: Blow-up and stabilization configurations.

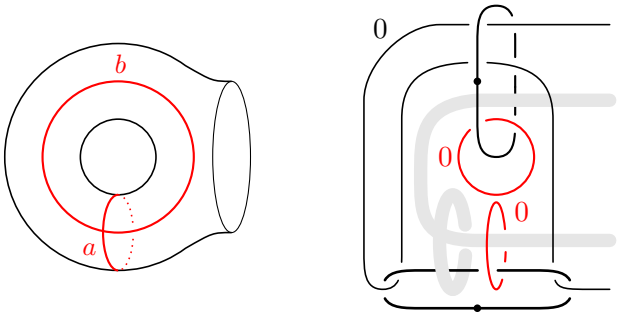


Figure 28: Excerpts of the surface diagram $\mathfrak{S} = (\Sigma; \dots, a, b)$ and a Kirby diagram for $Z_{\mathfrak{S}}$. In the latter, the omitted fold handles are supported in the shaded region.

Proof of Lemma 7.5. For simplicity we assume that a and b are the last vanishing cycles in $\mathfrak{S}$ – that is, $\mathfrak{S} = (\Sigma; \cdots, a, b)$ – and that they intersect transversely in one point. Let $T \subset \Sigma$ be a regular neighborhood of $a \cup b$ which is diffeomorphic to a one-holed torus as shown in the left of Figure 28. The right side of Figure 28 shows the corresponding excerpt of a practical Kirby diagram of $Z_{\mathfrak{S}}$ (where a and b have been normalized to the curves a_1', b_1 in Figure 19).

Let $N_{(a,b)}$ be the codimension 0 submanifold of $Z_{\mathfrak{S}}$ given by $T \times D^2$ with the fold handles attached along a and b. In the Kirby diagram in Figure 28, $N_{(a,b)}$ appears as the two dotted circles and their meridians, a cleaner picture is shown on the right of Figure 29. Obviously, $N_{(a,b)}$ is diffeomorphic to D^4. In order to prove (i) and (ii) we consider the 4–manifolds $N_{(a,\tau_a^{\pm 1}(b),b)}$ and $N_{(a,b,\tau_b^k(a),b)}$ obtained by attaching fold handles to $T \times D^2$ along the curves indicated in the subscripts, as usual in the order determined by the orientation of S^1. The blow-up and stabilization substitutions result in surface diagrams

$$\mathfrak{S}' = \big(\Sigma; \dots, a, \tau_a^{\pm 1}(b), b\big) \quad \text{and} \quad \mathfrak{S}'' = \big(\Sigma; \dots, a, b, \tau_b^k(a), b\big)$$

and it follows from the construction of $Z_{\mathfrak{S}'}$ and $Z_{\mathfrak{S}''}$ that these manifolds are obtained from $Z_{\mathfrak{S}}$ by removing $N_{(a,b)}$ and replacing it with $N_{(a,\tau_a^{\pm 1}(b),b)}$ and $N_{(a,b,\tau_b^k(a),b)}$, respectively. It remains to identify the latter pieces and for that purpose we use Kirby calculus. Recall that a blow-up with $\pm\mathbb{C}\mathrm{P}^2$ can be realized in a Kirby diagram by inserting an isolated unknot with framing ± 1, while for a stabilization with $\mathcal{S}_k$ one has to add an isolated Hopf link with framings $(0, k)$.

Figure 29 contains a Kirby diagram for $N_{(a,\tau_a^{-1}(b),b)}$ and the indicated 2–handle slides exhibit this manifold as the blow up $N_{(a,b)} \# \overline{\mathbb{C}\mathrm{P}^2}$. Note that the framing

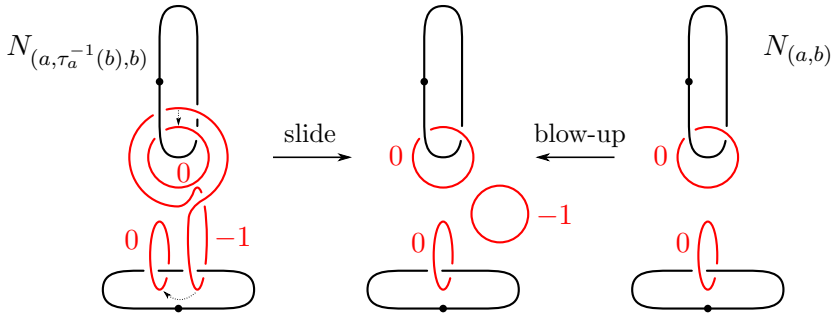


Figure 29: Understanding the blow-up substitution $(a, b) \rightsquigarrow (a, \tau_a^{-1}(b), b)$.

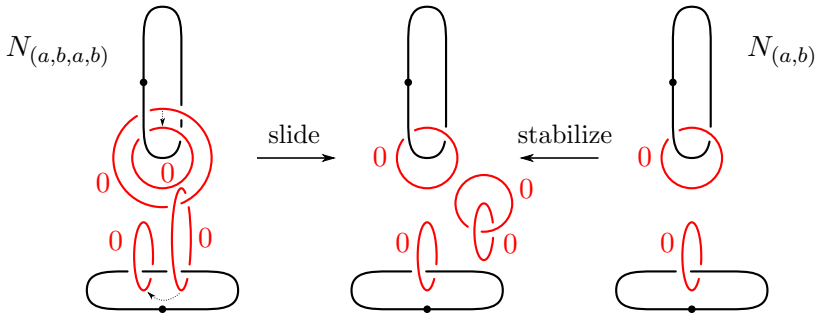


Figure 30: Understanding the stabilization substitution $(a, b) \rightsquigarrow (a, b, a, b)$.

coefficient of the fold handle corresponding to $\tau_a^{-1}(b)$ is -1 according to Example 6.5. For $N_{(a,\tau_a(b),b)}$ one finds a very similar Kirby diagram where $\tau_a(b)$ contributes a fold handle with framing coefficient $+1$ and the same arguments show that $N_{(a,\tau_a(b),b)} \cong N_{(a,b)} \# \mathbb{CP}^2$. It follows that the substitution $(a, b) \rightsquigarrow (a, \tau_a^{-1}(b), b)$ corresponds to a blow-up with $\pm\overline{\mathbb{CP}^2}$ and we succeeded in proving (i).

A proof of (ii) can be obtained similarly. In order to keep the Kirby diagrams simple, we only treat the case $k = 0$, that is, the substitution $(a, b) \rightsquigarrow (a, b, a, b)$. A Kirby diagram for $N_{(a,b,a,b)}$ is shown in Figure 30, and the indicated handle slides result in the diagram for $N_{(a,b)}$ alongside a $(0, 0)$–framed Hopf link. It follows that $N_{(a,b,a,b)} \cong N_{(a,b)} \#(S^2 \times S^2)$, as desired. In the general case, the curve $\tau_b^k(a)$ would lead to a slightly more complicated attaching circle and requires more slides. However, the argument is essentially the same: sliding over the fold handles corresponding to a and b results in an isolated Hopf link with framing $(0, k)$. An alternative proof of the general case is given in Lemma 7.8 and Example 7.10. □

Remark 7.6. It is interesting to observe that very similar pictures for blow-ups of 4–manifolds appear in the context of Heegaard Floer theory [56, Section 6]. Indeed, Ozsváth and Szabó represent blow-ups by certain Heegaard triples containing a punctured torus E together with three curves $\alpha, \beta, \gamma \subset E$ (see Figure 6, *loc. cit.*). Once the curves are ordered, they form a blow-up configuration in our sense. The author believes that this is not a coincidence, although the precise relationship between these pictures has not been understood yet.

7.2 Surgeries on Curves and Spheres

Another interesting class of substitutions was independently discovered by Hayano in [35, Lemma 6.13]. Consider a surface diagram $\mathfrak{S} = (\Sigma, \Gamma)$ and let $c \subset \Sigma$ be some vanishing cycles in $\mathfrak{S}$. If $\delta \subset \Sigma$ is a simple closed curves with $i(c, \delta) = 1$ (that is, a weakly dual curve for c), then we can perform the substitution $(c) \rightsquigarrow (c, \delta, c)$. Hayano shows that this substitution corresponds to a surgery on the curve $\delta \subset \Sigma \subset Z_{\mathfrak{S}}$ with respect to its *fiber framing*, that is, the framing in $Z_{\mathfrak{S}}$ induced by the canonical framing of δ in Σ and the framing of Σ in $Z_{\mathfrak{S}}$ as a regular fiber of $w_{\mathfrak{S}} \colon Z_{\mathfrak{S}} \to D^2$.

Remark 7.7. The attentive reader has probably noticed a small inaccuracy in the above discussion. Strictly speaking, we have to choose an orientation for δ to obtain the framing in Σ. However, different orientations yield diffeomorphic surgeries because the diffeomorphism of $S^1 \times S^2$ given by complex conjugation on the first factor has canonical extensions to both $S^1 \times D^3$ and $D^2 \times S^2$. Our sloppiness is therefore justified.

Comparing Hayano's substitution with equation (7.2), one immediately notices that the stabilization substitutions are special cases. We will say more about this in Example 7.10 below. This observation also paves the way for a minor generalization of Hayano's result which captures not only the fiber framed surgery but also the one with the opposite framing.[1]

Lemma 7.8. *(Surgery on curves) Let $\mathfrak{S} = (\Sigma, \Gamma)$ be a surface diagram and let $c \subset \Sigma$ be a vanishing cycle in $\mathfrak{S}$. If $\delta \subset \Sigma$ is a simple closed curve with $i(c, \delta) = 1$, then the substitution*

$$(c) \rightsquigarrow (c, \tau_c^k(\delta), c), \quad k \in \mathbb{Z}, \tag{7.3}$$

corresponds to the surgery on $\delta \subset \Sigma \subset Z_{\mathfrak{S}}$ with respect to the fiber framing when k is even and the opposite framing when k is odd.

Again, the proof relies on Kirby calculus. The following interpretation of surgery in terms of Kirby diagrams is well known but hard to find in the literature.

Remark 7.9 (Surgery in Kirby diagrams). Let δ be an embedded circle in an oriented 4–manifold M. Then we can draw a Kirby diagram for M which contains a dotted circle whose meridian is isotopic to δ as shown in the top left of Figure 31 (by taking a handle decomposition of M based on an $S^1 \times D^3$ neighborhood of δ). The gray ribbon indicates 2–handles that might link the dotted circle. In order to perform surgery on δ we equip δ with the k–framing in the Kirby diagram which canonically extends to a framing of δ in M; of course, the framing in M will only depend on the parity of k.[2] The surgery is then realized by the replacement shown in the top row of Figure 31 where two extra 2–handles appear, one is attached along δ with the k–framing and the other along a 0–framed meridian of δ. The latter 2–handle represents the embedded $D^2 \times S^2$ in the surgered manifold. In fact, the right hand side is easily seen to be diffeomorphic $D^2 \times S^2$ by canceling the 1–handle and the k–framed copy of δ after unlinking it from the gray ribbon, which acquires k full twists and now links the meridian of δ. Alternatively, one can also realize the surgery

[1] Recall that an embedded circle in an orientable 4–manifold always has trivial normal bundle and there are exactly two framings since $\pi_1(SO(3)) \cong \mathbb{Z}_2$.

[2] Again, strictly speaking, we need an orientation on δ. See Remark 7.7.

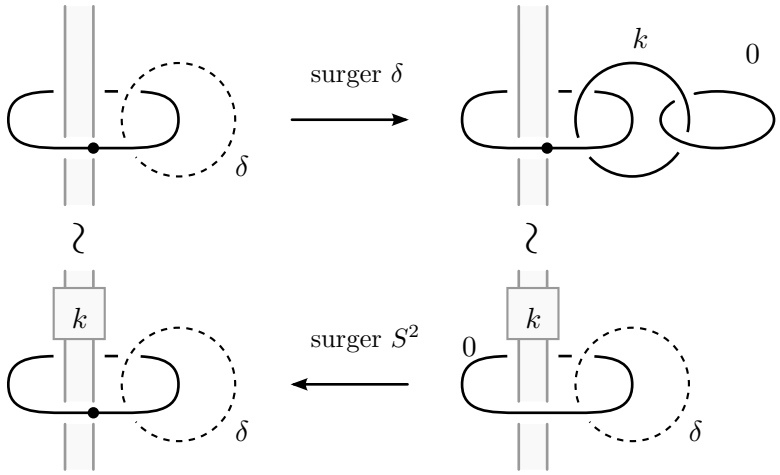


Figure 31: Surgery in terms of Kirby diagrams: k–framed surgery on δ and reverse surgery on a 2–sphere.

by going directly from the top left to the bottom right. This amounts to turning the dotted circle into a 0–framed unknot while putting in k full twists in the gray ribbon.

For the reverse surgery on $D^2 \times S^2$ we simply replace the 0–framed unknot with a dotted circle as shown in the bottom row of Figure 31. The equivalence of the two diagrams on the left is a standard trick in Kirby calculus (see [32, p.174f.]) and shows that we always recover M regardless of the value of k. In fact, whenever a Kirby diagram of a 4–manifold contains a 0–framed unknot that does not link any of the dotted circles, replacing it with a dotted circle represents the surgery on the 2–sphere given by union of the spanning disk and the core of the corresponding 2–handle. Moreover, the surgery on an arbitrary 2–sphere in M with trivial normal bundle can be realized in this way.

Proof of Lemma 7.8. We employ a similar localization strategy as in the proof of Lemma 7.5. We assume that c and δ intersect transversely in one point and fix a regular neighborhood $T \subset \Sigma$ of the union $c \cup \delta$ diffeomorphic to a one-holed torus. We denote by $N_{(c)}$ and $N_{(c,\tau^2(\delta),c)}$ the 4–manifolds obtained by attaching fold handles to $T \times D^2$ along the curves in the respective subscripts. It is then enough to show that $N_{(c,\tau_c^k(\delta),c)}$ is obtained from $N_{(c)}$ by surgery on δ.

We first consider the case $k = 0$ which is Hayano's original surgery substitution $(c) \rightsquigarrow (c, \delta, c)$. A Kirby diagram for $N_{(c)}$ is shown on the right of Figure 32. δ appears as a meridian for the upper 1–handle and the fiber framing agrees with the 0–framing in this diagram. Obviously, $N_{(c)}$ is diffeomorphic to $S^1 \times D^3$. On the left of Figure 32 we see a Kirby diagram for $N_{(c,\delta,c)}$ and the indicated 2–handle slide exhibits $N_{(c,\delta,c)}$ as the fiber framed surgery on δ in $N_{(c)}$ (compare Figure 31). In particular, $N_{(c,\delta,c)}$ is diffeomorphic to $D^2 \times S^2$.

Kirby diagrams for the cases $k = 1, 2$ are shown in Figures 33 and 34 and the indicated 2–handle slides combined with Figure 31 again prove the claim. The general pattern should also be clear by now. For arbitrary k the fold handle corresponding to $\tau_c^k(\delta)$ wraps around the lower 1–handle k times and becomes a k–framed meridian for the upper 1–handle after sliding k times over the lower left fold handle. Com-

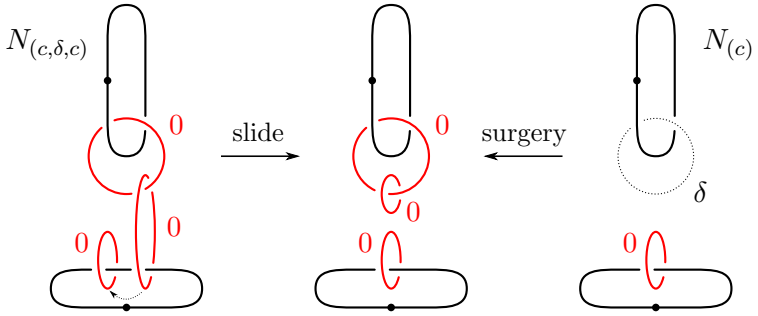


Figure 32: The substitution $(c) \rightsquigarrow (c, \delta, c)$ corresponds to surgery on δ with respect to the fiber framing.

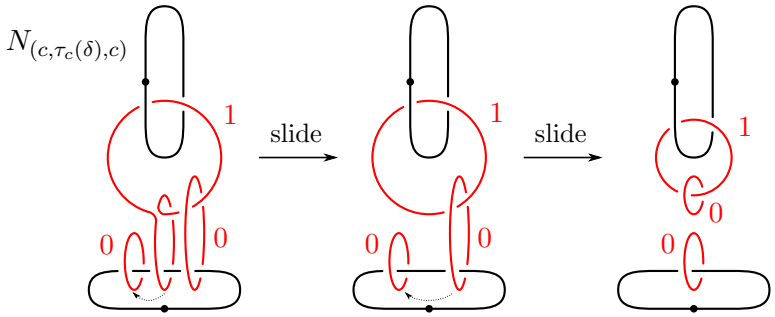


Figure 33: Simplifying $N_{(c,\tau_c(\delta),c)}$.

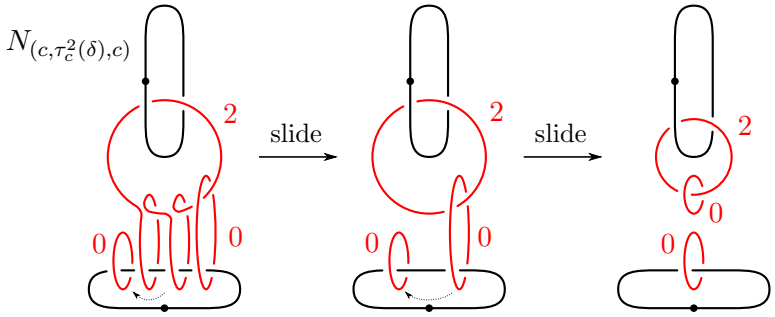


Figure 34: Simplifying $N_{(c,\tau_c^2(\delta),c)}$.

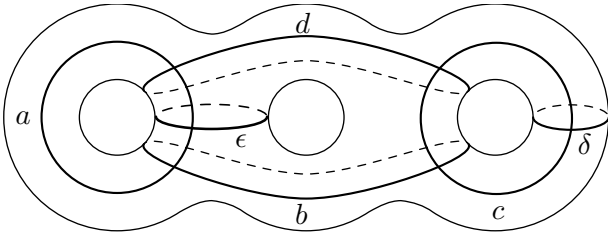


Figure 35: The curves in Σ_3 used in Example 7.11.

paring with Figure 31 shows that $N_{(c,\tau_c^k(\delta),c)}$ is a surgery of $N_{(c)}$ on δ with the fiber framing if k is even and the opposite framing if k is odd. □

Example 7.10. (Stabilization revisited) Recall the stabilization substitution from Lemma 7.5 of the form $(a,b) \rightsquigarrow (a,b,\tau_b^k(a),b)$. Of course, this is equivalent to the surgery substitution $(b) \rightsquigarrow (b,\tau_b^k(a),b)$ and thus corresponds to a surgery on the curve a in $Z_{\mathfrak{S}}$. But a is the attaching curve of a fold handle and it is therefore null-homotopic in $Z_{\mathfrak{S}}$. As is well known, a surgery on a null-homotopic curve results in a connected sum with either $S^2 \times S^2$ or $\mathbb{C}\mathrm{P}^2\#\overline{\mathbb{C}\mathrm{P}^2}$, depending on the framing. This also constitutes an alternative proof of part (ii) of Lemma 7.5.

Example 7.11 (A surface diagram of S^4)**.** The following constructions are due to Hayano [35, Example 6.10]. We consider the curves in Σ_3 shown in Figure 35. Hayano first uses Kirby calculus to show that $(\Sigma_3; a, b, c, d)$ is a surface diagram for $S^1 \times S^3 \# S^1 \times S^3$ and then performs surgery substitutions using the curves $\delta, \epsilon \subset \Sigma_3$. He shows that these curves generate the fundamental group so that each surgery kills one $S^1 \times S^3$ summand. This implies that the surface diagrams $(\Sigma_3; a, b, c, \delta, c, d)$ and $(\Sigma_3; a, \epsilon, a, b, c, \delta, c, d)$ represent $S^1 \times S^3$ and S^4, respectively.

Recall that a surgery on a circle leaves behind a framed 2–sphere in the surgered manifold and that the original manifold can be recovered by surgering on this 2–sphere. In the Kirby diagram interpretation of surgery in Figure 31, this 2–sphere is visible in either picture on the right as the obvious spanning disk of the 0–framed unknot hand sight capped off with the core of the corresponding 2–handle. Comparing with Figure 32, we see that the 2–sphere resulting from the surgery substitution $(c) \rightsquigarrow (c, \delta, c)$ appears as the obvious annulus spanning between the two 0–framed unknots that link the lower dotted circle capped off with the cores of the 2–handles. Moreover, we can even see this 2–sphere in the surface diagram after the substitution. We explain in how this works in a more general context.

Remark 7.12 (2–spheres in surface diagrams). Suppose that $\mathfrak{S} = (\Sigma; c_1, \ldots, c_l)$ is a surface diagram such that for some $i \neq j$ the vanishing cycles c_i and c_j are isotopic. Without loss of generality we can assume that they are disjoint and thus bound an annulus $A \subset \Sigma$. Observe that in $Z_{\mathfrak{S}}$ the curves c_i and c_j bound disks, namely the cores of the corresponding fold handles, so that A can be capped off to a 2–sphere in $Z_{\mathfrak{S}}$. Moreover, since the 2–handles are attached with the fiber framing, this 2–sphere has trivial normal bundle.

In other words, a repetition of vanishing cycles (up to isotopy) in a surface diagram represents an embedded 2–sphere with trivial self-intersection. In the situation

where a repetition occurs with only one vanishing cycle in between, we obtain the following converse of Lemma 7.8.

Lemma 7.13 (Surgery on spheres)**.** *A substitution of the form* $(c, \delta, c) \rightsquigarrow (c)$ *corresponds to a surgery on the 2–sphere represented by the two copies of* c*.*

A closely related construction to surgery on curves and spheres is the so called *Gluck twist*. This operation consists of removing an embedded copy of $S^2 \times D^2$ from a given 4–manifold M and re-gluing it with the non-trivial automorphism of $S^2 \times S^1$ corresponding to the generator of $\pi_1\big(\mathrm{SO}(3)\big)$. A different interpretation is as follows. One first performs surgery on the original $S^2 \times D^2$ which leaves behind a copy of $D^3 \times S^1$ in the surgered manifold. In particular, the circle $\{0\} \times S^1$ has a canonical framing and the surgery with this framing gives back M. The surgery with the other framing gives the Gluck twisted version of M. As a consequence of Lemmas 7.8 and 7.13, we immediately obtain the following.

Corollary 7.14 (Gluck twists)**.** *A substitution of the form* $(c, \delta, c) \rightsquigarrow (c, \tau_c(\delta), c)$ *corresponds to a Gluck twist on the 2–sphere represented by the two copies of* c*.*

We end this chapter with some remarks about the scope of Lemmas 7.8 and 7.13 and Corollary 7.14. Note that one frequent application of surgery on curves is to kill the fundamental group of a given 4–manifold. This process can be realized in terms of surface diagrams using Lemma 7.8. Indeed, given a surface diagram $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$ it follows from the handle decomposition of $Z_{\mathfrak{S}}$, that $\pi_1(Z_{\mathfrak{S}})$ is generated by simple closed curves in $\Sigma \subset Z_{\mathfrak{S}}$ (see Lemma 9.1 for more details). Moreover, an arbitrary simple closed curve $\delta \subset \Sigma$ can be transformed into a curve $\delta' \subset \Sigma$ with $i(\delta', c_i)$ by taking band sums with $c_{i+1} \subset \Sigma$. Since c_{i+1} is null-homotopic in $Z_{\mathfrak{S}}$, it follows that δ and δ' represent the same element of $\pi_1(Z_{\mathfrak{S}})$ and surgery on δ' is realizable by Lemma 7.8.

On the other hand, the use of Lemma 7.13 (andCorollary 7.14) is more limited since the 2–spheres must appear in a rather special way in $\mathfrak{S}$. For example, in the surface diagrams for $\Sigma_g \times S^2$ with only two curves discussed in Section 6.3.1, there is no repetition of vanishing cycles and it is not clear how to realize the surgery on $\{*\} \times S^2$ using Lemma 7.13. Also, when a vanishing cycle is repeated with more than one curve in between it is unclear what kind of substitution might realize the surgery on the corresponding 2–sphere.

Chapter 8

Manifolds with Genus One Surface Diagrams

In this chapter we address the question which closed 4–manifolds can be described by surface diagrams of genus one, the lowest possible genus. This means that we have to study surface diagrams that are supported on the torus T^2. Using the blow-up and stabilization substitutions discussed in Section 7.1 we obtain the following classification result which was already stated in the introduction on page 8.

Theorem 1.4. *A closed 4–manifold admits a surface diagram of genus one if and only if it is diffeomorphic to $kS^2 \times S^2$ or $m\mathbb{C}\mathrm{P}^2 \# n\overline{\mathbb{C}\mathrm{P}^2}$ where $k, m, n \geq 1$.*

The condition $k, m, n \geq 1$ means that S^4, $\mathbb{C}\mathrm{P}^2$, and $\overline{\mathbb{C}\mathrm{P}^2}$ are not in the list. In fact, it will be clear from the proof that all manifolds with genus one surface diagrams must contain an S^2–bundle over S^2.

Before going into the proof we make some general observations about genus one surface diagrams. First, note that monodromy is not an issue in the genus one case. In fact, genus one surface diagrams automatically have trivial monodromy since the mapping class group of the sphere is trivial. However, the price to pay is that we have to deal with a mild gluing ambiguity caused by $\pi_1\big(\mathrm{Diff}(S^2), \mathrm{id}\big) = \mathbb{Z}_2$. Second, and more importantly, many questions about simple closed curves on the torus (which are intractable for higher genus surfaces) reduce to problems in linear algebra by passing to homology. We will make use of the following facts. Let $a, b \subset T^2$ be two simple closed curves; we choose arbitrary orientations and denote the oriented curves by $\vec{a}$ and $\vec{b}$.

- a and b are isotopic if and only if $\vec{a} = \pm\vec{b} \in H_1(T^2)$.
- The geometric and algebraic intersection numbers satisfy

  $$i(a, b) = |\langle \vec{a}, \vec{b} \rangle|,$$

  that is, the inequality (2.6) is an equality.

Building on these observations we obtain the following result about the structure of genus one surface diagrams.

Lemma 8.1. *Any surface diagram of genus one of length at least three contains a blow-up or stabilization configuration.*

Proof. We first consider an arbitrary W–chain $\Gamma = (c_1, \dots, c_l)$ in T^2 of length $l \geq 3$ which is not necessarily a W–cycle, that is, we do not require c_l and c_1 to be weakly dual. We transfer the discussion to the level of homology by choosing orientations on the curves in Γ. In principle, we could choose random orientations on the curves but the following convention turns out to be convenient: we choose an arbitrary orientation on c_1 and orient the remaining curves by requiring that

$$\langle \vec{c}_i, \vec{c}_{i+1} \rangle = +1 \quad \text{for } i < l \tag{8.1}$$

where $\vec{c}_i$ denotes the oriented version of c_i. We can then consider each $\vec{c}_i$ as an element of $H_1(T^2)$ and using the notation

$$\sigma_i = \langle \vec{c}_1, \vec{c}_i \rangle \in \mathbb{Z}$$

the condition for Γ to be a W–cycle is equivalent to $|\sigma_l| = 1$. Next we observe that any two adjacent curves in Γ form a basis of $H_1(\Sigma)$ (which follows from equation (8.1)). In particular, for $i \geq 3$ we have expressions of the form

$$\vec{c}_i = k_i \vec{c}_{i-1} - \vec{c}_{i-2}, \quad k_i \in \mathbb{Z} \tag{8.2}$$

where the coefficient of $\vec{c}_{i-2}$ is determined by the instances of equation (8.1) involving $\vec{c}_i$, $\vec{c}_{i-1}$ and $\vec{c}_{i-2}$.

We claim that if $|k_i| \geq 2$ for all $i \geq 3$, then we have $|\sigma_{i+1}| > |\sigma_i|$ for all i. Note that $|\sigma_2| > |\sigma_1|$ is trivially satisfied because we have $\sigma_2 = \langle \vec{c}_1, \vec{c}_2 \rangle = 1$ by equation (8.1) and obviously $\sigma_1 = \langle \vec{c}_1, \vec{c}_1 \rangle = 0$. Now, equation (8.2) shows that for $i \geq 2$ we have $\sigma_{i+1} = k_{i+1}\sigma_i - \sigma_{i-1}$ which enables us to proceed by induction. We can estimate

$$\begin{aligned} |\sigma_{i+1}| &= |k_{i+1}\sigma_i - \sigma_{i-1}| \\ &\geq \big| |k_{i+1}||\sigma_i| - |\sigma_{i-1}| \big| \\ &= |k_{i+1}||\sigma_i| - |\sigma_{i-1}| > |\sigma_i| \end{aligned}$$

where we have used the reverse triangle inequality, the induction hypothesis that $|\sigma_i| > |\sigma_{i-1}|$, and the assumption that $|k_{i+1}| \geq 2$. In particular, if $|k_i| \geq 2$ for $i \geq 3$, then we must have $|\sigma_l| > |\sigma_2| = 1$ and Γ is not a W–cycle.

Now let Γ be a W–cycle so that (T^2, Γ) is a surface diagram. The above discussion shows that we must have $|k_i| \leq 1$ for some i. Assume first that $k_i = \pm 1$. To keep the notation clean we momentarily rename the relevant curves to

$$(\vec{c}_{i-2}, \vec{c}_{i-1}, \vec{c}_i) = (a, \xi, b). \tag{8.3}$$

The condition $k_i = \pm 1$ then translates into $b = \pm\xi - a$ and thus $\xi = \pm(a + b)$. Moreover, our orientation convention (8.1) implies $1 = \langle a, \xi \rangle = \pm\langle a, b \rangle$ so that $\langle a, b \rangle = \pm 1$. Combining this with the Picard–Lefschetz formula (Proposition 2.15) we obtain

$$\tau_a^{\pm 1}(b) = b \pm \langle a, b \rangle a = a + b = \pm\xi.$$

But this shows that (a, ξ, b) is a blow-up configuration (since orientations are irrelevant for that matter).

Lastly, suppose that we have $k_i = 0$ for some i which means that $\vec{c}_i = -\vec{c}_{i-2}$. This can only happen if Γ contains at least four curves, for if there were only three curves, then we would only have k_3 which would have to be ± 1 (by the definition of W–cycles). Moreover, possibly after a cyclic permutation we can assume that $i \geq 4$ so that we find a configuration of the form

$$\big(\vec{c}_{i-3}, \vec{c}_{i-2}, \vec{c}_{i-1}, \vec{c}_i\big) = \big(a,\, b,\, kb - a,\, -b\big) \tag{8.4}$$

within Γ where $a = \vec{c}_{i-3}$, $b = \vec{c}_{i-2}$ and $k = k_{i-1}$. Again, from the Picard-Lefschetz formula we see

$$kb - a = -\tau_b^{-k}(a)$$

which exhibits (8.4) as a stabilization configuration after forgetting orientations. □

It is now easy to prove Theorem 1.4 with an inductive argument.

Proof of Theorem 1.4. We first show that $k(S^2 \times S^2)$ and $m\mathbb{C}\mathrm{P}^2 \# n\overline{\mathbb{C}\mathrm{P}^2}$, $k, m, n \geq 1$, have genus one surface diagrams. Recall from Section 6.3.1 that both S^2–bundles over S^2 are described by the diagram $(T^2; a, b)$ of length two. Using Lemma 7.5 we can then add summands of the form $S^2 \times S^2$, $\mathbb{C}\mathrm{P}^2$ and $\overline{\mathbb{C}\mathrm{P}^2}$ at will.

Conversely, suppose that X is described by a genus one surface diagram $\mathfrak{S}$. According to Lemma 8.1, $\mathfrak{S}$ must contain a blow-up or stabilization configuration and we can use Lemma 7.5 to split off summands of the form $S^2 \times S^2$, $\mathbb{C}\mathrm{P}^2$ and $\overline{\mathbb{C}\mathrm{P}^2}$ while reducing the length of the surface diagram. By repeating this process, we eventually arrive at a diagram of length two which describes the S^2–bundles over S^2. □

Chapter 9

Homotopy Information in Surface Diagrams

Our next goal is to show how some basic information about the homotopy type of a given 4–manifold can be extracted from a surface diagram. We first discuss the fundamental group, homology, and the intersection form, and then go on to spin and spin^c structures. As applications, we obtain an obstruction for a surface diagram to have trivial monodromy (see Section 9.2) and a proof of Theorem 1.5 on which we elaborate further in Section 9.4.

Notation. In what follows we consider a surface diagram $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$ of genus g which may or may not have trivial monodromy. We choose arbitrary orientations on the vanishing cycles and use the notation $\vec{c}_i$ whenever the orientations are relevant. For brevity, we denote the descending simple wrinkled fibration associated to $\mathfrak{S}$ by $(Z, w) = (Z_{\mathfrak{S}}, w_{\mathfrak{S}})$ and we fix a boundary fiber Σ'. If $\mathfrak{S}$ happens to have trivial monodromy, the we choose a fiber preserving diffeomorphism $\varphi\colon \Sigma' \times S^1 \to \partial Z$ and denote the simple wrinkled fibration over S^2 obtained from closing off (Z, w) using φ by $(X, w) = (X^{\varphi}_{\mathfrak{S}}, w^{\varphi}_{\mathfrak{S}})$. We fix handle decompositions of Σ and Σ' with unique 0– and 2–handles and consider the corresponding handle decompositions of Z and X described in Section 6.1. We denote the attaching curve of the last 2–handle of X by

$$\kappa' = \varphi\big(\{p\} \times S^1\big) \subset \partial Z, \quad p \in \Sigma'$$

and we fix a closing curve

$$\kappa\colon S^1 \to \Sigma$$

as in Definition 6.2. Finally, the framing of κ' is determined by the choice of a tangent vector $v \in T_p\Sigma'$ and we also fix a framed lift $\boldsymbol{\kappa}\colon S^1 \to T\Sigma'$ of κ.

Recall that for $g \geq 3$ the choice of φ is essentially unique, and framed closing curves can in principle be extracted from $\mathfrak{S}$ although this can be complicated in practice (see page 75).

9.1 Fundamental Group, Homology, and Intersection Form

Here we take a look at the basic algebraic topology of Z and X from the perspective of their surface diagrams.

9.1.1 Fundamental Group and Euler Characteristic

We begin with some easy consequences of the handle decompositions of Z and X.

Lemma 9.1. *The inclusions $\Sigma \hookrightarrow Z$ and $\Sigma \hookrightarrow X$ induce isomorphisms*

$$\pi_1(Z) \cong \pi_1(\Sigma)/\langle\langle c_1, \dots c_l \rangle\rangle \quad \textit{and} \quad \pi_1(X) \cong \pi_1(\Sigma)/\langle\langle c_1, \dots c_l, \kappa \rangle\rangle$$

as well as

$$H_1(Z) \cong H_1(\Sigma)/\langle c_1, \dots c_l \rangle \quad \textit{and} \quad H_1(X) \cong H_1(\Sigma)/\langle c_1, \dots c_l, \kappa \rangle$$

where the brackets indicate (normal) subgroups generated by the enclosed elements.

Proof. The 2–handle attachments to $\Sigma \times D^2$ along c_i and κ kill the normal subgroups generated by these curves in $\pi_1(\Sigma \times D^2) \cong \pi_1(\Sigma)$. □

In particular, one can extract presentations for the fundamental groups of Z and X from the knowledge of $\mathfrak{S}$. We also see an emerging pattern. Information about Z is directly accessible from $\mathfrak{S}$ while information about X also depends on a closing curve which complicates the situation. This is true with only few exceptions and we will usually study Z first and discuss the passage to X separately. One exception where the closing curve is irrelevant is the Euler characteristic.

Lemma 9.2. *The Euler characteristics of Z and X are*

$$\chi(Z) = 2 - 2g + l \quad \textit{and} \quad \chi(X) = 6 - 4g + l. \tag{9.1}$$

Proof. This follows from the decompositions of X into Z and $\Sigma' \times D^2$, and Z into $\Sigma \times D^2$ and the fold handle cobordism. In the case of X the additivity of the Euler characteristic then implies

$$\chi(X) = \chi(Z) + 2 - 2(g-1) = (2-2g) + l + (4-2g)$$

which also contains the computation for Z. □

Remark 9.3. In particular, Euler characteristic agrees modulo 2 with the number of vanishing cycles which in turn is just the number of cusps of the associated simple wrinkled fibration. This is in accordance with a more general result of Thom which states that the number of cusps of a Boardman map from a manifold of arbitrary dimension to an orientable surface reduces modulo 2 to the Euler characteristic of the source (see [65, p.84]).

9.1.2 Second Homology and Intersection Form of Z

In order to compute $H_2(Z)$ we use the *handle complex* of Z, that is, the chain complex derived from the handle decomposition of Z (see Appendix A for more details). Since there are no handles of index three or higher, this takes the form

$$C_2(Z) \xrightarrow{\partial_2} C_1(Z) \xrightarrow{\partial_1} C_0(Z). \tag{9.2}$$

Moreover, the fixed handle decomposition of Σ allows us to identify $C_i(Z)$ with $H_i(\Sigma)$ for $i = 0, 1$ and ∂_1 must be the zero map. As for the 2–handles, if we denote by $V_{\mathfrak{S}}$ the free Abelian group generated by the vanishing cycles c_i, then $C_2(Z)$ can be identified with $H_2(\Sigma) \oplus V_{\mathfrak{S}}$. The first summand is generated by the fiber 2–handle which is a 2–cycle in the handle complex that represents the homology class of the fiber $[\Sigma] \in H_2(Z)$. We can describe the boundary map ∂_2 in this setting as follows. Observe that the orientations on the vanishing cycles give rise to a homomorphism

$$\begin{array}{rccc} \rho\colon & V_{\mathfrak{S}} & \longrightarrow & H_1(\Sigma) \\ & c_i & \mapsto & [\vec{c}_i] \end{array}$$

and by the definition of the boundary operator (and Remark 6.3) we have a commutative diagram

$$\begin{array}{ccc} C_2(Z) & \xrightarrow{\partial_2} & C_1(Z) \\ \cong\uparrow & & \uparrow\cong \\ H_2(\Sigma) \oplus V_{\mathfrak{S}} & \xrightarrow{\tilde{\rho}} & H_1(\Sigma) \end{array} \tag{9.3}$$

where $\tilde{\rho}(k[\Sigma], v) = \rho(v)$. It follows that $H_2(Z)$ is isomorphic to the direct sum of $H_2(\Sigma)$ and the group

$$K_{\mathfrak{S}} = \ker\big(\rho\colon V_{\mathfrak{S}} \to H_1(\Sigma)\big).$$

Put differently, $K_{\mathfrak{S}}$ is the subgroup of $V_{\mathfrak{S}}$ that records all the linear relations in $H_1(\Sigma)$ that hold among the oriented vanishing cycles. To summarize, we have proved:

Lemma 9.4. *$H_2(Z)$ is isomorphic to $\mathbb{Z} \oplus K_{\mathfrak{S}}$ where the first summand is generated by the homology class of the fiber $[\Sigma] \in H_2(Z)$.*

Remark 9.5. Alternatively, one can also appeal to the Mayer–Vietoris sequence for the decomposition of Z into $\Sigma \times D^2$ and the disjoint union of the fold handles. The intersection is given by the disjoint union of the attaching regions and its first homology is naturally identified with $V_{\mathfrak{S}}$. This results in an exact sequence

$$0 \longrightarrow H_2(\Sigma) \xrightarrow{\text{incl}_*} H_2(Z) \xrightarrow{\delta} V_{\mathfrak{S}} \xrightarrow{\rho} H_1(\Sigma) \xrightarrow{\text{incl}_*} H_1(Z) \longrightarrow 0$$

from which $H_1(Z)$ and $H_2(Z)$ are easily computed as in Lemmas 9.1 and 9.4.

Having described $H_2(Z)$ purely in terms of Z we now study how the intersection form of Z appears in this description. We define a symmetric bilinear form on $K_{\mathfrak{S}}$ as follows. For $\xi, \eta \in K_{\mathfrak{S}}$, written as $\xi = \sum_i \xi_i \vec{c}_i$ and $\eta = \sum_j \eta_j \vec{c}_j$, we define

$$Q_{\mathfrak{S}}(\xi, \eta) = \tfrac{1}{2} \sum_{i,j} \xi_i \eta_j \, \varepsilon_{ij} \langle \vec{c}_i, \vec{c}_j \rangle_\Sigma, \quad \epsilon_{ij} = \begin{cases} +1, & i > j \\ -1, & i < j \end{cases} \tag{9.4}$$

where $\langle \cdot, \cdot \rangle_\Sigma$ denotes the intersection pairing on $H_1(\Sigma)$. Note that $Q_{\mathfrak{S}}$ is indeed symmetric since both $\langle \vec{c}_i, \vec{c}_j \rangle_\Sigma$ and ϵ_{ij} change signs when i and j are switched. It can be thought of as a symmetrization of $\langle \cdot, \cdot \rangle_\Sigma$ which takes the order of the vanishing cycles into account.

Proposition 9.6. *Under the isomorphism $H_2(Z) \cong \mathbb{Z} \oplus K_{\mathfrak{S}}$ the intersection form Q_Z corresponds to $(0) \oplus Q_{\mathfrak{S}}$.*

Proof. Since Z has a handle decomposition with handles of index at most two, we can determine Q_Z from the linking information in a Kirby diagram for Z as described in Appendix A, see Proposition A.2 on page 120. Here we use the tactical Kirby diagrams derived from Figure 18 on page 77. As before, we identify the 2–chains $C_2(Z)$ in the handle complex with $\mathbb{Z} \oplus V_{\mathfrak{S}}$ and blur the distinction between the 2–handles and their attaching curves in the Kirby diagram.

According to Proposition A.2, Q_Z is isomorphic to the restriction of the linking form on $\mathbb{Z} \oplus V_{\mathfrak{S}}$, which we denote by lk, to the 2-cycles in the handle complex. To compute the linking form we first observe that the fiber 2–handle is unlinked from all the fold handles. Indeed, any fiber in the Kirby diagram that is not occupied by a fold handle provides a Seifert surface that is disjoint from the fold handles. Moreover, since the fiber 2–handle is 0–framed, the linking from splits off a trivial summand. To investigate the linking form on $V_{\mathfrak{S}}$ we appeal to Lemma 6.4 and also use its surrounding notation. If we orient the fold handles by identifying them with the oriented versions $\vec{c}_i$ of the vanishing cycles, then equation (6.2) gives

$$\mathrm{lk}(\vec{c}_i, \vec{c}_j) = \tfrac{1}{2}\epsilon_{ij}\langle \vec{c}_i, \vec{c}_j \rangle_\Sigma + \tfrac{1}{2}\sum_{k=1}^{g} \big[n_{a_k}(\vec{c}_i) n_{b_k}(\vec{c}_j) + n_{a_k}(\vec{c}_j) n_{b_k}(\vec{c}_i) \big]. \tag{9.5}$$

Now, the 2–cycles in the handle complex are given by $\mathbb{Z} \oplus K_{\mathfrak{S}}$ and the linking form vanishes on $\mathbb{Z}$ which is generated by the fiber 2–handle. Moreover, for $\xi \in K_{\mathfrak{S}}$ we have $n_{a_k}(\rho(\xi)) = n_{b_k}(\rho(\xi)) = 0$, so that the second term in equation (9.5) vanishes on $K_{\mathfrak{S}}$. Consequently, lk restricts to $Q_{\mathfrak{S}}$ on $K_{\mathfrak{S}}$ and the proof is finished. □

Second Homology via Domains

We can obtain another description of $H_2(Z)$ which parallels a construction used in Heegaard–Floer theory. Instead of considering the curves in $\mathfrak{S}$, we focus on their complement. By a slight abuse of notation, we write $\Gamma = \cup_i c_i \subset \Sigma$ and consider the connected components of $\Sigma \setminus \Gamma$. We call the closure of such a component an *elementary domain* and let $\mathcal{D}(\mathfrak{S})$ be the free Abelian group generated by all elementary domains; elements of $\mathcal{D}(\mathfrak{S})$ are simply called *domains*. By sending an elementary domain to its boundary we obtain a map

$$\partial \colon \mathcal{D}(\mathfrak{S}) \longrightarrow H_1(\Gamma).$$

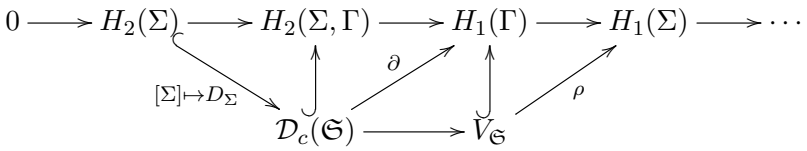
Note that $\mathcal{D}(\mathfrak{S})$ is naturally isomorphic to $H_2(\Sigma, \Gamma)$ and the map ∂ can also be interpreted as the connecting homomorphism in the homology sequence of the pair (Σ, Γ).

Definition 9.7. A domain $D \in \mathcal{D}(\mathfrak{S})$ is called *cyclic* if its boundary ∂D is a linear combination of the vanishing cycles c_i. The group of all cyclic domains is denoted by $\mathcal{D}_c(\mathfrak{S})$.

This definition is reminiscent of the "periodic domains" that are prominently featured in Heegaard–Floer theory (see [55, p.1040 ff.], for example). However, the term "periodic" has no apparent meaning in the present setting and we decided to use "cyclic" instead. This is further justified by the fact that cyclic domains correspond to 2–cycles in the handle complex of Z as seen in the proof of the following lemma.

Lemma 9.8. *The group $\mathcal{D}_c(\mathfrak{S})$ of cyclic domains is isomorphic to $H_2(Z)$.*

Proof. This is a consequence of the commutative diagram



which we shall now explain. The first row is simply an excerpt of the homology sequence of the pair (Σ,Γ). As for the second, we proceed from right to left. The first thing to note is that $V_\mathfrak{S}$ can naturally be considered as the subspace of $H_1(\Gamma)$ spanned by the classes $[\vec{c}_i] \in H_1(\Gamma)$. From this observation it becomes clear that ρ and ∂ can be factored as in the diagram. Moreover, as we already pointed out, the group of all domains is isomorphic to $H_2(\Sigma,\Gamma)$ so that $\mathcal{D}_c(\mathfrak{S})$ embeds therein. Finally, note that there is a special domain $D_\Sigma \in \mathcal{D}(\mathfrak{S})$ given by the sum of all elementary domains (each with multiplicity one). This domain generates the kernel of ∂ (in particular, it is cyclic) and it naturally corresponds to the image of the fundamental class of Σ in $H_2(\Sigma,\Gamma)$. Now, from the diagram above we can extract a short exact sequence

$$0 \longrightarrow H_2(\Sigma) \longrightarrow \mathcal{D}_c(\mathfrak{S}) \longrightarrow K_\mathfrak{S} \longrightarrow 0$$

which splits because $K_\mathfrak{S}$ is free. In fact, a concrete splitting $K_\mathfrak{S} \to \mathcal{D}_c(\mathfrak{S})$ is obtained by fixing an elementary domain $E \subset \Sigma$ and sending an element $\xi \in K_\mathfrak{S}$ to the unique cyclic domain with boundary ξ and multiplicity zero at E. It follows that $\mathcal{D}_c(\Sigma)$ is isomorphic to $H_2(\Sigma) \oplus K_\mathfrak{S}$ and thus to $H_2(Z)$ according to Lemma 9.4. □

Remark 9.9. A concrete, geometric isomorphism $\mathcal{D}(\mathfrak{S}) \to H_2(Z)$ can be constructed as follows. One can think of a cyclic domain $D \in \mathcal{D}_c(\mathfrak{S})$ as the image of an immersion $\iota\colon S \to \Sigma$ of a compact, oriented surface S whose boundary components are mapped to the $c_i \subset \Sigma$. If we cap off all boundary components of S with disks, then we obtain a closed, oriented surface $\hat{S}$ and, thinking of Σ as sitting inside Z, we can first perturb ι to an embedding $\tilde{\iota}\colon S \hookrightarrow Z$ and then further extend to an embedding of $\hat{\iota}\colon \hat{S} \hookrightarrow Z$ using the cores of the fold handles. Thus we obtain a homology class $H(D) = \hat{\iota}_*[\hat{S}] \in H_2(Z)$ and one can show that this construction gives rise to a well-defined isomorphism $H\colon \mathcal{D}_c(\mathfrak{S}) \to H_2(Z)$ by adapting the arguments in [55, Section 2.5].

The domain interpretation of $H_2(Z)$ certainly has its appeal, especially in the light of a potential relationship between surface diagrams and Heegaard–Floer theory. However, the curve interpretation given in Lemma 9.4 is better suited for our purposes since we are currently not aware of a good description of the intersection form in terms of domains.

9.1.3 Second Homology and Intersection Form of X

Now let us try to describe the homology of X. We begin with some remarks about the homology class of the fiber which turns out to play an important role. Note that the fiber class $[\Sigma] \in H_2(X)$ is an invariant of the homotopy class of $w\colon X \to S^2$. Since simple wrinkled fibrations can be found in all homotopy classes, it follows from the Pontrjagin–Thom construction that every class in $H_2(X)$ with trivial self-intersection can appear as a fiber class. In particular, while the fiber class is always primitive[1] in $H_2(Z)$ by Lemma 9.4, it can be torsion or divisible in $H_2(X)$. These properties of the fiber class can be related to closing curves as follows.

Lemma 9.10. *The following conditions are equivalent.*

(i) The fiber class $[\Sigma] \in H_2(X)$ is essential.

(ii) The attaching curve $\kappa' \subset \partial Z$ has finite order in $H_1(Z)$.

(iii) The closing curve satisfies $d[\kappa] \in \langle c_1, \dots, c_l \rangle \subset H_1(\Sigma)$ for some $d \geq 1$.

Proof. The equivalence of (ii) and (iii) follows from Lemma 9.1 and the observation that $\kappa' \subset \partial Z$ and $\kappa \subset \Sigma$ are homologous when considered as curves in Z. To see that (i) and (ii) are equivalent we will show that $d[\kappa'] = 0 \in H_1(Z)$ for some $d \geq 1$ if and only if $[\Sigma] \in H_2(X)$ has infinite order.

Assume first that $d[\kappa'] = 0$. Then the union of d parallel copies of κ' in ∂Z (parallel with respect to the framing of κ') bounds a properly embedded, oriented surface $(S, \partial S) \subset (Z, \partial Z)$ which we can cap off to a closed surface $\hat{S} \subset X$ using parallel copies of the core of the last 2–handle. By construction, this surface intersects Σ' transversely in d points of the same sign, and since Σ' and Σ are homologous in X, we see that $[\Sigma] \cdot [\hat{S}] = d$. In particular, $[\Sigma]$ has infinite order by the unimodularity of the intersection form of X.

Now suppose that the fiber class $[\Sigma] = [\Sigma'] \in H_2(X)$ has infinite order. We can essentially reverse the above argument. The unimodularity of the intersection form of X implies that there is a closed, oriented surface $\hat{S} \subset X$ pairing non-trivially with Σ', say $[\hat{S}] \cdot [\Sigma] = d \neq 0$. By general position, we can assume that $\hat{S}$ meets $\Sigma' \times D^2 \subset X$ in disks of the form $\{p\} \times D^2$. But then $S = \hat{S} \cap Z$ is properly embedded surface Z such that $\partial S \subset \partial Z$ consists of parallel copies of κ' which represent the class $d[\kappa'] \in H_1(Z)$. Thus $d[\kappa']$ is trivial in $H_1(Z)$. □

With these remarks in place, we return to the homology of X. Recall that $X = Z \cup_\varphi (\Sigma' \times D^2)$ where $\varphi\colon \Sigma' \times S^1 \to \partial Z$ is a fiber preserving diffeomorphism. In addition to the curve $\kappa' = \varphi(* \times S^1) \subset \partial Z$ we also consider tori of the form

$$\varphi(\gamma \times S^1) \subset \partial Z$$

where $\gamma \subset \Sigma'$ is a simple closed curve; these are commonly called *rim tori* in similar contexts. We denote the subgroup of $H_2(Z)$ generated by the rim tori by

$$\mathcal{R} = \varphi_*\big(H_1(\Sigma') \times [S^1]\big) \subset H_2(Z).$$

[1] We use the following standard terminology. Let H be an Abelian group. An element $e \in H$ is called *essential* if it is not torsion. We call $p \in H$ *primitive* if it is essential and *indivisible*, that is, it cannot be written as a non-trivial multiple of any other essential element. Any essential $e \in H$ can be written as dp for some primitive $p \in H$ and a uniquely determined $d \in Z$ with $d \geq 1$ called the *divisibility* of e.

Proposition 9.11. *If the fiber class $[\Sigma] \in H_2(X)$ is torsion, then the inclusion $Z \hookrightarrow X$ induces an isomorphism*

$$H_2(X) \cong H_2(Z)/\mathcal{R}.$$

If the fiber class is essential, then there is a short exact sequence

$$0 \longrightarrow H_2(Z)/\mathcal{R} \longrightarrow H_2(X) \longrightarrow \mathbb{Z} \longrightarrow 0. \tag{9.6}$$

Proof. The homology sequence of the pair (X, Z) provides an exact sequence

$$0 \to H_2(Z)/\delta H_3(X, Z) \to H_2(X) \to H_2(X, Z) \to H_1(Z) \to H_1(X) \to 0$$

and the claim follows from an inspection of the relative groups $H_k(X, Z)$ and the connecting homomorphisms. By excision, we have

$$H_k(X, Z) \cong H_k(\Sigma' \times D^2, \Sigma' \times S^1)$$

and the latter group is isomorphic to $H_{k-2}(\Sigma')$ via the cross product with the relative fundamental class $[D^2, S^1] \in H_2(D^2, S^1)$. It follows that $H_2(X, Z)$ is infinite cyclic, generated by $\big(\varphi(* \times D^2), \varphi(* \times S^1)\big) \subset (X, Z)$ and we have

$$\delta[\varphi(* \times D^2), \varphi(* \times S^1)] = [\varphi(* \times S^1)] = [\kappa'] \in H_1(Z).$$

Similarly, $H_3(X, Z)$ is generated by $\big(\varphi(\gamma \times D^2), \varphi(\gamma \times S^1)\big) \subset (X, Z)$ for simple closed curves $\gamma \subset \Sigma'$ which shows that $\delta H_3(X, Z)$ agrees with $\mathcal{R}$. □

As far as our goal to describe everything as directly as possible in terms of $\mathfrak{S}$ is concerned, Proposition 9.11 leaves something to be desired. In fact, even the knowledge of a closing curve is not enough since it only gives information about the fiber class. The main problem is the appearance of the rim tori whose relation to $\mathfrak{S}$ remains intangible. In principle, they should be visible in either description of $H_2(Z)$ in Lemmas 9.4 and 9.8, but it is not at all clear how they would appear. Hopefully, the future will shed some light on this problem.

However, we can still say something useful about the intersection form of X. To begin with, we can determine its signature.

Corollary 9.12. *The signature of X agrees with the signature of $Q_{\mathfrak{S}}$. Moreover, $\sigma(X)$ only depends on the homology classes $[\vec{c}_i] \in H_1(\Sigma)$.*

Proof. On the one hand, Novikov's additivity theorem shows that $\sigma(X)$ is the sum of $\sigma(Z)$ and $\sigma(\Sigma' \times D^2)$, but the latter is obviously zero. On the other hand, Proposition 9.6 shows that $\sigma(Z)$ agrees with the signature of $Q_{\mathfrak{S}}$. Lastly, the expression for $Q_{\mathfrak{S}}$ given in equation (9.4) only involves the homology classes of the vanishing cycles. □

As a consequence, we can compute the signature of X directly from $\mathfrak{S}$ without even knowing a closing curve. In fairness, we should mention that this still requires some work in practice. One has to choose a basis of $K_{\mathfrak{S}}$, evaluate $Q_{\mathfrak{S}}$ to obtain a matrix, and then compute its signature. If $K_{\mathfrak{S}}$ has high rank, then this process

is almost impossible to carry out by hand and the help of a computer becomes necessary. It would be more desirable to have a description of $\sigma(X)$ that bypasses the full computation of $Q_{\mathfrak{S}}$, ideally in the form of a closed formula. Unfortunately, we have not been able to obtain such a formula yet.

We now try to describe the intersection form of X. To state the results, we need some algebraic terminology which is discussed in more detail in Appendix A. More precisely, we consider the *radical* of Q_Z

$$\operatorname{rad}(Q_Z) = \{x \in H_2(Z) \,|\, Q_Z(a,b) = 0 \text{ for all } b \in H_2(Z)\},$$

and the form induced by Q_Z on the quotient $H_2(Z)/\operatorname{rad}(Q_Z)$. We call this the *reduced form* of Q_Z and denote it by Q_Z^{red}.

We follow the same strategy as in the computation of Q_Z and study the linking form on the 2–chains in the handle complex induced by a Kirby diagram. We take the Kirby diagram of Z used in the proof of Proposition 9.6 and complete it to a diagram for X by including the last 2–handle as explained in Section 6.2. Specifically, we draw the attaching curve by resolving the double points of a meridian for the fiber 2–handle followed by the image of κ in a fiber in the diagram above all the fold handles. Then we have

$$C_2(X) \cong \mathbb{Z}^2 \oplus V_{\mathfrak{S}}$$

where the two copies of $\mathbb{Z}$ are generated by the fiber 2–handle and the last 2–handle which we denote by F and L, respectively. As before, we denote the linking form on $C_2(X)$ by lk.

By definition, $\lambda = \operatorname{lk}(L,L)$ is the framing coefficient of the last 2–handle. Moreover, the discussion on page 80 shows that the last 2–handle links the fiber 2–handle algebraically and geometrically once. In particular, we have $\operatorname{lk}(L,F) = 1$. Finally, the linking between the last 2–handle and the fold handles is given by

$$\nu_i = \operatorname{lk}(L,\vec{c}_i) = \frac{1}{2}\left(\langle \kappa, \vec{c}_i \rangle_\Sigma + \sum_{i=1}^{g} \left[n_{a_i}(\kappa)\, n_{b_i}(\vec{c}_i) + n_{a_i}(\vec{c}_i)\, n_{b_i}(\kappa)\right]\right)$$

which follows from a similar count of crossings as in the proof of equation (6.2) in Lemma 6.4. Note that $\langle \kappa, \vec{c}_i \rangle_\Sigma$ appears with positive sign since the last 2–handle is attached above the fold handles. Altogether, the linking matrix takes the form

$$\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & \lambda & \nu_1 & \dots & \nu_l \\ 0 & \nu_1 & & & \\ \vdots & \vdots & & \operatorname{lk}(\vec{c}_i,\vec{c}_j) & \\ 0 & \nu_l & & & \end{pmatrix}$$

where $\operatorname{lk}(\vec{c}_i,\vec{c}_j)$ is given by equation (9.5).

Now let us look at the 2–cycles in the handle complex of X. As before, these contain F and $K_{\mathfrak{S}}$ but there might be an extra cycle caused by the last 2–handle. Observe that $\partial_2 L \in C_1(X)$ corresponds to $[\kappa] \in H_1(\Sigma)$ under the identification $C_1(X) \cong H_1(\Sigma)$ which follows just as in the case of the fold handles from the definition of ∂_2 and Remark 6.3. So a relation of the form

$$d[\kappa] = \sum_i \alpha_i [\vec{c}_i] \in H_1(\Sigma) \tag{9.7}$$

gives rise to a 2–cycle $dL - \sum_i \alpha_i \vec{c}_i$ in the handle complex. But by Lemma 9.10 such a relation exists if and only if the fiber class $[\Sigma] \in H_2(X)$ is essential. We thus arrive at the following conclusion.

Lemma 9.13. *If $[\Sigma] \in H_2(X)$ is torsion, then Q_X is isomorphic to $Q_{\mathfrak{S}}^{\mathrm{red}}$.*

Proof. In this situation the last 2–handle does not contribute a 2–cycle so that the 2–cycle group of X is the same as that of Z. According to Proposition 9.6 the restricted linking form is isomorphic to $\mathbb{Z} \oplus Q_{\mathfrak{S}}$ and Proposition A.5 implies that

$$Q_X \cong \big((0) \oplus Q_{\mathfrak{S}}\big)^{\mathrm{red}} \cong Q_{\mathfrak{S}}^{\mathrm{red}}.$$

Note that the passage to the reduced form is where the fiber class is killed (rationally) since it is contained in the radical of Q_Z. □

So if the fiber class is torsion, then the last 2–handle does not cause any problems and the intersection form of X can be computed solely in terms of $\mathfrak{S}$. In the case when the fiber is essential, we have gathered all the necessary information to compute Q_X. Given a relation of the form (9.7) we get an additional 2–cycle $C = dL - \sum_i \alpha_i c_i$ and the 2–cycle group is given by $\mathbb{Z}^2 \oplus K_{\mathfrak{S}}$ where the first summand is now generated by F and C (not L!). Unfortunately, the expressions for the linking of C with itself and elements of $K_{\mathfrak{S}}$ are not very enlightening. For example, the self-linking is given by

$$\gamma = \mathrm{lk}(C, C) = d^2\lambda + \sum_i \alpha_i^2 \mathrm{ff}(c_i) - 2d \sum_i \alpha_i \mathrm{lk}(L, \vec{c}_i)$$

where $\mathrm{ff}(c_i)$ is the framing coefficient of the corresponding fold handle. However, since the fiber 2–handle does not link the fold handles, we have $\mathrm{lk}(F, C) = d$ so that the linking form on the 2–cycles takes the form

$$\begin{pmatrix} 0 & d & 0 & \dots & 0 \\ d & \gamma & * & \dots & * \\ 0 & * & & & \\ \vdots & \vdots & & Q_{\mathfrak{S}} & \\ 0 & * & & & \end{pmatrix}$$

and in order to obtain the intersection form of X we have to divide out this form by its radical. In general, this is all we can say but there is one more special case where a more concise description is available, namely if $d = 1$. The proof of Lemma 9.10 shows that this is equivalent to the primitiveness of the fiber class. In this situation the unknown contributions in the above matrix can be removed by a base change and we obtain the following statement.

Lemma 9.14. *If $[\Sigma] \in H_2(X)$ is primitive, then Q_X is isomorphic to $\left(\begin{smallmatrix} 0 & 1 \\ 1 & \gamma \end{smallmatrix}\right) \oplus Q_{\mathfrak{S}}^{\mathrm{red}}$.*

Finally, we note that the discussion of the intersection form also gives us a little more information about the group $H_2(X)$. Indeed, the intersection form is defined on $H_2(X)/\mathrm{tors}$ which is related to the $K_{\mathfrak{S}}^{\mathrm{red}} = K_{\mathfrak{S}}/\mathrm{rad}(Q_{\mathfrak{S}})$ as follows.

Corollary 9.15. *Depending on whether $[\Sigma] \in H_2(X)$ is torsion or essential we have*

$$H_2(X)/\mathrm{tors} \cong K_{\mathfrak{S}}^{\mathrm{red}} \quad \textit{or} \quad H_2(X)/\mathrm{tors} \cong K_{\mathfrak{S}}^{\mathrm{red}} \oplus \mathbb{Z}^2.$$

9.2 Betti Numbers and an Obstruction for Trivial Monodromy

We now derive formulas for the Betti numbers of X and use them to produce an obstruction for surface diagrams to have trivial monodromy. The formulas involve the following quantities.

- $\mathrm{rk}(\mathfrak{S})$: the rank of the subgroup $\langle c_1, \dots, c_l \rangle \subset H_1(\Sigma)$, called the *rank* of $\mathfrak{S}$
- $\mathrm{rk}(Q_{\mathfrak{S}})$: the rank of the symmetric bilinear form $Q_{\mathfrak{S}}$ defined in equation (9.4)
- $\delta \in \{0, 1\}$: defined to be zero if $[\Sigma] \in H_2(X)$ is torsion and one otherwise.

Lemma 9.16. *The Betti numbers of X are given by*

$$b_1(X) = 2g(\Sigma) - \mathrm{rk}(\mathfrak{S}) - (1 - \delta)$$
$$b_2(X) = \mathrm{rk}(Q_{\mathfrak{S}}) + 2\delta.$$

Proof. Recall from Lemma 9.1 that

$$H_1(X) \cong H_1(\Sigma)/\langle c_1, \dots, c_l, \kappa \rangle.$$

Moreover, it follows from Lemma 9.10 that $\delta = 0$ if and only if κ is linearly independent of the vanishing cycles. The formula for $b_1(X)$ follows. The claim about $b_2(X)$ follows from Corollary 9.15 since $\mathrm{rk}(Q_{\mathfrak{S}}) = \mathrm{rk}(K_{\mathfrak{S}}^{\mathrm{red}})$. □

Corollary 9.17. *If $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$ has trivial monodromy, then*

$$l = 2\,\mathrm{rk}(\mathfrak{S}) + \mathrm{rk}(Q_{\mathfrak{S}}) - 2.$$

In particular, we have $l \geq 2\,\mathrm{rk}(\mathfrak{S}) - 2$.

Proof. According to equation (9.1) we have

$$\chi(X) = 2 - 2b_1(X) + b_2(X) = 6 - 4g(\Sigma) + l.$$

Inserting the expressions for $b_1(X)$ and $b_2(X)$ from Lemma 9.16 and solving for l finishes the proof. □

So roughly speaking, if $\mathfrak{S}$ is supposed to describe a closed 4–manifold, then only every second curve in $\mathfrak{S}$ is allowed to generate a new class in $H_1(\Sigma)$. We illustrate this obstruction for trivial monodromy in a simple example.

Example 9.18. Consider the surface diagram $\mathfrak{C}_g = (\Sigma_g; c_1, \dots, c_{2g})$, $g \geq 2$, shown in Figure 36. This very symmetric diagram is obtained from a non-separating chain of length $2g - 1$ in Σ_g by adding one extra curve c_{2g} which produces a W–cycle. In particular, we have $\mathrm{rk}(\mathfrak{C}_g) = 2g-1$ and Corollary 9.17 shows that $\mathfrak{C}_g$ must have non-trivial monodromy for $g \geq 3$. On the other hand, $\mathfrak{C}_2$ agrees with the diagram $\mathfrak{Ch}_{2,4}$ from Example 6.9 where we saw that it actually has trivial monodromy.

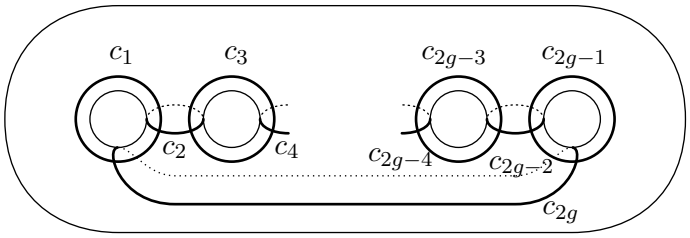


Figure 36: The surface diagram $\mathfrak{C}_g$ from Example 9.18.

9.3 Spin and Spin^c Structures

So far we have focused on computations of homotopy invariants of Z and X. We now shift our attention toward structures on their tangent bundles, namely spin and spin^c structures. Recall that a spin structure on an oriented Riemannian manifold M of dimension n can be considered as a lift of the classifying map $M \to \mathrm{BSO}(n)$ of the tangent bundle to $\mathrm{BSpin}(n)$. In this picture two spin structures are isomorphic if their corresponding lifts are homotopic. We denote the set of isomorphism classes of spin structures by $\mathrm{Spin}(M)$. We say that M is *spin* if it admits a spin structure which is equivalent to the vanishing of the second Stiefel–Whitney class $w_2(M) \in H^2(M;\mathbb{Z}_2)$. Moreover, if M is spin, then $\mathrm{Spin}(M)$ is a torsor over $H^1(M;\mathbb{Z}_2)$, that is, it admits a free and transitive action of this group. More precisely, for $\xi, \xi' \in \mathrm{Spin}(M)$ their difference is measured by a class $\delta_M(\xi', \xi) \in H^1(M;\mathbb{Z}_2)$, and for fixed ξ the map $\delta_M(\cdot, \xi)$ is a bijection between $\mathrm{Spin}(M)$ and $H^1(M;\mathbb{Z}_2)$.

The discussion of spin^c structures follows the same pattern. In the definition one only has to replace $\mathrm{Spin}(n)$ with the group $\mathrm{Spin}^c(n) = (\mathrm{Spin}(n) \times S^1)/\mathbb{Z}_2$. It is known that M admits a spin^c structure if and only if $w_2(M)$ is the mod 2 reduction of a class in $H^2(M;\mathbb{Z})$, and if so, then the set of equivalence classes of spin^c structures $\mathrm{Spin}^c(M)$ is a torsor over $H^2(M;\mathbb{Z})$. The difference of $\mathfrak{s}, \mathfrak{s}' \in \mathrm{Spin}^c(M)$ is measured by a class $\Delta_M(\mathfrak{s}', \mathfrak{s}) \in H^2(M;\mathbb{Z})$. An extra feature is that each $\mathfrak{s} \in \mathrm{Spin}^c(M)$ comes equipped with a complex line bundle $\det(\mathfrak{s})$, the so called *determinant line bundle*, whose first Chern class is denoted by $c_1(\mathfrak{s})$.

While spin structures fundamental objects in manifold topology, spin^c structures are especially important in the 4–dimensional context. It is well known that spin^c structures exist on all oriented 4–manifolds and that they are closely tied to the topology of 4–manifolds. For example, for a closed 4–manifold X the set $\{c_1(\mathfrak{s}) \,|\, \mathfrak{s} \in \mathrm{Spin}^c(X)\}$ is Poincaré dual to the set of characteristic elements[2] for the intersection pairing. Moreover, spin^c structures are basic ingredients in the definitions of the delicate invariants coming from Seiberg–Witten theory and Heegaard–Floer theory.

9.3.1 Spin Structures and Surface Diagrams

We first discuss how spin structures can be described in terms of surface diagrams. Our approach is essentially the same as that taken by Stipsicz in [60] who characterizes spin structures on (total spaces of) Lefschetz fibrations in terms of their vanishing cycles by analyzing the handle decomposition induced by Lefschetz fibrations.

[2] $\xi \in H_2(X)$ is called *characteristic* if $x \cdot x \equiv \xi \cdot x \mod 2$ for al $x \in H_2(X)$.

Since the structure of the handle decompositions obtained from simple wrinkled fibrations only differs in the framings, the arguments in [60] can be translated to this setting almost word by word.

As a preparation, we need to recall some more generalities about spin structures. As before, let M be an oriented Riemannian manifold. If $K \subset M$ is an oriented submanifold with trivial normal bundle, then a trivialization of the normal bundle induces a restriction map $\mathrm{Spin}(M) \to \mathrm{Spin}(K)$. In particular, if K has codimension ≤ 1, then the normal bundle is canonically trivialized (by the orientation and Riemannian metric of M) and spin structures on M have well-defined restrictions to K. Now suppose that $M = M_1 \cup_N M_2$ where $M_i \subset M$ are codimension 0 submanifolds with common boundary $\partial M_i = N$. Then any $\xi \in \mathrm{Spin}(M)$ restricts to $\xi_i = \xi|_{M_i} \in \mathrm{Spin}(M_i)$, $i = 1, 2$, such that $\xi_1|_N = \xi_2|_N$ where N is equipped with the boundary orientation of M_1. Moreover, if $\xi' \in \mathrm{Spin}(M)$ satisfies $\xi'|_{M_i} = \xi|_{M_i}$, $i = 1, 2$, then the difference class $\delta_M(\xi', \xi)$ must be contained in the image of the connecting homomorphism $\delta\colon H^0(N; \mathbb{Z}_2) \to H^1(M; \mathbb{Z}_2)$ of the Mayer–Vietoris sequence of the decomposition $M = M_1 \cup_N M_2$. In particular, if N is connected, then this image is trivial so that $\xi \in \mathrm{Spin}(M)$ is uniquely determined by its restrictions to M_i. Conversely, any pair $\xi_i \in \mathrm{Spin}(M_i)$ with $\xi_1|_N = \xi_2|_N$ can be glued together to a spin structure on M and the gluing is unique if N is connected.

Finally, we need to be familiar with spin structures in low dimensions. First of all, in dimension one it is enough to understand the circle, which admits two spin structures for each orientation: there is the *trivial* spin structure induced from the (unique) spin structure on the disk, and another one which does not extend over any surface bounded by the circle. In order to understand spin structures on an oriented surface Σ we argue as follows. Given $\xi \in \mathrm{Spin}(\Sigma)$ there is a well defined map

$$q_\xi\colon H_1(\Sigma; \mathbb{Z}_2) \to \mathbb{Z}_2$$

which has the property that for a homology class represented by an oriented simple closed curve $c \subset \Sigma$ we have

$$q_\xi([c]) = \begin{cases} 1 & \text{if } \xi|_c \text{ is non-trivial} \\ 0 & \text{if } \xi|_c \text{ is trivial} \end{cases}$$

and satisfies the equation

$$q_\xi(x + y) = q_\xi(x) + q_\xi(y) + \langle x, y\rangle_2 \tag{9.8}$$

where $\langle \cdot, \cdot \rangle_2$ is the intersection form on $H_1(\Sigma; \mathbb{Z}_2)$. Functions from $H_1(\Sigma; \mathbb{Z}_2)$ to $\mathbb{Z}_2$ satisfying equation (9.8) are called quadratic refinements of the intersection form of Σ, or for brevity just a *quadratic refinements* on Σ. It is well known that the sending ξ to q_ξ gives a bijective correspondence between spin structures and quadratic refinements on Σ [38].

After these preliminaries, we return to the notation set up in the beginning of the chapter on page 101 and attempt to describe spin structure on Z in terms of $\mathfrak{S}$.

Proposition 9.19. *Spin structures on Z correspond bijectively to those on Σ whose restriction to each vanishing cycle c_i in $\mathfrak{S}$ is trivial. Moreover, Z is spin if and only*

if any relation of the form $c_{i_1} + \cdots + c_{i_r} = 0 \in H_1(\Sigma; \mathbb{Z}_2)$ *implies*

$$\sum_{k<l} \langle c_{i_k}, c_{i_l} \rangle_2 = 0. \tag{9.9}$$

As mentioned above, the proof is essentially the same as that of [60, Theorem 1.1] with only minor adjustments. For the reader's convenience we repeat the arguments.

Proof. We think of Σ as a subspace of Z and try to extend a given spin structure $\xi \in \mathrm{Spin}(\Sigma)$ to all of Z. Obviously, ξ extends uniquely to $\Sigma \times D^2$ and it remains to understand when we can extend further across the fold handles. Since the fold handle corresponding to c_i is attached with fiber framing, it follows that the unique spin structure on the handle induces the trivial spin structure on c_i. Consequently, ξ extends across the fold handles if and only if it restricts to the trivial spin structure on each c_i. Conversely, every spin structure on Z restricts to Σ and is trivial on all vanishing cycles because these spin bound the cores of the fold handles.

For the second claim we work with quadratic refinements. Note that for any quadratic refinement q on Σ satisfies

$$q(c_{i_1} + \cdots + c_{i_r}) = q(c_{i_1}) + \ldots q(c_{i_r}) + \sum_{k<l} \langle c_{i_k}, c_{i_l} \rangle_2 \tag{9.10}$$

for an arbitrary sum of vanishing cycles. This follows from iterating equation (9.8). Now, if Z is spin and $\xi \in \mathrm{Spin}(\Sigma)$ is the restriction of a spin structure on Z, then the discussion above shows that the corresponding quadratic refinement satisfies $q_\xi(c_i) = 0$ which implies the condition (9.9). On the other hand, if condition (9.9) is satisfied, then we can define a quadratic refinement on Σ as follows. Let $V \subset H_1(\Sigma; \mathbb{Z}_2)$ be the subspace generated by the vanishing cycles and let V' be a complementary subspace. We first choose a basis for V consisting of vanishing cycles $c_{i_1}, \ldots, c_{i_r}$ and define $q(c_{i_j}) = 0$. Then we choose a basis for V' and define q randomly on the basis elements. Finally, we extend q to $H_1(\Sigma)$ according to equation (9.8) to obtain a quadratic refinement. In order to see that the corresponding spin structure on Σ extends to Z, we have to show that for every vanishing cycle c_j that was not among the a basis elements we still have $q(c_j) = 0$. But this follows from condition (9.9) and equation (9.10). □

The condition (9.9) has another interpretation. According to Wu's formula, for a 4–manifold M the second Stiefel–Whitney class $w_2(M)$ evaluates on an element of $H_2(M; \mathbb{Z}_2)$ as the mod 2 self-intersection. In particular, M is spin if and only if all mod 2 self-intersections vanish. Now, an easy adaption of the proofs of Lemma 9.4 and Proposition 9.6 to the mod 2 setting shows that

(a) the left hand side of equation (9.9) is the mod 2 self intersection of a class in $H_2(Z; \mathbb{Z}_2)$ determined by the relation $c_{i_1} + \cdots + c_{i_r} = 0 \in H_1(\Sigma; \mathbb{Z}_2)$, and

(b) $H_2(Z; \mathbb{Z}_2)$ is generated by such classes and the class of the fiber $\Sigma \subset Z$ (which obviously has trivial self-intersection).

This point of view also indicates how to pass to X.

Lemma 9.20. *X is spin if and only if Z is spin and either $[\Sigma]_2 = 0 \in H_2(X; \mathbb{Z}_2)$ or there is a class $x \in H_2(X; \mathbb{Z}_2)$ with $(x \cdot [\Sigma]_2)_2 = 1$ and $(x \cdot x)_2 = 0$.*

The following proof is inspired by that of [60, Theorem 1.3].

Proof. If X is spin, then the other conditions are obviously satisfied. Conversely, assume first that Z is spin and that Σ is trivial in $H_2(X;\mathbb{Z}_2)$. Then the homology sequence of the pair (X,Z) shows that the map $H_2(Z;\mathbb{Z}_2) \to H_2(X;\mathbb{Z}_2)$ is surjective. In particular, all mod 2 self-intersections in X can be computed in Z and must therefore be even. So X is spin. Now suppose that Z is spin but $[\Sigma]_2 = [\Sigma']_2$ is non-trivial in $H_2(X;\mathbb{Z}_2)$. Then we can find a (possibly non-orientable) surface $F \subset X$ that intersects Σ' transversely in one point. A neighborhood N of $\Sigma' \cup F$ can be identified with a plumbing of D^2–bundles and its mod 2 intersection form is represented by the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & n \end{smallmatrix}\right)$ where n is the mod 2 self-intersection of F. In particular, N is spin if and only $n = 0$. Moreover, we can assume that the complement $X \setminus N$ is a subset of Z so that every spin structure on Z induces one on ∂N. Now, since the intersection form of N is unimodular (over $\mathbb{Z}_2$), an investigation of the cohomology sequence of the pair $(N, \partial N)$ shows that $H^1(N;\mathbb{Z}_2) \to H^1(\partial N;\mathbb{Z}_2)$ is surjective. But this implies that, if N is spin, then every spin structure on ∂N can be extended over N. □

Remark 9.21. Similar results to Proposition 9.19 and Lemma 9.20 have also been obtained independently by Hayano [34, Corollary 5.3] using the handle decompositions coming from broken Lefschetz fibrations.

Again, the question whether or not the fiber $\Sigma \subset X$ is trivial in $H_2(X;\mathbb{Z}_2)$ is related to properties of a closing curve $\kappa \subset \Sigma$ by a mod 2 version of Lemma 9.10.

Lemma 9.22. *The fiber $\Sigma \subset X$ is trivial in $H_2(X;\mathbb{Z}_2)$ if and only if the closing curve $\kappa \subset \Sigma$ is linearly independent of the vanishing cycles in $H_1(\Sigma;\mathbb{Z}_2)$.*

However, if the fiber is non-zero in $H_2(X;\mathbb{Z}_2)$, then the mod 2 square of a dual remains mysterious.

9.3.2 Spin^c Structures and Simple Wrinkled Fibrations

Next we focus on spin^c structures. Here we take a slightly different approach and establish a relation between spin^c structures and simple wrinkled fibrations over S^2. This is done by a minor modification of a construction which was originally used by Perutz [58, p.1500] in the context of broken Lefschetz fibrations. In fact, elaborating on Perutz's ideas, we show in Appendix B how spin^c structures interact with a rather general class of singular fibration structures on closed 4–manifolds.

We briefly summarize the results of Appendix B in the special case of a simple wrinkled fibration $w\colon X \to S^2$. The first thing to note is that w restricts to a submersion outside the critical circle $\mathcal{C}_w$ and thus induces a spin^c structure $\mathfrak{s}_w$ on $X \setminus \mathcal{C}_w$. It follows from Lemma B.4 that $\mathfrak{s}_w$ does not extend across $\mathcal{C}_w$ so that there is no canonical spin^c structure induced by w. But $\mathfrak{s}_w$ can still be used to parametrize spin^c structures on X as follows. Every $\mathfrak{s} \in \mathrm{Spin}^c(X)$ can be compared to $\mathfrak{s}_w$ after restricting to $X \setminus \mathcal{C}_w$. The difference is measured by a class in $H^2(X \setminus \mathcal{C}_w)$ whose Poincaré dual we denote by $\tau_w(\mathfrak{s}) \in H_2(X, \mathcal{C}_w)$. This construction gives rise to the so called *Taubes map*

$$\tau_w\colon \mathrm{Spin}^c(X) \longrightarrow H_2(X, \mathcal{C}_w).$$

In Proposition B.1 we show that τ_w is injective and that its image consists of those elements in $H_2(X, \mathcal{C}_w)$ that are mapped to $[\mathcal{C}_w] \in H_1(\mathcal{C}_w)$ by the connecting homomorphism $\delta\colon H_2(X, \mathcal{C}_w) \to H_1(\mathcal{C}_w)$ of the homology sequence of the pair $(X, \mathcal{C}_w)$. Here we use the orientation on $\mathcal{C}_w$ obtained by lifting the orientation of the critical image $w(\mathcal{C}_w) \subset S^2$ considered as the boundary of the higher genus region. In more geometric terms, the Taubes map identifies spin^c structures on X with homology classes of "Seifert surfaces" for $\mathcal{C}_w$ in X, that is, oriented surfaces with oriented boundary $\mathcal{C}_w$. This observation together with the fact that all smooth 4–manifolds admit spin^c structures leads to the following surprising conclusion.

Lemma 9.23. *Let $w\colon X \to S^2$ be a simple wrinkled fibration. Then the critical circle $\mathcal{C}_w \subset X$ is null-homologous.*

Now suppose that $\mathfrak{S}$ is a surface diagram for (X, w). If we want to describe spin^c structures on X in terms of $\mathfrak{S}$, then the above discussion tells us to understand how the critical circle $\mathcal{C}_w$ and its Seifert surfaces appear in $\mathfrak{S}$. Unfortunately, this is not as straightforward as one might hope and remains an open problem at the time of writing.

9.4 Smooth 4–Manifolds and Torelli Groups

Using what we have learned about homotopy information in surface diagrams, we can finally prove our last main result.

Theorem 1.5. *Let $w\colon X \to S^2$ be a simple wrinkled fibration with surface diagram $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$. If X is simply connected and $[\Sigma] = 0 \in H_2(X)$, then the homeomorphism type of X is determined by the homology classes $[c_i] \in H_1(\Sigma)$.*

Note that the condition on the fiber does not put any restrictions on X. Indeed, according to Theorem 4.2, the homotopy class of constant maps $X \to S^2$ contains simple wrinkled fibrations and their fibers are clearly null-homologous in X.

Proof. According to the theorems of Freedman, Donaldson and Serre stated in the introduction, the homeomorphism type of X is determined by its Euler characteristic, signature, and type (that is, whether X is spin or not). As we have seen in equation (9.1) and Corollary 9.12, $\chi(X)$ depends only on the genus of Σ and the number of curves in $\mathfrak{S}$, while $\sigma(X)$ agrees with the signature of $Q_{\mathfrak{S}}$ which, in turn, only depends on the homology classes of the vanishing cycles. As for the type, since Σ is null-homologous in X, Lemma 9.20 shows that X is spin if and only if Z is spin. But this depends only on the mod 2 reductions $[c_i]_2 \in H_2(\Sigma; \mathbb{Z}_2)$ according to Proposition 9.19. □

What makes Theorem 1.5 interesting is that the diffeomorphism type of X a priori depends on the *isotopy classes* of the vanishing cycles. In particular, it is conceivable that a homologous substitution, that is, changing some vanishing cycles in $\mathfrak{S}$ within their homology classes, might produce an exotic copy of X.

Definition 9.24. Two surface diagrams $(\Sigma; c_1, \dots, c_l)$ and $(\Sigma; c_1', \dots, c_l')$ are called *homologous* if $[c_i] = [c_i'] \in H_1(\Sigma)$ (for some choice of orientations) for $i = 1, \dots, l$.

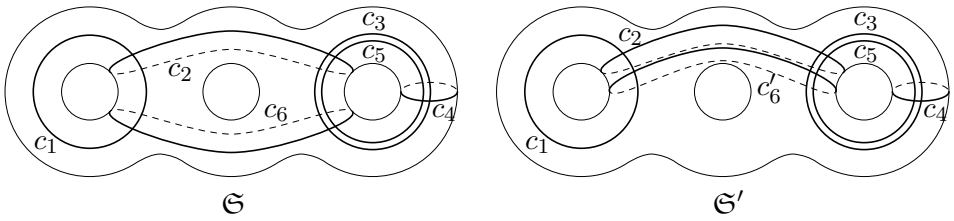


Figure 37: Homologous surface diagrams.

On a different note, the difference between homology and isotopy for simple closed curves in Σ is measured by the so called *Torelli group* $\mathcal{I}(\Sigma)$ – this is the subgroup of $\mathrm{Mod}(\Sigma)$ consisting of all mapping classes whose action on homology is trivial. Indeed, two non-separating simple closed curves in Σ are homologous if and only if they can be mapped onto each other by an element of $\mathcal{I}(\Sigma)$ (see [24, Proposition 6.14]). The Torelli groups can be considered as the non-linear and more mysterious parts of mapping class groups and they are an active area of research. Not surprisingly, their structure is quite complicated; for example, it is not known whether the $\mathcal{I}(\Sigma_g)$ is finitely presented for $g \geq 3$.[3] With a good amount of optimism, one can take Theorem 1.5 as a pointer that there might be a relation between Torelli groups and exotic smooth structures on 4–manifolds. But as usual in 4–manifold topology, there are subtleties. First, it is unlikely that the monodromy of $\mathfrak{S}$ depends only on the homology classes of the vanishing cycles and homologous substitutions could destroy trivial monodromy. (It is interesting to note that such a failure of trivial monodromy would happen below the radar of Corollary 9.17.) Second, even if some particular substitution preserves the trivial monodromy condition, and thus actually produces a new closed 4–manifold, it might change the fundamental group as in Example 9.25 below. Nevertheless, if one manages to control the monodromy and the fundamental group, then homologous substitutions can produce exotic copies.

Example 9.25. We consider the two surface diagrams

$$\mathfrak{S} = (\Sigma_3; c_1, c_2, c_3, c_4, c_5, c_6) \quad \text{and} \quad \mathfrak{S}_{\log} = (\Sigma_3; c_1, c_2, c_3, c_4, c_5, c_6')$$

shown in Figure 37. Obviously, these diagrams are homologous. Indeed, $\mathfrak{S}_{\log}$ is obtained from $\mathfrak{S}$ by substituting c_6 with c_6' and these curves bound a two-holed torus $T \subset \Sigma_3$. Moreover, both diagrams are known to have trivial monodromy and, since the genus is three, they describe unique closed 4–manifolds. In Example 7.11 we saw that $\mathfrak{S}$ represents $S^1 \times S^3$, while $\mathfrak{S}_{\log}$ appeared in Example 6.9 under the name $\mathfrak{Ch}_{3,4}$ (the double of a non-separating chain of length 4 in Σ_3) and it describes $T^2 \times S^2$. In particular, we see that the fundamental groups are different.

However, this substitution is interesting for another reason. If we write $T^2 \times S^2$ as $S^1 \times (S^1 \times S^3)$, then a relation between $T^2 \times S^2$ and $S^1 \times S^3$ becomes apparent. Indeed, since $S^1 \times S^2$ is obtained from S^3 by Dehn surgery on a 0–framed unknot $U \subset S^3$, we see that $T^2 \times S^2$ is obtained from $S^1 \times S^3$ by "S^1 times Dehn surgery"; this operation is also known as a *logarithmic transformation (of multiplicity* 0*)* along the embedded torus $S^1 \times U \subset S^1 \times S^3$ (see [32, p.310]). Another embedded torus $\widehat{T} \subset S^1 \times S^3$ is visible in the surface diagram $\mathfrak{S}$. It is obtained

[3] $\mathcal{I}(\Sigma_2)$ is a countably generated free group [49] while $\mathcal{I}(T^2)$ is trivial [24, Theorem 2.5].

by capping off the holed torus T bounded by c_2 and c_4 using the cores of the fold handles. Moreover, it follows from Proposition 9.6 that $\widehat{T}$ has trivial self-intersection ($\widehat{T}$ is represented by $c_2 + c_4 \in K_{\mathfrak{S}}$ and the right hand side of equation (9.4) vanishes its square since c_2 and c_4 are disjoint.) We suspect that the tori $\widehat{T}$ and $S^1 \times U$ are isotopic and that the substitution $(c_6) \rightsquigarrow (c_6')$ corresponds to the logarithmic transformation on $\widehat{T}$. Unfortunately, we have been unable to prove this so far.

Altogether, it seems feasible to realize some logarithmic transformations as homologous substitutions. Since it is well known that logarithmic transformations can produce exotic copies (see [32, Ch. 3.3], for example), this is one possible approach to construct exotic 4–manifolds via homologous substitutions. We plan to continue this train of thought in future research.

Appendix

Appendix A

Kirby Diagrams and Intersection Forms

In this appendix we show how the intersection form of a 4–manifold can be determined from a Kirby diagram. This material is probably well known among experts but we could not find an account in the literature that covers the most general situation with the presence of 1– and 3–handles. We assume basic knowledge of handlebody theory and Kirby diagrams and we use the same notation as in Section 2.1.

Let X be a 4–dimensional handlebody and let $(\mathcal{U}, \mathcal{L})$ be a Kirby diagram for the 2–skeleton $X^{\leq 2}$. We orient the cores of all handles and consider the handle complex. The orientations of the cores induce orientations for all components of $\mathcal{U}$ and $\mathcal{L}$ so that we have well defined *linking numbers*. Recall that the linking number of two disjoint, oriented knots $K, L \subset S^3$ is an integer $\mathrm{lk}(K, L) \in \mathbb{Z}$ which has several different descriptions. We will use the following two:

- If Σ is a Seifert surface for K, then $\mathrm{lk}(K, L) = [\Sigma] \cdot [L]$ (the intersection number of Σ and L computed in $S^3 \setminus K$). (see [32, Proposition 4.5.5])
- The group $H_1(S^3 \setminus K)$ is generated by a right-handed meridian μ for K and we have $[L] = \mathrm{lk}(K, L)[\mu] \in H_1(S^3 \setminus K)$. (see [32, Definition 4.5.1])

The general idea is then to relate the intersection form of X to linking information in the Kirby diagram, more precisely to the *linking form* of $\mathcal{L}$. This is the symmetric bilinear form on the free Abelian group generated by the components of $\mathcal{L}$ defined by the matrix of linking numbers $\mathrm{lk}(L_i, L_j)$ where $\mathrm{lk}(L_i, L_i) = f_i$. The following well known fact is a prime example for this connection between 4–dimensional topology and knot theory (see [32, Proposition 4.5.11]).

Proposition A.1. *Let X be a 4–dimensional handlebody without 1– and 3–handles and let $\mathcal{L}$ be the framed link which constitutes the Kirby diagram of X. Then the intersection form of X is isomorphic to the linking form of $\mathcal{L}$.*

Proof (sketch): Since there are no 1– and 3–handles, the handle complex has trivial differentials and $H_2(X)$ is isomorphic to the 2–chains $C_2(X)$. Moreover, since the 2–handles are in one to one correspondence with the components of $\mathcal{L}$, the linking form can be considered as a form on the 2–chains of the handle complex. Now, each

component L_i bounds a Seifert surface Σ_i in S^3 which can be capped off to a closed surface $\widehat{\Sigma}_i$ in X by adding the core of the 2–handle attached along L_i. The thus obtained surfaces, or rather their homology classes, constitute a basis of $H_2(X)$ and their intersection numbers can be related to linking numbers by carefully pushing the Seifert surfaces into D^4. For more details see [32, Proposition 4.5.11]. □

In the general situation where X might have 1– and 3–handles, the linking form of $\mathcal{L}$ can still be considered as a form

$$\mathrm{lk}\colon C_2(X) \times C_2(X) \longrightarrow \mathbb{Z}$$

but the relation to the intersection form becomes less obvious. This is mostly because the computation of $H_2(X)$ from the handle complex

$$C_3(X) \xrightarrow{\partial_3} C_2(X) \xrightarrow{\partial_2} C_1(X)$$

becomes non-trivial. We first establish a generalization of Proposition A.1 which allows the presence of 1–handles.

Proposition A.2. *Let X be a 4–dimensional handlebody with Kirby diagram $(\mathcal{U}, \mathcal{L})$. If there are no 3–handles, then the intersection form of X is isomorphic to the linking form of $\mathcal{L}$ restricted to the 2–cycles $\ker \partial_2 \subset C_2(X)$.*

For the lack of a better name, we will call the restriction of lk to $\ker \partial_2$ the *restricted linking form* of $\mathcal{L}$. For the proof of Proposition A.2 we will need two lemmas that relate the algebra of the handle complex to the geometry of the Kirby diagram. The first one gives a description of the differential $\partial_2\colon C_2(X) \to C_1(X)$ in terms of linking numbers.

Lemma A.3. *Using the canonical identification of $C_2(X)$ and $C_1(X)$ with the free Abelian groups generated by the components of $\mathcal{L}$ and $\mathcal{U}$ we have*

$$\partial_2 L_i = \sum_j \mathrm{lk}(L_i, U_j) U_j.$$

Proof. By definition, the coefficient of U_j in $\partial_2 L_i$ is given by the intersection number of the belt sphere of the 1–handle corresponding to U_j and the attaching sphere of the 2–handle corresponding to L_i computed in $\partial X^{\leq 1}$. But the belt sphere appears in the Kirby diagram as a spanning disk for U_j while the attaching sphere is L_i itself. So the relevant intersection number agrees with $\mathrm{lk}(L_i, U_j)$. □

We also need a geometric interpretation of what it means for a formal sum $\xi = \sum_i k_i L_i \in C_2(X)$ to be a 2–cycle in the handle complex. For that purpose, we construct a framed link $\mathcal{L}(\xi)$ in $S^3 \setminus \mathcal{U}$ by taking k_i parallel copies of L_i with respect to the framing and giving each parallel copy the same framing as L_i. In fact, it is technically more convenient to fix a thin tubular neighborhood $\nu\mathcal{U}$ of $\mathcal{U}$ and to construct $\mathcal{L}(\xi)$ in the complement of $\nu\mathcal{U}$.

Lemma A.4. *For $\xi = \sum_i k_i L_i \in C_2(X)$ the following statements are equivalent:*

(i) ξ is a 2–cycle in the handle complex, that is, $\partial_2 \xi = 0$.

(ii) $\mathcal{L}(\xi)$ *is null-homologous in* $S^3 \setminus \nu\mathcal{U}$.

(iii) $\mathcal{L}(\xi)$ *is null-homologous in* $\partial X^{\leq 1}$.

Proof. We first argue that (*ii*) and (*iii*) are equivalent, which follows from the interpretation of $\partial X^{\leq 1}$ as 0–surgery on $\mathcal{U}$ (as explained on page 15). In detail, there is an inclusion $S^3 \setminus \nu\mathcal{U} \hookrightarrow \partial X^{\leq 1}$ which is known to induce a surjection $H_1(S^3 \setminus \nu\mathcal{U}) \twoheadrightarrow H_1(\partial X^{\leq 1})$, and since $\mathcal{U}$ is an unlink, this map is actually an isomorphism (see [32, Proposition 5.3.11]). To establish the equivalence of (*i*) and (*ii*) we note that $H_1(S^3 \setminus \nu\mathcal{U})$ is free Abelian, generated by right-handed meridians of the components of $\mathcal{U}$ – let μ_j be such a meridian for U_j. Using the meridian description of linking numbers it is easy to see that

$$[L_i] = \sum_j \operatorname{lk}(L_i, U_j)[\mu_j] \in H_1(S^3 \setminus \nu\mathcal{U})$$

and Lemma A.3 combined with the fact that $[\mathcal{L}(\xi)] = \sum_i k_i[L_i] \in H_1(S^3 \setminus \nu\mathcal{U})$ implies the claim. □

Proof of Proposition A.2. The proof uses the same ideas as that of Proposition A.1 but has some additional twists (compare [32, Proposition 4.5.11]). Since there are no 3–handles, the handle complex shows that $H_2(X)$ is isomorphic to $\ker \partial_2$. To obtain a concrete isomorphism, we fist have to choose a basis $\xi_1, \dots, \xi_k$ of $\ker \partial_2$. Next, according to Lemma A.4 the links $\mathcal{L}(\xi_j)$ are null-homologous in $S^3 \setminus \nu\mathcal{U}$ and thus bound Seifert surfaces $\Sigma_i \subset S^3 \setminus \nu\mathcal{U}$ – for convenience, we assume that the $\mathcal{L}(\xi_i)$ are pairwise disjoint and intersect the Σ_j transversely. Since the parallel copies of the L_i in $\mathcal{L}(\xi_j)$ were obtained using the framing of L_i, we can cap them off with parallel copies of the core in the corresponding 2–handle to obtain a closed surface $\widehat{\Sigma}_i$ in X. The map sending ξ_i to $[\widehat{\Sigma}_i]$ then gives an isomorphism $\ker \partial_2 \to H_2(X)$.

It remains to show that $[\widehat{\Sigma}_i] \cdot [\widehat{\Sigma}_j] = \operatorname{lk}(\xi_i, \xi_j)$. For $i \neq j$ we perturb $\widehat{\Sigma}_j$ as follows. We take a collar of the form $\partial X^{\leq 1} \times [0,1]$, push Σ_j into $S^3 \setminus \nu\mathcal{U} \times \{\varepsilon\}$ for some small ϵ, add the cylinder $\mathcal{L}(\xi_j) \times [0, \epsilon]$, and cap off with the cores of the 2–handles. The result is a closed surface $\widehat{\Sigma}'_j \subset X$ isotopic to $\widehat{\Sigma}_j$ and transverse to $\widehat{\Sigma}_i$. Counting intersection points, we see that $[\widehat{\Sigma}_i] \cdot [\widehat{(\Sigma_j)}] = [\Sigma_i] \cdot [\mathcal{L}(\xi_j)]$ where the right hand side is an intersection number in $S^3 \setminus \mathcal{L}(\xi_i)$. But the expression $[\Sigma_i] \cdot [\mathcal{L}(\xi_j)]$ also computes the linking number of $\mathcal{L}(\xi_i)$ and $\mathcal{L}(\xi_j)$)[1] which is just $\operatorname{lk}(\xi_i, \xi_j)$. For $i = j$ we have to use a slightly more complicated perturbation. Instead of using the cylinder $\mathcal{L}(\xi_j) \times [0, \epsilon]$, we choose an isotopy that pushes $\mathcal{L}(\xi_j)$ off itself according to the framing and take the trace of this isotopy in $\Sigma^3 \setminus \nu\mathcal{U} \times [0, \epsilon]$. Then we cap off to a closed surface $\widehat{\Sigma}'_j$ using copies of the cores of the 2–handles, which are disjoint from those used in the construction of $\widehat{\Sigma}_j$. The same arguments as above now show that $[\widehat{\Sigma}_j] \cdot [\widehat{\Sigma}_j] = \operatorname{lk}(\xi_j, \xi_j)$. □

Finally, we turn to the case where X is also allowed to have 3–handles, but we require that X is closed. In order to describe the intersection form of X we need some purely algebraic terminology. If H is an Abelian group of finite rank equipped

[1] Note that this does not immediately follow from the expression of linking numbers in terms of Seifert surfaces for knots. The generalization to Seifert surfaces for links is given in Cromwell's book [12, Theorem 5.7.3].

with a symmetric bilinear form Q, then the *radical* of Q is the subgroup of H defined by

$$\operatorname{rad}(Q) = \{v \in V \mid q(v,w) = 0 \text{ for all } w \in V\}\,.$$

Furthermore, Q descends to a form on $H^{\mathrm{red}} = H/\operatorname{rad}(Q)$ which we denote by Q^{red} and call the *reduced* form of Q. Note that the radical has the property that if $kh \in \operatorname{rad}(Q)$ for some $k \neq 0$, then h itself must be contained in $\operatorname{rad}(Q)$. In particular, $\operatorname{rad}(Q)$ contains the torsion of H and H^{red} is torsion free. Also, the radical of Q^{red} is obviously trivial, so that Q^{red} is non-degenerate. As an example, consider the intersection pairing $\widetilde{Q}_X$ of a closed 4–manifold X. Then by Poincaré duality the radical is precisely the torsion subgroup of $H_2(X)$ and $\widetilde{Q}_X^{\mathrm{red}}$ is by definition just the intersection form Q_X.

Proposition A.5. *Let X be a closed 4–dimensional handlebody with Kirby diagram $(\mathcal{U}, \mathcal{L})$. Then the intersection form Q_X is isomorphic to the reduced restricted linking form of $\mathcal{L}$, that is, the reduced form of* $\mathrm{lk}\colon \ker \partial_2 \times \ker \partial_2 \to \mathbb{Z}$.

Proof. The idea is to relate the intersection forms of X and its 2–skeleton $X^{\leq 2}$. For the sake of cleaner notation we let $V = X^{\leq 2}$. We first observe that the inclusion $V \hookrightarrow X$ induces a surjection $H_2(V) \to H_2(X)$, which can either be seen from the handle complex or from the homology sequence of the pair (X, V)[2]. Since the intersection pairing is natural under codimension 0 embeddings, the following purely algebraic lemma applies.

Lemma A.6. *Let $\varphi\colon \widetilde{H} \to H$ be a surjective homomorphism of Abelian groups. If Q is a symmetric bilinear form on H and $\widetilde{Q} = \varphi^* Q$ is its pullback to $\widetilde{H}$, then $\widetilde{Q}^{\mathrm{red}}$ and Q^{red} are isomorphic via an isomorphism induced by φ.*

Proof. It follows from the surjectivity of φ and the definition of $\widetilde{Q}$ that

$$\operatorname{rad}(\widetilde{Q}) = \varphi^{-1}\big(\operatorname{rad}(Q)\big)$$

so that φ induces a homomorphism $\varphi^{\mathrm{red}}\colon \widetilde{H}^{\mathrm{red}} \to H^{\mathrm{red}}$ which is again surjective and preserves the reduced forms. Moreover, since the kernel of φ is contained in the radical of $\widetilde{Q}$, φ^{red} is also injective. □

In the situation above, this means that Q_X is isomorphic to Q_V^{red}, but since $V = X^{\leq 2}$ has no 3–handles, Q_V is isomorphic to the restricted linking form of $\mathcal{L}$ by Proposition A.2. This finishes the proof of Proposition A.5 □

[2] Excision and Poincaré duality imply that $H_2(X, V) \cong H^2(X^{\geq 3})$ which vanishes since $X^{\geq 3}$ is homotopy equivalent to a wedge of circles.

Appendix B

Spin^c Structures and the Taubes Map

In this appendix we will explain the details behind a short remark in a paper of Perutz [58, p.1500], which gives a geometric description of the set of Spin^c structures on a given 4–manifold in the presence of a broken Lefschetz fibration. We will work in a slightly more general setting and allow our maps to have fold and cusp singularities, both indefinite and definite, as well as Lefschetz and achiral Lefschetz singularities. Note that we do not put any restriction on how the critical locus is mapped into the base since this turns out to be irrelevant for the discussion. For the lack of a better name we shall call such maps *singular fibrations*.

For simplicity we only consider the case when X is a closed, oriented 4–manifold, although a generalization to the case when X has non-empty boundary can be worked out as well. If $f\colon X \to B$ is such a singular fibration and $\mathcal{S} \subset X$ denotes the 1–dimensional part of the critical locus of f (consisting of the folds and cusps), then the relation to Spin^c structures is given by the so called *Taubes map*

$$\tau_f\colon \mathrm{Spin}^c(X) \to H_2(X, \mathcal{S})$$

defined in equation (B.9) below. Recall that our convention is to orient $\mathcal{S}$ so that the normal bundle of a fold arc in B is oriented in the direction of increasing Euler characteristic of the fibers. Our goal is to prove the following.

Proposition B.1. *The Taubes map $\tau_f\colon \mathrm{Spin}^c(X) \to H_2(X, \mathcal{S})$ is injective with image $\partial^{-1}([\mathcal{S}])$ where $\delta\colon H_2(X, \mathcal{S}) \to H_1(\mathcal{S})$ is the connecting homomorphism of the homology sequence of the pair $(X, \mathcal{S})$.*

In more geometric terms, this means that spin^c structures on X correspond to homology classes of *Seifert surfaces* for $\mathcal{S}$ in X, that is, orientable surfaces in X bounded by $\mathcal{S}$. A similar idea first appeared in a paper of Taubes [64] (in the closely related context of near-symplectic 4–manifolds) and was later exploited by Perutz [58] in his approach to construct smooth invariants of 4–manifolds from broken Lefschetz fibrations – Perutz also introduced the name Taubes map. The author would like to thank David Gay for bringing this aspect of [58] to his attention and Tim Perutz for explaining his reference to [64].

Partial Spin^c Structures of Singular Fibrations

As in Section 9.3, we use the classifying space interpretation of spin^c structures. We begin by discussing some further particulars. Let M be an oriented n–manifold, implicitly equipped with a Riemannian metric. Recall that the set $\mathrm{Spin}^c(M)$ of (isomorphism classes of) spin^c structures has a free and transitive action of $H^2(M)$. We identify $H^2(M)$ with the set of isomorphism classes of complex line bundles via the first Chern class and denote the action of $\mathcal{L} \to M$ on $\mathfrak{s} \in \mathrm{Spin}^c(M)$ by $\mathfrak{s} \otimes \mathcal{L}$. The effect of this action on the determinant line bundles is given by $\det(\mathfrak{s} \otimes \mathcal{L}) = \det(\mathfrak{s}) \otimes \mathcal{L}^2$. In this language, the *difference map*

$$\Delta_M \colon \mathrm{Spin}^c(M) \times \mathrm{Spin}^c(M) \to H^2(M)$$

takes the form $\Delta_M(\mathfrak{s} \otimes \mathcal{L}, \mathfrak{s}) = c_1(\mathcal{L})$. The difference map is natural in the sense that for a smooth map $\phi \colon M \to N$ (for which pulling back spin^c structures makes sense) and $\mathfrak{s}, \mathfrak{s}' \in \mathrm{Spin}^c(N)$ we have $\Delta_M(\phi^*\mathfrak{s}, \phi^*\mathfrak{s}') = \phi^* \Delta_N(\mathfrak{s}, \mathfrak{s}')$.

Now let X be a closed, oriented 4–manifold. Given a singular fibration $f \colon X \to B$ we decompose the critical locus as

$$\mathcal{C}_f = \mathcal{P} \cup \mathcal{S}$$

where $\mathcal{P}$ is the finite set of Lefschetz and achiral Lefschetz singularities, and $\mathcal{S}$ is the 1–dimensional submanifold formed by the folds and cusps. In other words, $\mathcal{C}_f$ is a submanifold of *dimension at most one*; this terminology will be used from now on. We denote the *vertical distribution* by $\mathcal{V}^f = \ker(df) \subset TX$ and we take a *horizontal distribution* $\mathcal{H}$ given as the orthogonal complement of $\mathcal{V}^f$ with respect to some Riemannian metric on X which we fix once and for all. The trivial but important observation that f is a submersion outside of its critical locus implies that both $\mathcal{V}$ and $\mathcal{H}$ have constant rank 2 on $X \setminus \mathcal{C}_f$. Moreover, for a regular point $p \in X$ the fibers $\mathcal{H}_p$ and $\mathcal{V}^f_p$ are oriented since $df_p \colon \mathcal{H}_p \to T_{f(p)}B$ is an isomorphism and the orientation of B pulls back to $\mathcal{H}$, and $\mathcal{V}^f_p$ is oriented by requiring that $T_pX = \mathcal{V}^f_p \oplus \mathcal{H}_p$ be an oriented sum. This means that we have a splitting

$$T(X \setminus \mathcal{C}_f) = \mathcal{V}|_{X \setminus \mathcal{C}_f} \oplus \mathcal{H}|_{X \setminus \mathcal{C}_f} \tag{B.1}$$

of $T(X \setminus \mathcal{C}_f)$ into two oriented 2–plane fields. Since oriented 2–plane fields can be identified with complex line bundles, we have exhibited an almost complex structure J_f on $X \setminus \mathcal{C}_f$. More precisely, the splitting (B.1) (and the metric on X) reduce the structure group of $T(X \setminus \mathcal{C}_f)$ to $\mathrm{SO}(2) \times \mathrm{SO}(2)$ and, using the standard identification of $\mathrm{SO}(2)$ with $\mathrm{U}(1)$, we obtain a reduction to $\mathrm{U}(1) \times \mathrm{U}(1) \subset \mathrm{U}(2)$ which gives an almost complex structure. Moreover, using the standard embedding[1] $\mathrm{U}(2) \hookrightarrow \mathrm{Spin}^c(4)$ we obtain a Spin^c structure $\tilde{\mathfrak{s}}_f \in \mathrm{Spin}^c(X \setminus \mathcal{C}_f)$ whose determinant line bundle is given by

$$\det(\tilde{\mathfrak{s}}_f) = \Lambda^2_{\mathbb{C}}\big(T(X \setminus \mathcal{C}_f), J_f\big) \cong \mathcal{V}|_{X \setminus \mathcal{C}_f} \otimes_{\mathbb{C}} \mathcal{H}|_{X \setminus \mathcal{C}_f}. \tag{B.2}$$

We want to investigate whether J_f or $\tilde{\mathfrak{s}}_f$ can be extended across $\mathcal{C}_f$ or at least parts thereof.

[1] See [44, p.392f.], for example.

Extending Spin^c Structures and Line Bundles

Abstracting the above situation, we consider a closed, oriented 4–manifold X together with a partial Spin^c structure $\mathfrak{s}$ defined in the complement of a submanifold $A \subset X$ of dimension at most one. Since complex line bundles are certainly more accessible than spin^c structures, it would be advantageous if the problem of extending $\mathfrak{s}$ across A reduced to merely extending the determinant line bundle $\det(\mathfrak{s})$. Fortunately, this is indeed the case in the situation at hand.

Proposition B.2. *Let X be a closed, oriented 4–manifold, $A \subset X$ a compact submanifold of dimension at most one, and $\mathfrak{s} \in \mathrm{Spin}^c(X \setminus A)$. Then the following statements are equivalent.*

(i) $\mathfrak{s}$ *can be extended across* A.

(ii) $\det(\mathfrak{s})$ *extends across* A.

(iii) $c_1(\mathfrak{s})$ *is contained in the image of the map* $H^2(X) \to H^2(X \setminus A)$.

Furthermore, if either $\mathfrak{s}$ *or* $\det(\mathfrak{s})$ *extend across* A*, then the extensions are unique.*

Proof. Obviously, (ii) is equivalent to (iii), and (i) implies (ii). It remains to show the implication $(ii) \Rightarrow (i)$. So we assume that $\det(\mathfrak{s})$ can be extended to a complex line bundle $\mathcal{L} \to X$ and try to extend $\mathfrak{s}$. Note that the extension of $\det(\mathfrak{s})$ is unique since the map $H^2(X) \to H^2(X \setminus A)$ is injective. Indeed, the cohomology sequence for pairs gives

$$\cdots \longrightarrow H^2(X, X \setminus A) \longrightarrow H^2(X) \longrightarrow H^2(X \setminus A) \longrightarrow \cdots$$

and by Poincaré duality we have $H^2(X, X \setminus A) \cong H_2(A) = 0$.

Now, the extension problem for $\mathfrak{s}$ is equivalent to the relative lifting problem

$$\begin{array}{ccc} \mathrm{BSpin}^c(4) & \longrightarrow & \mathrm{BSO}(4) \times \mathbb{C}\mathrm{P}^\infty \\ {\scriptstyle \mathfrak{s}}\uparrow & \nearrow & \uparrow{\scriptstyle TX \times \mathcal{L}} \\ X \setminus A & \longrightarrow & X \end{array} \tag{B.3}$$

where the dotted arrow indicates the missing piece. Note that we have identified $\mathfrak{s}$, TX and $\mathcal{L}$ with their respective classifying maps for brevity of notation. This problem can be attacked using obstruction theory (see for example [14, Section 7]).

The map $\mathrm{BSpin}^c(4) \to \mathrm{BSO}(4) \times \mathbb{C}\mathrm{P}^\infty$ is induced from the standard group homomorphism $\mathrm{Spin}^c(4) \to \mathrm{SO}(4) \times S^1$ which is known to have kernel $\mathbb{Z}_2$.[2] It follows that the map $\mathrm{BSpin}^c(4) \to \mathrm{BSO}(4) \times \mathbb{C}\mathrm{P}^\infty$ is a Serre fibration with homotopy fiber $\mathrm{B}\mathbb{Z}_2 \simeq \mathbb{R}\mathrm{P}^\infty \simeq K(\mathbb{Z}_2, 1)$. The relevant obstructions thus take values in the homology groups

$$H^{i+1}\big(X, X \setminus A\,;\, \pi_i(\mathbb{R}\mathrm{P}^\infty)\big)$$

which obviously vanish for $i \neq 1$. Moreover, for $i = 1$ it follows from Poincaré duality and the fact that A has dimension at most one that

$$H^2\big(X, X \setminus A\,;\, \pi_1(\mathbb{R}\mathrm{P}^\infty)\big) \cong H_2(A; \mathbb{Z}_2) = 0.$$

[2] To be precise, recall that $\mathrm{Spin}^c(4) = \mathrm{Spin}(4) \times_{\mathbb{Z}_2} S^1$ where $(a, \xi) \sim (-a, -\xi)$, and that the map to $\mathrm{SO}(4) \times S^1$ sends a class $[a, \xi] \in \mathrm{Spin}^c(4)$ to the pair $(\rho(a), \xi^2)$ where $\rho\colon \mathrm{Spin}(4) \to \mathrm{SO}(4)$ the standard universal covering map. The kernel of $\mathrm{Spin}^c(4) \to \mathrm{SO}(4) \times S^1$ is generated by $[1, -1] = [-1, 1]$.

Hence, the lifting problem (B.3) can always be solved. Moreover, its solutions can be parametrized by $H^1(X, X \setminus A; \mathbb{Z}_2)$ which also vanishes for dimension reasons. □

Having reduced the problem of extending Spin^c structures to extending their determinant line bundles, the next step is to try and extend line bundles across submanifolds of dimension at most one.

Lemma B.3. *Let X and A be as before and let $\mathcal{L} \to X \setminus A$ be a complex line bundle.*

(i) If A is a finite set of points, then $\mathcal{L}$ can be uniquely extended across A.

(ii) If A is an oriented circle, then there is an integer valued obstruction $\mathfrak{o}(\mathcal{L}, A)$ for extending $\mathcal{L}$ whose vanishing guarantees a unique extension.

Proof. If we identify complex line bundles with their first Chern classes, then the cohomology sequence of the pair $(X, X \setminus A)$, Poincaré duality, and the fact that A has dimension at most one provide an exact sequence

$$0 \longrightarrow H^2(X) \xrightarrow{i^*} H^2(X \setminus A) \xrightarrow{\delta} H^3(X, X \setminus A) \cong H_1(A).$$

If A is zero dimensional, then i^* is an isomorphism and the first claim follows. On the other hand, if A is an oriented circle, then we have

$$\mathrm{PD}\left(\delta c_1(\mathcal{L})\right) = -\mathfrak{o}(\mathcal{L}, A)\,[A] \in H_1(A) \tag{B.4}$$

for some $\mathfrak{o}(\mathcal{L}, A) \in \mathbb{Z}$ and $\mathcal{L}$ extends across A if and only if $\mathfrak{o}(\mathcal{L}, A) = 0$. Moreover, since i^* is still injective, the extension is unique. □

The obstruction $\mathfrak{o}(\mathcal{L}, A)$ encountered in the case when A is an oriented circle has a geometric interpretation. Since X is oriented, the normal bundle of A must be trivial. In particular, if $\bar{\nu}A$ is a closed tubular neighborhood of A (with interior νA), then $\partial\bar{\nu}A$ is a trivial S^2–bundle over A and $H_2(\partial\bar{\nu}A)$ is generated by the homology class $[S]$ of a sphere fiber. (Let us be precise about orientations: $\bar{\nu}A$ inherits an orientation from X and we orient $\partial\bar{\nu}A$ as the boundary of $\bar{\nu}A$ by the *outward normal first* convention. Moreover, given the orientation of A and $\partial\bar{\nu}A$ the *fiber first* convention specifies an orientation on S.)

Lemma B.4. *In the above notation, the obstruction $\mathfrak{o}(\mathcal{L}, A)$ agrees with the Euler number of the restriction $\mathcal{L}|_S$. In particular, $\mathcal{L}$ extends across A if and only if $\mathcal{L}|_S$ is trivial.*

Proof. We first note that the Euler number of $\mathcal{L}|_S$ is given by $\langle c_1(\mathcal{L}|_S), [S]\rangle$. Moreover, we compute

$$\begin{aligned}\langle c_1(\mathcal{L}|_S), [S]\rangle_S &= \langle c_1(\mathcal{L}|_{\partial\bar{\nu}A}), \mathrm{PD}[S] \cap [\partial\bar{\nu}A]\rangle_{\partial\bar{\nu}A} \\ &= \langle c_1(\mathcal{L}|_{\partial\bar{\nu}A}) \cup \mathrm{PD}[S], [\partial\bar{\nu}A]\rangle_{\partial\bar{\nu}A} \\ &= -[S] \cdot \mathrm{PD}\, c_1(\mathcal{L}|_{\partial\bar{\nu}A}).\end{aligned}$$

Next, we observe that we have the following commutative squares:

$$\begin{array}{ccc} H^2(X \setminus \nu A) & \xrightarrow{incl^*} & H^2(\partial\bar{\nu}A) \\ \downarrow \mathrm{PD} & & \downarrow \mathrm{PD} \\ H_2(X \setminus \nu A, \partial\bar{\nu}A) & \xrightarrow{\partial} & H_1(\partial\bar{\nu}A) \end{array} \qquad \begin{array}{ccc} H^2(X \setminus A) & \xrightarrow{\delta} & H^3(X, X \setminus A) \\ \downarrow \mathrm{PD} & & \downarrow \mathrm{PD} \\ H_2(X, A) & \xrightarrow{\partial} & H_1(A) \end{array}$$

Note that $H_1(\partial\bar{\nu}A)$ is generated by a class $[A']$ where A' is a parallel push off of A; moreover, our orientation conventions imply that $[S]\cdot[A']=1$. Using this and the left square we see that

$$\begin{aligned}\partial\,\mathrm{PD}\,c_1(\mathcal{L}|_{X\setminus\nu A}) &= \mathrm{PD}\,c_1(\mathcal{L}|_{\partial\bar{\nu}A})\\ &= \big([S]\cdot c_1(\mathcal{L}|_{\partial\bar{\nu}A})\big)[A']\\ &= -\langle c_1(\mathcal{L}|_S),[S]\rangle[A']\in H_1(\partial\bar{\nu}A)\end{aligned}\tag{B.5}$$

and the right square shows that

$$\partial\,\mathrm{PD}\,c_1(\mathcal{L}) = \mathrm{PD}\,\delta c_1(\mathcal{L}) = -\mathfrak{o}(\mathcal{L},A)[A]\in H_1(A).\tag{B.6}$$

In order to compare the expressions in equations (B.5) and (B.6) we consider the diagram

$$\begin{array}{ccccc} H^2(X\setminus A) & & \xrightarrow{\quad\cong\quad} & & H^2(X\setminus\nu A)\\ \downarrow{\scriptstyle\mathrm{PD}} & & & & \downarrow{\scriptstyle\mathrm{PD}}\\ H_2(X,A) & \xrightarrow{\cong} & H_2(X,\bar{\nu}A) & \xleftarrow{\cong} & H_2(X\setminus\nu A,\partial\bar{\nu}A)\\ \downarrow{\scriptstyle\partial} & & \downarrow{\scriptstyle\partial} & & \downarrow{\scriptstyle\partial}\\ H_1(A) & \xrightarrow{\cong} & H_1(\bar{\nu}A) & \xleftarrow{\cong} & H_1(\partial\bar{\nu}A)\end{array}$$

where all horizontal maps are induced by inclusions. The lower squares commute by the naturality of the homology sequence of a pair while the upper square is commutative, essentially by definition. Finally, since both $[A]\in H_1(A)$ and $[A']\in H_1(\partial\bar{\nu}A)$ are mapped to the same element in $H_1(\bar{\nu}A)$, the commutativity of the diagram together with equations (B.5) and (B.6) implies $\mathfrak{o}(\mathcal{L},A)=\langle c_1(\mathcal{L}|_S),[S]\rangle$. □

The Proof of Proposition B.1

With these remarks in place we return to a singular fibration $f\colon X\to B$. As before we write $\mathcal{C}_f=\mathcal{P}\cup\mathcal{S}$ with $\mathcal{P}$ and $\mathcal{S}$ the 0– and 1–dimensional parts of the critical locus. By Proposition B.2 and Lemma B.3 the Spin^c structure $\tilde{\mathfrak{s}}_f\in\mathrm{Spin}^c(X\setminus\mathcal{C}_f)$ can be uniquely extended across $\mathcal{P}$ to a Spin^c structure $\mathfrak{s}_f\in\mathrm{Spin}^c(X\setminus\mathcal{S})$. However, a further extension is not possible.

Lemma B.5. *For all connected components $\mathcal{S}_i$ of $\mathcal{S}$ we have $\mathfrak{o}\big(\det(\mathfrak{s}_f),\mathcal{S}_i\big)=2$.*

Proof. Since any component of $\mathcal{S}$ contains a fold point, it is enough to prove the statement for the fold models, that is, we can assume that

$$f=\mathrm{F}_\pm\colon\mathbb{R}^4\to\mathbb{R}^2,\quad (t,x,y,z)\mapsto(t,x^2+y^2\pm z^2).$$

In this case we have $\mathcal{C}_f=\mathcal{S}=\mathbb{R}\times\{(0,0,0)\}\subset\mathbb{R}^4$ and as a normal sphere fiber at the origin we take $S=\{0\}\times\{x^2+y^2+z^2=1\}$. A quick calculation shows that the vertical distribution $\mathcal{V}=\ker(df)$ is given by the orthogonal complement of the vector fields $(1,0,0,0)$ and $(0,x,y,\pm z)$ with respect to the Euclidean metric. Conversely, this shows that the horizontal distribution $\mathcal{H}$ – which was defined as the

orthogonal complement of $\mathcal{V}$ – is canonically trivialized outside the critical locus. In particular, we see from equation (B.2) that

$$\det(\mathfrak{s}_f) \cong \mathcal{V}|_{X\setminus\mathcal{C}_f} \otimes_{\mathbb{C}} \mathcal{H}|_{X\setminus\mathcal{C}_f} \cong \mathcal{V}|_{X\setminus\mathcal{C}_f}.$$

So by Lemma B.4 it remains to compute the Euler number of $\mathcal{V}|_S$. Note that the Euler number of $\mathcal{V}|_S$ is given by the total index of section with only isolated zeros. Such a section is given by the vector field $(0, y, -x, 0)$ which has two isolated zeros at the poles $p_\pm = (0, 0, 0, \pm 1)$ of S. Since this vector field is also tangent to S, the total index must be $\pm\chi(S) = \pm 2$ and we have to show that the sign is positive.

This is a tedious exercise in orientation bookkeeping and we first have to set up some notation. We will have to keep track of two independent signs appearing in $\mathrm{F}_\pm$ and $p_\pm$, and in order to tell them apart we let $p_\epsilon = (0, 0, 0, \epsilon)$ where $\epsilon = \pm 1$. Note that $\ker(d\mathrm{F}_\pm)$ is tangent to S at p_ϵ and it is enough to compare the orientations of $\ker(d\mathrm{F}_\pm|_{p_\epsilon})$ and $T_{p_\epsilon}S^\pm$, where $S^\pm$ is the oriented version of S with respect to $\mathrm{F}_\pm$. We also let $\mathcal{V}_p^\pm = \ker(d\mathrm{F}_\pm|_p)$ and $\mathcal{H}_p^\pm = \ker(d\mathrm{F}_\pm|_p)^\perp$. As a last piece of notation, we consider an orientation of a real vector space E as elements $\mathrm{or}(E) \in \Lambda^{\dim E}E$. For example, we orient the source and target of $\mathrm{F}_\pm$ by

$$\begin{aligned}\mathrm{or}(\mathbb{R}^4) &= \partial_t \wedge \partial_x \wedge \partial_y \wedge \partial_z\\ \mathrm{or}(\mathbb{R}^2) &= \partial_u \wedge \partial_y\end{aligned}$$

where (t, x, y, z) and (u, v) are the Cartesian coordinates on $\mathbb{R}^4$ and $\mathbb{R}^2$, respectively. We first determine the orientation of $\mathcal{V}_\epsilon^\pm$. Observe that for a general point $p = (t, x, y, z)$ we have

$$\mathcal{H}_p^\pm = \left\langle \partial_t, x\partial_x + y\partial_y \pm z\partial_z \right\rangle$$

and the spanning vectors are mapped as follows:

$$d\mathrm{F}_\pm|_p(\partial_t) = \partial_u \quad \text{and} \quad d\mathrm{F}_\pm|_p(x\partial_x + y\partial_y \pm z\partial_z) = (x^2 + y^2 + z^2)\partial_v.$$

The fiber first convention requires that $d\mathrm{F}_\pm\big(\mathrm{or}(\mathcal{H}_{p_\epsilon}^\pm)\big) = \mathrm{or}(\mathbb{R}^2) = \partial_u \wedge \partial_v$ so that

$$\mathrm{or}(\mathcal{H}_{p_\epsilon}^\pm) = \pm\epsilon\partial_t \wedge \partial_z,$$

and furthermore $\mathrm{or}(\mathcal{V}_{p_\epsilon}^\pm) \wedge \mathrm{or}(\mathcal{H}_{p_\epsilon}^\pm) = \mathrm{or}(\mathbb{R}^4)$ which shows

$$\mathrm{or}(\mathcal{V}_{p_\epsilon}^\pm) = \pm\epsilon\partial_x \wedge \partial_y. \tag{B.7}$$

Now we turn to $T_{p_\epsilon}S^\pm$. Let $\mathcal{C}^\pm$ be the critical locus of $\mathcal{F}_\pm$. Of course, the set $\mathbb{C}^\pm$ is independent of the sign, but it turns out that the orientations are different. Indeed, for F_+ the region $\{v < 0\}$ has empty fibers while the fibers over $\{v > 0\}$ are 2–spheres. So the Euler characteristic increases in direction of ∂_v and the critical image of oriented by $-\partial_u$. It follows that $\mathcal{C}^+$ is oriented by $-\partial_t$. Similarly, for F_- the fibers over $\{v < 0\}$ and $\{v > 0\}$ are pairs of disks and annuli, respectively, so that $\mathcal{C}^-$ is oriented by ∂_t. It follows from the fiber first convention for the normal bundle of $\mathcal{C}^\pm$ that

$$\mathrm{or}(\mathbb{R}^4) = \mathrm{or}(\{0\} \times \mathbb{R}^3) \wedge \mathrm{or}(T_0\mathcal{C}^\pm) = \pm\partial_t \wedge \mathrm{or}(\{0\} \times \mathbb{R}^3)$$

from which we see that

$$\mathrm{or}(\{0\} \times \mathbb{R}^3) = \pm\partial_x \wedge \partial_y \wedge \partial_z.$$

Finally, we can use the outward normal first convention to determine $\mathrm{or}(T_{p_\epsilon} S^{\pm})$. The outward normal to $S^{\pm}$ at p_ϵ is $\epsilon\partial_z$, so that

$$\mathrm{or}(\{0\} \times \mathbb{R}^3) = \epsilon\partial_z \wedge \mathrm{or}(T_{p_\epsilon} S^{\pm}) = \pm\partial_x \wedge \partial_y \wedge \partial_z$$

and thus

$$\mathrm{or}(T_{p_\epsilon} S^{\pm}) = \pm\epsilon\partial_x \wedge \partial_y. \tag{B.8}$$

Comparing equations (B.7) and (B.8) we see that we are done. □

Remark B.6. Let us briefly comment on extending the almost complex structure J_f defined over $X \setminus \mathcal{C}_f$ by the splitting (B.1). Although $\widetilde{K}_f$ extends across all Lefschetz points as a line bundle it does not necessarily extend as a 2–plane field, that is, the extended line bundle may not embed into the tangent bundle anymore. It is well known that such an extension is only possible for Lefschetz singularities, but not for achiral ones (see [32, Chapter 8.4]). As a consequence, the splitting (B.1) as well as J_f only extend across Lefschetz singularities. In the presence of achiral Lefschetz singularities, $\mathfrak{s}_f$ does not come from an almost complex structure anymore.

To summarize the discussion above, whenever a singular fibration has folds or cusps there is *no* canonical spin^c structure. This might seem disheartening at first sight, but it turns out that there is at least a way to parametrize Spin^c structures on X. Using the difference map and the canonical Spin^c structure $\mathfrak{s}_f$ on $X \setminus \mathcal{S}$ we can define the anticipated *Taubes map*

$$\tau_f \colon \mathrm{Spin}^c(X) \to H_2(X, \mathcal{S})$$

associated to f by the formula

$$\tau_f(\mathfrak{s}) = \mathrm{PD}\left(\Delta_{X\setminus\mathcal{S}}(\mathfrak{s}|_{X\setminus\mathcal{S}}, \mathfrak{s}_f)\right). \tag{B.9}$$

We can now prove Proposition B.1 which stated that τ_f is injective and that $\tau_f(\mathfrak{s})$ is mapped to $[\mathcal{S}]$ under the connecting homomorphism $\partial \colon H_2(X, \mathcal{S}) \to H_1(\mathcal{S})$.

Proof of B.1. For the proof of injectivity observe that τ_f is the composition of three maps: restriction of Spin^c structures from X to $X \setminus \mathcal{S}$, taking the difference with $\mathfrak{s}_f$, and Poincaré duality. Since the latter two maps are bijections, it suffices to show that the restriction map $\mathrm{Spin}^c(X) \to \mathrm{Spin}^c(X \setminus \mathcal{S})$ is injective. By the naturality of the difference map under restrictions, this is equivalent to the injectivity of the map $j^* \colon H^2(X) \to H^2(X \setminus \mathcal{S})$ induced by the inclusion $X \setminus \mathcal{S} \overset{j}{\hookrightarrow} X$. This follows from the cohomology sequence of the pair $(X, X \setminus \mathcal{S})$; the relevant excerpt reads

$$\cdots \to H^2(X, X \setminus \mathcal{S}) \to H^2(X) \to H^2(X \setminus \mathcal{S}) \to \cdots$$

and the left term vanishes since it is Poincaré dual to $H_2(\mathcal{S})$ and $\mathcal{S}$ is 1–dimensional.

It remains to determine the image of τ_f in $H_2(X,\mathcal{S})$. Note that $H^2(X)$ acts on $H_2(X,\mathcal{S})$ by sending $x \in H_2(X,\mathcal{S})$ and $\xi \in H^2(X)$ to $x + j_* \operatorname{PD}(\xi)$ and the long exact homology sequence of the pair $(X,\mathcal{S})$

$$H_2(\mathcal{S}) = 0 \longrightarrow H_2(X) \xrightarrow{j_*} H_2(X,\mathcal{S}) \xrightarrow{\partial} H_1(\mathcal{S}) \to \cdots$$

shows that this action is free. It follows from the definitions that τ_f is $H^2(X)$ equivariant which shows that the image of τ_f is some coset of $H^2(X)$ in $H_2(X,\mathcal{S})$. Moreover, these cosets are parametrized by their image under ∂ in $H_1(\mathcal{S})$. So altogether it is enough to compute $\partial\tau_f(\mathfrak{s})$ for one $\mathfrak{s} \in \operatorname{Spin}^c(X)$. Let $\mathcal{L}$ be the line bundle over $X \setminus \mathcal{S}$ determined by $\mathfrak{s}|_{X\setminus\mathcal{S}} \cong \mathfrak{s}_f \otimes \mathcal{L}$. Then by definition $\tau_f(\mathfrak{s})$ is Poincaré dual to $c_1(\mathcal{L})$ and the arguments leading up to equation (B.6) show that

$$\partial\tau_f(\mathfrak{s}) = \partial \operatorname{PD}(c_1(\mathcal{L})) = -\sum_i \mathfrak{o}(\mathcal{L},\mathcal{S}_i)[\mathcal{S}_i] \tag{B.10}$$

where $\mathcal{S}_i$ are the connected components of $\mathcal{S}$. In order to compute $\mathfrak{o}(\mathcal{L},\mathcal{S}_i)$ we observe that, on the one hand, we have

$$\mathfrak{o}\big(\det(\mathfrak{s}_f \otimes \mathcal{L}), \mathcal{S}_i\big) = \mathfrak{o}\big(\det(\mathfrak{s}|_{X\setminus\mathcal{S}}), \mathcal{S}_i\big) = 0$$

since $\mathfrak{s}|_{X\setminus\mathcal{S}}$ obviously extends across $\mathcal{S}_i$. On the other hand, the interpretation of $\mathfrak{o}$ as an Euler number (Lemma B.4) shows that $\mathfrak{o}$ is additive under tensor products of line bundles and we get

$$\begin{aligned}\mathfrak{o}\big(\det(\mathfrak{s}_f \otimes \mathcal{L}), \mathcal{S}_i\big) &= \mathfrak{o}\big(\det(\mathfrak{s}_f) \otimes \mathcal{L}^2, \mathcal{S}_i\big) \\ &= \mathfrak{o}\big(\det(\mathfrak{s}_f), \mathcal{S}_i\big) + 2\mathfrak{o}\big(\mathcal{L}, \mathcal{S}_i\big).\end{aligned}$$

But by Lemma B.5 we have $\mathfrak{o}\big(\det(\mathfrak{s}_f), \mathcal{S}_i\big) = 2$ so that $\mathfrak{o}(\mathcal{L},\mathcal{S}_i) = -1$ and it follows from equation (B.10) that $\partial\tau_f(\mathfrak{s}) = \sum_i [\mathcal{S}_i] = [\mathcal{S}]$, which finishes the proof. □

Appendix C

Cancellation in 3–Dimensional Morse Theory

In this appendix we address the following question about Morse functions in the special case of orientable 3–manifolds.

> *Given a Morse function with a canceling pair of critical points, can the cancellation be realized uniquely in a suitable sense?*

This question was studied by Cerf [13, Chapitre III] with an emphasis on dimensions six and higher, but some of his methods work in all dimensions. We obtain an affirmative answer to the above question for orientable 3–manifolds in Theorem C.3 below. We also give a 4–dimensional interpretation of Theorem C.3 in Corollary C.5 which is relevant for our proof of the annular correspondence in Chapter 5.

We begin by introducing some notation and terminology. Let $(W; V_0, V_1)$ be an orientable 3–dimensional cobordism, that is, W is an orientable 3–manifold with $\partial W = V_0 \amalg V_1$. We consider smooth maps

$$f\colon W \longrightarrow [0,1]$$

with the properties that $V_i = f^{-1}(i)$ for $i = 0, 1$ and that f has no critical points on the boundary; for brevity we usually write I for the unit interval $[0, 1]$. These maps form a subspace of $C^\infty(W, I)$ denoted by $\mathcal{F}$. Moreover, by a *homotopy* in $\mathcal{F}$ (emanating from f) we mean a 1–parameter family of maps $f_t \in \mathcal{F}$, $t \in I$, (with $f_0 = f$) which is smooth in the sense that the map $(w, t) \mapsto f_t(w)$ is smooth. We consider the set of all homotopies in $\mathcal{F}$ as a subset of $C^\infty(W \times I, I)$ and equip it with the subspace topology.

Now let $f \in \mathcal{F}$ be a Morse function. Suppose there is a pair of critical points p and q of index k and $k+1$, respectively, such that $f(p) < f(q)$ and no other critical point maps to the interval $[f(p), f(q)]$. The choice of a Riemannian metric on W gives rise to the ascending manifold $\widetilde{A}$ of p and descending manifold $\widetilde{D}$ of q. We fix an intermediate level set $\Sigma = f^{-1}(\sigma)$, where $f(p) < \sigma < f(q)$, and denote by

$$a = \widetilde{A} \cap \Sigma \quad \text{and} \quad d = \widetilde{D} \cap \Sigma$$

the ascending and descending spheres in Σ, which are embedded spheres in Σ with $\dim(a) = 2 - k$ and $\dim(d) = k$. We also consider truncated versions of

the ascending and descending manifolds

$$A = \widetilde{A} \cap \{f \leq \sigma\} \quad \text{and} \quad D = \widetilde{D} \cap \{f \leq \sigma\}$$

which constitute embedded disks in W with $\partial A = a$ and $\partial D = d$.

Definition C.1. Two critical points p and q as above are called a *canceling pair* if there is a Riemannian metric such that a and d intersect transversely in one point.

It is well known that if p and q form a canceling pair, then they can be removed from f by a homotopy in $\mathcal{F}$. We have to understand in more detail what kind of homotopies realize this cancellation (or *death*) of p and q. Homotopies in $\mathcal{F}$ were studied in great detail by Cerf [13] and we briefly recall their basic structure.

The starting point is a natural stratification of $\mathcal{F}$ by a notion of codimension. Roughly, a function has codimension k if it exhibits singular behavior that can be avoided by small perturbations in $(k-1)$–parameter families in $\mathcal{F}$ but not in k–parameter families. We are thus only interested in functions with codimension ≤ 1. The codimension 0 functions are the *excellent* Morse functions, that is, Morse functions which are injective on their critical points. These form an open and dense subset of $\mathcal{F}$. In codimension 1 one finds two types of functions: Morse functions with one critical value of multiplicity two, and functions with one degenerate critical point whose Hessian has 1–dimensional kernel. In the 3–dimensional context the degenerate critical points can be modeled on cubic polynomials of the form $x^3 \pm y^2 \pm z^2$ (where the signs are understood to be independent). It then follows from standard transversality arguments that the set of homotopies $(g_t)_{t\in I}$ in $\mathcal{F}$ such that g_t is a Morse function for all but finitely many t where g_t has codimension 1 is open and dense in the space of all homotopies. For the lack of a better name we will call these *generic homotopies*. Moreover, for the passage through a cubic degeneracy a local model is given by the 1–parameter family of functions on $\mathbb{R}^3$

$$C_t(x, y, z) = x^3 + 3tx \pm y^2 \pm z^2 \tag{C.1}$$

which has two non-degenerate critical points of adjacent index for $t < 0$, one cubic critical point for $t = 0$, and no critical point for $t > 0$.

Returning to our Morse function $f \in \mathcal{F}$ with the canceling pair of critical points $p, q \in \mathcal{C}_f$, it is clear that every homotopy that cancels p and q must pass through a cubical degeneracy. We are interested in homotopies that realize the cancellation in a minimal way.

Definition C.2. Let $f \in \mathcal{F}$ be a Morse function and let p and q be a canceling pair of critical points. A *path of death*[1] for p and q is a generic homotopy $(f_t)_{t\in I}$ in $\mathcal{F}$ emanating from $f = f_0$ with the following properties:

(a) There is a small neighborhood U of $[f(p), f(q)]$ without further critical values such that f_t agrees with f outside $f^{-1}(U)$ for all t.

(b) (f_t) passes through exactly one codimension 1 function which has a cubical degeneracy in which p and q collide and then disappear.

We denote the space of all paths of death for p and q by $\mathfrak{Death}(f; p, q)$.

[1] called "chemin de mort" in [13]

Using this terminology we can state the main result of this appendix.

Theorem C.3 (Uniqueness of deaths)**.** *Let $f \in \mathcal{F}$ be a Morse function defined on an orientable 3–dimensional cobordism W. If $p, q \in \mathcal{C}_f$ is a canceling pair of critical points, then $\mathfrak{Death}(f; p, q)$ is non-empty and connected.*

As mentioned before, the fact that $\mathfrak{Death}(f; p, q)$ is non-empty is well known and we will concentrate on its connectivity. Using the machinery developed by Cerf [13], the question whether $\mathfrak{Death}(f; p, q)$ is connected can be translated into a question about embeddings of ascending and descending spheres as follows. We fix a Riemannian metric such that a and d intersect transversely in one point. We consider the space

$$\mathcal{E} = \mathrm{Emb}(d, \Sigma) \subset C^\infty(d, \Sigma)$$

of embeddings of d into Σ and the subspace

$$\mathcal{E}_1 = \{\phi \in \mathcal{E} \mid \#(\phi(d) \pitchfork a) = 1\}$$

of all embeddings of d whose images intersect a transversely in a single point (both equipped with the subspace topology induced by $C^\infty(d, \Sigma)$). Note that $\mathcal{E}$ and $\mathcal{E}_1$ have preferred base points given by the inclusion $d \subset \Sigma$ which we also denote by $d \in \mathcal{E}_1 \subset \mathcal{E}$ by a slight abuse of notation. In Lemme 2 and the proof of Proposition 4 in III.2.4 of [13, p.72ff.], Cerf proves the following result.

Lemma C.4 (Cerf)**.** *If $\pi_1(\mathcal{E}, \mathcal{E}_1; d) = 1$, then $\mathfrak{Death}(f; p, q)$ is connected.*

We include a brief outline of the arguments.

Proof (sketch). Cerf first defines a notion of *elementary* paths of death, denoted by $\mathfrak{El}$, which are defined using embeddings of (a compactly supported version of) the model $C_{-1} = x^3 - 3x \pm y^2 \pm z^2$ into f. He then shows that each connected component of $\mathfrak{Death}(f; p, q)$ contains an elementary path so that it suffices to show that $\mathfrak{El}$ is connected. More or less by definition there is a surjection $\mathcal{P} \to \mathfrak{El}$, where $\mathcal{P}$ is the space of embeddings of C_{-1} into f, so that it suffices to show that $\mathcal{P}$ is connected. To that end, Cerf observes that an element of $\mathcal{P}$ gives rise to a pair of truncated ascending and descending manifolds for p and q, and he considers the space $\mathcal{N}$ of such pairs. The map $\mathcal{P} \to \mathcal{N}$ turns out to be a locally trivial fiber bundle with connected fibers, leading to the last reduction to showing that $\mathcal{N}$ is connected. Note that $\mathcal{N}$ has a preferred base point, namely the truncated ascending and descending spheres (A, D) of f with respect to our fixed metric. Finally, the connection to $\mathcal{E}$ and $\mathcal{E}_1$ is established by the above cited Lemme 2 which states that

$$\pi_j\big(\mathcal{N}; (A, D)\big) \cong \pi_{j+1}(\mathcal{E}, \mathcal{E}_1; d).$$

To summarize, we have a surjection $\pi_0(\mathcal{P}) \to \pi_0\big(\mathfrak{Death}(f; p, q)\big)$ and an injection $\pi_0(\mathcal{P}) \to \pi_1(\mathcal{E}, \mathcal{E}_1; d)$. The lemma follows. □

In fact, Lemma C.4 holds in all dimensions and the connectivity of $\mathfrak{Death}(f; p, q)$ can always be deduced from a sufficiently good understanding of embeddings of spheres of complementary dimension in Σ. But one can imagine that the structure of $\pi_1(\mathcal{E}, \mathcal{E}_1; d)$ can be quite complicated in general. Luckily, in the 3–dimensional context Σ is a surface and embeddings of spheres in surfaces are well understood.

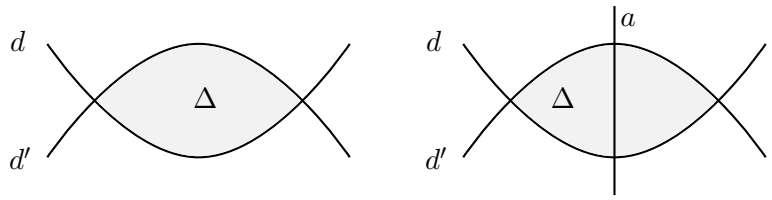


Figure 38: The two possible types of bigons for b and c.

Proof of Theorem C.3. By Lemma C.4 it is enough to show that $\pi_1(\mathcal{E}, \mathcal{E}_1; d) = 1$. According to the homotopy sequence of the pair $(\mathcal{E}, \mathcal{E}_1)$

$$\cdots \to \pi_1(\mathcal{E}_1; d) \to \pi_1(\mathcal{E}; d) \to \pi_1(\mathcal{E}, \mathcal{E}_1; d) \to \pi_0(\mathcal{E}_1) \to \pi_0(\mathcal{E})$$

this is equivalent to the following claims:

(a) $\pi_0(\mathcal{E}_1) \to \pi_0(\mathcal{E})$ is injective

(b) $\pi_1(\mathcal{E}_1; d) \to \pi_1(\mathcal{E}; d)$ is surjective

Recall that $d \cong S^k$ and $a \cong S^{2-k}$ where $k \in \{0, 1, 2\}$ is the index of p. We first discuss the extremal cases when $k \in \{0, 2\}$. Since these are interchanged by replacing f with $-f$, it is enough to treat $k = 2$. In this situation we find that a is a 2–sphere, and thus appears as a connected component of Σ, while d consists of two points, one of which lies on a. The conditions (a) and (b) are almost trivially satisfied.

It remains to treat the case $k = 1$ where both a and d are simple closed curves in Σ. For simplicity we assume that Σ is connected and has genus g, the general case is only notationally more complicated. Note that $g \geq 1$ since no two curves on a sphere can intersect transversely in one point.

Proof of (a): Suppose that $b, c \in \mathcal{E}_1$ are contained in the same component of $\mathcal{E}$. We have to show that we can connect b and c within $\mathcal{E}_1$. We first observe that if b and c have the same image in Σ, then they differ by an orientation preserving diffeomorphism of b. But b is diffeomorphic to S^1, so that $\mathrm{Diff}^+(d)$ is homotopy equivalent to $\mathrm{SO}(2) \cong S^1$ and therefore connected. So not only can c and b be connected through elements of $\mathcal{E}_1$, but in fact through elements with the same image. We can thus assume that b and c have different images and it is also safe to blur the distinction between embeddings and their images.

Next suppose that b and c are disjoint. Since they are isotopic, they must bound and annulus in Σ and it follows from the assumptions that a intersects this annulus in a properly embedded arc connecting the two boundary components. Using the annulus as a guide, it is easy to construct an isotopy connecting b and c through curves that intersect a in a single point, as desired.

Finally, if $b \cap c \neq \emptyset$, then we claim that c can be moved within $\mathcal{E}_1$ to become disjoint from b, reducing to the disjoint case discussed above. By general position we can assume that b and c are transverse and that the triple intersection $a \cap b \cap c$ is empty. It follows from the discussion surrounding Proposition 2.12 on page 23 that c can be pushed off b by a finite sequence of Whitney moves across bigons and we have to argue that the Whitney moves can be performed within $\mathcal{E}_1$. Suppose that b and c form a bigon $\Delta \subset \Sigma$. Then Δ is either disjoint from a or a passes through Δ

in a properly embedded arc connecting the boundary parts $\partial\Delta \setminus d$ and $\partial\Delta \setminus d$ (see Figure 38). In both cases, the Whitney trick can be performed within $\mathcal{E}_1$.

Proof of (b): We begin by describing some obvious elements in $\pi_1(\mathcal{E}; d)$. If we choose an identification $d \cong S^1$, we obtain an S^1 action on d which rotates the image of d in Σ. One full rotation gives rise to a loop in $\mathcal{E}$ based at d and represents an element of $\pi_1(\mathcal{E}; d)$ which is contained in the image of $\pi_1(\mathcal{E}_1)$ since the rotations of d preserve its image. Moreover, if Σ has genus one, then we can choose a diffeomorphism $T^2 \xrightarrow{\cong} \Sigma$ that identifies d with $S^1 \times \{1\}$ and a with $\{1\} \times S^1$. This gives rise to an additional rotation of d along the a direction and another element of $\pi_1(\mathcal{E}; d)$. Although these rotation do not preserve the image of d, they still move d through embeddings whose images intersect a transversely in one point, so that the all elements of $\pi_1(\mathcal{E}; d)$ obtained by rotations lie in the image of $\pi_1(\mathcal{E}_1)$. The upshot of this discussion is that in order to prove (b) it suffices to show that

$$\pi_1(\mathcal{E}) \cong \begin{cases} \mathbb{Z} & \text{for } g \geq 1 \\ \mathbb{Z}^2 & \text{for } g = 1 \end{cases} \tag{C.2}$$

generated by the rotations along d and a described above.

To see this we follow Ivanov [37, p.533ff.] and note that the group $\mathcal{D} = \mathrm{Diff}(\Sigma)$ acts on $\mathcal{E}$ and which gives rise to map $\mathrm{Diff}(\Sigma) \to \mathcal{E}$ sending $\phi \in \mathrm{Diff}(\Sigma)$ to $\phi|_d$. By the change of coordinates principle (Proposition 2.17) this map is surjective and, in fact, it is a Serre fibration with fiber $\mathcal{D}_d = \mathrm{Diff}_d(\Sigma)$, the diffeomorphisms of Σ that restrict to the identity on d (see [37, Theorem 2.6A]). We can thus try to determine $\pi_1(\mathcal{E}; d)$ from the exact sequence of homotopy groups

$$\cdots \to \pi_1(\mathcal{D}_b; \mathrm{id}) \to \pi_1(\mathcal{D}; \mathrm{id}) \to \pi_1(\mathcal{E}; d) \to \pi_0(\mathcal{D}_b) \to \pi_0(\mathcal{D}) \to \pi_0(\mathcal{E}).$$

The success of this approach rests on the following two observations:

- The group $\pi_1(\mathcal{D}_b; \mathrm{id})$ is trivial since $\mathcal{D}_b$ can be identified with the diffeomorphism group of a compact surface with boundary and thus has contractible components by Theorem 2.18(ii). Indeed, let $\widetilde{\Sigma}$ be the surface obtained by compactifying $\Sigma \setminus d$ with two copies of d as boundary components. Then $\mathcal{D}_d$ is naturally identified with $\mathrm{Diff}_\partial(\widetilde{\Sigma})$.
- The map $\pi_0(\mathcal{D}_b) \to \pi_0(\mathcal{D})$ can be identified with the so called *inclusion homomorphism* $\mathrm{Mod}(\Sigma^\circ) \to \mathrm{Mod}(\Sigma)$, where $\Sigma^\circ = \Sigma \setminus \nu d$ and νd is a tubular neighborhood of d in Σ, which is induced by extending elements of $\mathrm{Diff}^+_\partial(\Sigma^\circ)$ to Σ by the identity on νd. In fact, the inclusion homomorphism factors through $\pi_0(\mathcal{D}_d) \to \pi_0(\mathcal{D})$ via an isomorphism $\mathrm{Mod}(\Sigma^\circ) \xrightarrow{\cong} \pi_0(\mathcal{D}_d)$. To see this, note that we also have an inclusion $\Sigma^\circ \hookrightarrow \widetilde{\Sigma}$ and $\widetilde{\Sigma}$ retracts onto Σ° by a smooth isotopy supported in a collar neighborhood. We thus have isomorphism $\mathrm{Diff}_\partial(\Sigma^\circ) \cong \mathrm{Diff}_\partial(\widetilde{\Sigma}) \cong \mathrm{Diff}_d(\Sigma)$

From this point on we have to distinguish the cases when Σ has genus one or higher.

Case 1: $\Sigma \cong T^2$. In this situation Σ° is an annulus whose core is isotopic to d in Σ. In particular, the inclusion homomorphism is injective since it maps the Dehn twist about the core (which generates $\mathrm{Mod}(\Sigma^\circ) \cong \mathbb{Z}$) to the Dehn twist about d and the latter has infinite order in $\mathrm{Mod}(\Sigma)$. The long exact homotopy sequence thus

shows that $\pi_1(\mathcal{D};\mathrm{id}) \to \pi_1(\mathcal{E};d)$ is an isomorphism. Moreover, according to Theorem 2.18(i), an identification $\Sigma \cong T^2$ provides an isomorphism $\pi_1(T^2) \cong \pi_1(\mathcal{D};\mathrm{id})$. The composition gives an isomorphism

$$\pi_1(T^2) \xrightarrow{\cong} \pi_1(\mathcal{E};d)$$

and it follows from the definition that the generators of $\pi_1(T^2)$, which are identified with a and d, are mapped to the rotation loops in $\pi_1(\mathcal{E};d)$.

Case 2: $g > 1$. By Theorem 2.18(i) we have $\pi_1(\mathcal{D};\mathrm{id}) = 1$ so that $\pi_1(\mathcal{E};d)$ corresponds to the kernel of the inclusion homomorphism $\mathrm{Mod}(\Sigma^\circ) \to \mathrm{Mod}(\Sigma)$. And according to [24, Theorem 3.18] the inclusion homomorphism has infinite cyclic kernel generated by $(\tau_{d^-})^{-1}\tau_{d^+}$ where $d^\pm$ are the two boundary components of Σ°. If we choose representatives $T_\pm$ for $\tau_{d^\pm}$, then $T_-^{-1}T_+$ is isotopic to the identity when considered as an element of $\mathrm{Diff}^+(\Sigma)$. Moreover, a choice of an isotopy to the identity gives rise to a well defined element of $\pi_1(\mathcal{E};d)$ which is easily identified with the rotation loop along d. This finishes the proof of Theorem C.3. □

Lastly, we want to give 4–dimensional interpretation of paths of death and discuss a consequence of Theorem C.3. Given a smooth 1–parameter family of functions $f_t\colon W \to I$, $t \in I$, we consider the map

$$F\colon W \times I \to I \times I, \quad F(p,t) = \big(f_t(p), t\big)$$

which we call the *trace* of (f_t). Note that the trace F has a 4–dimensional source and a 2–dimensional target. Thus we are back in the setting of Section 2.2 where we discussed the structure of smooth maps between manifolds of these dimensions. Note that the critical locus of F is just the trace of critical points of the maps f_t; the trace of the critical values in $I \times I$ is commonly known as the *Cerf graphic* of the family (f_t). A simple but important observation is that the above construction applied to the constant family of maps $f_t = \pm x^2 \pm y^2 \pm z^2$ recovers the local models for fold singularities (equation (2.3)), and for the standard model for deaths $C_t(x,y,z) = x^3 + 3t \pm y^2 \pm z^2$ we get the cusp models (equation (2.4)). This shows that for a generic homotopy (f_t) in $\mathcal{F}$ the map f has only fold and cusp singularities, where the cusps occur for those maps f_t of codimension 1 with a cubical degeneracy, and it is easy to see that the normal crossing condition in Theorem 2.9 is also satisfied for F. In particular, as pointed out in Remark 2.11, it follows the trace of a generic homotopy is a stable map. For example, if (f_t) is a path of death, then F has one arc of folds emanating from $W \times \{0\}$ for each critical point of $f = f_0$ and the two fold arcs corresponding to p and q eventually run into a cusp while all others persist until $W \times \{1\}$. Moreover, F is injective on its critical locus (since double point would require another codimension 1 function) so that F is stable.

We now discuss a consequence of Theorem C.3 that was used in the proof of Lemma 5.8 during the construction of annular simple wrinkled fibrations with prescribed generalized surface diagrams. Let Y be a closed, orientable 3–manifold equipped with a circle valued Morse function $f\colon Y \to S^1$. Since the definitions of canceling pairs, paths of death, as well as the arguments in the proof of Theorem C.3 are of local nature with respect to the target space, the whole discussion carries over

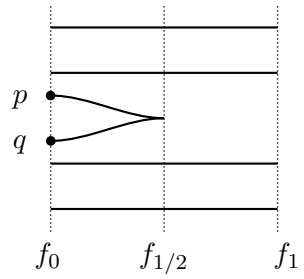


Figure 39: The Cerf graphic of a path of death.

to this new setting. We identify S^1 with $[0,1]/\{0,1\}$ and denote the coordinate on $[0,1]$ by θ. We assume that $f\colon Y \to S^1$ has critical points $\{p_i, q_i\}_{i=1}^N$ with

$$0 < f(p_1) < f(q_1) < \cdots < f(p_N) < f(q_N) < 1 \tag{C.3}$$

such that p_i and q_i form a canceling pair for each i. We would like to cancel all critical points simultaneously. To that end, we consider the space $\mathfrak{Death}(f)$ of homotopies emanating from f for which there are intermediate levels $f(q_i) < \theta_i < f(p_{i+1})$ such that the restriction to $f^{-1}[\theta_{i-1}, \theta_i]$ is a path of death in the sense of Definition C.2. The following result shows that the trace of such a path of death for f is unique in a suitable.

Corollary C.5. *Let $f\colon Y \to S^1$ be as above and let $(f_{0,t}), (f_{1,t}) \in \mathfrak{Death}(f)$ be two paths of death for f. Then the corresponding traces $F_0, F_1\colon Y \times I \to S^1 \times I$ are equivalent via diffeomorphisms $\Phi \in \mathrm{Diff}(Y \times I)$ and $\Psi \in \mathrm{Diff}(S^1 \times I)$, that is,*

$$F_1 = \Phi \circ F_0 \circ \Psi^{-1},$$

such that $\Phi|_{Y\times\{0\}} = \mathrm{id}_{Y\times\{0\}}$ and $\Psi|_{S^1\times\{0\}} = \mathrm{id}_{S^1\times\{0\}}$

Proof. It is an easy consequence of Theorem C.3 that $\mathfrak{Death}(f)$ is connected. Hence, we can find a 2–parameter family of functions $f_{s,t}\colon Y \to S^1$ such that for fixed s we have $(f_{s,t})_{t\in I} \in \mathfrak{Death}(f)$ and the traces F_s give rise to a homotopy from F_0 to F_1. As we noted above, each F_s is a stable map and the stability implies the existence of $\Phi_s \in \mathrm{Diff}(Y \times I)$ and $\Psi_s \in \mathrm{Diff}(S^1 \times I)$ such that

$$F_s = \Psi_s \circ F_0 \circ \Phi_s^{-1}. \tag{C.4}$$

Moreover, according to [47, Theorem 3] the diffeomorphisms Ψ_s and Φ_s can be chosen with smooth dependence on s. In particular, we see that F_0 and F_1 are equivalent via Φ_1 and Ψ_1. However, these maps do not necessarily restrict to the identity on $Y \times \{0\}$ and $S^1 \times \{0\}$ and we have to remedy this fact.

Let $\Phi_{s,0} \in \mathrm{Diff}(Y)$ and $\Psi_{s,0} \in \mathrm{Diff}(S^1)$ be the diffeomorphisms given by restricting Φ_s and Ψ_s to $Y \times \{0\}$ and $S^1 \times \{0\}$, respectively. By construction the restriction of F_s to $Y \times \{0\}$ is naturally identified with $f_{s,0}$ which, in turn, agrees with f by definition. Thus for all s we have

$$f = \Psi_{s,0} \circ f \circ \Phi_{s,0}^{-1}. \tag{C.5}$$

For each $\delta > 0$ we define diffeomorphisms of $Y \times [0,\delta]$ and $S^1 \times [0,\delta]$ given by

$$\widetilde{\Phi}_\delta(y,t) = \big(\Phi^{-1}_{1-t/\delta,0}(y), t\big) \quad \text{and} \quad \widetilde{\Psi}_\delta(\theta,t) = \big(\Psi^{-1}_{1-t/\delta,0}(\theta), t\big)$$

which, according to equation (C.5), make the following square commutative:

$$\begin{array}{ccc} Y \times [0,\delta] & \xrightarrow{\widetilde{\Phi}_\delta} & Y \times [0,\delta] \\ \downarrow{\scriptstyle f\times\mathrm{id}} & & \downarrow{\scriptstyle f\times\mathrm{id}} \\ S^1 \times [0,\delta] & \xrightarrow{\widetilde{\Psi}_\delta} & S^1 \times [0,\delta] \end{array}$$

Moreover, the restrictions of $\widetilde{\Phi}_\delta$ to the boundary of $Y \times [0,\delta]$ satisfies

$$\widetilde{\Phi}_\delta|_{Y\times\{0\}} = \Phi_1^{-1}|_{Y\times\{0\}} \quad \text{and} \quad \widetilde{\Phi}_\delta|_{Y\times\{\delta\}} = \mathrm{id} \tag{C.6}$$

and the analogous properties hold for $\widetilde{\Psi}_\delta$. Our strategy is to identify F_1 with $f \times \mathrm{id}$ in some collar neighborhood of $Y \times \{0\}$ of the form $Y \times [0,\delta]$ and to modify Φ_1 and Ψ_1 in this collar using $\widetilde{\Phi}_\delta$ and $\widetilde{\Psi}_\delta$.

To obtain such a collar, we recall that excellent Morse functions are also stable maps. In particular, $f = f_{s,0} \colon Y \to S^1$ is stable and there must be some $\epsilon > 0$ such that $f_{s,t}$ is stable for $t \le \epsilon$. Reasoning as above we obtain $\phi_{s,t} \in \mathrm{Diff}(Y)$ and $\psi_{s,t} \in \mathrm{Diff}(S^1)$ with $\phi_{s,0} = \mathrm{id}_Y$ and $\psi_{s,0} = \mathrm{id}_{S^1}$ for all s such that

$$f_{s,t} = \psi_{s,t} \circ f_{s,0} \circ \phi_{s,t}^{-1} = \psi_{s,t} \circ f \circ \phi_{s,t}^{-1}.$$

We now define diffeomorphisms of $Y \times [0,\epsilon]$ and $S^1 \times [0,\epsilon]$

$$\widetilde{\phi}_s(y,t) = (\phi_{s,t}(y), t) \quad \text{and} \quad \widetilde{\psi}_s(\theta,t) = (\psi_{s,t}(\theta), t)$$

and for each s we obtain a commutative diagram:

$$\begin{array}{ccccc} Y \times [0,\epsilon] & \xrightarrow{\widetilde{\phi}_s} & Y \times [0,\epsilon] & \xrightarrow{\widetilde{\Phi}_\epsilon} & Y \times [0,\epsilon] \\ {\scriptstyle F_s}\downarrow & & \downarrow{\scriptstyle f\times\mathrm{id}} & & \downarrow{\scriptstyle f\times\mathrm{id}} \\ S^1 \times [0,\epsilon] & \xrightarrow{\widetilde{\psi}_s} & S^1 \times [0,\epsilon] & \xrightarrow{\widetilde{\Psi}_\epsilon} & S^1 \times [0,\epsilon] \end{array}$$

Note that by equation (C.6) the composition $\widetilde{\phi}_s \circ \widetilde{\Phi}_\epsilon \circ \widetilde{\phi}_s^{-1}$ restricts to the identity on $Y \times \{\epsilon\}$ and thus extends to all of $Y \times I$. Similarly, $\widetilde{\psi}_s \circ \widetilde{\Psi}_\epsilon \circ \widetilde{\psi}_s^{-1}$ extends to $S^1 \times I$ and we have

$$F_s = \big(\widetilde{\psi}_s \circ \widetilde{\Psi}_\epsilon \circ \widetilde{\psi}_s^{-1}\big) \circ F_s \circ \big(\widetilde{\phi}_s \circ \widetilde{\Phi}_\epsilon \circ \widetilde{\phi}_s^{-1}\big)^{-1}. \tag{C.7}$$

In particular, combining equations (C.4) and (C.7) for $s = 1$ we get

$$\begin{aligned} F_1 &= \big(\widetilde{\psi}_1 \circ \widetilde{\Psi}_\epsilon \circ \widetilde{\psi}_1^{-1}\big) \circ F_1 \circ \big(\widetilde{\phi}_1 \circ \widetilde{\Phi}_\epsilon \circ \widetilde{\phi}_1^{-1}\big)^{-1} \\ &= \underbrace{\big(\widetilde{\psi}_1 \circ \widetilde{\Psi}_\epsilon \circ \widetilde{\psi}_1^{-1}\big) \circ \Psi_1}_{\Psi} \circ F_0 \circ \underbrace{\Phi_1^{-1} \circ \big(\widetilde{\phi}_1 \circ \widetilde{\Phi}_\epsilon \circ \widetilde{\phi}_1^{-1}\big)^{-1}}_{\Phi^{-1}}. \end{aligned}$$

Now, recall that $\widetilde{\phi}_1$ is the identity on $Y \times \{0\}$ while $\widetilde{\Phi}_\epsilon$ is inverse to Φ_1 by equation (C.6). It follows that $\Phi \in \mathrm{Diff}(Y \times I)$ defined above restricts to the identity on $Y \times \{0\}$ and the same arguments show that $\Psi \in \mathrm{Diff}(S^1 \times I)$ restricts to the identity on $S^1 \times \{0\}$. This finishes the proof. □

Bibliography

[1] S. Akbulut and Ç. Karakurt, *Every 4-manifold is BLF*, J. Gökova Geom. Topol. GGT **2** (2008), 83–106.

[2] V. I. Arnold, S. M. Gusein-Zade, and A. N. Varchenko, *Singularities of differentiable maps. Volume 1*, Modern Birkhäuser Classics, Birkhäuser/Springer, New York, 2012. Reprint of the 1985 edition.

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

[4] R. İ. Baykur, *Existence of broken Lefschetz fibrations*, Int. Math. Res. Not. IMRN (2008), Art. ID rnn 101, 15.

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

[6] R. İ. Baykur and S. Kamada, *Classification of broken Lefschetz fibrations with small fiber genera* (2010), available at `http://de.arxiv.org/abs/1010.5814v2`.

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

[8] S. Behrens and K. Hayano, *Vanishing Cycles and Homotopies of Wrinkled Fibrations* (2012), preprint, available at `http://arxiv.org/abs/1210.5948v2`.

[9] J. M. Boardman, *Singularities of differentiable maps*, Inst. Hautes Études Sci. Publ. Math. **33** (1967), 21–57.

[10] G. E. Bredon, *Topology and geometry*, Graduate Texts in Mathematics, vol. 139, Springer-Verlag, New York, 1993.

[11] T. Bröcker and K. Jänich, *Introduction to differential topology*, Cambridge University Press, Cambridge, 1982. Translated from the German by C. B. Thomas and M. J. Thomas.

[12] P. R. Cromwell, *Knots and links*, Cambridge University Press, Cambridge, 2004.

[13] 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).

[14] J. F. Davis and P. Kirk, *Lecture notes in algebraic topology*, Graduate Studies in Mathematics, vol. 35, American Mathematical Society, Providence, RI, 2001.

[15] S. K. Donaldson, *An application of gauge theory to four-dimensional topology*, J. Differential Geom. **18** (1983), no. 2, 279–315.

[16] ———, *Lefschetz pencils on symplectic manifolds*, J. Differential Geom. **53** (1999), no. 2, 205–236.

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

[18] B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov, *Modern geometry—methods and applications. Part II*, Graduate Texts in Mathematics, vol. 104, Springer-Verlag, New York, 1985. The geometry and topology of manifolds; Translated from the Russian by Robert G. Burns.

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

[20] C. J. Earle and A. Schatz, *Teichmüller theory for surfaces with boundary*, J. Differential Geometry **4** (1970), 169–185.

[21] Y. Eliashberg and N. M. Mishachev, *Wrinkling of smooth mappings and its applications. I*, Invent. Math. **130** (1997), no. 2, 345–369.

[22] H. Endo and Y. Z. Gurtas, *Lantern relations and rational blowdowns*, Proc. Amer. Math. Soc. **138** (2010), no. 3, 1131–1142.

[23] H. Endo, T. E. Mark, and J. Van Horn-Morris, *Monodromy substitutions and rational blow-downs*, J. Topol. **4** (2011), no. 1, 227–253.
[24] B. Farb and D. Margalit, *A Primer on Mapping Class Groups*, Princeton Mathematical Series, vol. 49, Princeton University Press, Providence, RI, 2011.
[25] M. H. Freedman, *The topology of four-dimensional manifolds*, J. Differential Geom. **17** (1982), no. 3, 357–453.
[26] M. H. Freedman and F. Quinn, *Topology of 4-manifolds*, Princeton Mathematical Series, vol. 39, Princeton University Press, Princeton, NJ, 1990.
[27] D. T. Gay and R. C. Kirby, *Constructing Lefschetz-type fibrations on four-manifolds*, Geom. Topol. **11** (2007), 2075–2115.
[28] ———, *Indefinite Morse 2-functions; broken fibrations and generalizations* (2011), preprint, available at `http://arxiv.org/abs/1102.0750`.
[29] ———, *Fiber-connected, indefinite Morse 2-functions on connected n-manifolds*, Proc. Natl. Acad. Sci. USA **108** (2011), no. 20, 8122–8125.
[30] ———, *Reconstructing 4-manifolds from Morse 2-functions* (2012), preprint, available at `http://arxiv.org/abs/1202.3487`.
[31] M. Golubitsky and V. Guillemin, *Stable mappings and their singularities*, Springer-Verlag, New York, 1973. Graduate Texts in Mathematics, Vol. 14.
[32] R. E. Gompf and A. I. Stipsicz, *4-manifolds and Kirby calculus*, Graduate Studies in Mathematics, vol. 20, American Mathematical Society, Providence, RI, 1999.
[33] K. Hayano, *On genus-1 simplified broken Lefschetz fibrations*, Algebr. Geom. Topol. **11** (2011), no. 3, 1267–1322.
[34] ———, *A note on sections of broken Lefschetz fibrations*, Bull. Lond. Math. Soc. **44** (2012), no. 4, 823–836.
[35] ———, *Modification rule of monodromies in R_2-move* (2012), preprint, available at `http://arxiv.org/abs/1203.4299`.
[36] N. V. Ivanov, *Subgroups of Teichmüller modular groups*, Translations of Mathematical Monographs, vol. 115, American Mathematical Society, Providence, RI, 1992.
[37] ———, *Mapping class groups*, Handbook of geometric topology, North-Holland, Amsterdam, 2002, pp. 523–633.
[38] D. Johnson, *Spin structures and quadratic forms on surfaces*, J. London Math. Soc. (2) **22** (1980), no. 2, 365–373.
[39] A. Kas, *On the handlebody decomposition associated to a Lefschetz fibration*, Pacific J. Math. **89** (1980), no. 1, 89–104.
[40] R. C. Kirby, P. Melvin, and P. Teichner, *Cohomotopy sets of 4-manifolds* (2012), available at `http://de.arxiv.org/abs/1203.1608v1`.
[41] M. Korkmaz and A. I. Stipsicz, *Lefschetz fibrations on 4-manifolds*, Handbook of Teichmüller theory. Vol. II, IRMA Lect. Math. Theor. Phys., vol. 13, Eur. Math. Soc., Zürich, 2009, pp. 271–296.
[42] K. Lamotke, *The topology of complex projective varieties after S. Lefschetz*, Topology **20** (1981), no. 1, 15–51.
[43] F. Laudenbach and V. Poénaru, *A note on 4-dimensional handlebodies*, Bull. Soc. Math. France **100** (1972), 337–344.
[44] H. B. Lawson Jr. and M.-L. Michelsohn, *Spin geometry*, Princeton Mathematical Series, vol. 38, Princeton University Press, Princeton, NJ, 1989.
[45] Y. Lekili, *Wrinkled fibrations on near-symplectic manifolds*, Geom. Topol. **13** (2009), no. 1, 277–318. Appendix B by R. İnanç Baykur.
[46] H. I. Levine, *The singularities, S^q_1*, Illinois J. Math. **8** (1964), 152–168.
[47] J. N. Mather, *Stability of C^∞ mappings. II. Infinitesimal stability implies stability*, Ann. of Math. (2) **89** (1969), 254–291.
[48] Y. Matsumoto, *Lefschetz fibrations of genus two—a topological approach*, Topology and Teichmüller spaces (Katinkulta, 1995), World Sci. Publ., River Edge, NJ, 1996, pp. 123–148.
[49] G. Mess, *The Torelli groups for genus 2 and 3 surfaces*, Topology **31** (1992), no. 4, 775–790.
[50] J. Milnor, *Lectures on the h-cobordism theorem*, Notes by L. Siebenmann and J. Sondow, Princeton University Press, Princeton, N.J., 1965.
[51] B. Moishezon, *Complex surfaces and connected sums of complex projective planes*, Lecture Notes in Mathematics, Vol. 603, Springer-Verlag, Berlin, 1977. With an appendix by R. Livne.

[52] J. M. Montesinos, *Heegaard diagrams for closed 4-manifolds*, Geometric topology (Proc. Georgia Topology Conf., Athens, Ga., 1977), Academic Press, New York, 1979, pp. 219–237.

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

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

[55] P. Ozsváth and Z. Szabó, *Holomorphic disks and topological invariants for closed three-manifolds*, Ann. of Math. (2) **159** (2004), no. 3, 1027–1158.

[56] ———, *Holomorphic triangles and invariants for smooth four-manifolds*, Adv. Math. **202** (2006), no. 2, 326–400.

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

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

[59] J.-P. Serre, *Formes bilinéaires symétriques entières à discriminant* ± 1, Séminaire Henri Cartan, 1961/62, Exp. 14-15, Secrétariat mathématique, Paris, 1961/1962, pp. 16.

[60] A. I. Stipsicz, *Spin structures on Lefschetz fibrations*, Bull. London Math. Soc. **33** (2001), no. 4, 466–472.

[61] C. H. Taubes, *Seiberg Witten and Gromov invariants for symplectic 4-manifolds*, First International Press Lecture Series, vol. 2, International Press, Somerville, MA, 2000. Edited by Richard Wentworth.

[62] ———, *The structure of pseudo-holomorphic subvarieties for a degenerate almost complex structure and symplectic form on* $S^1 \times B^3$, Geom. Topol. **2** (1998), 221–332.

[63] ———, *The geometry of the Seiberg-Witten invariants*, Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998), 1998, pp. 493–504.

[64] ———, *Seiberg-Witten invariants and pseudo-holomorphic subvarieties for self-dual, harmonic 2-forms*, Geom. Topol. **3** (1999), 167–210.

[65] R. Thom, *Les singularités des applications différentiables*, Ann. Inst. Fourier, Grenoble **6** (1955), 43–87.

[66] M. Usher, *The Gromov invariant and the Donaldson-Smith standard surface count*, Geom. Topol. **8** (2004), 565–610.

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

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

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

[70] L. C. Wilson, *Corrections to: "Nonopenness of the set of Thom-Boardman maps"*, Pacific J. Math. **85** (1979), no. 2, 501–502.

Summary

The broad context of this thesis is the topology of smooth, 4–dimensional manifolds. The central idea is to study certain maps from 4–manifolds to surfaces known as *simple wrinkled fibrations*, which can be described combinatorially in terms of curve configurations on surfaces, so called *surface diagrams*. These notions were introduced by Williams [67, 68] who also showed that all closed, oriented 4–manifolds admit simple wrinkled fibrations over S^2 and that the corresponding surface diagrams determine the topology of the 4–manifolds. In principle, this allows to translate the complexities of smooth 4–manifolds into 2–dimensional problems. We investigate how simple wrinkled fibrations and surface diagrams interact with the topology of the underlying 4–manifolds. The results of this thesis have partially already appeared in the author's article [7].

The thesis has two more or less independent parts. The first one is concerned with foundational questions about simple wrinkled fibrations, surface diagrams, and their relation. We define simple wrinkled fibrations over arbitrary base surfaces and show how to reduce the general case to the study of certain simple wrinkled fibration over the annulus which we call *annular*. As a combinatorial counterpart we introduce *generalized* surface diagrams and establish the following result.

Theorem. *There is a bijective correspondence between annular simple wrinkled fibrations and generalized surface diagrams (up to suitable notions of equivalence).*

From this we deduce precise correspondence results for simple wrinkled fibrations over D^2 and S^2. These improve the aforementioned result of Williams [67] in that they not only recover a 4–manifold from a surface diagram but also the simple wrinkled fibration that gave rise to the surface diagram. Along the way, we also exhibit handle decompositions induced by simple wrinkled fibrations and explain how they are encoded in surface diagrams.

In the second part we focus on the interplay between the combinatorics of surface diagrams and the topology of the corresponding 4–manifolds. Among other things, we give surface diagram interpretations of certain cut-and-paste operations on 4–manifolds. These include connected sums with $S^2 \times S^2$ and $\mathbb{C}\mathrm{P}^2$ (with either orientation) as well as surgeries on curves and 2–spheres. Using these techniques, we classify the closed 4–manifolds that can be described by surface diagrams of the lowest possible genus.

Theorem. *A closed, oriented 4–manifold admits a genus one surface diagram if and only if it is diffeomorphic to* $kS^2 \times S^2$ *or* $m\mathbb{C}\mathrm{P}^2 \# n\overline{\mathbb{C}\mathrm{P}^2}$ *where* $k, m, n \geq 1$.

We then investigate how surface diagrams encode homotopy information. We discuss the fundamental group, homology, the intersection form, as well as spin and spin^c structures. Our main application is the following.

Theorem. *Let $w\colon X \to S^2$ be a simple wrinkled fibration with surface diagram $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$. If X is simply connected and the fibers of w are null-homologous in X, then the homeomorphism type of X is determined by the homology classes $[c_i] \in H_1(\Sigma)$.*

Note that the condition on the fibers does not put restrictions on X since it is known that all maps $X \to S^2$ are homotopic to simple wrinkled fibrations. In particular, we can work with the homotopy class of the constant map. An interesting observation is that the *diffeomorphism type* of X a priori depends on the *isotopy classes* of the curves $c_i \subset \Sigma$. Furthermore, the difference between isotopy and homology for curves on a surface is measured in terms of the *Torelli group* of the surface, which is the non-linear and more mysterious part of the mapping class group. This suggests the possibility of a relation between exotic smooth structures on 4–manifolds and Torelli groups.