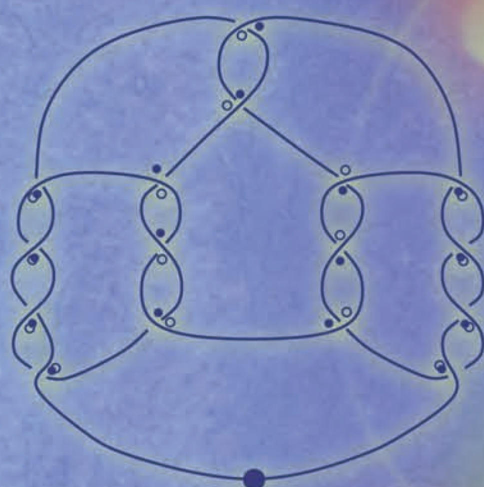
Clay Mathematics Proceedings
Volume 5

Floer Homology, Gauge Theory, and Low-Dimensional Topology

Proceedings of the
Clay Mathematics Institute
2004 Summer School

Alfréd Rényi Institute of Mathematics
Budapest, Hungary

June 5–26, 2004



David A. Ellwood
Peter S. Ozsváth
András I. Stipsicz
Zoltán Szabó
Editors

American Mathematical Society
Clay Mathematics Institute

Floer Homology, Gauge Theory, and Low-Dimensional Topology

Clay Mathematics Proceedings
Volume 5

Floer Homology, Gauge Theory, and Low-Dimensional Topology

Proceedings of the
Clay Mathematics Institute
2004 Summer School

Alfréd Rényi Institute of Mathematics
Budapest, Hungary

June 5–26, 2004

David A. Ellwood
Peter S. Ozsváth
András I. Stipsicz
Zoltán Szabó
Editors

American Mathematical Society

Clay Mathematics Institute

2000 *Mathematics Subject Classification.* Primary 57R17, 57R55, 57R57, 57R58, 53D05, 53D40, 57M27, 14J26.

The cover illustrates a Kinoshita-Terasaka knot (a knot with trivial Alexander polynomial), and two Kauffman states. These states represent the two generators of the Heegaard Floer homology of the knot in its topmost filtration level. The fact that these elements are homologically non-trivial can be used to show that the Seifert genus of this knot is two, a result first proved by David Gabai.

Library of Congress Cataloging-in-Publication Data

Clay Mathematics Institute. Summer School (2004 : Budapest, Hungary)
Floer homology, gauge theory, and low-dimensional topology : proceedings of the Clay Mathematics Institute 2004 Summer School, Alfréd Rényi Institute of Mathematics, Budapest, Hungary, June 5–26, 2004 / David A. Ellwood . . . [et al.], editors.
p. cm. — (Clay mathematics proceedings, ISSN 1534-6455 ; v. 5)
ISBN 0-8218-3845-8 (alk. paper)
1. Low-dimensional topology—Congresses. 2. Symplectic geometry—Congresses. 3. Homology theory—Congresses. 4. Gauge fields (Physics)—Congresses. I. Ellwood, D. (David), 1966– II. Title. III. Series.

QA612.14.C55 2004
514′.22—dc22 2006042815

Published by the American Mathematical Society, Providence, RI,
for the Clay Mathematics Institute, Cambridge, MA.
Printed in the United States of America.

∞ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at `http://www.ams.org/`
Visit the Clay Mathematics Institute home page at `http://www.claymath.org/`

10 9 8 7 6 5 4 3 2 1 11 10 09 08 07 06

Contents

Introduction

The Clay Mathematical Institute hosted its 2004 Summer School on *Floer homology, gauge theory, and low–dimensional topology* at the Alfréd Rényi Institute of Mathematics in Budapest, Hungary. The aim of this school was to bring together students and researchers in the rapidly developing crossroads of gauge theory and low–dimensional topology. In part, the hope was to foster dialogue across closely related disciplines, some of which were developing in relative isolation until fairly recently. The lectures centered on several topics, including Heegaard Floer theory, knot theory, symplectic and contact topology, and Seiberg–Witten theory. This volume is based on lecture notes from the school, some of which were written in close collaboration with assigned teaching assistants. The lectures have revised the choice of material somewhat from that presented at the school, and the topics have been organized to fit together in logical categories. Each course consisted of two to five lectures, and some had associated problem sessions in the afternoons.

Mathematical gauge theory studies connections on principal bundles, or, more precisely, the solution spaces of certain partial differential equations for such connections. Historically, these equations have come from mathematical physics. Gauge theory as a tool for studying topological properties of four–manifolds was pioneered by the fundamental work of Simon Donaldson in the early 1980's. Since the birth of the subject, it has retained its close connection with symplectic topology, a subject whose intricate structure was illuminated by Mikhail Gromov's introduction of pseudo–holomorphic curve techniques, also introduced in the early 1980's. The analogy between these two fields of study was further underscored by Andreas Floer's construction of an infinite–dimensional variant of Morse theory that applies in two *a priori* different contexts: either to define symplectic invariants for pairs of Lagrangian submanifolds of a symplectic manifold (the so–called *Lagrangian Floer homology*), providing obstructions to disjoining the submanifolds through Hamiltonian isotopies, or to give topological invariants for three–manifolds (the so–called *instanton Floer homology*), which fit into a framework for calculating Donaldson's invariants for smooth four–manifolds.

In the mid–1990's, gauge–theoretic invariants for four–manifolds underwent a dramatic change with the introduction of a new set of partial differential equations introduced by Nathan Seiberg and Edward Witten in their study of string theory. Very closely connected with the underlying geometry of the four–manifolds over which they are defined, the Seiberg–Witten equations lead to four–manifold invariants which are in many ways much easier to work with than the anti–self–dual Yang–Mills equations which Donaldson had studied. The introduction of the new invariants led to a revolution in the field of smooth four–manifold topology.

Highlights in four–manifold topology from this period include the deep theorems of Clifford Taubes about the differential topology of symplectic four–manifolds. These give an interpretation of some of Gromov's invariants for symplectic manifolds in terms of the Seiberg–Witten invariants of the underlying smooth four–manifold. Another striking consequence of the new invariants was a quick, elegant proof by Kronheimer and Mrowka of a conjecture by Thom, stating that the algebraic curves in the complex projective plane minimize genus in their homology class. The invariants were also used particularly effectively in work of Ron Fintushel and Ron Stern, who discovered several operations on smooth four–manifolds, for which the Seiberg–Witten invariants transform in a predictable manner. These operations include *rational blow–downs*, where the neighborhood of a certain chain of spheres is replaced by a space with vanishing second homology, and also *knot surgery*, for which the Alexander polynomial of a knot is reflected in the Seiberg–Witten invariants of a corresponding four–manifold. These operations can be used to construct a number of smooth four–manifolds with interesting properties.

In an attempt to better understand the somewhat elusive gauge theoretic invariants, a different construction was given by Peter Ozsváth and Zoltán Szabó. They formulated an invariant for three– and four–manifolds which takes as its starting point a Lagrangian Floer homology associated to Heegaard diagrams for three–manifolds. The resulting "Heegaard Floer homology" theory is conjecturally isomorphic to Seiberg–Witten theory, but more topological and combinatorial in its flavor and correspondingly easier to work with in certain contexts. Moreover, this theory has benefitted a great deal from an array of contemporary results rendering various analytical and geometric structures in a more topological and combinatorial form, such as Donaldson's introduction of "Lefschetz pencils" in the symplectic category and Giroux's correspondence between open book decompositions and contact structures.

The two lecture series of Ozsváth and Szabó in the first section of this volume provide a leisurely introduction to Heegaard Floer theory. The first lecture series (the lectures given by Szabó at the Summer School) start with the basic notions, and move on to the constructions of the primary variants of Floer homology groups and maps between them. These lectures also cover basics of a corresponding Heegaard Floer homology invariant for knots. The second lecture series (given by Ozsváth) gives a rapid proof of one of the basic calculational tools of the subject, the surgery exact triangle, and its immediate applications. Special emphasis is placed on a Dehn surgery characterization of the unknot, a result whose proof is outlined in these lectures. Section 1 concludes with the lecture notes from Goda's course. Whereas Heegaard diagrams correspond to real–valued Morse theory in three dimensions, in these lectures, Goda considers circle–valued Morse theory for link complements. He uses this theory to give obstructions to a knot being fibered.

The main theme in Section 2 is contact geometry and its interplay with Floer homology. The lectures of John Etnyre give a detailed account of open book decompositions and contact structures, and the Giroux correspondence. The proof of the Giroux correspondence is followed by some applications of this theory, including an embedding theorem for weak symplectic fillings, which turned out to be a crucial step in many of the recent developments of the subject, including the verification of *Property P* by Kronheimer and Mrowka. The definition of the contact invariant in Heegaard Floer theory (resting on the above mentioned Giroux correspondence) is discussed in the lecture notes of András Stipsicz, together with a short discussion on contact surgeries. Results regarding existence of tight contact structures on various 3–manifolds and their fillability properties are also given. A similar application of the contact invariants is described in the paper of Paolo Lisca and András Stipsicz, with the use of minimum machinery required in the proof. A different type of Floer homology (called *contact homology*) is studied in Tobias Ekholm's paper. A classical result of Gromov states that any exact Lagrangian immersion into $\mathbb{C}^n$ has at least one double–point. Ekholm generalizes this result, using Floer homology to give estimates on the minimum number of double–points of an exact Legendrian immersion into some Euclidean space.

Section 3 discusses symplectic geometry and Seiberg–Witten invariants. Ron Fintushel's lectures give an introduction to Seiberg–Witten invariants and the knot surgery construction. The lectures give a thorough discussion of how the Seiberg–Witten invariants transform under the knot surgery operation. Applications include exotic embeddings of surfaces in smooth four–manifolds. Ron Stern's contribution describes the current state of art in the classification of smooth 4–manifolds, and collects a number of intriguing questions and problems which can motivate further results in the subject. The paper of Jongil Park provides new applications of the rational blow–down construction, which led him to discover symplectic 4–manifolds homeomorphic but not diffeomorphic to rational surfaces with small Euler characteristic. Tian–Jun Li studies symplectic 4–manifolds systematically using the generalization of the notion of the holomorphic Kodaira dimension κ to this category. After the discussion of the $\kappa = -\infty$ case, the state of the art for $\kappa = 0$ is described, where a reasonably nice classification scheme is expected. The contribution of Denis Auroux also addresses the problem of understanding symplectic 4–manifolds, but from a completely different point of view. In this case the manifolds are presented as branched covers of the complex projective plane along certain curves, and the discussion centers on the possibility of getting symplectic invariants from topological properties of these branch sets. The volume concludes with Ivan Smith's contribution, where the author reviews basics about symplectic fibrations, leading him (in a joint project with Paul Seidel) to knot invariants defined using symplectic topology and Floer homology, conjecturally recapturing the celebrated knot invariants of Khovanov.

It is hoped that this volume will give the reader a sampling of these many new and exciting developments in low–dimensional topology and symplectic geometry. Before commencing with the mathematics, we would like to pause to thank some of the many people who have contributed in one way or another to this volume. We would like to thank Arthur Greenspoon for a meticulous proofreading of this text. We would like to thank the Clay Mathematical Institute for making this program possible, through both their financial support and their enthusiasm; special mention goes to Vida Salahi for her careful and diligent work in bringing this volume to print. Next, we thank the staff at the Rényi Institute for helping to create a conducive environment for the Summer School. We would like to thank the lecturers for giving clear, accessible accounts of their research, and we are also grateful to their course assistants, who helped make these courses run smoothly. Finally, we thank

the many students and young researchers whose remarkable energy and enthusiasm helped to make the conference a success.

David Ellwood, Peter Ozsváth, András Stipsicz, Zoltán Szabó

October 2005

Heegaard Floer Homology and Knot Theory

Clay Mathematics Proceedings
Volume 5, 2006

An Introduction to Heegaard Floer Homology

Peter Ozsváth and Zoltán Szabó

1. Introduction

The aim of these notes is to give an introduction to Heegaard Floer homology for closed oriented 3-manifolds [**31**]. We will also discuss a related Floer homology invariant for knots in S^3 [**29**], [**34**].

Let Y be an oriented closed 3-manifold. The simplest version of Heegaard Floer homology associates to Y a finitely generated Abelian group $\widehat{HF}(Y)$. This homology is defined with the help of Heegaard diagrams and Lagrangian Floer homology. Variants of this construction give related invariants $HF^+(Y)$, $HF^-(Y)$, $HF^\infty(Y)$.

While its construction is very different, Heegaard Floer homology is closely related to Seiberg-Witten Floer homology [**10, 15, 17**], and instanton Floer homology [**3, 4, 7**]. In particular it grew out of our attempt to find a more topological description of Seiberg-Witten theory for three-manifolds.

2. Heegaard decompositions and diagrams

Let Y be a closed oriented three-manifold. In this section we describe decompositions of Y into more elementary pieces, called handlebodies.

A *genus g handlebody* U is diffeomorphic to a regular neighborhood of a bouquet of g circles in $\mathbb{R}^3$; see Figure 1. The boundary of U is an oriented surface of genus g. If we glue two such handlebodies together along their common boundary, we get a closed 3-manifold

$$Y = U_0 \cup_\Sigma U_1$$

oriented so that Σ is the oriented boundary of U_0. This is called a Heegaard decomposition for Y.

2.1. Examples. The simplest example is the (genus 0) decomposition of S^3 into two balls. A similar example is given by taking a tubular neighborhood of the unknot in S^3. Since the complement is also a solid torus, we get a genus 1 Heegaard decomposition of S^3.

2000 *Mathematics Subject Classification.* 57R58, 57M27.

PO was partially supported by NSF Grant Number DMS 0234311.
ZSz was partially supported by NSF Grant Number DMS 0406155 .

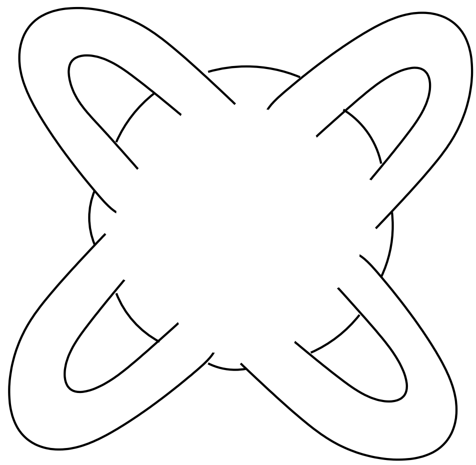


FIGURE 1. A handlebody of genus 4.

Other simple examples are given by lens spaces. Take

$$S^3 = \{(z,w) \in \mathbb{C}^2 |\ |z^2| + |w|^2 = 1\}$$

Let $(p,q) = 1$, $1 \leq q < p$. The lens space $L(p.q)$ is given by dividing out S^3 by the free $\mathbb{Z}/p$ action

$$f : (z,w) \longrightarrow (\alpha z, \alpha^q w),$$

where $\alpha = e^{2\pi i/p}$. Clearly $\pi_1(L(p,q)) = \mathbb{Z}/p$. Note also that the solid tori $U_0 = \{|z| \leq \frac{1}{2}\ \}$, $U_1 = \{|z| \geq \frac{1}{2}\}$ are preserved by the action, and their quotients are also solid tori. This gives a genus 1 Heegaard decomposition of $L(p,q)$.

2.2. Existence of Heegaard decompositions. While the small genus examples might suggest that 3-manifolds that admit Heegaard decompositions are special, in fact the opposite is true:

THEOREM 2.1. *([**39**]) Let Y be an oriented closed three-dimensional manifold. Then Y admits a Heegaard decomposition.*

PROOF. Start with a triangulation of Y. The union of the vertices and the edges gives a graph in Y. Let U_0 be a small neighborhood of this graph. In other words replace each vertex by a ball, and each edge by a solid cylinder. By definition U_0 is a handlebody. It is easy to see that $Y - U_0$ is also a handlebody, given by a regular neighborhood of a graph on the centers of the triangles and tetrahedra in the triangulation. □

2.3. Stabilizations. It follows from the above proof that the same three-manifold admits lots of different Heegaard decompositions. In particular, given a Heegaard decomposition $Y = U_0 \cup_\Sigma U_1$ of genus g, we can define another decomposition of genus $g+1$ by choosing two points in Σ and connecting them by a small unknotted arc γ in U_1. Let U_0' be the union of U_0 and a small tubular neighborhood N of γ. Similarly let $U_1' = U_1 - N$. The new decomposition

$$Y = U_0' \cup_{\Sigma'} U_1'$$

is called the *stabilization* of $Y = U_0 \cup_\Sigma U_1$. Clearly $g(\Sigma') = g(\Sigma) + 1$. For an easy example note that the genus 1 decomposition of S^3 described earlier is the stabilization of the genus 0 decomposition.

According to a theorem of Singer [**39**], any two Heegaard decompositions can be connected by stabilizations (and destabilizations):

THEOREM 2.2. *Let (Y, U_0, U_1) and (Y, U_0', U_1') be two Heegaard decompositions of Y of genus g and g' respectively. Then for k large enough the $(k-g')$-fold stabilization of the first decomposition is diffeomorphic to the $(k-g)$-fold stabilization of the second decomposition.*

2.4. Heegaard diagrams. In view of Theorem 2.2, if we find an invariant for Heegaard decompositions with the property that it does not change under stabilization, then this is in fact a three-manifold invariant. For example the Casson invariant [**1, 37**] is defined in this way. However, for the definition of Heegaard Floer homology we need some additional information which is given by diagrams.

Let us start with a handlebody U of genus g.

DEFINITION 2.3. A set of attaching circles $(\gamma_1, ..., \gamma_g)$ for U is a collection of closed embedded curves in $\Sigma_g = \partial U$ with the following properties

- The curves γ_i are disjoint from each other.
- $\Sigma_g - \gamma_1 - \cdots - \gamma_g$ is connected.
- The curves γ_i bound disjoint embedded disks in U.

REMARK 2.4. The second property in the above definition is equivalent to the property that $([\gamma_1], ..., [\gamma_g])$ are linearly independent in $H_1(\Sigma, \mathbb{Z})$.

DEFINITION 2.5. Let (Σ_g, U_0, U_1) be a genus g Heegaard decomposition for Y. A *compatible Heegaard diagram* is given by Σ_g together with a collection of curves $\alpha_1, ..., \alpha_g, \beta_1, ..., \beta_g$ with the property that $(\alpha_1, ..., \alpha_g)$ is a set of attaching circles for U_0 and $(\beta_1, ..., \beta_g)$ is a set of attaching circles for U_1.

REMARK 2.6. A Heegaard decomposition of genus $g > 1$ admits lots of different compatible Heegaard diagrams.

In the opposite direction any diagram $(\Sigma_g, \alpha_1, ..., \alpha_g, \beta_1, ..., \beta_g)$ where the α and β curves satisfy the first two conditions in Definition 2.3 determines uniquely a Heegaard decomposition and therefore a 3-manifold.

2.5. Examples. It is helpful to look at a few examples. The genus 1 Heegaard decomposition of S^3 corresponds to a diagram $(\Sigma_1, \alpha, \beta)$ where α and β meet transversely in a unique point. $S^1 \times S^2$ corresponds to $(\Sigma_1, \alpha, \alpha)$.

The lens space $L(p, q)$ has a diagram $(\Sigma_1, \alpha, \beta)$ where α and β intersect at p points and in a standard basis $x, y \in H_1(\Sigma_1) = \mathbb{Z} \oplus \mathbb{Z}$, $[\alpha] = y$ and $[\beta] = px + qy$.

Another example is given in Figure 2. Here we think of S^2 as the plane together with the point at infinity. In the picture the two circles on the left are identified, or equivalently we glue a handle to S^2 along these circles. Similarly we identify the two circles on the right side of the picture. After this identification the two horizontal lines become closed circles α_1 and α_2. As for the two β curves, β_1 lies in the plane and β_2 goes through both handles once.

DEFINITION 2.7. We can define a one-parameter family of Heegaard diagrams by changing the right side of Figure 2. For $n > 0$ instead of twisting around the

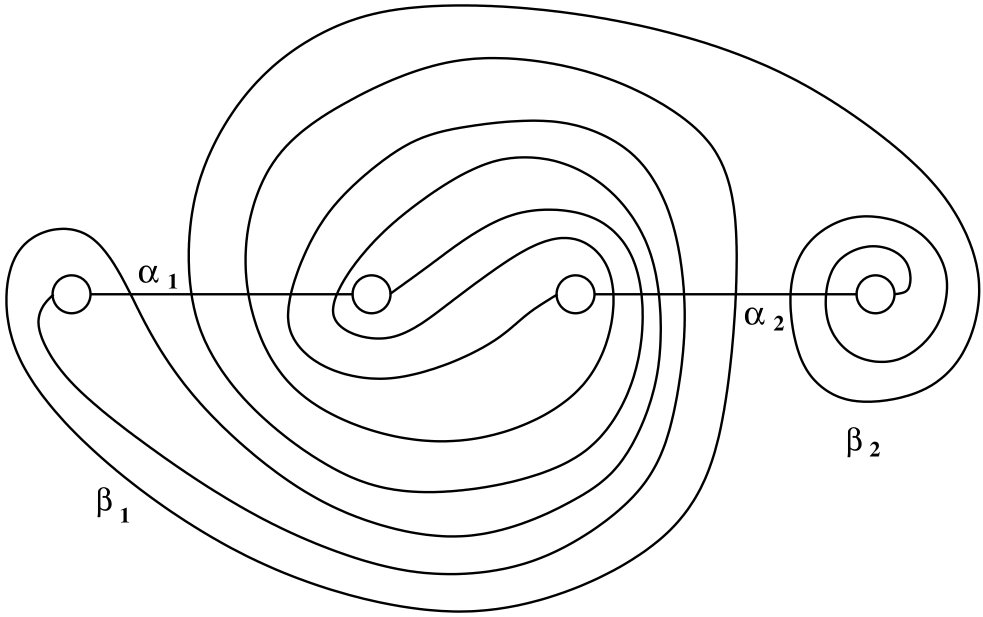


FIGURE 2. A genus 2 Heegaard diagram.

right circle twice as in the picture, twist n times. When $n < 0$, twist $-n$ times in the opposite direction. Let Y_n denote the corresponding three-manifold.

2.6. Heegaard moves. While a Heegaard diagram is a good way to describe Y, the same three-manifold has lots of different diagrams. There are three basic moves on diagrams that do not change the underlying three-manifold. These are *isotopy*, *handle-slide* and *stabilization*. The first two moves can be described for attaching circles $\gamma_1, ..., \gamma_g$ for a given handlebody U:

An isotopy moves $\gamma_1, ..., \gamma_g$ in a one-parameter family in such a way that the curves remain disjoint.

For a handle-slide, we choose two of the curves, say γ_1 and γ_2, and replace γ_1 with γ_1' provided that γ_1' is any simple, closed curve which is disjoint from the $\gamma_1, \ldots, \gamma_g$ with the property that γ_1', γ_1 and γ_2 bound an embedded pair of pants in $\Sigma - \gamma_3 - \ldots - \gamma_g$ (see Figure 3 for a genus 2 example).

PROPOSITION 2.8. *(*[**38**]*) Let U be a handlebody of genus g, and let $(\alpha_1, ..., \alpha_g)$, $(\alpha_1', ..., \alpha_g')$ be two sets of attaching circles for U. Then the two sets can be connected by a sequence of isotopies and handle-slides.*

The stabilization move is defined as follows. We enlarge Σ by making a connected sum with a genus 1 surface $\Sigma' = \Sigma \# E$ and replace $\{\alpha_1, ..., \alpha_g\}$ and $\{\beta_1, ..., \beta_g\}$ by $\{\alpha_1, \ldots, \alpha_{g+1}\}$ and $\{\beta_1, \ldots, \beta_{g+1}\}$ respectively, where α_{g+1} and β_{g+1} are a pair of curves supported in E, meeting transversally in a single point. Note that the new diagram is compatible with the stabilization of the original decomposition.

Combining Theorem 2.2 and Proposition 2.8 we get the following

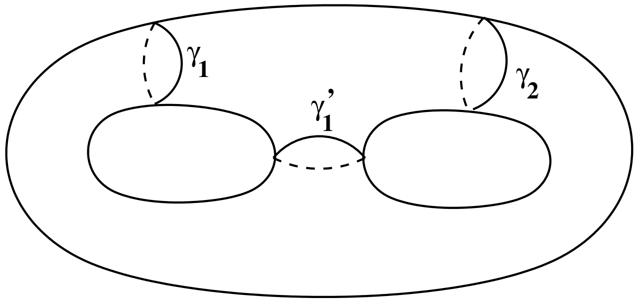


FIGURE 3. Handlesliding γ_1 over γ_2.

THEOREM 2.9. *Let Y be a closed oriented 3-manifold. Let*

$$(\Sigma_g, \alpha_1, ..., \alpha_g, \beta_1, ..., \beta_g), \quad (\Sigma_{g'}, \alpha'_1, ..., \alpha'_{g'}, \beta'_1, ..., \beta'_{g'})$$

be two Heegaard diagrams of Y. Then by applying sequences of isotopies, handle-slides and stabilizations we can change the above diagrams so that the new diagrams are diffeomorphic to each other.

2.7. The basepoint. In later sections we will also look at pointed Heegaard diagrams $(\Sigma_g, \alpha_1, ..., \alpha_g, \beta_1, ..., \beta_g, z)$, where the basepoint $z \in \Sigma_g$ is chosen in the complement of the curves

$$z \in \Sigma_g - \alpha_1 - ... - \alpha_g - \beta_1 - ... - \beta_g.$$

There is a notion of pointed Heegaard moves. Here we also allow isotopy for the basepoint. During isotopy we require that z is disjoint from the curves. For the pointed handle-slide move we require that z is not in the pair of pants region where the handle-slide takes place. The following is proved in [**31**].

PROPOSITION 2.10. *Let z_1 and z_2 be two basepoints. Then the pointed Heegaard diagrams*

$$(\Sigma_g, \alpha_1, ..., \alpha_g, \beta_1, ..., \beta_g, z_1) \text{ and } (\Sigma_g, \alpha_1, ..., \alpha_g, \beta_1, ..., \beta_g, z_2)$$

can be connected by a sequence of pointed isotopies and handle-slides.

3. Morse functions and Heegaard diagrams

In this section we study a Morse-theoretic approach to Heegaard decompositions. In Morse theory, see [**20**], [**21**], one studies smooth functions on n-dimensional manifolds $f : M^n \to \mathbb{R}$. A point $P \in Y$ is a critical point of f if for some coordinate system $(x_1, ..., x_n)$ around P, $\frac{\partial f}{\partial x_i} = 0$ for $i = 1, ..., n$. At a critical point the Hessian matrix $H(P)$ is given by the second partial derivatives $H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$. A critical point P is called non-degenerate if $H(P)$ is non-singular. This notion is independent of the choice of coordinate system.

DEFINITION 3.1. The function $f : M^n \to \mathbb{R}$ is called a Morse function if all the critical points are non-degenerate.

Now suppose that f is a Morse function and P is a critical point. Since $H(P)$ is symmetric, it induces an inner product on the tangent space. The dimension of a maximal negative-definite subspace is called the index of P. In other words we can diagonalize $H(P)$ over the reals, and $\text{index}(P)$ is the number of negative entries in the diagonal.

Clearly a local minimum of f has index 0, while a local maximum has index n. The local behavior of f around a critical point is studied in [**20**]:

PROPOSITION 3.2. *([**20**]) Let P be an index i critical point of f. Then there is a diffeomorphism h between a neighborhood U of $0 \in \mathbb{R}^n$ and a neighborhood U' of $P \in M^n$ so that*

$$f \circ h = -\sum_{j=1}^{i} x_j^2 + \sum_{j=i+1}^{n} x_j^2.$$

For us it will be beneficial to look at a special class of Morse functions:

DEFINITION 3.3. A Morse function f is called self-indexing if for each critical point P we have $f(P) = \text{index}(P)$.

PROPOSITION 3.4. [**20**] *Every smooth n-dimensional manifold M admits a self-indexing Morse function. Furthermore, if M is connected and has no boundary, then we can choose f so that it has unique index 0 and index n critical points.*

The following exercises can be proved by studying how the level sets $f^{-1}((\infty, t])$ change when t goes through a critical value.

EXERCISE 3.5. If $f\colon Y \longrightarrow [0,3]$ is a self-indexing Morse function on Y with one minimum and one maximum, then f induces a Heegaard decomposition with Heegaard surface $\Sigma = f^{-1}(3/2)$, and handlebodies $U_0 = f^{-1}[0,3/2]$, $U_1 = f^{-1}[3/2,3]$.

EXERCISE 3.6. Show that if Σ has genus g, then f has g index one and g index two critical points.

Let us denote the index 1 and 2 critical points of f by $P_1,...,P_g$ and $Q_1,...,Q_g$ respectively.

LEMMA 3.7. *The Morse function and a Riemannian metric on Y induces a Heegaard diagram for Y.*

PROOF. Take the gradient vector field ∇f of the Morse function. For each point $x \in \Sigma$ we can look at the gradient trajectory of $\pm\nabla f$ that goes through x. Let α_i denote the set of points that flow down to the critical point P_i and let β_i correspond to the points that flow up to Q_i. It follows from Proposition 3.2 and the fact that f is self-indexing that α_i, β_i are simple closed curves in Σ. It is also easy to see that $\alpha_1,...,\alpha_g$ and $\beta_1,...,\beta_g$ are attaching circles for U_0 and U_1 respectively. It follows that this is a Heegaard diagram of Y compatible with the given Heegaard decomposition. $\square$

4. Symmetric products and totally real tori

To a pointed Heegaard diagram $(\Sigma_g, \alpha_1,...,\alpha_g, \beta_1,...,\beta_g, z)$ we can associate certain configuration spaces that will be used in later sections in the definition of Heegaard Floer homology. Our ambient space is

$$\text{Sym}^g(\Sigma_g) = \Sigma_g \times \cdots \times \Sigma_g / S_g\ ,$$

where S_g denotes the symmetric group on g letters. In other words $\text{Sym}^g(\Sigma_g)$ consists of unordered g-tuples of points in Σ_g where the same points can appear more than once. Although S_g does not act freely, $\mathrm{Sym}^g(\Sigma_g)$ is a smooth manifold. Furthermore a complex structure on Σ_g induces a complex structure on $\mathrm{Sym}^g(\Sigma_g)$.

The topology of symmetric products of surfaces is studied in [**16**].

PROPOSITION 4.1. *Let Σ be a genus g surface. Then*

$$\pi_1(\mathrm{Sym}^g(\Sigma)) \cong H_1(\mathrm{Sym}^g(\Sigma)) \cong H_1(\Sigma).$$

PROPOSITION 4.2. *Let Σ be a surface of genus $g > 2$. Then*

$$\pi_2(\mathrm{Sym}^g(\Sigma)) \cong \mathbb{Z}.$$

The generator of $S \in \pi_2(\mathrm{Sym}^g(\Sigma))$ can be constructed in the following way: Let $\tau : \Sigma \longrightarrow \Sigma$ be an orientation preserving involution with the property that $\Sigma/\tau = S^2$. (such a map is called a hyperelliptic involution). Then $(y, \tau(y), z, ..., z)$ is a sphere representing S. An explicit calculation gives

LEMMA 4.3. *Let $S \in \pi_2(\mathrm{Sym}^g(\Sigma))$ be the positive generator as above. Then*

$$\langle c_1(\mathrm{Sym}^g(\Sigma_g)), [S]\rangle = 1$$

REMARK 4.4. The small genus examples can be understood as well. When $g = 1$ we get a torus and π_2 is trivial. $\mathrm{Sym}^2(\Sigma_2)$ is diffeomorphic to the real four-dimensional torus blown up at one point. Here π_2 is large but after dividing by the action of $\pi_1(\mathrm{Sym}^2(\Sigma_2))$ we get a group π_2' satisfying

$$\pi_2'(\mathrm{Sym}^2(\Sigma_2)) \cong \mathbb{Z}$$

with the generator S as before. $\langle c_1, [S]\rangle = 1$ still holds.

EXERCISE 4.5. Compute $\pi_2(\mathrm{Sym}^2(\Sigma_2)$.

4.1. Totally real tori, and V_z. Inside $\mathrm{Sym}^g(\Sigma_g)$ our attaching circles induce a pair of smoothly embedded, g-dimensional tori

$$\mathbb{T}_\alpha = \alpha_1 \times ... \times \alpha_g \quad \text{and} \quad \mathbb{T}_\beta = \beta_1 \times ... \times \beta_g \ .$$

More precisely $\mathbb{T}_\alpha$ consists of those g-tuples of points $\{x_1, ..., x_g\}$ for which $x_i \in \alpha_i$ for $i = 1, ..., g$.

These tori enjoy a certain compatibility with any complex structure on $\mathrm{Sym}^g(\Sigma)$ induced from Σ:

DEFINITION 4.6. Let (Z, J) be a complex manifold, and $L \subset Z$ be a submanifold. Then L is called *totally real* if none of its tangent spaces contains a J-complex line, i.e. $T_\lambda L \cap JT_\lambda L = (0)$ for each $\lambda \in L$.

EXERCISE 4.7. Let $\mathbb{T}_\alpha \subset \mathrm{Sym}^g(\Sigma)$ be the torus induced from a set of attaching circles $\alpha_1, ..., \alpha_g$. Then $\mathbb{T}_\alpha$ is a totally real submanifold of $\mathrm{Sym}^g(\Sigma)$ (for any complex structure induced from Σ).

The basepoint z also induces a subspace that we use later:

$$V_z = \{z\} \times \mathrm{Sym}^{g-1}(\Sigma_g),$$

which has complex codimension 1. Note that since z is in the complement of the α and β curves, V_z is disjoint from $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$.

We finish the section with the following problems.

Exercise 4.8. Show that
$$\frac{H_1(\mathrm{Sym}^g(\Sigma))}{H_1(\mathbb{T}_\alpha)\oplus H_1(\mathbb{T}_\beta)} \cong \frac{H_1(\Sigma)}{[\alpha_1],...,[\alpha_g],[\beta_1],...,[\beta_g]} \cong H_1(Y;\mathbb{Z}).$$

Exercise 4.9. Compute $H_1(Y_n,\mathbb{Z})$ for the three-manifolds Y_n in Definition 2.7.

5. Disks in symmetric products

Let $\mathbb{D}$ be the unit disk in $\mathbb{C}$. Let e_1, e_2 be the arcs in the boundary of $\mathbb{D}$ with $\mathrm{Re}(z)\geq 0$, $\mathrm{Re}(z)\leq 0$ respectively.

Definition 5.1. Given a pair of intersection points $\mathbf{x},\mathbf{y}\in\mathbb{T}_\alpha\cap\mathbb{T}_\beta$, a Whitney disk connecting $\mathbf{x}$ and $\mathbf{y}$ is a continuos map
$$u:\ \mathbb{D}\longrightarrow \mathrm{Sym}^g(\Sigma_g)$$
with the properties that $u(-i)=\mathbf{x}$, $u(i)=\mathbf{y}$, $u(e_1)\subset\mathbb{T}_\alpha$, $u(e_2)\subset\mathbb{T}_\beta$. Let $\pi_2(\mathbf{x},\mathbf{y})$ denote the set of homotopy classes of maps connecting $\mathbf{x}$ and $\mathbf{y}$.

The set $\pi_2(\mathbf{x},\mathbf{y})$ is equipped with a certain multiplicative structure. Note that there is a way to splice spheres to disks:
$$\pi_2'(\mathrm{Sym}^g(\Sigma))*\pi_2(\mathbf{x},\mathbf{y})\longrightarrow\pi_2(\mathbf{x},\mathbf{y}).$$
Also, if we take a disk connecting $\mathbf{x}$ to $\mathbf{y}$, and one connecting $\mathbf{y}$ to $\mathbf{z}$, we can glue them to get a disk connecting $\mathbf{x}$ to $\mathbf{z}$. This operation gives rise to a multiplication
$$*\colon \pi_2(\mathbf{x},\mathbf{y})\times\pi_2(\mathbf{y},\mathbf{z})\longrightarrow\pi_2(\mathbf{x},\mathbf{z}).$$

5.1. An obstruction. Let $\mathbf{x},\mathbf{y}\in\mathbb{T}_\alpha\cap\mathbb{T}_\beta$ be a pair of intersection points. Choose a pair of paths $a\colon[0,1]\longrightarrow\mathbb{T}_\alpha$, $b\colon[0,1]\longrightarrow\mathbb{T}_\beta$ from $\mathbf{x}$ to $\mathbf{y}$ in $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$ respectively. The difference $a-b$ gives a loop in $\mathrm{Sym}^g(\Sigma)$.

Definition 5.2. Let $\epsilon(\mathbf{x},\mathbf{y})$ denote the image of $a-b$ in $H_1(Y,Z)$ under the map given by Exercise 4.8. Note that $\epsilon(\mathbf{x},\mathbf{y})$ is independent of the choice of the paths a and b.

It is obvious from the definition that if $\epsilon(\mathbf{x},\mathbf{y})\neq 0$ then $\pi_2(\mathbf{x},\mathbf{y})$ is empty. Note that ϵ can be calculated in Σ, using the identification between $\pi_1(\mathrm{Sym}^g(\Sigma))$ and $H_1(\Sigma)$. Specifically, writing $\mathbf{x}=\{x_1,\ldots,x_g\}$ and $\mathbf{y}=\{y_1,\ldots,y_g\}$, we can think of the path $a\colon[0,1]\longrightarrow\mathbb{T}_\alpha$ as a collection of arcs in $\alpha_1\cup\ldots\cup\alpha_g\subset\Sigma$ whose boundary is given by $\partial a=y_1+\ldots+y_g-x_1-\ldots-x_g$; similarly, the path $b\colon[0,1]\longrightarrow\mathbb{T}_\beta$ can be viewed as a collection of arcs in $\beta_1\cup\ldots\cup\beta_g\subset\Sigma$ whose boundary is given by $\partial b=y_1+\ldots+y_g-x_1-\ldots-x_g$. Thus, the difference $a-b$ is a closed one-cycle in Σ, whose image in $H_1(Y;\mathbb{Z})$ is the difference $\epsilon(\mathbf{x},\mathbf{y})$ defined above.

Clearly ϵ is additive, in the sense that
$$\epsilon(\mathbf{x},\mathbf{y})+\epsilon(\mathbf{y},\mathbf{z})=\epsilon(\mathbf{x},\mathbf{z}).$$

Definition 5.3. Partition the intersection points of $\mathbb{T}_\alpha\cap\mathbb{T}_\beta$ into equivalence classes, where $\mathbf{x}\sim\mathbf{y}$ if $\epsilon(\mathbf{x},\mathbf{y})=0$.

Exercise 5.4. Take a genus 1 Heegaard diagram of $L(p,q)$, and isotop α and β so that they have only p intersection points. Show that all the intersection points lie in different equivalence classes.

Exercise 5.5. In the genus 2 example of Figure 2 find all the intersection points in $\mathbb{T}_\alpha\cap\mathbb{T}_\beta$, (there are 18 of them), and partition the points into equivalence classes (there are 2 equivalence classes).

5.2. Domains. In order to understand topological disks in $\mathrm{Sym}^g(\Sigma_g)$ it is helpful to study their "shadow" in Σ_g.

DEFINITION 5.6. Let $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$. For any point $w \in \Sigma$ which is in the complement of the α and β curves let

$$n_w : \ \pi_2(\mathbf{x}, \mathbf{y}) \longrightarrow \mathbb{Z}$$

denote the algebraic intersection number

$$n_w(\phi) = \#\phi^{-1}(\{w\} \times \mathrm{Sym}^{g-1}(\Sigma_g)).$$

Note that since $V_w = \{w\} \times \mathrm{Sym}^{g-1}(\Sigma_g)$ is disjoint from $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$, n_w is well-defined.

DEFINITION 5.7. Let $D_1, \ldots, D_m$ denote the closures of the components of $\Sigma - \alpha_1 - \ldots - \alpha_g - \beta_1 - \ldots - \beta_g$. Given $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$ the *domain associated to* ϕ is the formal linear combination of the regions $\{D_i\}_{i=1}^m$:

$$\mathcal{D}(\phi) = \sum_{i=1}^{m} n_{z_i}(\phi) D_i,$$

where $z_i \in D_i$ are points in the interior of D_i. If all the coefficients $n_{z_i}(\phi) \geq 0$, then we write $\mathcal{D}(\phi) \geq 0$.

EXERCISE 5.8. Let $\mathbf{x}, \mathbf{y}, \mathbf{p} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$, $\phi_1 \in \pi_2(\mathbf{x}, \mathbf{y})$ and $\phi_2 \in \pi_2(\mathbf{y}, \mathbf{p})$. Show that

$$\mathcal{D}(\phi_1 * \phi_2) = \mathcal{D}(\phi_1) + \mathcal{D}(\phi_2).$$

Similarly

$$\mathcal{D}(S * \phi) = \mathcal{D}(\phi) + \sum_{i=1}^{n} D_i \ ,$$

where S denotes the positive generator of $\pi_2(\mathrm{Sym}^g(\Sigma_g))$.

The domain $\mathcal{D}(\phi)$ can be regarded as a two-chain. In the next exercise we study its boundary.

EXERCISE 5.9. Let $\mathbf{x} = (x_1, ..., x_g)$, $\mathbf{y} = (y_1, ..., y_g)$ where

$$x_i \in \alpha_i \cap \beta_i, \quad y_i \in \alpha_i \cap \beta_{\sigma^{-1}(i)}$$

and σ is a permutation. For $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$, show that

- The restriction of $\partial\mathcal{D}(\phi)$ to α_i is a one-chain with boundary $y_i - x_i$.
- The restriction of $\partial\mathcal{D}(\phi)$ to β_i is a one-chain with boundary $x_i - y_{\sigma(i)}$.

REMARK 5.10. Informally the above result says that $\partial(\mathcal{D}(\phi))$ connects x to y on the α curves and y to x on the β curves.

EXERCISE 5.11. Take the genus 2 examples is of Figure 4. Find disks ϕ_1 and ϕ_2 with $\mathcal{D}(\phi_1) = D_1$ and $\mathcal{D}(\phi_2) = D_2$.

DEFINITION 5.12. Let $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$. If a formal sum

$$\mathcal{A} = \sum_{i=1}^{n} a_i D_i$$

satisfies that $\partial\mathcal{A}$ connects $\mathbf{x}$ to $\mathbf{y}$ along the α curves and connects $\mathbf{y}$ to $\mathbf{x}$ along the β curves, we will say that $\partial\mathcal{A}$ connects $\mathbf{x}$ to $\mathbf{y}$.

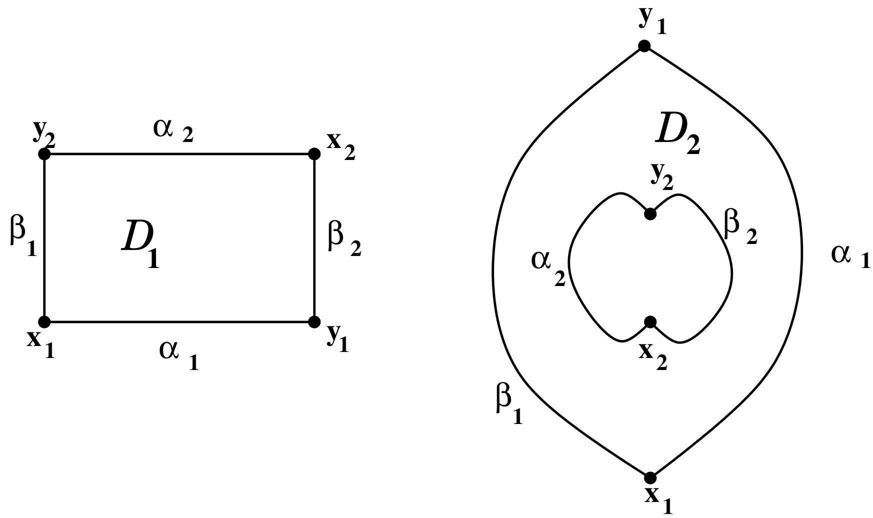


FIGURE 4. Domains of disks in $\mathrm{Sym}^2(\Sigma)$.

When $g > 1$ the argument in Exercise 5.9 can be reversed:

PROPOSITION 5.13. *Suppose that* $g > 1$, $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$. *If* $\mathcal{A}$ *connects* $\mathbf{x}$ *to* $\mathbf{y}$ *then there is a homotopy class* $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$ *with*

$$\mathcal{D}(\phi) = \mathcal{A}$$

Furthermore if $g > 2$ *then* ϕ *is uniquely determined by* $\mathcal{A}$.

As an easy corollary we have the following

PROPOSITION 5.14. [**31**] *Suppose* $g > 2$. *For each* $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$, *if* $\epsilon(\mathbf{x}, \mathbf{y}) \neq 0$, *then* $\pi_2(\mathbf{x}, \mathbf{y})$ *is empty; otherwise,*

$$\pi_2(\mathbf{x}, \mathbf{y}) \cong \mathbb{Z} \oplus H^1(Y, \mathbb{Z}).$$

REMARK 5.15. When $g = 2$ we can define $\pi_2'(\mathbf{x}, \mathbf{y})$ by modding out $\pi_2(\mathbf{x}, \mathbf{y})$ with the relation: ϕ_1 is equivalent to ϕ_2 if $\mathcal{D}(\phi_1) = \mathcal{D}(\phi_2)$. For $\epsilon(\mathbf{x}, \mathbf{y}) = 0$ we have

$$\pi_2'(\mathbf{x}, \mathbf{y}) \cong \mathbb{Z} \oplus H^1(Y, \mathbb{Z}).$$

Note that working with π_2' is the same as working with *homology classes* of disks, and for simplifying notation this is the approach used in [**23**].

6. Spin^c-structures

In order to refine the discussion about the equivalence classes encountered in the previous section we will need the notion of Spin^c structures. These structures can be defined in every dimension. For three-dimensional manifolds it is convenient to use a reformulation of Turaev [**40**].

Let Y be an oriented closed 3-manifold. Since Y has trivial Euler characteristic, it admits nowhere vanishing vector fields.

DEFINITION 6.1. Let v_1 and v_2 be two nowhere vanishing vector fields. We say that v_1 is *homologous* to v_2 if there is a ball B in Y with the property that $v_1|_{Y-B}$

is homotopic to $v_2|_{Y-B}$. This gives an equivalence relation, and we define the space of Spin^c structures over Y as nowhere vanishing vector fields modulo this relation.

We will denote the space of Spin^c structures over Y by $\mathrm{Spin}^c(Y)$.

6.1. Action of $H^2(Y,\mathbb{Z})$ on $\mathrm{Spin}^c(Y)$. Fix a trivialization τ of the tangent bundle TY. This gives a one-to-one correspondence between vector fields v over Y and maps f_v from Y to S^2.

DEFINITION 6.2. Let μ denote the positive generator of $H^2(S^2,\mathbb{Z})$. Define

$$\delta^\tau(v) = f_v^*(\mu) \in H^2(Y,\mathbb{Z})$$

EXERCISE 6.3. Show that δ^τ induces a one-to-one correspondence between $\mathrm{Spin}^c(Y)$ and $H^2(Y,\mathbb{Z})$.

The map δ^τ is independent of the the trivialization if $H_1(Y,\mathbb{Z})$ has no two-torsion. In the general case we have a weaker property:

EXERCISE 6.4. Show that if v_1 and v_2 are a pair of nowhere vanishing vector fields over Y, then the difference

$$\delta(v_1,v_2) = \delta^\tau(v_1) - \delta^\tau(v_2) \in H^2(Y,\mathbb{Z})$$

is independent of the trivialization τ, and

$$\delta(v_1,v_2) + \delta(v_2,v_3) = \delta(v_1,v_3).$$

This gives an action of $H^2(Y,\mathbb{Z})$ on $\mathrm{Spin}^c(Y)$. If $a \in H^2(Y,\mathbb{Z})$ and $v \in \mathrm{Spin}^c(Y)$ we define $a+v \in \mathrm{Spin}^c(Y)$ by the property that $\delta(a+v,v) = a$. Similarly for $v_1, v_2 \in \mathrm{Spin}^c(Y)$, we let $v_1 - v_2$ denote $\delta(v_1,v_2)$.

There is a natural involution on the space of Spin^c structures which carries the homology class of the vector field v to the homology class of $-v$. We denote this involution by the map $\mathfrak{s} \mapsto \bar{\mathfrak{s}}$.

There is also a natural map

$$c_1 \colon \mathrm{Spin}^c(Y) \longrightarrow H^2(Y,\mathbb{Z}),$$

the first Chern class. This is defined by $c_1(\mathfrak{s}) = \mathfrak{s} - \bar{\mathfrak{s}}$. It is clear that $c_1(\bar{\mathfrak{s}}) = -c_1(\mathfrak{s})$.

6.2. Intersection points and Spin^c structures. Now we are ready to define a map

$$s_z : \ \mathbb{T}_\alpha \cap \mathbb{T}_\beta \longrightarrow \mathrm{Spin}^c(Y),$$

which will be a refinement of the equivalence classes given by $\epsilon(\mathbf{x},\mathbf{y})$.

Let f be a Morse function on Y compatible with the attaching circles $\alpha_1,...,\alpha_g$, $\beta_1,...,\beta_g$. Then each $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ determines a g-tuple of trajectories for ∇f connecting the index one critical points to index two critical points. Similarly z gives a trajectory connecting the index zero critical point with the index three critical point. Deleting tubular neighborhoods of these $g+1$ trajectories, we obtain the complement of disjoint union of balls in Y where the gradient vector field ∇f does not vanish. Since each trajectory connects critical points of different parities, the gradient vector field has index 0 on all the boundary spheres, so it can be extended as a nowhere vanishing vector field over Y. According to our definition of Spin^c-structures the homology class of the nowhere vanishing vector field obtained in this manner gives a Spin^c structure. Let us denote this element by $s_z(\mathbf{x})$. The following is proved in [**31**].

LEMMA 6.5. *Let $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$. Then we have*

$$s_z(\mathbf{y}) - s_z(\mathbf{x}) = \mathrm{PD}[\epsilon(\mathbf{x}, \mathbf{y})]. \tag{1}$$

In particular $s_z(\mathbf{x}) = s_z(\mathbf{y})$ if and only if $\pi_2(\mathbf{x}, \mathbf{y})$ is non-empty.

EXERCISE 6.6. Let $(\Sigma_1, \alpha, \beta)$ be a genus 1 Heegaard diagram of $L(p, 1)$ so that α and β have p intersection points. Using this diagram $\Sigma_1 - \alpha - \beta$ has p components. Choose a point z_i in each region. Show that for any $x \in \alpha \cap \beta$, we have

$$s_{z_i}(x) \neq s_{z_j}(x)$$

for $i \neq j$.

7. Holomorphic disks

A complex structure on Σ induces a complex structure on $\mathrm{Sym}^g(\Sigma_g)$. For a given homotopy class $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$ let $\mathcal{M}(\phi)$ denote the moduli space of holomorphic representatives of ϕ. Note that in order to guarantee that $\mathcal{M}(\phi)$ is smooth, in Lagrangian Floer homology one has to use appropriate perturbations, see [**8**], [**9**], [**11**].

The moduli space $\mathcal{M}(\phi)$ admits an $\mathbb{R}$ action. This corresponds to the group of complex automorphisms of the unit disk that preserve i and $-i$. It is easy to see that this group is isomorphic to $\mathbb{R}$. For example using the Riemann mapping theorem change the unit disk to the infinite strip $[0, 1] \times i\mathbb{R} \subset \mathbb{C}$, where e_1 corresponds to $1 \times i\mathbb{R}$ and e_2 corresponds to $0 \times i\mathbb{R}$. Then the automorphisms preserving e_1 and e_2 correspond to the vertical translations. Now if $u \in \mathcal{M}(\phi)$ then we could precompose u with any of these automorphisms and get another holomorphic disk. Since in the definition of the boundary map we would like to count holomorphic disks we will divide $\mathcal{M}(\phi)$ by the above $\mathbb{R}$ action, and define the unparametrized moduli space

$$\widehat{\mathcal{M}}(\phi) = \frac{\mathcal{M}(\phi)}{\mathbb{R}}.$$

It is easy to see that the $\mathbb{R}$ action is free except in the case when ϕ is the homotopy class of the constant map ($\phi \in \pi_2(\mathbf{x}, \mathbf{x})$, with $\mathcal{D}(\phi) = 0$). In this case $\mathcal{M}(\phi)$ is a single point corresponding to the constant map.

The moduli space $\mathcal{M}(\phi)$ has an expected dimension called the Maslov index $\mu(\phi)$, see [**35**], which corresponds to the index of an elliptic operator. The Maslov index has the following significance: If we vary the almost complex structure of $\mathrm{Sym}^g(\Sigma_g)$ in an n-dimensional family, the corresponding parametrized moduli space has dimension $n + \mu(\phi)$ around solutions that are smoothly cut out by the defining equation. The Maslov index is additive:

$$\mu(\phi_1 * \phi_2) = \mu(\phi_1) + \mu(\phi_2)$$

and for the homotopy class of the constant map μ is equal to zero.

LEMMA 7.1. *([**31**]) Let $S \in \pi_2'(\mathrm{Sym}^g(\Sigma))$ be the positive generator. Then for any $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$, we have that*

$$\mu(\phi + k[S]) = \mu(\phi) + 2k.$$

PROOF. It follows from the excision principle for the index that attaching a topological sphere Z to a disk changes the Maslov index by $2\langle c_1, [Z]\rangle$ (see [**18**]). On the other hand for the positive generator S we have $\langle c_1, [S]\rangle = 1$. $\square$

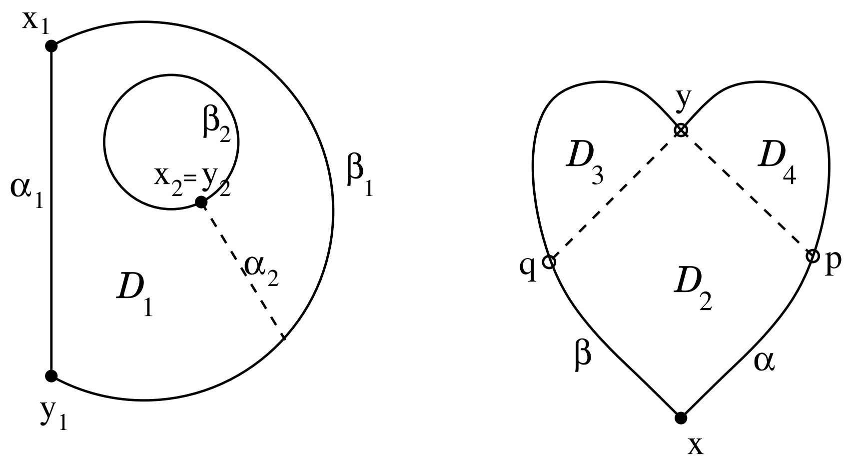


FIGURE 5.

COROLLARY 7.2. *If $g = 2$ and $\phi, \phi' \in \pi_2(\mathbf{x}, \mathbf{y})$ satisfies*

$$\mathcal{D}(\phi) = \mathcal{D}(\phi')$$

then $\mu(\phi) = \mu(\phi')$. In particular μ is well-defined on $\pi_2'(\mathbf{x}, \mathbf{y})$.

LEMMA 7.3. *If $\mathcal{M}(\phi)$ is non-empty, then $\mathcal{D}(\phi) \geq 0$.*

PROOF. Let us choose a reference point z_i in each region $\mathcal{D}_i$. Since V_{z_i} is a subvariety, a holomorphic disk is either contained in it (which is excluded by the boundary conditions) or it must meet it non-negatively. □

By studying energy bounds, orientations and Gromov limits we prove in [**31**]

THEOREM 7.4. *There is a family of (admissible) perturbations with the property that if $\mu(\phi) = 1$ then $\widehat{\mathcal{M}}(\phi)$ is a compact oriented zero dimensional manifold. When $g = 2$, the same result holds for $\phi \in \pi_2'(\mathbf{x}, \mathbf{y})$ as well.*

7.1. Examples. The space of holomorphic disks connecting $\mathbf{x}, \mathbf{y}$ can be given an alternate description, using only maps between one-dimensional complex manifolds.

LEMMA 7.5. *([**31**]) Given any holomorphic disk $u \in \mathcal{M}(\phi)$, there is a g-fold branched covering space $p \colon \widehat{\mathbb{D}} \longrightarrow \mathbb{D}$ and a holomorphic map $\widehat{u} \colon \widehat{\mathbb{D}} \longrightarrow \Sigma$, with the property that for each $z \in \mathbb{D}$, $u(z)$ is the image under $\widehat{u}$ of the pre-image $p^{-1}(z)$.*

EXERCISE 7.6. Let ϕ_1, ϕ_2 be homotopy classes in Figure 4, with $\mathcal{D}(\phi_1) = D_1$, $\mathcal{D}(\phi_2) = D_2$. Also let $\phi_0 \in \pi_2(\mathbf{y}, \mathbf{x})$ be a class with $\mathcal{D}(\phi_0) = -D_1$. Show that $\mu(\phi_1) = 1$, $\mu(\phi_2) = 0$ and $\mu(\phi_0) = -1$.

For additional examples see Figure 5. The left example is in the second symmetric product and $x_2 = y_2$. The right example is in the first symmetric product, the α and β curves intersect each other in 4 points. Let ϕ_3, ϕ_4 be classes with $\mathcal{D}(\phi_3) = D_1$, $\mathcal{D}(\phi_4) = D_2 + D_3 + D_4$.

EXERCISE 7.7. Show that $\mu(\phi_3) = 1$ and $\mu(\phi_4) = 2$.

Exercise 7.8. Use the Riemman mapping theorem to show that $\widehat{\mathcal{M}}(\phi_4)$ is homeomorphic to an open interval $\mathcal{I}$.

Exercise 7.9. Study the limit of $u_i \in \mathcal{I}$ as u_i approaches one of the ends in $\mathcal{I}$. Show that the limit corresponds to a decomposition

$$\phi_4 = \phi_5 * \phi_6, \text{ or } \phi_4 = \phi_7 * \phi_8,$$

where $\mathcal{D}(\phi_5) = D_2 + D_4$, $\mathcal{D}(\phi_6) = D_3$, $\mathcal{D}(\phi_7) = D_2 + D_3$ and $\mathcal{D}(\phi_8) = D_4$.

8. The Floer chain complexes

In this section we will define the various chain complexes corresponding to $\widehat{HF}$, HF^+, HF^- and HF^∞.

We start with the case when Y is a rational homology 3-sphere. Let $(\Sigma, \alpha_1, ..., \alpha_g, \beta_1, ..., \beta_g, z)$ be a pointed Heegaard diagram of genus $g > 0$ for Y. Choose a Spin^c structure $t \in \mathrm{Spin}^c(Y)$.

Let $\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, t)$ denote the free Abelian group generated by the points in $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ with $s_z(\mathbf{x}) = t$. This group can be endowed with a relative grading

$$\mathrm{gr}(\mathbf{x}, \mathbf{y}) = \mu(\phi) - 2n_z(\phi), \tag{2}$$

where ϕ is any element $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$, and μ is the Maslov index.

In view of Proposition 5.14 and Lemma 7.1, this integer is independent of the choice of homotopy class $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$.

Definition 8.1. Choose a perturbation as in Theorem 7.4. For $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ and $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$ let us define $c(\phi)$ to be the signed number of points in $\widehat{\mathcal{M}}(\phi)$ if $\mu(\phi) = 1$. If $\mu(\phi) \neq 1$ let $c(\phi) = 0$.

Let

$$\partial \colon \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, t) \longrightarrow \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, t)$$

be the map defined by

$$\partial \mathbf{x} = \sum_{\{\mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta,\ \phi \in \pi_2(\mathbf{x}, \mathbf{y}) \big| s_z(\mathbf{y}) = t,\ n_z(\phi) = 0\}} c(\phi) \cdot \mathbf{y}$$

By analyzing the Gromov compactification of $\widehat{\mathcal{M}}(\phi)$ for $n_z(\phi) = 0$ and $\mu(\phi) = 2$ it is proved in [**31**] that $(\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, t), \partial)$ is a chain complex; i.e. $\partial^2 = 0$.

Definition 8.2. The Floer homology groups $\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, t)$ are the homology groups of the complex $(\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, t), \partial)$.

One of the main results of [**31**] is that the homology group $\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, t)$ is independent of the Heegaard diagram, the basepoint and the other choices in the definition (complex structures, perturbations). After analyzing the effect of isotopies, handle-slides and stabilizations, it is proved in [**31**] that under pointed isotopies, pointed handle-slides, and stabilizations we get chain homotopy equivalent complexes $\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, t)$. This together with Theorem 2.9, and Proposition 2.10 implies:

Theorem 8.3. *([**31**]) Let $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z)$ and $(\Sigma', \boldsymbol{\alpha}', \boldsymbol{\beta}', z')$ be pointed Heegaard diagrams of Y, and $t \in \mathrm{Spin}^c(Y)$. Then the homology groups $\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, t)$ and $\widehat{HF}(\boldsymbol{\alpha}', \boldsymbol{\beta}', t)$ are isomorphic.*

Using the above theorem we can at last define $\widehat{HF}$:

$$\widehat{HF}(Y,t) = \widehat{HF}(\boldsymbol{\alpha},\boldsymbol{\beta},t).$$

8.1. $CF^\infty(Y)$**.** The definition in the previous section uses the basepoint z in a special way: in the boundary map we only count holomorphic disks that are disjoint from the subvariety V_z.

Now we study a chain complex where all the holomorphic disks are used (but we still record the intersection number with V_z):

Let $CF^\infty(\boldsymbol{\alpha},\boldsymbol{\beta},t)$ be the free Abelian group generated by pairs $[\mathbf{x},i]$ where $s_z(\mathbf{x}) = t$, and $i \in \mathbb{Z}$ is an integer. We give the generators a relative grading defined by

$$\mathrm{gr}([\mathbf{x},i],[\mathbf{y},j]) = \mathrm{gr}(\mathbf{x},\mathbf{y}) + 2i - 2j.$$

Let

$$\partial \colon CF^\infty(\boldsymbol{\alpha},\boldsymbol{\beta},t) \longrightarrow CF^\infty(\boldsymbol{\alpha},\boldsymbol{\beta},t)$$

be the map defined by

$$\partial[\mathbf{x},i] = \sum_{\mathbf{y}\in\mathbb{T}_\alpha\cap\mathbb{T}_\beta} \sum_{\phi\in\pi_2(\mathbf{x},\mathbf{y})} c(\phi)\cdot[\mathbf{y}, i - n_z(\phi)]. \tag{3}$$

There is an isomorphism U on $CF^\infty(\boldsymbol{\alpha},\boldsymbol{\beta},t)$ given by

$$U([\mathbf{x},i]) = [\mathbf{x}, i-1]$$

that decreases the grading by 2.

It is proved in [**30**] that for rational homology three-spheres, $HF^\infty(Y,t)$ is always isomorphic to $\mathbb{Z}[U,U^{-1}]$. So clearly this is not an interesting invariant. Luckily the base-point z together with Lemma 7.3 induces a filtration on $CF^\infty(\boldsymbol{\alpha},\boldsymbol{\beta},t)$ and that produces more subtle invariants.

8.2. $CF^+(\boldsymbol{\alpha},\boldsymbol{\beta},t)$ **and** $CF^-(\boldsymbol{\alpha},\boldsymbol{\beta})$**.** Let $CF^-(\boldsymbol{\alpha},\boldsymbol{\beta},t)$ denote the subgroup of $CF^\infty(\boldsymbol{\alpha},\boldsymbol{\beta},t)$ which is freely generated by pairs $[\mathbf{x},i]$, where $i<0$. Let $CF^+(\boldsymbol{\alpha},\boldsymbol{\beta},t)$ denote the quotient group

$$CF^\infty(\boldsymbol{\alpha},\boldsymbol{\beta},t)/CF^-(\boldsymbol{\alpha},\boldsymbol{\beta},t)$$

Lemma 8.4. *The group $CF^-(\boldsymbol{\alpha},\boldsymbol{\beta},t)$ is a subcomplex of $CF^\infty(\boldsymbol{\alpha},\boldsymbol{\beta},t)$, so we have a short exact sequence of chain complexes:*

$$0 \longrightarrow CF^-(\boldsymbol{\alpha},\boldsymbol{\beta},t) \xrightarrow{\ \iota\ } CF^\infty(\boldsymbol{\alpha},\boldsymbol{\beta},t) \xrightarrow{\ \pi\ } CF^+(\boldsymbol{\alpha},\boldsymbol{\beta},t) \longrightarrow 0.$$

Proof. If $[\mathbf{y},j]$ appears in $\partial([\mathbf{x},i])$ then there is a homotopy class $\phi(\mathbf{x},\mathbf{y})$ with $\mathcal{M}(\phi)$ non-empty, and $n_z(\phi) = i-j$. According to Lemma 7.3 we have $\mathcal{D}(\phi)\geq 0$ and in particular $i \geq j$. $\square$

Clearly, U restricts to an endomorphism of $CF^-(\boldsymbol{\alpha},\boldsymbol{\beta},t)$ (which lowers degree by 2), and hence it also induces an endomorphism on the quotient $CF^+(\boldsymbol{\alpha},\boldsymbol{\beta},t)$.

Exercise 8.5. There is a short exact sequence

$$0 \longrightarrow \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta},t) \xrightarrow{\ \iota\ } CF^+(\boldsymbol{\alpha},\boldsymbol{\beta},t) \xrightarrow{\ U\ } CF^+(\boldsymbol{\alpha},\boldsymbol{\beta},t) \longrightarrow 0,$$

where $\iota(\mathbf{x}) = [\mathbf{x},0]$.

Definition 8.6. The Floer homology groups $HF^+(\boldsymbol{\alpha}, \boldsymbol{\beta}, t)$ and $HF^-(\boldsymbol{\alpha}, \boldsymbol{\beta}, t)$ are the homology groups of $(CF^+(\boldsymbol{\alpha}, \boldsymbol{\beta}, t), \partial)$ and $(CF^-(\boldsymbol{\alpha}, \boldsymbol{\beta}, t), \partial)$ respectively.

It is proved in [**31**] that the chain homotopy equivalences under pointed isotopies, handle-slides and stabilizations for $\widehat{CF}$ can be lifted to filtered chain homotopy equivalences on CF^∞ and in particular the corresponding Floer homologies are unchanged. This allows us to define

$$HF^\pm(Y, t) = HF^\pm(\boldsymbol{\alpha}, \boldsymbol{\beta}, t).$$

8.3. Three manifolds with $b_1(Y) > 0$. When $b_1(Y)$ is positive, then there is a technical problem due to the fact that $\pi_2(\mathbf{x}, \mathbf{y})$ is larger. In the definition of the boundary map we now have infinitely many homotopy classes with Maslov index 1. In order to get a finite sum we have to prove that only finitely many of these homotopy classes support holomorphic disks. This is achieved through the use of special Heegaard diagrams together with the positivity property of Lemma 7.3, see [**31**]. With this said, the constructions from the previous subsections apply and give the Heegaard Floer homology groups. The only difference is that when the image of $c_1(t)$ in $H^2(Y, \mathbb{Q})$ is non-zero, the Floer homologies no longer have relative $\mathbb{Z}$ grading.

9. A few examples

We study Heegaard Floer homology for a few examples. To simplify things we deal with homology three-spheres. Here $H_1(Y, \mathbb{Z}) = 0$ so there is a unique Spin^c-structure. In [**25**] we show how to use maps on $HF^\pm$ induced by smooth cobordisms to lift the relative grading to an absolute grading.

For $Y = S^3$ we can use a genus 1 Heegaard diagram. Here α and β intersect each other in a unique point $\mathbf{x}$. It follows that CF^+ is generated by $[\mathbf{x}, i]$ with $i \geq 0$. Since $\mathrm{gr}[\mathbf{x}, i] - \mathrm{gr}[\mathbf{x}, i-1] = 2$, the boundary map is trivial so $HF^+(S^3)$ is isomorphic to $\mathbb{Z}[U, U^{-1}]/\mathbb{Z}[U]$ as a $\mathbb{Z}[U]$ module. The absolute grading is determined by

$$\mathrm{gr}([\mathbf{x}, 0]) = 0.$$

A large class of homology three-spheres is provided by Brieskorn spheres: Recall that if p, q, and r are pairwise relatively prime integers, then the Brieskorn variety $V(p, q, r)$ is the locus

$$V(p, q, r) = \{(x, y, z) \in \mathbb{C}^3 \big| x^p + y^q + z^r = 0\}$$

Definition 9.1. The Brieskorn sphere $\Sigma(p, q, r)$ is the homology sphere obtained by $V(p, q, r) \cap S^5$ (where $S^5 \subset \mathbb{C}^3$ is the standard 5-sphere).

The simplest example is the Poincare sphere $\Sigma(2, 3, 5)$.

Exercise 9.2. Show that the diagram in Definition 2.7 with $n = 3$ is a Heegaard diagram for $\Sigma(2, 3, 5)$.

Unfortunately, in this Heegaard diagram there are lots of generators (21) and computing the Floer chain complex directly is not an easy task. Instead of this direct approach one can establish exact sequences between the Heegaard Floer homology groups of 3-manifolds modified by surgeries along knots. In [**25**] we use these surgery exact sequences to prove

PROPOSITION 9.3.

$$HF_k^+(\Sigma(2,3,5)) = \begin{cases} \mathbb{Z} & \textit{if } k \textit{ is even and } k \geq 2 \\ 0 & \textit{otherwise} \end{cases}$$

Moreover,

$$U\colon HF_{k+2}^+(\Sigma(2,3,5)) \longrightarrow HF_k^+(\Sigma(2,3,5))$$

is an isomorphism for $k \geq 2$.

This means that as a relatively graded $Z[U]$ module $HF^+((\Sigma(2,3,5))$ is isomorphic to $HF^+(S^3)$, but the absolute grading still distinguishes them.

Another example is provided by $\Sigma(2,3,7)$. (Note that this three manifold corresponds to the $n=5$ diagram when we switch the role of the α and β circles.)

PROPOSITION 9.4.

$$HF_k^+(\Sigma(2,3,7)) = \begin{cases} \mathbb{Z} & \textit{if } k \textit{ is even and } k \geq 0 \\ \mathbb{Z} & \textit{if } k = -1 \\ 0 & \textit{otherwise} \end{cases} \tag{4}$$

For a description of $HF^+(\Sigma(p,q,r))$ see [**27**], and also [**22**], [**36**].

10. Knot Floer homology

In this section we study a version of Heegaard Floer homology that can be applied to knots in three-manifolds. Here we will restrict our attention to knots in S^3. For a more general discussion see [**29**] and [**34**].

Let us consider a Heegaard diagram $(\Sigma_g, \alpha_1, ..., \alpha_g, \beta_1, ..., \beta_g)$ for S^3 equipped with two basepoints w and z. This data gives rise to a knot in S^3 by the following procedure. Connect w and z by a curve a in $\Sigma_g - \alpha_1 - ... - \alpha_g$ and also by another curve b in $\Sigma_g - \beta_1 - ... - \beta_g$. By pushing a and b into U_0 and U_1 respectively, we obtain a knot $K \subset S^3$. We call the data $(\Sigma_g, \boldsymbol{\alpha}, \boldsymbol{\beta}, w, z)$ a two-pointed Heegaard diagram compatible with the knot K.

A Morse theoretic interpretation can be given as follows. Fix a metric on Y and a self-indexing Morse function so that the induced Heegaard diagram is $(\Sigma_g, \alpha_1, ..., \alpha_g, \beta_1, ..., \beta_g)$. Then the basepoints w, z give two trajectories connecting the index 0 and index 3 critical points. Joining these arcs together gives the knot K.

LEMMA 10.1. *Every knot can be represented by a two-pointed Heegaard diagram.*

PROOF. Fix a height function h on K so that it has only two critical points, m and m' with $h(m) = 0$ and $h(m') = 3$. Now extend h to a self-indexing Morse function from $K \subset Y$ to Y so that the index 1 and 2 critical points are disjoint from K, and let z and w be the two intersection points of K with the Heegaard surface $\tilde{h}^{-1}(3/2)$. $\square$

A straightforward generalization of $\widehat{CF}$ is the following.

DEFINITION 10.2. Let K be a knot in S^3 and $(\Sigma_g, \alpha_1, ..., \alpha_g, \beta_1, ..., \beta_g, z, w)$ be a compatible two-pointed Heegaard diagram. Let $C(K)$ be the free abelian group generated by the intersection points $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$. For a generic choice of almost-complex structures let $\partial_K : C(K) \longrightarrow C(K)$ be given by

$$\partial_K(\mathbf{x}) = \sum_{\mathbf{y}} \sum_{\{\phi \in \pi_2(\mathbf{x},\mathbf{y}) | \mu(\phi)=1,\ n_z(\phi)=n_w(\phi)=0\}} c(\phi) \cdot \mathbf{y} \tag{5}$$

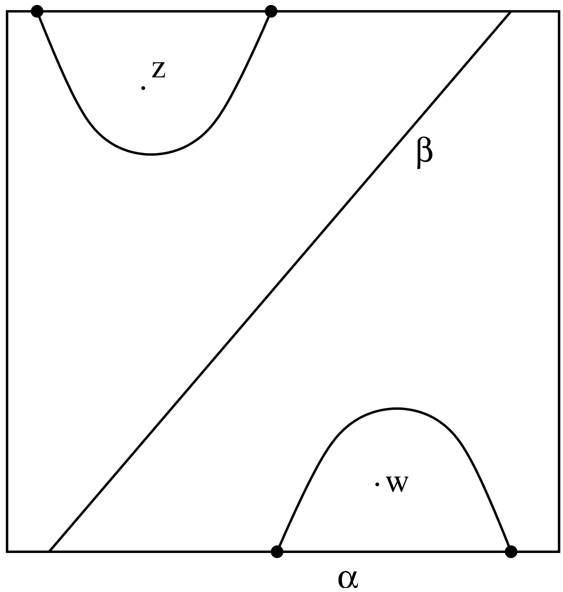


FIGURE 6.

PROPOSITION 10.3. *([**29**], [**34**]) $(C(K), \partial_K)$ is a chain complex. Its homology $H(K)$ is independent of the choice of two-pointed Heegaard diagrams representing K and the almost-complex structures.*

10.1. Examples. For the unknot U we can use the standard genus 1 Heegaard diagram of S^3, and get $H(U) = \mathbb{Z}$.

EXERCISE 10.4. Take the two-pointed Heegaard diagram in Figure 6. Show that the corresponding knot is the trefoil $T_{2,3}$.

EXERCISE 10.5. Find all the holomorphic disks in Figure 6, and show that $H(T_{2,3})$ has rank 3.

10.2. A bigrading on $C(K)$. For $C(K)$ we define two gradings. These correspond to functions:

$$F,\ G : \mathbb{T}_\alpha \cap \mathbb{T}_\beta \longrightarrow \mathbb{Z}.$$

We start with F:

DEFINITION 10.6. For $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ let

$$f(\mathbf{x}, \mathbf{y}) = n_z(\phi) - n_w(\phi),$$

where $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$.

EXERCISE 10.7. Show that for $\mathbf{x}, \mathbf{y}, \mathbf{p} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ we have

$$f(\mathbf{x}, \mathbf{y}) + f(\mathbf{y}, \mathbf{p}) = f(\mathbf{x}, \mathbf{p}).$$

EXERCISE 10.8. Show that f can be lifted uniquely to a function $F \colon \mathbb{T}_\alpha \cap \mathbb{T}_\beta \longrightarrow \mathbb{Z}$ satisfying the relation

$$F(\mathbf{x}) - F(\mathbf{y}) = f(\mathbf{x}, \mathbf{y}), \tag{6}$$

and the additional symmetry

$$\#\{\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta \big| F(\mathbf{x}) = i\} \equiv \#\{\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta \big| F(\mathbf{x}) = -i\} \pmod 2$$

for all $i \in \mathbb{Z}$.

The other grading comes from the Maslov grading.

DEFINITION 10.9. For $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ let

$$g(\mathbf{x}, \mathbf{y}) = \mu(\phi) - 2n_w(\phi),$$

where $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$.

In order to lift g to an absolute grading we use the one-pointed Heegaard diagram $(\Sigma_g, \boldsymbol{\alpha}, \boldsymbol{\beta}, w)$. This is a Heegaard diagram of S^3. It follows that the homology of $\widehat{CF}(\mathbb{T}_\alpha, \mathbb{T}_\beta, w)$ is isomorphic to $\mathbb{Z}$. Using the normalization that this homology is supported in grading zero we get a function

$$G \colon \mathbb{T}_\alpha \cap \mathbb{T}_\beta \longrightarrow \mathbb{Z}$$

that associates to each intersection points its absolute grading in $\widehat{CF}(\mathbb{T}_\alpha, \mathbb{T}_\beta, w)$. It also follows that $G(\mathbf{x}) - G(\mathbf{y}) = g(\mathbf{x}, \mathbf{y})$.

DEFINITION 10.10. Let $C_{i,j}$ denote the free Abelian group generated by those intersection points $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ that satisfy

$$i = F(\mathbf{x}), \quad j = G(\mathbf{x}).$$

The following is straightforward:

LEMMA 10.11. *For a two-pointed Heegaard diagram corresponding to a knot K in S^3 decompose $C(K)$ as*

$$C(K) = \bigoplus_{i,j} C_{i,j}.$$

Then $\partial_K(C_{i,j})$ is contained in $C_{i,j-1}$.

As a corollary we can decompose $H(K)$:

$$H(K) = \bigoplus_{i,j} H_{i,j}(K).$$

Since the chain homotopy equivalences of $C(K)$ induced by (two-pointed) Heegaard moves respect both gradings it follows that $H_{i,j}(K)$ is also a knot invariant.

11. Kauffman states

When studying knot Floer homology it is natural to consider a special diagram where the intersection points correspond to Kauffman states.

Let K be a knot in S^3. Fix a projection for K. Let $v_1, ..., v_n$ denote the double points in the projection. If we forget the pattern of over and under crossings in the diagram we get an immersed circle C in the plane.

Fix an edge e which appears in the closure of the unbounded region A in the planar projection. Let B be the region on the other side of the marked edge.

DEFINITION 11.1. ([**14**]) A Kauffman state (for the projection and the distinguished edge e) is a map that associates for each double point v_i one of the four corners in such a way that each component in $S^2 - C - A - B$ gets exactly one corner.

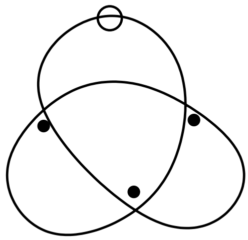


FIGURE 7.

FIGURE 8. The definition of $a(c_i)$ for both kinds of crossings.

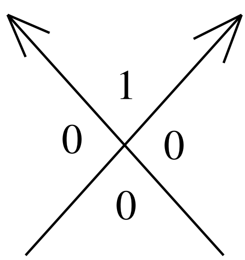


FIGURE 9. The definition of $B(c_i)$.

Let us write a Kauffman state as $(c_1, ..., c_n)$, where c_i is a corner for v_i.

For an example see Figure 7 that shows the Kauffman states for the trefoil. In that picture the black dots denote the corners, and the white circle indicates the marking.

EXERCISE 11.2. Find the Kauffman states for the $T_{2,2n+1}$ torus knots (using a projection with $2n + 1$ double points).

11.1. Kauffman states and Alexander polynomial. The Kauffman states could be used to compute the Alexander polynomial for the knot K. Fix an orientation for K. Then for each corner c_i we define $a(c_i)$ by the formula in Figure 8, and $B(c_i)$ by the formula in Figure 9.

THEOREM 11.3. *(*[**14**]*) Let K be a knot in S^3, and fix an oriented projection of K with a marked edge. Let $\mathcal{K}$ denote the set of Kauffman states for the projection.*

Then the polynomial

$$\sum_{c\in\mathcal{K}}\prod_{i=1}^{n}(-1)^{B(c_i)}T^{a(c_i)}$$

is equal to the symmetrized Alexander polynomial $\Delta_K(T)$ of K.

12. Kauffman states and Heegaard diagrams

PROPOSITION 12.1. *Let K be a knot in S^3. Fix a knot projection for K together with a marked edge. Then there is a Heegaard diagram for K, where the generators are in one-to-one correspondence with the Kauffman states of the projection.*

PROOF. Let C be the immersed circle as before. A regular neighborhood $nd(C)$ is a handlebody of genus $n+1$. Clearly $S^3-nd(C)$ is also a handlebody, so we get a Heegaard decomposition of S^3. Let Σ be the oriented boundary of $S^3-nd(C)$. This will be the Heegaard surface. The complement of C in the plane has $n+2$ components. For each region, except for A, we associate an α curve, which is the intersection of the region with Σ. It is easy to see $\Sigma-\alpha_1-...-\alpha_{n+1}$ is connected and all α_i bound disjoint disks in $S^3-nd(C)$.

Fix a point in the edge e and let β_{n+1} be the meridian for K around this point. The curves $\beta_1,...,\beta_n$ correspond to the double points $v_1,...,v_n$, see Figure 10. As for the basepoints, choose w and z on the two sides of β_{n+1}. There is a small arc connecting z and w. This arc is in the complement of the α curves. We can also choose a long arc from w to z in the complement of the β curves that travels along the knot K. It follows that this two-pointed Heegaard diagram is compatible with K.

In order to see the relation between $\mathbb{T}_\alpha\cap\mathbb{T}_\beta$ and Kauffman states note that in a neighborhood of each v_i, there are at most four intersection points of β_i with circles corresponding to the four regions which contain v_i, see Figure 10. Clearly these intersection points are in one-to-one correspondence with the corners. This property together with the observation that β_{n+1} intersects only the α curve of region B finishes the proof. □

13. A combinatorial formula

In this section we describe $F(\mathbf{x})$ and $G(\mathbf{x})$ in terms of the knot projection. Both of these functions will be given as a state sum over the corners of the corresponding Kauffman state. For a given corner c_i we use $a(c_i)$ and $b(c_i)$, where $a(c_i)$ is given as before, see Figure 8, and $b(c_i)$ is defined in Figure 11. Note that $b(c_i)$ and $B(c_i)$ are congruent modulo 2. The following result is proved in [**26**].

THEOREM 13.1. *Fix an oriented knot projection for K together with a distinguished edge. Let us fix a two-pointed Heegaard diagram for K as above. For $\mathbf{x}\in\mathbb{T}_\alpha\cap\mathbb{T}_\beta$ let $(c_1,...,c_n)$ be the corresponding Kauffman state. Then we have*

$$F(\mathbf{x})=\sum_{i=1}^{n}a(c_i)\qquad G(\mathbf{x})=\sum_{i=1}^{n}b(c_i).$$

EXERCISE 13.2. Compute $H_{i,j}$ for the trefoil, see Figure 7, and more generally for the $T_{2,2n+1}$ torus knots.

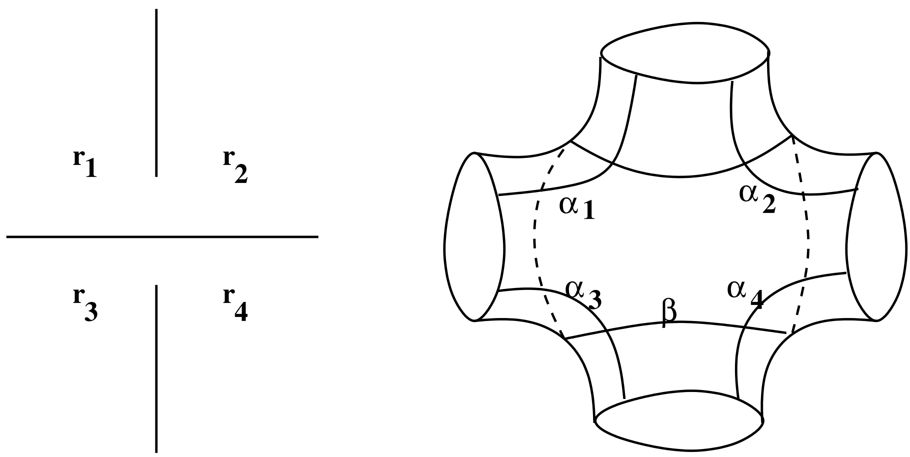


FIGURE 10. Special Heegaard diagram for knot crossings. At each crossing as pictured on the left, we construct a piece of the Heegaard surface on the right (which is topologically a four-punctured sphere). The curve β is the one corresponding to the crossing on the left; the four arcs $\alpha_1, ..., \alpha_4$ will close up.

FIGURE 11. Definition of $b(c_i)$.

13.1. The Euler characteristic of knot Floer homology. As an obvious consequence of Theorem 13.1 we have the following

THEOREM 13.3.

$$\sum_i \sum_j (-1)^j \cdot \mathrm{rk}(H_{i,j}(K)) \cdot T^i = \Delta_K(T). \tag{7}$$

It is interesting to compare this with [**1**], [**19**], and [**6**].

13.2. Computing knot Floer homology for alternating knots. It is clear from the above formulas that if K has an alternating projection, then $F(\mathbf{x}) - G(\mathbf{x})$ is independent of the choice of state $\mathbf{x}$. It follows that if we use the chain complex associated to this Heegaard diagram, then there are no differentials in the knot Floer homology, and indeed, its rank is determined by its Euler characteristic. Indeed, by calculating the constant, we get the following result, proved in [**26**]:

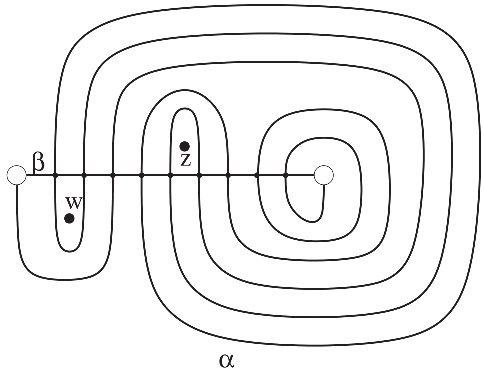


FIGURE 12.

THEOREM 13.4. *Let $K \subset S^3$ be an alternating knot in the three-sphere, write its symmetrized Alexander polynomial as*

$$\Delta_K(T) = \sum_{i=-n}^{n} a_i T^i$$

and let $\sigma(K)$ denote its signature. Then, $H_{i,j}(K) = 0$ for $j \neq i + \frac{\sigma(K)}{2}$, and

$$H_{i,i+\sigma(K)/2} \cong \mathbb{Z}^{|a_i|}.$$

We see that knot Floer homology is relatively simple for alternating knots. For general knots, however, the computation is more subtle because it involves counting holomorphic disks. In the next section we study more examples.

14. More computations

For knots that admit two-pointed genus 1 Heegaard diagrams computing knot Floer homology is relatively straightforward. In this case we study holomorphic disks in the torus. For an interesting example see Figure 12. The two empty circles are glued along a cylinder, so that no new intersection points are introduced between the curve α (the darker curve) and β (the lighter, horizontal curve).

EXERCISE 14.1. Compute the Alexander polynomial of K in Figure 12.

EXERCISE 14.2. Compute the knot Floer homology of K in Figure 12.

Another special class is given by Berge knots [**2**]. These are knots that admit lens space surgeries.

THEOREM 14.3. *([**24**]) Suppose that $K \subset S^3$ is a knot for which there is a positive integer p so that p surgery along K is a lens space. Then, there is an increasing sequence of integers*

$$n_{-m} < ... < n_m$$

with the property that $n_s = -n_{-s}$, with the following significance. For $-m \leq s \leq m$ we let

$$\delta_i = \begin{cases} 0 & \text{if } s = m \\ \delta_{s+1} - 2(n_{s+1} - n_s) + 1 & \text{if } m - s \text{ is odd} \\ \delta_{s+1} - 1 & \text{if } m - s > 0 \text{ is even,} \end{cases}$$

Then for each s with $|s| \leq m$ we have

$$H_{n_s,\delta_s}(K) = \mathbb{Z}$$

Furthermore, for all other values of i, j we have $H_{i,j}(K) = 0$.

For example the right-handed (p, q) torus knots admit lens space surgeries with slopes $pq \pm 1$, so the above theorem gives a quick computation for $H_{i,j}(T_{p,q})$.

14.1. Relationship with the genus of K. A knot $K \subset S^3$ can be realized as the boundary of an embedded, orientable surface in S^3. Such a surface is called a Seifert surface for K, and the minimal genus of any Seifert surface for K is called its *Seifert genus*, denoted $g(K)$. Clearly $g(K) = 0$ if and only if K is the unknot. The following theorem is proved in [**28**].

THEOREM 14.4. *For any knot $K \subset S^3$, let*

$$\deg H_{i,j}(K) = \max\{i \in \mathbb{Z} \big| \oplus_j H_{i,j}(K) \neq 0\}$$

denote the degree of the knot Floer homology. Then

$$g(K) = \deg H_{i,j}(K).$$

In particular knot Floer homology distinguishes every non-trivial knot from the unknot.

For more results on computing knot Floer homology see [**33**], [**34**] [**29**] [**12**], [**32**], [**13**], and [**5**].

References

[1] S. Akbulut and J. D. McCarthy. *Casson's invariant for oriented homology 3-spheres*, volume 36 of *Mathematical Notes*. Princeton University Press, Princeton, NJ, 1990. An exposition.

[2] J. O. Berge. Some knots with surgeries giving lens spaces. Unpublished manuscript.

[3] P. Braam and S. K. Donaldson. Floer's work on instanton homology, knots, and surgery. In H. Hofer, C. H. Taubes, A. Weinstein, and E. Zehnder, editors, *The Floer Memorial Volume*, number 133 in Progress in Mathematics, pages 195–256. Birkhäuser, 1995.

[4] S. K. Donaldson. *Floer homology groups in Yang-Mills theory*, volume 147 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2002. With the assistance of M. Furuta and D. Kotschick.

[5] E. Eftekhary. Knot Floer homologies for pretzel knots. math.GT/0311419.

[6] R. Fintushel and R. J. Stern. Knots, links, and 4-manifolds. *Invent. Math.*, 134(2):363–400, 1998.

[7] A. Floer. An instanton-invariant for 3-manifolds. *Comm. Math. Phys.*, 119:215–240, 1988.

[8] A. Floer. The unregularized gradient flow of the symplectic action. *Comm. Pure Appl. Math.*, 41(6):775–813, 1988.

[9] A. Floer, H. Hofer, and D. Salamon. Transversality in elliptic Morse theory for the symplectic action. *Duke Math. J*, 80(1):251–29, 1995.

[10] K. A. Frøyshov. The Seiberg-Witten equations and four-manifolds with boundary. *Math. Res. Lett*, 3:373–390, 1996.

[11] K. Fukaya, Y-G. Oh, K. Ono, and H. Ohta. *Lagrangian intersection Floer theory—anomaly and obstruction.* Kyoto University, 2000.

[12] H. Goda, H. Matsuda, and T. Morifuji. Knot Floer homology of $(1, 1)$-knots. math.GT/0311084.

[13] M. Hedden. On knot Floer homology and cabling. math.GT/0406402.

[14] L. H. Kauffman. *Formal knot theory.* Number 30 in Mathematical Notes. Princeton University Press, 1983.

[15] P. B. Kronheimer and T. S. Mrowka. Floer homology for Seiberg-Witten Monopoles. In preparation.

[16] I. G. MacDonald. Symmetric products of an algebraic curve. *Topology*, 1:319–343, 1962.
[17] M. Marcolli and B-L. Wang. Equivariant Seiberg-Witten Floer homology. dg-ga/9606003.
[18] D. McDuff and D. Salamon. *J-holomorphic curves and quantum cohomology*. Number 6 in University Lecture Series. American Mathematical Society, 1994.
[19] G. Meng and C. H. Taubes. $\underline{\text{SW}}$=Milnor torsion. *Math. Research Letters*, 3:661–674, 1996.
[20] J. Milnor. *Morse theory*. Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51. Princeton University Press, Princeton, N.J., 1963.
[21] J. Milnor. *Lectures on the h-cobordism theorem*. Princeton University Press, 1965. Notes by L. Siebenmann and J. Sondow.
[22] A. Némethi. On the Ozsváth-Szabó invariant of negative definite plumbed 3-manifolds. math.GT/0310083.
[23] P. S. Ozsváth and Z. Szabó. Lectures on Heegaard Floer homology. preprint.
[24] P. S. Ozsváth and Z. Szabó. On knot Floer homology and lens space surgeries. math.GT/0303017.
[25] P. S. Ozsváth and Z. Szabó. Absolutely graded Floer homologies and intersection forms for four-manifolds with boundary. *Adv. Math.*, 173(2):179–261, 2003.
[26] P. S. Ozsváth and Z. Szabó. Heegaard Floer homology and alternating knots. *Geom. Topol.*, 7:225–254, 2003.
[27] P. S. Ozsváth and Z. Szabó. On the Floer homology of plumbed three-manifolds. *Geometry and Topology*, 7:185–224, 2003.
[28] P. S. Ozsváth and Z. Szabó. Holomorphic disks and genus bounds. *Geom. Topol.*, 8:311–334, 2004.
[29] P. S. Ozsváth and Z. Szabó. Holomorphic disks and knot invariants. *Adv. Math.*, 186(1):58–116, 2004.
[30] P. S. Ozsváth and Z. Szabó. Holomorphic disks and three-manifold invariants: properties and applications. *Ann. of Math. (2)*, 159(3):1027–1158, 2004.
[31] P. S. Ozsváth and Z. Szabó. Holomorphic disks and topological invariants for closed three-manifolds. *Ann. of Math. (2)*, 159(3):1159–1245, 2004.
[32] P. S. Ozsváth and Z. Szabó. Knot Floer homology, genus bounds, and mutation. *Topology Appl.*, 141(1-3):59–85, 2004.
[33] J. A. Rasmussen. Floer homology of surgeries on two-bridge knots. *Algebr. Geom. Topol.*, 2:757–789, 2002.
[34] J. A. Rasmussen. *Floer homology and knot complements*. PhD thesis, Harvard University, 2003. math.GT/0306378.
[35] J. Robbin and D. Salamon. The Maslov index for paths. *Topology*, 32(4):827–844, 1993.
[36] R. Rustamov. Calculation of Heegaard Floer homology for a class of Brieskorn spheres. math.SG/0312071, 2003.
[37] N. Saveliev. *Lectures on the topology of* 3*-manifolds*. de Gruyter Textbook. Walter de Gruyter & Co., Berlin, 1999. An introduction to the Casson invariant.
[38] M. Scharlemann and A. Thompson. Heegaard splittings of (surface) $\times I$ are standard. *Math. Ann.*, 295(3):549–564, 1993.
[39] J. Singer. Three-dimensional manifolds and their Heegaard diagrams. *Trans. Amer. Math. Soc.*, 35(1):88–111, 1933.
[40] V. Turaev. Torsion invariants of Spin^c-structures on 3-manifolds. *Math. Research Letters*, 4:679–695, 1997.

Department of Mathematics, Columbia University, New York 10025
E-mail address: `petero@math.columbia.edu`

Department of Mathematics, Princeton University, New Jersey 08544
E-mail address: `szabo@math.princeton.edu`

Clay Mathematics Proceedings
Volume 5, 2006

Lectures on Heegaard Floer Homology

Peter Ozsváth and Zoltán Szabó

Abstract. These are notes for a lecture series on Heegaard Floer homology. Their aim is to study the surgery long exact sequence for these invariants, which relates the Heegaard Floer homology groups of three-manifolds which differ by surgeries along a knot. We sketch here a proof of this result, and give some of its applications. In fact, the primary application we focus on is the Dehn surgery classification of the unknot.

These are notes for the second lecture course on Heegaard Floer homology in the Clay Mathematics Institute Budapest Summer School in June 2004, taught by the first author. Although some of the topics covered in that course did not make it into these notes (specifically, the discussion of "knot Floer homology" which instead is described in the lecture notes for the first course, cf. [**44**]), the central aim has remained largely the same: we have attempted to give a fairly direct path towards some topological applications of the surgery long exact sequence in Heegaard Floer homology. Specifically, the goal was to sketch with the minimum amount of machinery necessary a proof of the Dehn surgery characterization of the unknot, first established in a collaboration with Peter Kronheimer, Tomasz Mrowka, and the authors. (This problem was first solved in [**29**] using Seiberg-Witten gauge theory, rather than Heegaard Floer homology; the approach outlined here can be found in [**39**].)

In Lecture 1, the surgery exact triangle is stated, and some of its immediate applications are given. In Lecture 2, it is proved. Lecture 3 concerns the maps induced by smooth cobordisms between three-manifolds. This is the lecture containing the fewest technical details – though most of those can be found in [**34**]. In Lecture 4, we show how the exact triangle, together with properties of the maps appearing in it, lead to a proof of the Dehn surgery classification of the unknot.

An attempt has been made to keep the discussion as simple as possible. For example, in these notes we avoid the use of "twisted coefficients". This comes at a price: as a result, we do not develop the necessary machinery required to handle knots with genus one. It is hoped that the reader's interest will be sufficiently piqued to study the original papers to fill in this gap. There are also a number of exercises scattered throughout the text, in topics ranging from homological algebra

2000 *Mathematics Subject Classification.* 53D, 57R.

PSO was supported by NSF grant number DMS 0234311.
ZSz was supported by NSF grant number DMS 0107792.

and elementary conformal mapping to low-dimensional topology. The reader is strongly encouraged to think through these exercises; some of the proofs in the text rely on them. At the conclusion of each lecture, there is a discussion on further reading on the material.

Several thanks are in order. First of all, we would like to thank the Clay Mathematics Institute for making this conference possible. Second, we would like to thank Peter Kronheimer, Tomasz Mrowka, and András Stipsicz for fruitful interactions and collaborations. Finally, we would like to thank Matthew Hedden and Shaffiq Welji for taking very helpful notes during the lectures, and for giving valuable feedback on an early draft of these notes.

1. Introduction to the surgery exact triangle

The exact triangle is a key calculational tool in Heegaard Floer homology. It relates the Heegaard Floer homology groups of three-manifolds obtained by surgeries along a framed knot in a closed, oriented three-manifold. Before stating the result precisely, we review some aspects of Heegaard Floer homology briefly, and then some of the topological constructions involved.

1.1. Background on Heegaard Floer groups: notation. Recall that Heegaard Floer homology is an Abelian group associated to a three-manifold, equipped with a Spin^c structure $\mathfrak{t} \in \mathrm{Spin}^c(Y)$. It comes in several variants.

Let $(\Sigma, \{\alpha_1, ...\alpha_g\}, \{\beta_1, ..., \beta_g\}, z)$ be a Heegaard diagram for Y, where here $\boldsymbol{\alpha} = \{\alpha_1, ..., \alpha_g\}$ and $\boldsymbol{\beta} = \{\beta_1, ..., \beta_g\}$ are attaching circles for two handlebodies bounded by Σ, and $z \in \Sigma - \alpha_1 - ... - \alpha_g - \beta_1 - ... - \beta_g$ is a reference point.

Form the g-fold symmetric product $\mathrm{Sym}^g(\Sigma)$, and let $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$ be the tori

$$\mathbb{T}_\alpha = \alpha_1 \times ... \times \alpha_g \quad \text{and} \quad \mathbb{T}_\beta = \beta_1 \times ... \times \beta_g.$$

The simplest version of Heegaard Floer homology is the homology groups of a chain complex generated by the intersection points of $\mathbb{T}_\alpha$ with $\mathbb{T}_\beta$: $\widehat{CF}(Y) = \bigoplus_{\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta} \mathbb{Z}\mathbf{x}$. This is endowed with a differential

$$\partial \mathbf{x} = \sum_{\mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta} \sum_{\{\phi \in \pi_2(\mathbf{x},\mathbf{y}) \big| \mu(\phi)=1, n_z(\phi)=0\}} \# \left(\frac{\mathcal{M}(\phi)}{\mathbb{R}} \right) \mathbf{y}.$$

Here, $\pi_2(\mathbf{x}, \mathbf{y})$ denotes the space of homology classes of Whitney disks connecting $\mathbf{x}$ and $\mathbf{y}$ [1], $n_z(\phi)$ denotes the algebraic intersection number of a representative of ϕ with the codimension-two submanifold $\{z\} \times \mathrm{Sym}^{g-1}(\Sigma) \subset \mathrm{Sym}^g(\Sigma)$, $\mathcal{M}(\phi)$ denotes the moduli space of pseudo-holomorphic representatives of ϕ, and $\mu(\phi)$ denotes the expected dimension of that moduli space, its Maslov index. Also, $\# \left(\frac{\mathcal{M}(\phi)}{\mathbb{R}} \right)$ is an appropriately signed count of points in the quotient of $\mathcal{M}(\phi)$ by the natural $\mathbb{R}$ action defined by automorphisms of the domain. To avoid a distracting discussion of signs, we sometimes change to the base ring $\mathbb{Z}/2\mathbb{Z}$, where now this coefficient is simply the parity of the number of points in $\mathcal{M}(\phi)/\mathbb{R}$. The loss of generality coming with this procedure is irrelevant for the topological applications appearing later in these lecture notes.

[1] In the case where $g(\Sigma) > 2$, we have that $\pi_2(\mathrm{Sym}^g(\Sigma)) \cong \mathbb{Z}$, and hence the distinction between homotopy and homology classes of Whitney disks disappears.

There is an obstruction to connecting $\mathbf{x}$ and $\mathbf{y}$ by a Whitney disk, which leads to a splitting of the above chain complex according to Spin^c structures over Y, induced from a partitioning of $\mathbb{T}_\alpha \cap \mathbb{T}_\beta$ according to Spin^c structures, $\widehat{CF}(Y) = \bigoplus_{\mathfrak{t}\in\mathrm{Spin}^c(Y)} \widehat{CF}(Y,\mathfrak{t})$. The homology groups of $\widehat{CF}(Y,\mathfrak{t})$, $\widehat{HF}(Y,\mathfrak{t})$, are topological invariants of Y and the Spin^c structure $\mathfrak{t}$.

There are other versions of these groups, taking into account more of the homology classes $\phi \in \pi_2(\mathbf{x},\mathbf{y})$. Specifically, we consider the boundary operator

$$\partial \mathbf{x} = \sum_{\mathbf{y}\in\mathbb{T}_\alpha\cap\mathbb{T}_\beta} \sum_{\{\phi\in\pi_2(\mathbf{x},\mathbf{y})\,\big|\,\mu(\phi)=1\}} \#\left(\frac{\mathcal{M}(\phi)}{\mathbb{R}}\right)\cdot U^{n_z(\phi)}\mathbf{y},$$

where U is a formal variable. This can be thought of as acting on either the free $\mathbb{Z}[U]$-module generated by intersection points of $\mathbb{T}_\alpha \cap \mathbb{T}_\beta$ ($CF^-(Y,\mathfrak{t})$), or the free $\mathbb{Z}[U,U^{-1}]$-module generated by these same intersection points ($CF^\infty(Y,\mathfrak{t})$), or the module with one copy of $\mathcal{T}^+ = \mathbb{Z}[U,U^{-1}]/U\cdot\mathbb{Z}[U]$ for each intersection point ($CF^+(Y,\mathfrak{t})$). Note also that when the first Betti number of Y, $b_1(Y)$, is non-zero, special "admissible" Heegaard diagrams must be used to ensure the necessary finiteness properties for the sums defining the boundary maps. Once this is done, the homology groups of the chain complexes $HF^-(Y,\mathfrak{t})$, $HF^\infty(Y,\mathfrak{t})$, and $HF^+(Y,\mathfrak{t})$ are topological invariants of Y equipped with its Spin^c structure $\mathfrak{t}$.

For instance, when working with $\widehat{HF}$ and HF^+ for a three-manifold with $b_1(Y) > 0$, we need the following notions.

Definition 1.1. Let $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z)$ be a pointed Heegaard diagram. The attaching curves divide Σ into a collection of components $\{\mathcal{D}_i\}_{i=1}^n$, one of which contains the distinguished point z. Let $P = \sum_i n_i \cdot \mathcal{D}_i$ be a two-chain in Σ. Its boundary can be written as a sum of subarcs of the α_i and the β_j. The two-chain P is called a *periodic domain* its local multiplicity at z vanishes and if for each i the segments of α_i appear with the same multiplicity. (More informally, we express this condition by saying that the boundary of P can be represented as a sum of the α_i and the β_j.) A Heegaard diagram is said to be *weakly admissible* if all the non-trivial periodic domains have both positive and negative local multiplicities.

Exercise 1.2. Identify the group of periodic domains (where the group law is given by addition of two-chains) with $H_2(Y;\mathbb{Z})$.

Weakly admissible Heegaard diagrams can be found for any three-manifold, and the groups $\widehat{HF}(Y,\mathfrak{t})$ and $HF^+(Y,\mathfrak{t})$ are the homology groups of the chain complexes $\widehat{CF}(Y,\mathfrak{t})$ and $CF^+(Y,\mathfrak{t})$ associated to such a diagram. For more details, and also a stronger notion of admissibility which gives HF^- and HF^∞, see for example Subsection 4.2.2 of [**41**]).

Exercise 1.3. Show that, with coefficients in $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$, $\widehat{HF}(S^1\times S^2) \cong \mathbb{F}\oplus\mathbb{F}$. Note that there is a also a Heegaard diagram for $S^1\times S^2$ for which $\mathbb{T}_\alpha\cap\mathbb{T}_\beta = \emptyset$ (but of course this diagram is not weakly admissible). *Hint:* draw a genus one Heegaard diagram for $S^2\times S^1$.

Exercise 1.4. Let M be a module over the ring $\mathbb{Z}[U]$. Let M_U denote its localization $M_U = M \otimes_{\mathbb{Z}[U]} \mathbb{Z}[U,U^{-1}]$.

(1) Show that the kernel of the natural map $M \longrightarrow M_U$ consists of the submodule of $m \in M$ such that there is an $n \geq 0$ with $U^n \cdot m = 0$.
(2) Let C be a chain complex of free modules over the ring $\mathbb{Z}[U]$. Show that there is a natural isomorphism $H_*(C_U) \cong H_*(C)_U$.

If C is a chain complex of free $\mathbb{Z}[U]$-modules, we have natural short exact sequences

$$0 \longrightarrow C \longrightarrow C_U \longrightarrow C_U/C \longrightarrow 0,$$

and

$$0 \longrightarrow C/UC \longrightarrow C_U/C \xrightarrow{U} C_U/C \longrightarrow 0,$$

both of which are functorial under chain maps between complexes over $\mathbb{Z}[U]$.

(3) Show that if a chain map $f\colon C \longrightarrow C'$ of free $\mathbb{Z}[U]$-modules induces an isomorphism on $H_*(C/UC) \longrightarrow H_*(C'/UC')$, then it induces isomorphisms

$$H_*(C) \cong H_*(C'), \qquad H_*(C_U) \cong H_*(C'_U), \qquad H_*(C_U/C) \cong H_*(C'_U/C')$$

as well. Indeed, if $g\colon C_U/C \longrightarrow C'/C'_U$ is a map of $\mathbb{Z}[U]$-complexes (not necessarily induced from a map from C to C'), then there is an induced map $\widehat{g}\colon C/UC \longrightarrow C'/UC'$, and if $\widehat{g}$ induces an isomorphism on homology, then so does g.
(4) Suppose that there is some d so that $\mathrm{Ker}\ U^d = \mathrm{Ker}\ U^{d+1}$ on $H_*(C)$ (as is the case, for example, if C is a finitely generated complex of $\mathbb{Z}[U]$ modules). Show then that $H_*(C_U) \longrightarrow H_*(C_U/C)$ is surjective if and only if the map $U\colon H_*(C_U/C) \longrightarrow H_*(C_U/C)$ is.
(5) Show that $H_*(C_U/C) \neq 0$ if and only if $H_*(C/UC) \neq 0$.

The relevance of the above exercises is the following: $CF^-(Y,\mathfrak{t})$ is a chain complex of free $\mathbb{Z}[U]$-modules, and $CF^\infty(Y,\mathfrak{t})$ $CF^+(Y,\mathfrak{t})$, and $\widehat{CF}(Y,\mathfrak{t})$ are the associated complexes $CF^-(Y,\mathfrak{t})_U$, $CF^-(Y,\mathfrak{t})_U/CF^-(Y,\mathfrak{t})$, and $CF^-(Y,\mathfrak{t})/UCF^-(Y,\mathfrak{t})$ respectively. In particular, we have two functorially assigned long exact sequences

$$\text{(1)} \qquad \ldots \longrightarrow HF^-(Y,\mathfrak{t}) \xrightarrow{\ell_*} HF^\infty(Y,\mathfrak{t}) \xrightarrow{q_*} HF^+(Y,\mathfrak{t}) \longrightarrow \ldots$$

and

$$\text{(2)} \qquad \ldots \longrightarrow \widehat{HF}(Y,\mathfrak{t}) \longrightarrow HF^+(Y,\mathfrak{t}) \xrightarrow{U} HF^+(Y,\mathfrak{t}) \longrightarrow \ldots$$

(both of which are natural under chain maps $CF^-(Y,\mathfrak{t}) \longrightarrow CF^-(Y',\mathfrak{t}')$).

1.2. Background: $\mathbb{Z}/2\mathbb{Z}$ gradings. Heegaard Floer homology is a relatively $\mathbb{Z}/2\mathbb{Z}$-graded group. To describe this, fix arbitrary orientations on $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$, and give $\mathrm{Sym}^g(\Sigma)$ its induced orientation from Σ. At each intersection point $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$, we can then define a local intersection number $\iota(\mathbf{x})$ by the rule that the complex orientation on $T_{\mathbf{x}}\mathrm{Sym}^g(\Sigma)$ is $\iota(\mathbf{x}) \in \{\pm 1\}$ times the induced orientation from $T_{\mathbf{x}}\mathbb{T}_\alpha \oplus T_{\mathbf{x}}\mathbb{T}_\beta$. As is familiar in differential topology (compare [**33**]), we can define the algebraic intersection number of $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$ by the formula

$$\#(\mathbb{T}_\alpha \cap \mathbb{T}_\beta) = \sum_{\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta} \iota(\mathbf{x}).$$

The overall sign of this depends on the choice of orientations of $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$, but once this is decided, the intersection number depends only on the induced homology classes of $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$.

We can think about the intersection number directly in terms of the Heegaard surface as follows. Fix orientations on all the curves $\{\alpha_i\}_{i=1}^g$ and $\{\beta_i\}_{i=1}^g$ (these in turn induce orientations on the tori $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$). In this case, $\#(\mathbb{T}_\alpha \cap \mathbb{T}_\beta)$ is the determinant of the $g \times g$ matrix formed from the algebraic intersection of α_i and β_j (with $i, j \in \{1, ..., g\}$).

Exercise 1.5. Let $(\Sigma, \{\alpha_1, ..., \alpha_g\}, \{\beta_1, ..., \beta_g\})$ be a Heegaard diagram for a closed, oriented three-manifold Y. Show that there is a corresponding CW-complex structure on Y with one zero-cell, one three-cell, g one-cells $\{a_i\}_{i=1}^g$, and g two-cells $\{b_i\}_{i=1}^g$. Show that the only non-trivial boundary operator $\partial\colon C_2 \longrightarrow C_1$ has the form

$$\partial b_i = \sum_{i=1}^{g} \#(\alpha_i \cap \beta_j) a_j.$$

Choose orientations for the $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$ as above. Then there is a corresponding splitting of $\widehat{CF}(Y)$ into two summands,

$$\widehat{CF}(Y) = \bigoplus_{i \in \mathbb{Z}/2\mathbb{Z}} \widehat{CF}_i(Y), \tag{3}$$

where here $\widehat{CF}_i(Y)$ is generated by intersection points $\mathbf{x}$ with $\iota(\mathbf{x}) = (-1)^i$. Note that although $\iota(\mathbf{x})$ depends on the (arbitrarily chosen) orientations of $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$, if $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ are two different intersection points, it is easy to see that the product $\iota(\mathbf{x}) \cdot \iota(\mathbf{y})$ is independent of this choice. In fact, according to standard properties of the Maslov index (see for example [**46**]), if $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$

$$\iota(\mathbf{x}) \cdot \iota(\mathbf{y}) = (-1)^{\mu(\phi)},$$

where $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$ is any homology class of Whitney disk. Thus, the boundary map reverses the splitting from Equation (3), i.e. we have that

$$\partial\colon \widehat{CF}_i(Y) \longrightarrow \widehat{CF}_{i+1}(Y)$$

(thinking of $i \in \mathbb{Z}/2\mathbb{Z}$). It is a straightforward consequence of this that there is also a $\mathbb{Z}/2\mathbb{Z}$ splitting of the homology:

$$\widehat{HF}(Y) = \bigoplus_{i \in \mathbb{Z}/2\mathbb{Z}} \widehat{HF}_i(Y),$$

where here $\widehat{HF}_i(Y)$ is represented by cycles supported in $\widehat{CF}_i(Y)$. An element of $\widehat{HF}(Y)$ which is supported in $\widehat{HF}_i(Y)$ for some $i \in \mathbb{Z}/2\mathbb{Z}$ is said to be *homogeneous*.

Now according to standard properties of the Euler characteristic, we have that

$$\chi(\widehat{HF}_*(Y)) = \mathrm{rk}(\widehat{HF}_0(Y)) - \mathrm{rk}(\widehat{HF}_1(Y)) = \mathrm{rk}(\widehat{CF}_0(Y)) - \mathrm{rk}(\widehat{CF}_1(Y));$$

and it is also clear from the definitions that

$$\chi(\widehat{CF}_*(Y)) = \#(\mathbb{T}_\alpha \cap \mathbb{T}_\beta).$$

Indeed, the latter intersection number can also be interpreted in terms of homological data, as follows.

Lemma 1.6. *Given a three-manifold Y, let $|H_1(Y;\mathbb{Z})|$ denote the integer defined as follows. If the number of elements n in $H_1(Y;\mathbb{Z})$ is finite, then $|H_1(Y;\mathbb{Z})| = n$; otherwise, $|H_1(Y;\mathbb{Z})| = 0$. Then,*

$$\chi(\widehat{HF}(Y)) = \pm|H_1(Y;\mathbb{Z})|.$$

In fact, if Y is a three-manifold and $\mathfrak{t} \in \mathrm{Spin}^c(Y)$, then

$$\chi(\widehat{HF}(Y,\mathfrak{t})) = \begin{cases} \pm 1 & \text{if } H_1(Y;\mathbb{Z}) \text{ is finite} \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Proof. The identification of $\chi(\widehat{HF}(Y))$ with $|H_1(Y;\mathbb{Z})|$ is a direct consequence of the above discussion and Exercise 1.5. Now, Equation (4) amounts to the fact that $\chi(\widehat{HF}(Y,\mathfrak{t}))$ is independent of the choice of $\mathfrak{t}$. This is a consequence of the fact that $\chi(\widehat{HF}(Y,\mathfrak{t}))$ is independent of the choice of basepoint (i.e. by varying the basepoint, the generators of the chain complex and their $\mathbb{Z}/2\mathbb{Z}$ gradings remain the same), whereas the Spin^c depends on this choice. $\square$

We can use Lemma 1.6 to lift the relative $\mathbb{Z}/2\mathbb{Z}$ grading on $\widehat{HF}(Y)$ to an absolute grading, provided that $H_1(Y;\mathbb{Z})$ is finite: the $\mathbb{Z}/2\mathbb{Z}$ grading is pinned down by the convention that $\chi(\widehat{HF}(Y))$ is positive. (In fact, this $\mathbb{Z}/2\mathbb{Z}$ grading can be naturally generalized to all closed three-manifolds, cf. Section 10.4 of [**40**].)

There are refinements of this $\mathbb{Z}/2\mathbb{Z}$ grading in the presence of additional structure. For example, a rational homology three-sphere is a three-manifold with finite $H_1(Y;\mathbb{Z})$ (equivalently, $H_*(Y;\mathbb{Q}) \cong H_*(S^3;\mathbb{Q})$). For a rational homology three-sphere, if $\mathbf{x},\mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ can be connected by some $\phi \in \pi_2(\mathbf{x},\mathbf{y})$, then in fact the quantity

$$\mathrm{gr}(\mathbf{x},\mathbf{y}) = \mu(\phi) - 2n_z(\phi) \tag{5}$$

is independent of the choice of ϕ (depending only on $\mathbf{x}$ and $\mathbf{y}$). Correspondingly, we can use $\mathrm{gr}(\mathbf{x},\mathbf{y})$ to define relative $\mathbb{Z}$ gradings on the Heegaard Floer homology groups, by defining the grading of the generator $U^i \cdot \mathbf{x}$ minus the grading of $U^j \cdot \mathbf{y}$ to be $\mathrm{gr}(\mathbf{x},\mathbf{y}) - 2(i-j)$. This relative $\mathbb{Z}$ grading can be lifted to an absolute $\mathbb{Q}$ grading, as discussed in Lecture 3.

There is one additional basic property of Heegaard Floer homology which we will need, and that is the conjugation symmetry. The set of Spin^c structures over Y admits an involution, written $\mathfrak{t} \mapsto \overline{\mathfrak{t}}$. It is always true that

$$HF^\circ(Y,\mathfrak{t}) \cong HF^\circ(Y,\overline{\mathfrak{t}}) \tag{6}$$

(for any of the variants $HF^\circ = \widehat{HF}$, HF^-, HF^∞, or HF^+).

1.3. ***L*-spaces.** An L-space is a rational homology three-sphere whose Heegaard Floer homology is as simple as possible.

Exercise 1.7. Prove that the following conditions on Y are equivalent:

- $\widehat{HF}(Y)$ is a free Abelian group with rank $|H^2(Y;\mathbb{Z})|$
- $HF^-(Y)$ is a free $\mathbb{Z}[U]$-module with rank $|H^2(Y;\mathbb{Z})|$
- $HF^\infty(Y)$ is a free $\mathbb{Z}[U,U^{-1}]$ module of rank $|H^2(Y;\mathbb{Z})|$, and the map

  $$U\colon HF^+(Y) \longrightarrow HF^+(Y)$$

  is surjective.

In fact, the hypothesis that $HF^\infty(Y)$ is a free $\mathbb{Z}[U, U^{-1}]$-module of rank $|H^2(Y;\mathbb{Z})|$ holds for any rational homology three-sphere (cf. Theorem 10.1 of [**40**]); but we do not require this result for our present purposes.

A three-manifold satisfying any of the hypotheses of Exercise 1.7 is called an *L-space*. Note that any lens space is an L-space. (This can be seen by drawing a genus one Heegaard diagram for $L(p,q)$, for which the two circles α and β meet transversally in p points.)

1.4. Statement of the surgery exact triangle. Let K be a knot in a closed, oriented three-manifold Y. Let $\mathrm{nd}(K)$ denote a tubular neighborhood of K, so that $M = Y - \mathrm{nd}(K)$ is a three-manifold with torus boundary. The *meridian* for K in Y is a primitive homology class in ∂M which lies in the kernel of the natural map

$$H_1(\partial M) = H_1(\partial \mathrm{nd}(K)) \longrightarrow H_1(\mathrm{nd}(K)).$$

Such a homology class can be represented by a homotopically non-trivial, simple, closed curve in the boundary of M which bounds a disk in $\mathrm{nd}(K)$. The homology (or isotopy) class of the meridian is uniquely specified up to multiplication by ± 1 by this property. A *longitude* for K is a homology class in $H_1(\partial M)$, with the property that the algebraic intersection number of $\#(\mu \cap \lambda) = -1$, where here ∂M is oriented as the boundary of M. Unlike the meridian, the homology (or isotopy) class of a longitude is not uniquely determined by this property. In fact, the set of longitudes for K is of the form $\{\lambda + n\cdot\mu\}_{n\in\mathbb{Z}}$. A *framed knot* $K \subset Y$ is a knot, together with a choice of longitude λ. When $K \subset Y$ is a knot with framing λ, we can form the new three-manifold $Y_\lambda(K)$ obtained by attaching a solid torus to M, in such a way that λ bounds a disk in the new solid torus. This three-manifold is said to be obtained from Y by λ-framed surgery along K.

It might seem arbitrary to restrict attention to longitudes. After all, if γ is any homotopically non-trivial, simple closed curve in ∂M, we can form a three-manifold which is a union of M and a solid torus, attached so that γ bounds a disk in the solid torus. (This more general operation is called *Dehn filling.*) However, if we restrict attention to longitudes, then there is not only a three-manifold, but also a canonical four-manifold $W_\lambda(K)$ consisting of a single two-handle attached to $[0,1]\times Y$ along $\{1\}\times Y$ with the framing specified by λ, giving a cobordism from Y to $Y_\lambda(K)$.

Exercise 1.8. Note that if $K \subset Y$ is a null-homologous knot (e.g. any knot in S^3), then there is a unique longitude λ for K which is null-homologous in $Y - \mathrm{nd}(K)$. This longitude is called the *Seifert framing* for $K \subset Y$. Show that for this choice of framing, the first Betti number of $Y_\lambda(K)$ is one; more generally,

$$H_1(Y_{p\cdot\mu+q\cdot\lambda}(K);\mathbb{Z}) \cong \mathbb{Z}/p\mathbb{Z}.$$

The three-manifold $Y_{p\cdot\mu+q\cdot\lambda}(K)$ is typically denoted $Y_{p/q}(K)$, where here $p/q \in \mathbb{Q}$.

Exercise 1.9. Let $K \subset S^3$ be a knot, equipped with its Seifert framing, and let $r \in \mathbb{Q}$ be any rational number. Show that $S^3_r(K) \cong -S^3_{-r}(\overline{K})$, where here $\overline{K}$ denotes the mirror of K (i.e. given a knot projection of K, $\overline{K}$ has a knot projection where all the over-crossings have been replaced by under-crossings), and the orientation on $S^3_r(K)$ is taken to be the one it inherits from S^3.

Fix a closed, oriented three-manifold Y, and let K be a framed knot in Y (i.e. a knot with a choice of longitude λ). Let $Y_0 = Y_0(K)$ denote the three-manifold obtained from λ-framed surgery on Y along K, and let $Y_1 = Y_1(K)$ denote the three-manifold obtained from $(\mu+\lambda)$-framed surgery on Y along K. We call the ordered triple (Y, Y_0, Y_1) a *triad* of three-manifolds.

This relationship between Y, Y_0, and Y_1 is symmetric under a cyclic permutation of the three three-manifolds. Indeed, it is not difficult to see that (Y, Y_0, Y_1) fit into a triad if and only if there is a single oriented three-manifold M with torus boundary, and three simple, closed curves γ, γ_0, and γ_1 in ∂M with

$$\#(\gamma \cap \gamma_0) = \#(\gamma_0 \cap \gamma_1) = \#(\gamma_1 \cap \gamma) = -1, \tag{7}$$

so that Y resp. Y_0 resp. Y_1 are obtained from M by attaching a solid torus along the boundary with meridian γ resp. γ_0 resp. γ_1.

EXAMPLE 1.10. Let $K \subset S^3$ be a knot in S^3 equipped with its Seifert framing, cf. Exercise 1.8. Then the three-manifolds S^3, $S^3_p(K)$ and $S^3_{p+1}(K)$ form a triad for any integer p. More generally, given relatively prime integers p_1 and q_1, we can find p_2 and q_2 so that $p_1q_2 - q_1p_2 = 1$. Then, writing $p_3 = p_1 + p_2$ and $q_3 = q_1 + q_2$, we have that $S^3_{p_1/q_1}(K)$, $S^3_{p_2/q_2}(K)$, and $S^3_{p_3/q_3}(K)$ fit into a triad.

Another natural example of triads appears in skein theory for links.

Let $L \subset S^3$ be a link. The *branched double-cover of* L, $\Sigma(L)$ is the three-manifold which admits an orientation-preserving involution whose quotient is S^3, so that the fixed point set of the involution is identified with $L \subset S^3$. The three-manifold $\Sigma(L)$ is uniquely determined by L.

Fix a generic planar projection of L, and let x denote a crossing for this planar projection. There are two naturally associated links L_0 and L_1 which are obtained by resolving the crossing x. These two resolutions are pictured in Figure 1. Note that if we begin with a knot, and fix a crossing, then one of its resolutions will also be a knot, but the other will be a two-component link.

EXERCISE 1.11. Show that the three-manifolds $\Sigma(L)$, $\Sigma(L_0)$, and $\Sigma(L_1)$ form a triad. *Hint:* Use the fact that the branched double-cover of the three-ball branched along two disjoint arcs is a solid torus.

We have set up the relevant topology necessary to state the surgery exact triangle:

THEOREM 1.12. *(Theorem 9.12 of* [**40**]*) Let Y, Y_0, and Y_1 be three three-manifolds which fit into a triad then there are long exact sequences which relate*

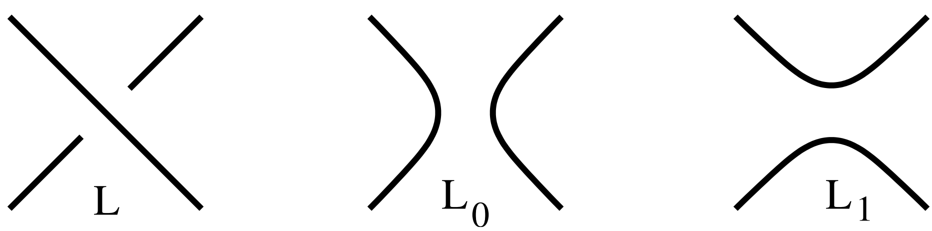


FIGURE 1. Resolutions. Given a link with a crossing as labeled in L above, we have two "resolutions" L_0 and L_1, obtained by replacing the crossing by the two simplifications pictured above.

their Heegaard Floer homologies (thought of as modules over $\mathbb{Z}[U]$*):*

$$... \longrightarrow \widehat{HF}(Y) \xrightarrow{\widehat{F}} \widehat{HF}(Y_0) \xrightarrow{\widehat{F}_0} \widehat{HF}(Y_1) \xrightarrow{\widehat{F}_1} ...$$

and

$$... \longrightarrow HF^+(Y) \xrightarrow{F^+} HF^+(Y_0) \xrightarrow{F_0^+} HF^+(Y_1) \xrightarrow{F_1^+} ...$$

All of the above maps respect the relative $\mathbb{Z}/2\mathbb{Z}$ gradings, in the sense that each map carries homogeneous elements to homogeneous elements.

We return to the proof of Theorem 1.12 in Lecture 2. In Lecture 3, we interpret the maps appearing in the long exact sequences as maps induced by the canonical two-handle cobordisms from Y to Y_0, Y_0 to Y_1, and Y_1 to Y. We focus now on some immediate applications. First, we use Theorem 1.12 to find examples of L-spaces.

Exercise 1.13. Suppose that Y, Y_0, and Y_1 are three three-manifolds which fit into a triad. For some cyclic reordering (Y, Y_0, Y_1), we can arrange that

$$|H_1(Y)| = |H_1(Y_0)| + |H_1(Y_1)|,$$

in the notation of Lemma 1.6.

The following is a quick application of Theorem 1.12 for $\widehat{HF}$:

Exercise 1.14. Let (Y, Y_0, Y_1) be a triad of rational homology three-spheres, ordered so that

$$|H_1(Y)| = |H_1(Y_0)| + |H_1(Y_1)|.$$

If Y_0 and Y_1 are L-spaces, then so is Y. *Hint:* Apply Theorem 1.12 and Lemma 1.6.

Exercise 1.14 provides a large number of examples of L-spaces.

For example, if $K \subset S^3$ is a knot in S^3 with the property that $S^3_r(K)$ is an L-space for some rational number $r > 0$ (with respect to the Seifert framing), then $S^3_s(K)$ is also an L-space for all $s > r$. This follows from Exercise 1.14, combined with Example 1.10. Concretely, if K is the (p, q) torus knot, then $S^3_{pq-1}(K)$ is a lens space, and hence, applying this principle, we see that in fact $S^3_r(K)$ is an L-space for all $r \geq pq - 1$.

There are other knots which admit lens space surgeries, which give rise to infinitely many interesting L-spaces. For example, if K is the $(-2, 3, 7)$ pretzel knot (cf. Figure 2), then $S^3_{18}(K) \cong L(18, 5)$ and $S^3_{19}(K) \cong L(19, 8)$ (cf. [**13**]).

Let $L \subset S^3$ be a link, and fix a generic projection of L. This projection gives a four-valent planar graph, which divides the plane into regions. These regions can be given a checkerboard coloring: we color them black and white so that two regions with the same color never meet along an edge. Thus, at each vertex there are always two (not necessarily distinct) black regions which meet. The *black graph* $\mathcal{B}(L)$ is the graph whose vertices correspond to the black regions, and whose edges correspond to crossings for the original projection, connecting the two black regions which meet at the corresponding crossing. (Strictly speaking, the black graph $\mathcal{B}(L)$ depends on a projection of L, but we do not record this dependence in the notation.) See Figure 3 for an illustration.

A knot or link projection is called *alternating* if, as we traverse each component of the link, the crossings of the projections alternate between over- and under-crossings. A knot which admits an alternating projection is simply called an *alternating knot*.

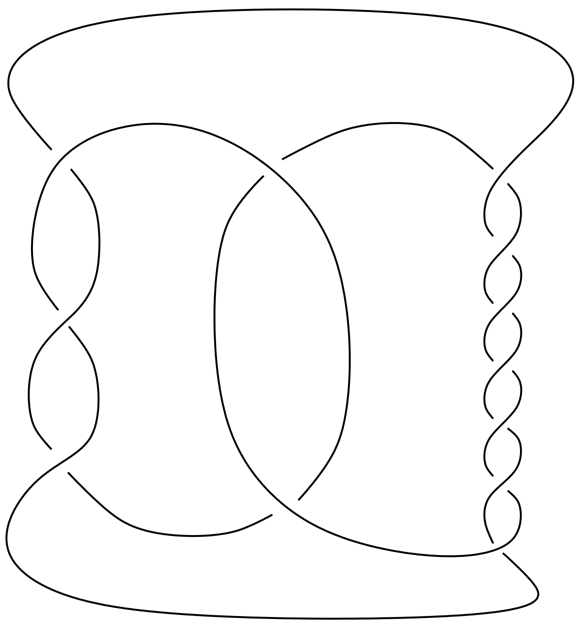


FIGURE 2. The $(-2,3,7)$ pretzel knot. Surgery on this knot with coefficients 18 and 19 give the lens spaces $L(18,5)$ and $L(19,8)$ respectively.

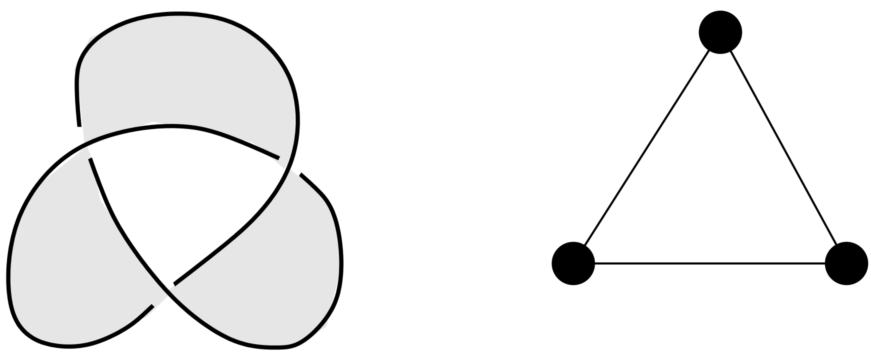


FIGURE 3. BlackGraph. We have illustrated at the left a checkerboard coloring of a projection of the trefoil; at the right, we have illustrated its corresponding "black graph".

PROPOSITION 1.15. *Let K be an alternating knot or, more generally, a link which admits an alternating, connected projection; then its branched double-cover $\Sigma(K)$ is an L-space.*

PROOF. We claim that if K is an alternating link with connected, alternating projection, and we can choose a crossing with the property that K_0 and K_1 both have connected projections, then the projections of K_0 and K_1 remain alternating, and moreover

$$|H_1(\Sigma(K))| = |H_1(\Sigma(K_0))| + |H_1(\Sigma(K_1))|. \tag{8}$$

This follows from two observations: first, it is a standard result in knot theory (see for example Chapter 9 of [**31**]) that for any link K, $|H_1(\Sigma(K))| = |\Delta_K(-1)|$. Second, if K is an alternating link with a connected, alternating projection, then $|\Delta_K(-1)|$ is the number of maximal subtrees of the black graph of that projection, cf. [**2**]. (Note that $|\Delta_K(-1)|$ for an arbitrary link can be interpreted as a signed count of maximal subtrees of $\mathcal{B}(L)$; but for an alternating projection, the signs are all $+1$.)

Returning to Equation (8), note that the black graph of K_0 and K_1 can be obtained from the black graph of K by either deleting or contracting the edge e corresponding to the given crossing; thus, the maximal subtrees of $\mathcal{B}(K_0)$ correspond to the maximal subtrees of $\mathcal{B}(K)$ which contain e, while the maximal subtrees of $\mathcal{B}(K_1)$ correspond to the maximal subtrees of $\mathcal{B}(K)$ which do not contain e. Equation (8) now follows at once from the expression of $|H_1(\Sigma(K))|$ for an alternating link (with connected projection) in terms of the number of maximal subtrees.

Recall also that a connected, alternating projection for link is called *reduced*, if for each crossing, either resolution is connected. If a connected, alternating projection of a link is not reduced, then we can always find a reduced projection as well. This is constructed inductively: if there is a crossing one of whose resolutions disconnected the projection, then it can be eliminated by twisting half the projection (to obtain a new connected, alternating projection with one fewer crossing).

The proposition now follows from induction on $|H_1(\Sigma(K))|$. Take a reduced projection of K. Either K represents the unknot, whose branched double-cover is S^3 (this is the basic case), or there is a crossing, neither of whose resolutions disconnects the connected, alternating projection. Thus, Equation (8) holds, and in particular, $0 < |H_1(\Sigma(K_i))| < |H_1(\Sigma(K))|$ for $i = 1, 2$. Thus, in view of the inductive hypothesis, we verify the inductive step by applying Exercise 1.14. $\square$

1.5. An application to Dehn surgery on knots in S^3. Note that for $K \subset S^3$, $H^2(S^3_0(K);\mathbb{Z}) \cong \mathbb{Z}$, and hence we can identify $\mathrm{Spin}^c(S^3_0(K)) \cong \mathbb{Z}$ (this is done by taking the first Chern class of $\mathfrak{s} \in \mathrm{Spin}^c(S^3_0(K))$, dividing it by two, and using a fixed isomorphism $H^2(S^3_0(K);\mathbb{Z}) \cong \mathbb{Z}$). We will correspondingly think of the decomposition of $HF^+(S^3_0(K))$ as indexed by integers,

$$HF^+(S^3_0(K)) = \bigoplus_{i\in\mathbb{Z}} HF^+(S^3_0(K), i).$$

COROLLARY 1.16. *Suppose that $K \subset S^3$ is a knot with $\widehat{HF}(S^3_{+1}(K)) \cong \widehat{HF}(S^3)$ (as $\mathbb{Z}/2\mathbb{Z}$-graded Abelian groups). Then $\widehat{HF}(S^3_0(K), i) = 0$ for all $i \neq 0$.*

PROOF. The long exact sequence from Theorem 1.12 ensures that $\widehat{HF}(S^3_0(K))$ must be either $\mathbb{Z}/m\mathbb{Z}$ for some m (which can be ruled out by other properties of Heegaard Floer homology, but this is is not necessary for our present purposes) or $\widehat{HF}(S^3_0(K)) \cong \mathbb{Z}^2$. Our goal is to understand in which Spin^c structure this group is supported. In order to be consistent with the Euler characteristic calculation, (Equation (4)) we must have that $\widehat{HF}(S^3_0(K), s) = 0$ for all but at most one s. But the conjugation symmetry $\widehat{HF}(S^3_0(K), t) \cong \widehat{HF}(S^3_0(K), -t)$ for all t ensures that in fact $\widehat{HF}(S^3_0(K), t) = 0$ for all $t \neq 0$. $\square$

The above corollary is particularly powerful when it is combined with a theorem from [**39**] (sketched in the proof of Theorem 4.2 below), according to which

Heegaard Floer homology of the zero-surgery detects the genus of K. Combining these results, we get that, if K is a knot with $\widehat{HF}(S^3) \cong \widehat{HF}(S^3_{+1}(K))$ as $\mathbb{Z}/2\mathbb{Z}$-graded Abelian groups, then either K is the unknot, or the Seifert genus of K is one. This claim should be compared with a theorem of Gordon and Luecke [**24**] which states that if $S^3 \cong S^3_{+1}(K)$, then K is the unknot. It is not a strict consequence of that result, since there are three-manifolds $Y \not\cong S^3$ with $\widehat{HF}(S^3) \cong \widehat{HF}(Y)$, such as the Poincaré homology three-sphere P, cf. [**37**]. Note that $+1$ surgery on the right-handed trefoil gives this three-manifold.

Note that any three-manifold Y which is a connected sum of several copies of P (with either orientation) has $\widehat{HF}(S^3) \cong \widehat{HF}(Y)$ (as $\mathbb{Z}/2\mathbb{Z}$-graded Abelian groups), and it is a very interesting question whether there are any other three-manifolds with this property. We return to generalizations and refinements of Corollary 1.16 in Lecture 4.

1.6. Further remarks. Heegaard Floer homology fits into a general framework of a (3+1)-dimensional topological quantum field theory. The first non-trivial theory which appears to possess this kind of structure is the instanton theory for four-manifolds, defined by Simon Donaldson [**8**], coupled with its associated three-manifold invariant, defined by Andreas Floer [**15**], [**7**]. Floer's instanton homology has not yet been constructed for all three-manifolds, but it can be defined for three-manifolds with some additional algebro-topological assumptions. For instance, it is defined in the case where $H_1(Y;\mathbb{Z}) = 0$. In a correspondingly more restricted setting, Floer noticed the existence of an exact triangle, see [**16**] and also [**1**].

A number of other instances of exact triangles have since appeared in several other variants of Floer homology, including Seidel's exact sequence for Lagrangian Floer homology, cf. [**48**], and another exact triangle [**29**] which holds for the Seiberg-Witten monopole Floer homology defined by Kronheimer and Mrowka, cf. [**26**].

L-spaces are of interest to three-manifold topologists, since these are three-manifolds which admit no taut foliations, cf. [**29**], [**39**]. Hyperbolic three-manifolds which admit no taut foliations were first constructed in [**47**], see also [**3**].

2. Proof of of the exact triangle.

We sketch here a proof of Theorem 1.12. To avoid issues with signs and orientations, we will restrict attention to coefficients in $\mathbb{Z}/2\mathbb{Z}$. We also focus on the case of $\widehat{HF}$ for simplicity, returning to HF^+ in Subsection 2.5.

Let $K \subset Y$ be a knot with framing λ. Then, we can find a compatible Heegaard diagram. Specifically, we can assume that K is an unknotted knot in the β-handlebody, meeting the attaching disk belonging to β_1 transversally in one point, and disjoint from all the other attaching disks for the β_i with $i > 1$. Thus, $(\Sigma, \{\alpha_1, ..., \alpha_g\}, \{\beta_1, ..., \beta_g\}, z)$ is a pointed Heegaard diagram for Y, and β_1 is a meridian for K. There is also a curve γ_1 which represents the framing λ for K, so that if we replace β_1 by γ_1, and let γ_i be an isotopic translate of β_i for $i > 1$, then $(\Sigma, \{\alpha_1, ..., \alpha_g\}, \{\gamma_1, ..., \gamma_g\}, z)$ is a pointed Heegaard diagram for $Y_\lambda(K)$. Similarly, we can find an embedded curve δ_g representing $\mu + \lambda$, so that if we let δ_i be an isotopic translate of β_i for $i > 1$, then $(\Sigma, \{\alpha_1, ..., \alpha_g\}, \{\delta_1, ..., \delta_g\}, z)$ is a pointed Heegaard diagram representing $Y_{\mu+\lambda}(K)$.

With this understood, we choose a more symmetrical notation $Y_{\alpha\beta}$, $Y_{\alpha\gamma}$, $Y_{\alpha\delta}$ to represent the three-manifolds Y, $Y_\lambda(K)$, and $Y_{\mu+\lambda}(K)$ respectively. Also, $Y_{\beta\gamma}$

denotes the three-manifold described by the Heegaard diagram $(\Sigma, \{\beta_1, ..., \beta_g\}, \{\gamma_1, ..., \gamma_g\})$, and $Y_{\gamma\delta}$ and $Y_{\delta\beta}$ are defined similarly. The reason we chose isotopic translates of the β_i to be our γ_i and δ_i (when $i > 1$) was to ensure that all the tori $\mathbb{T}_\alpha$, $\mathbb{T}_\beta$, $\mathbb{T}_\gamma$, $\mathbb{T}_\delta$ meet transversally in $\mathrm{Sym}^g(\Sigma)$.

EXERCISE 2.1. Show that $Y_{\beta\gamma} \cong Y_{\gamma\delta} \cong Y_{\delta\beta} \cong \#^{g-1}(S^1 \times S^2)$.

Before defining the maps appearing in the exact triangle, we allow ourselves a digression on holomorphic triangles. Counts of holomorphic triangles play a prominent role in Lagrangian Floer homology, cf. [**32**], [**5**], [**19**].

2.1. Holomorphic triangles. A *pointed Heegaard triple*

$$(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, z)$$

is an oriented two-manifold Σ, together with three g-tuples of attaching circles $\boldsymbol{\alpha} = \{\alpha_1, ..., \alpha_g\}$,$\boldsymbol{\beta} = \{\beta_1, ..., \beta_g\}$,$\boldsymbol{\gamma} = \{\gamma_1, ..., \gamma_g\}$ for handlebodies, and a choice of reference point

$$z \in \Sigma - \alpha_1 - ... - \alpha_g - \beta_1 - ... - \beta_g - \gamma_1 - ... - \gamma_g.$$

In the preceding discussion, we constructed the Heegaard triple of a framed link.

Let Δ denote the two-simplex, with vertices $v_\alpha, v_\beta, v_\gamma$ labeled clockwise, and let e_i denote the edge v_j to v_k, where $\{i, j, k\} = \{\alpha, \beta, \gamma\}$. Fix $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$, $\mathbf{y} \in \mathbb{T}_\beta \cap \mathbb{T}_\gamma$, $\mathbf{w} \in \mathbb{T}_\alpha \cap \mathbb{T}_\gamma$. Consider the map

$$u \colon \Delta \longrightarrow \mathrm{Sym}^g(\Sigma)$$

with the boundary conditions that $u(v_\gamma) = \mathbf{x}$, $u(v_\alpha) = \mathbf{y}$, and $u(v_\beta) = \mathbf{w}$, and $u(e_\alpha) \subset \mathbb{T}_\alpha$, $u(e_\beta) \subset \mathbb{T}_\beta$, $u(e_\gamma) \subset \mathbb{T}_\gamma$. Such a map is called a *Whitney triangle connecting* $\mathbf{x}$*,* $\mathbf{y}$*, and* $\mathbf{w}$. We let $\pi_2(\mathbf{x}, \mathbf{y}, \mathbf{w})$ denote the space of homology classes of Whitney triangles connecting $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{w}$.

Given $z \in \Sigma - \alpha_1 - ... - \alpha_g - \beta_1 - ... - \beta_g - \gamma_1 - ... - \gamma_g$, the algebraic intersection of a Whitney triangle with $\{z\} \times \mathrm{Sym}^{g-1}(\Sigma)$ descends to a well-defined map on homology classes

$$n_z \colon \pi_2(\mathbf{x}, \mathbf{y}, \mathbf{w}) \longrightarrow \mathbb{Z}.$$

This intersection number is additive in the following sense. Letting $\mathbf{x}' \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ and $\phi \in \pi_2(\mathbf{x}', \mathbf{x})$ and $\psi \in \pi_2(\mathbf{x}', \mathbf{y}, \mathbf{w})$, we can juxtapose ϕ and ψ to construct a new Whitney triangle $\psi * \phi \in \pi_2(\mathbf{x}', \mathbf{y}, \mathbf{w})$. Clearly,

$$n_z(\psi * \phi) = n_z(\psi) + n_z(\phi).$$

Also, if $n_z(\psi)$ is negative, then the homology class ψ supports no pseudo-holomorphic representative (for suitably chosen almost-complex structures).

Indeed, the homology class of a Whitney triangle $\psi \in \pi_2(\mathbf{x}, \mathbf{y}, \mathbf{w})$ determines a two-chain in Σ, just as homology classes of Whitney disks give rise to two-chains. The two-chain can be thought of as a sum of closures of the regions in

$$\Sigma - \alpha_1 - ... - \alpha_g - \beta_1 - ... - \beta_g - \gamma_1 - ... - \gamma_g,$$

where the multiplicity assigned to some region R is $n_x(\psi)$, where $x \in R$ is an interior point. We can generalize the notions of periodic domain and weak admissibility to this context:

Definition 2.2. A *triply-periodic domain* P for $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, z)$ is a two-chain whose local multiplicity at z is zero and whose boundary is a linear combination of one-cycles chosen from the α_i, β_j, and γ_k. The set of triply periodic domains is naturally an Abelian group, denoted $\mathcal{P}$.

Definition 2.3. A triple $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, z)$ is called *weakly admissible* if all the non-zero triply-periodic domains have both positive and negative local multiplicities.

Exercise 2.4. Suppose that Y is a rational homology three-sphere, and $K \subset Y$ is a knot with framing λ with the property that $Y_{\alpha\gamma}$ is a rational homology three-sphere. For the corresponding Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, z)$, find the dimension of the space of triply-periodic domains (in terms of the genus of Σ). Show that, after a sequence of isotopies, we can always arrange that this Heegaard triple is weakly admissible.

Suppose that $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, z)$ is weakly admissible. Then, we construct a map

$$\widehat{f}_{\alpha\beta\gamma} \colon \widehat{CF}(Y_{\alpha\beta}) \otimes \widehat{CF}(Y_{\beta\gamma}) \longrightarrow \widehat{CF}(Y_{\alpha\gamma})$$

by the formula

$$\widehat{f}_{\alpha\beta\gamma}(\mathbf{x} \otimes \mathbf{y}) = \sum_{\mathbf{w} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta} \sum_{\{\psi \in \pi_2(\mathbf{x},\mathbf{y},\mathbf{w}) \big| n_z(\psi)=0=\mu(\psi)\}} \#(\mathcal{M}(\psi)) \cdot \mathbf{w}. \tag{9}$$

Note that if $\psi, \psi' \in \pi_2(\mathbf{x}, \mathbf{y}, \mathbf{w})$ both have $n_z(\psi) = n_z(\psi') = 0$, then $\mathcal{D}(\psi) - \mathcal{D}(\psi')$ can be thought of as a triply-periodic domain. In fact, if $\pi_2(\mathbf{x}, \mathbf{y}, \mathbf{w})$ is non-empty, we can fix some $\psi_0 \in \pi_2(\mathbf{x}, \mathbf{y}, \mathbf{w})$ with $n_z(\psi) = 0$; then there is an isomorphism

$$\pi_2(\mathbf{x}, \mathbf{y}, \mathbf{w}) \cong \mathbb{Z} \oplus \mathcal{P},$$

defined by

$$\psi \mapsto n_z(\psi) \oplus (\mathcal{D}(\psi) - \mathcal{D}(\psi_0)).$$

It is not difficult to see that weak admissibility ensures that for any fixed $\mathbf{x}, \mathbf{y}, \mathbf{w}$, there are only finitely many $\psi \in \pi_2(\mathbf{x}, \mathbf{y}, \mathbf{w})$ with $n_z(\psi) = 0$ and $\mathcal{D}(\psi) \geq 0$. In particular this guarantees finiteness of the sum appearing in Equation (9).

Modifying the usual proof that $\partial^2 = 0$, we have the following:

Proposition 2.5. *The map $\widehat{f}_{\alpha\beta\gamma}$ defined above determines a chain map, where the tensor product appearing in the domain of Equation* (9) *is given its usual differential*

$$\partial(\mathbf{x} \otimes \mathbf{y}) = (\partial \mathbf{x}) \otimes \mathbf{y} + \mathbf{x} \otimes (\partial \mathbf{y}).$$

Sketch of Proof. The idea is to consider ends of one-dimensional moduli spaces of pseudo-holomorphic representatives of $\psi \in \pi_2(\mathbf{x}, \mathbf{y}, \mathbf{w})$. Such moduli spaces have three types of ends. For example, there is an end where a pseudo-holomorphic Whitney disk connecting $\mathbf{x}$ to $\mathbf{x}'$ (for some other $\mathbf{x}' \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$) is juxtaposed with a pseudo-holomorphic Whitney triangle connecting $\mathbf{x}', \mathbf{y}, \mathbf{w}$. The number of such ends corresponds to the $\mathbf{w}$-component of $\widehat{f}_{\alpha\beta\gamma}((\partial \mathbf{x}) \otimes \mathbf{y})$. There are two other types of ends, where Whitney disks bubble off at the $\mathbb{T}_\beta \cap \mathbb{T}_\gamma$ (representing the $\mathbf{w}$-component of $\widehat{f}_{\alpha\beta\gamma}(\mathbf{x} \otimes (\partial \mathbf{y})))$ resp. $\mathbb{T}_\alpha \cap \mathbb{T}_\gamma$ corner (representing the $\mathbf{w}$ component of $\partial \widehat{f}_{\alpha\beta\gamma}(\mathbf{x} \otimes \mathbf{y})$. □

In particular, we obtain an induced map on homology

$$\widehat{F}_{\alpha\beta\gamma} \colon \widehat{HF}(Y_{\alpha\beta}) \otimes \widehat{HF}(Y_{\beta\gamma}) \longrightarrow \widehat{HF}(Y_{\alpha\gamma}).$$

The maps induced by counting holomorphic triangles satisfy an associativity law, stating that if we start with four g-tuples of attaching circles $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$, then

$$\widehat{F}_{\alpha\gamma\delta}(\widehat{F}_{\alpha\beta\gamma}(\cdot\otimes\cdot)\otimes\cdot) = \widehat{F}_{\alpha\beta\delta}(\cdot\otimes\widehat{F}_{\beta\gamma\delta}(\cdot\otimes\cdot)) \tag{10}$$

as maps

$$\widehat{HF}(Y_{\alpha\beta})\otimes\widehat{HF}(Y_{\beta\gamma})\otimes\widehat{HF}(Y_{\gamma\delta}) \longrightarrow \widehat{HF}(Y_{\alpha\delta}).$$

We give a more precise version presently. Let $\square$ denote the "rectangle": unit disk whose boundary is divided into four arcs (topologically closed intervals) labeled e_α, e_β, e_γ, and e_δ (in clockwise order). The justification for calling this a rectangle is given in the following:

Exercise 2.6. Let $\square$ be any rectangle in the above sense, with the conformal structure induced from $\mathbb{C}$. Show that there is a pair of real numbers w and h and a unique holomorphic identification

$$\theta\colon \square \longrightarrow [0,w]\times[0,h]$$

carrying e_α to $[0,w]\times\{h\}$, e_β to $\{0\}\times[0,h]$, e_γ to $[0,w]\times\{0\}$, and e_δ to $\{w\}\times[0,h]$. Indeed, the ratio w/h is uniquely determined by the conformal structure on $\square$.

The above exercise can be interpreted in the following manner: the space of conformal structures on $\square$ is identified with $\mathbb{R}$ under the map which takes a fixed conformal structure to the real number $\log(w)-\log(h)$ in the above uniformization.

A map $\varphi\colon \square \longrightarrow \mathrm{Sym}^g(\Sigma)$ which carries e_α, e_β, e_γ, and e_δ to $\mathbb{T}_\alpha$, $\mathbb{T}_\beta$, $\mathbb{T}_\gamma$, and $\mathbb{T}_\delta$ respectively is called a *Whitney rectangle*. For fixed $\mathbf{x}\in\mathbb{T}_\alpha\cap\mathbb{T}_\beta$, $\mathbf{y}\in\mathbb{T}_\beta\cap\mathbb{T}_\gamma$, $\mathbf{w}\in\mathbb{T}_\gamma\cap\mathbb{T}_\delta$, and $\mathbf{p}\in\mathbb{T}_\alpha\cap\mathbb{T}_\delta$, spaces of Whitney rectangles can be collected into homology classes $\pi_2(\mathbf{x},\mathbf{y},\mathbf{w},\mathbf{p})$. Let $\mathcal{M}(\varphi)$ denote the moduli space of pseudo-holomorphic representatives of φ, with respect to any conformal structure on the domain $\square$.

Given a pointed Heegaard quadruple $(\Sigma,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},z)$, define a map

$$\widehat{h}_{\alpha\beta\gamma\delta}\colon \widehat{CF}(Y_{\alpha\beta})\otimes\widehat{CF}(Y_{\beta\gamma})\otimes\widehat{CF}(Y_{\gamma\delta}) \longrightarrow \widehat{CF}(Y_{\alpha\delta})$$

by the formula

$$\widehat{h}_{\alpha\beta\gamma\delta}(\mathbf{x}\otimes\mathbf{y}\otimes\mathbf{w}) = \sum_{\mathbf{p}\in\mathbb{T}_\alpha\cap\mathbb{T}_\delta}\ \sum_{\{\varphi\in\pi_2(\mathbf{x},\mathbf{y},\mathbf{w},\mathbf{p})\,\big|\,\mu(\varphi)=-1,n_z(\varphi)=0\}} \#\mathcal{M}(\varphi)\cdot\mathbf{p}.$$

Again, to ensure that this formula has the required finiteness properties, we need the quadruple to satisfy a weak admissibility hypothesis analogous to Definition 2.3.

Theorem 2.7. *The map $\widehat{h}_{\alpha\beta\gamma\delta}$ determines a chain homotopy between the maps*

$$\widehat{f}_{\alpha\gamma\delta}(\widehat{f}_{\alpha\beta\gamma}(\cdot\otimes\cdot)\otimes\cdot) \quad \textit{and} \quad \widehat{f}_{\alpha\beta\delta}(\cdot\otimes\widehat{f}_{\beta\gamma\delta}(\cdot\otimes\cdot));$$

i.e. for all $\xi\in\widehat{CF}(Y_{\alpha\beta})$, $\eta\in\widehat{CF}(Y_{\beta\gamma})$, $\zeta\in\widehat{CF}(Y_{\gamma\delta})$, we have that

$$\partial\widehat{h}_{\alpha\beta\gamma\delta}(\xi\otimes\eta\otimes\zeta)+\widehat{h}_{\alpha\beta\gamma\delta}(\partial(\xi\otimes\eta\otimes\zeta)) = \widehat{f}_{\alpha\gamma\delta}(\widehat{f}_{\alpha\beta\gamma}(\xi\otimes\eta)\otimes\zeta)+\widehat{f}_{\alpha\beta\delta}(\xi\otimes\widehat{f}_{\beta\gamma\delta}(\eta\otimes\zeta)).$$

Sketch of Proof. We wish to consider moduli spaces of pseudo-holomorphic Whitney rectangles with formal dimension one. Some ends of these moduli spaces are modeled on flowlines breaking off at the corners, but there is another type of end not encountered before in the counts of trianges, arising from the non-compactness of $\mathcal{M}(\square)\cong\mathbb{R}$. As this parameter goes to $\pm\infty$, the corresponding rectangle breaks up conformally into a pair of triangles meeting at a vertex (in two different ways, depending on which end we are considering), as illustrated in Figure 4. This is how a count of holomorphic squares induces a chain homotopy between two different compositions of holomorphic triangle counts. □

For more details on the associativity in Lagrangian Floer homology, compare [**32**] [**5**] [**19**].

EXERCISE 2.8. Deduce Equation (10) from Theorem 2.7.

2.2. Maps in the exact sequence. We are now ready to define the maps appearing in the exact sequence. Since $Y_{\beta\gamma} \cong \#^{g-1}(S^2 \times S^1)$, we have that $\widehat{HF}(Y_{\beta\gamma}) \cong \Lambda^* H^1(\#^{g-1}(S^2 \times S^1); \mathbb{Z}/2\mathbb{Z})$. As such, its top-dimensional group is isomorphic to $\mathbb{Z}/2\mathbb{Z}$. Let $\widehat{\Theta}_{\beta\gamma}$ denote this generator. The map $\widehat{F}$ is defined by

$$\widehat{F}(\xi) = \widehat{F}_{\alpha\beta\gamma}(\xi \otimes \widehat{\Theta}_{\beta\gamma}).$$

In fact, we can exhibit a Heegaard diagram for $\#^{g-1}(S^2 \times S^1)$ for which all the differentials are trivial, and hence $\widehat{\Theta}_{\beta\gamma}$ is represented by an intersection point of $\mathbb{T}_\beta \cap \mathbb{T}_\gamma$. By a slight abuse of notation, we also denote this intersection point by the symbol $\widehat{\Theta}_{\beta\gamma}$.

We can define the maps $\widehat{F}_0$ and $\widehat{F}_1$ analogously;

$$\widehat{F}_0(\eta) = \widehat{F}_{\alpha\gamma\delta}(\eta \otimes \widehat{\Theta}_{\gamma\delta}), \quad \text{and} \quad \widehat{F}_1(\zeta) = \widehat{F}_{\alpha\delta\beta}(\zeta \otimes \widehat{\Theta}_{\delta\beta}),$$

where here $\widehat{\Theta}_{\gamma\delta}$ and $\widehat{\Theta}_{\delta\beta}$ are generators for the top-dimensional non-trivial groups in $\widehat{HF}(Y_{\gamma\delta}) \cong \widehat{HF}(Y_{\delta\beta}) \cong \Lambda^* H^1(S^2 \times S^1)$.

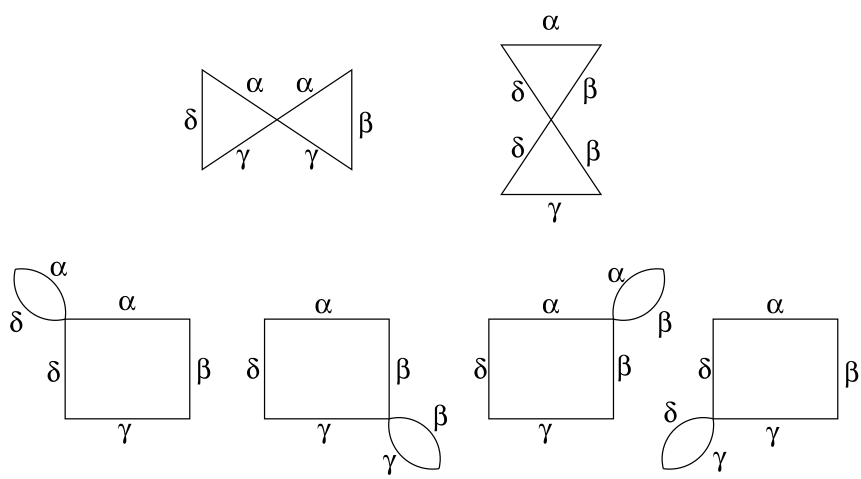


FIGURE 4. Degenerate rectangles. We have illustrated here a schematic diagram for the degenerations of pseudo-holomorphic rectangles. Edges are marked with the corresponding torus they are mapped to. (Conformal moduli for rectangles appearing in this figure are arbitrary.)

To prove Theorem 1.12, we must verify that Ker $\widehat{F}_0 = \text{Im } \widehat{F}$. As a first step, we would like to prove that Im $\widehat{F}_0 \subseteq$ Ker $\widehat{F}$, i.e. $\widehat{F}_0 \circ \widehat{F} = 0$. To this end, note that

$$\widehat{F}_0 \circ \widehat{F}(\xi) = \widehat{F}_{\alpha\gamma\delta}(\widehat{F}_{\alpha\beta\gamma}(\xi \otimes \widehat{\Theta}_{\beta\gamma}) \otimes \widehat{\Theta}_{\gamma\delta}) = \widehat{F}_{\alpha\beta\delta}(\xi \otimes \widehat{F}_{\beta\gamma\delta}(\widehat{\Theta}_{\beta\gamma} \otimes \widehat{\Theta}_{\gamma\delta})), \tag{11}$$

so it suffices to prove that

$$\widehat{F}_{\beta\gamma\delta}(\widehat{\Theta}_{\beta\gamma} \otimes \widehat{\Theta}_{\gamma\delta}) = 0,$$

which in turn follows from a model calculation.

Exercise 2.9. Let β, γ, δ be three straight curves in the torus Σ as above, and let $\widehat{\Theta}_{\beta\gamma}$, $\widehat{\Theta}_{\gamma\delta}$, $\widehat{\Theta}_{\delta\beta}$ denote the three intersection points. Prove that $\pi_2(\widehat{\Theta}_{\beta\gamma}, \widehat{\Theta}_{\gamma\delta}, \widehat{\Theta}_{\delta\beta}) = \{\psi_k^{\pm}\}_{k=1}^{\infty}$, where $\mu(\psi_k^{\pm}) = 0$, $n_z(\psi_k^{\pm}) = \frac{k(k-1)}{2}$, and each $\psi_k^{\pm}$ has a unique holomorhpic representative. *Hint:* Lift to the universal cover of Σ.

Proposition 2.10. *The are exactly two homology classes of $\psi \in \pi_2(\widehat{\Theta}_{\beta\gamma}, \widehat{\Theta}_{\gamma\delta}, \widehat{\Theta}_{\delta\beta})$ with $\mathcal{D}(\psi) \geq 0$, $n_z(\psi) = 0$, and $\mu(\psi) = 0$. For either homology class ψ,*

$$\#\mathcal{M}(\psi) \equiv 1 \pmod 2.$$

Proof. In the case where $g(\Sigma) = 1$, we appeal to Exercise 2.9.

In the general case, we can decompose the Heegaard surface $\Sigma = E_1 \# ... \# E_g$ as a connected sum of g tori, with each β_i, γ_i, and δ_i supported inside E_i. For $i > 1$, the summand E_i with its three curves is homeomorphic to the one pictured in Figure 5, while for E_1, it is the case considered in Exercise 2.9. In this case, any homology class $\psi \in \pi_2(\widehat{\Theta}_{\beta\gamma}, \widehat{\Theta}_{\gamma\delta}, \widehat{\Theta}_{\delta\beta})$ with $n_z(\psi) = 0$ decomposes as a product of homology classes $\psi_i \in \pi_2(\beta_i, \gamma_i, \delta_i)$ for E_i. It is easy to see that there are precisely two homology class ψ in $\pi_2(\widehat{\Theta}_{\beta\gamma}, \widehat{\Theta}_{\gamma\delta}, \widehat{\Theta}_{\delta\beta})$ with $\mathcal{D}(\psi) \geq 0$ and $n_z(\psi) = 0$. These homology classes are obtained by taking the product of $g - 1$ copies of the distinguished homology classes from Figure 5 in the E_i summand for $i > 1$, and one copy of either ψ_1^+ or ψ_1^- from Exercise 2.9 in the E_1 summand. ☐

The fact that

$$\widehat{f}_{\beta\gamma\delta}(\widehat{\Theta}_{\beta\gamma} \otimes \widehat{\Theta}_{\gamma\delta}) = 0 \tag{12}$$

is a quick consequence of Proposition 2.10. Thus, from the associativity of the maps induced by holomorphic triangles (cf. Equation (11) above), it follows that $\widehat{F}_0 \circ \widehat{F} = 0$. The other double composites $\widehat{F}_1 \circ \widehat{F}_0$ and $\widehat{F} \circ \widehat{F}_1$ also vanish by a symmetrical argument.

Thus, we have verified that the sequence of maps on $\widehat{HF}$ appearing in the statement of Theorem 1.12 form a chain complex. It remains to verify that the chain complex has trivial homology. To this end, we find it useful to make a digression into some homological algebra.

2.3. Some homological algebra.

We begin with some terminology.

Let A_1 and A_2 be a pair of chain complexes of vector spaces over the field $\mathbb{Z}/2\mathbb{Z}$ (though the discussion here could again be given over $\mathbb{Z}$, with more attention paid to signs). A chain map

$$\phi \colon A_1 \longrightarrow A_2$$

is called a *quasi-isomorphism* if the induced map on homology is an isomorphism.

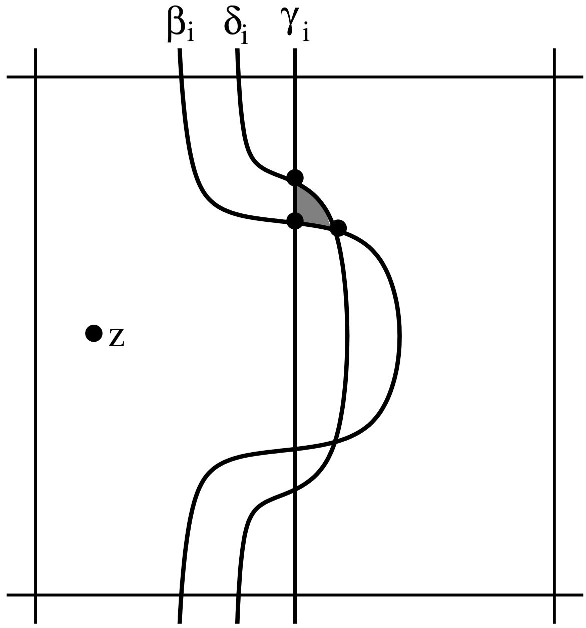


Figure 5. Other factors of the holomorphic triangle. We have illustrated here a Heegaard triple, where γ_i, β_i and δ_i are small isotopic translates of one another. The unique homology class of triangles $\pi_2(\widehat{\Theta}_{\beta\gamma}, \widehat{\Theta}_{\gamma\delta}, \widehat{\Theta}_{\delta\beta})$ with $n_z(\psi) = 0$ and $\mathcal{D}(\psi) \geq 0$ is indicated by the shading.

Recall that if we have a chain map between chain complexes $f_1 \colon A_1 \longrightarrow A_2$, we can form its mapping cone $M(f_1)$, whose underlying module is the direct sum $A_1 \oplus A_2$, endowed with the differential

$$\partial = \begin{pmatrix} \partial_1 & 0 \\ f_1 & \partial_2 \end{pmatrix},$$

where here ∂_i denotes the differential for the chain complex A_i. Recall that there is a short exact sequence of chain complexes

$$0 \longrightarrow A_2 \xrightarrow{\iota} M(f_1) \xrightarrow{\pi} A_1 \longrightarrow 0. \tag{13}$$

Exercise 2.11. Show that the short exact sequence from Equation (13) induces a long exact sequence in homology, for which the connecting homomorphism is the map on homology induced by f_1.

Exercise 2.12. Verify naturality of the mapping cylinder in the following sense. Suppose that we have a diagram of chain complexes

$$\begin{tikzcd} A_1 \arrow[r, "f_1"] \arrow[d, "\psi_1"'] & A_2 \arrow[d, "\psi_2"] \\ B_1 \arrow[r, "g_1"] & B_2 \end{tikzcd}$$

which commutes up to homotopy; then there is an induced map

$$m(\psi_1,\psi_2)\colon M(f_1) \longrightarrow M(g_1)$$

which fits into the following diagram, where the rows are exact and the squares are homotopy-commutative:

$$\begin{tikzcd}
0 \arrow[r] & A_2 \arrow[r] \arrow[d, "\psi_2"'] & M(f_1) \arrow[r] \arrow[d, "m(\psi_1,\psi_2)"'] & A_1 \arrow[r] \arrow[d, "\psi_1"] & 0 \\
0 \arrow[r] & B_2 \arrow[r] & M(g_1) \arrow[r] & B_1 \arrow[r] & 0.
\end{tikzcd}$$

LEMMA 2.13. *Let* $\{A_i\}_{i=1}^{\infty}$ *be a collection of chain maps and let*

$$\{f_i\colon A_i \longrightarrow A_{i+1}\}_{i\in\mathbb{Z}}$$

be a collection of chain maps satisfying the following two properties:

(1) $f_{i+1}\circ f_i$ *is chain homotopically trivial, by a chain homotopy*

$$H_i\colon A_i \longrightarrow A_{i+2}$$

(2) *the map*

$$\psi_i = f_{i+2}\circ H_i + H_{i+1}\circ f_i\colon A_i \longrightarrow A_{i+3}$$

is a quasi-isomorphism.

Then, $H_*(M(f_i)) \cong H_*(A_{i+2})$.

EXERCISE 2.14. Show that the hypotheses of Lemma 2.13 imply that ψ_i is a chain map. Then supply a proof of Lemma 2.13. *Hint:* Construct chain maps $M(f_i) \longrightarrow A_{i+2}$ and $A_i \longrightarrow M(f_{i+1})$, and use the five-lemma to prove that they induce isomorphisms on homology.

2.4. Completion of the proof of Theorem 1.12 for $\widehat{HF}$ with $\mathbb{Z}/2\mathbb{Z}$ coefficents. Continuing notation from before, let $Y_{\alpha\beta}$, $Y_{\alpha\gamma}$ $Y_{\alpha\delta}$ describe Y, Y_0, and Y_1 respectively, and so that the remaining three-manifolds $Y_{\beta\gamma}$, $Y_{\gamma\delta}$, $Y_{\delta\beta}$ describe $\#^{g-1}(S^2\times S^1)$. Indeed, to fit precisely with the hypotheses of Lemma 2.13, we choose infinitely many copies of the g-tuples $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$ (denoted $\boldsymbol{\beta}^{(i)}$, $\boldsymbol{\gamma}^{(i)}$, $\boldsymbol{\delta}^{(i)}$ for $i\in\mathbb{Z}$), all of which are generic exact Hamiltonian perturbations of one another.

Let A_{3i+1}, A_{3i+2} and A_{3i+3} represent $\widehat{CF}(Y_0)$, $\widehat{CF}(Y_1)$ and $\widehat{CF}(Y)$ respectively, only now we use the various translates of the $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, and $\boldsymbol{\beta}$; in particular A_{3i+1} is the Floer complex $\widehat{CF}(Y_{\alpha\gamma^{(i)}})$.

We have already verified Hypothesis (1) in the discussion of Subsection 2.2. For example, the null-homotopy $H_i\colon A_{3i} \longrightarrow A_{3i+2}$ is given by the map

$$H_i(\xi) = \widehat{h}_{\alpha,\beta^{(i)},\gamma^{(i)},\delta^{(i)}}(\xi\otimes\widehat{\Theta}_{\beta^{(i)}\gamma^{(i)}}\otimes\widehat{\Theta}_{\gamma^{(i)}\delta^{(i)}})$$

gotten by counting pseudo-holomorphic rectangles.

It remains to verify Hypothesis (2) of Lemma 2.13. It is useful to have the following:

DEFINITION 2.15. An $\mathbb{R}$-filtration of a group G is a sequence of subgroups indexed by $r\in\mathbb{R}$, so that

- $G_r \subseteq G_s$ if $r\leq s$ and
- $G = \cup_{r\in\mathbb{R}} G_r$.

This induces a partial ordering on G. If $x, y \in G$, we say $x < y$ if $x \in G_r$, while $y \notin G_r$.

DEFINITION 2.16. The *area filtration* on $\widehat{CF}(Y_{\alpha\beta})$ is the $\mathbb{R}$-filtration defined as follows. Fix $\mathbf{x}_0 \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$, and define a function

$$\mathcal{F}\colon \mathbb{T}_\alpha \cap \mathbb{T}_\beta \longrightarrow \mathbb{R}$$

gotten by taking

$$\mathcal{F}(\mathbf{x}) = \mathcal{A}(\mathcal{D}(\phi)) - 2n_z(\phi)\cdot\mathcal{A}(\Sigma),$$

where here $\phi \in \pi_2(\mathbf{x}_0, \mathbf{x})$ is any homotopy class connecting $\mathbf{x}$ and $\mathbf{y}$, $\mathcal{A}(R)$ denotes the signed area of some region R in Σ, with respect to a fixed area form over Σ.

In the case where $b_1(Y_{\alpha\beta}) > 0$, in order for the area filtration to be well-defined, we must use an area form over Σ with the property that $\mathcal{A}(P) = 0$ for each periodic domain. Such an area form can be found for any weakly admissible diagram.

LEMMA 2.17. *If $\boldsymbol{\beta}'$ is a sufficiently small perturbation of $\boldsymbol{\beta}$, and $\widehat{\Theta}_{\beta\beta'}$ denotes the canonical top-dimensional homology class in $\widehat{HF}(Y_{\beta\beta'})$, then the chain map*

$$\widehat{CF}(Y_{\alpha\beta}) \longrightarrow \widehat{CF}(Y_{\alpha\beta'})$$

defined by

$$\xi \mapsto \widehat{f}_{\alpha\beta\beta'}(\xi \otimes \widehat{\Theta}_{\beta\beta'})$$

induces an isomorphism in homology.

PROOF. We perform the perturbation so that β_i' and β_i intersect transversally in two points, and indeed, so that the signed area of the region between β_i and β_i' is zero.

If each β_i' is sufficiently close to the corresponding β_i, then for each $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$, there is a corresponding closest point $\mathbf{x}' \in \mathbb{T}_\alpha \cap \mathbb{T}_{\beta'}$. This closest point map induces a group isomorphism

$$\iota\colon \widehat{CF}(Y_{\alpha\beta}) \longrightarrow \widehat{CF}(Y_{\alpha\beta'}).$$

Note that for each $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$, there is a canonical smallest triangle $\psi \in \pi_2(\mathbf{x}, \widehat{\Theta}_{\beta\beta'}, \mathbf{x}')$ which admits a unique holomorphic representative (by the Riemann mapping theorem). By taking sufficiently nearby translates β_i' of the β_i, we can arrange for the area of this triangle to be smaller than the areas of any homotopy classes of $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$ for any $\mathbf{x}$ and $\mathbf{y}$ either in $\mathbb{T}_\alpha \cap \mathbb{T}_\beta$ or in $\mathbb{T}_\alpha \cap \mathbb{T}_{\beta'}$.

This map perhaps does not quite agree with the chain map

$$f(\xi) = \widehat{f}_{\alpha\beta\beta'}(\xi \otimes \widehat{\Theta}_{\beta\beta'}).$$

However, for each $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$, the element $f(\mathbf{x}) - \iota(\mathbf{x})$ can be written as a linear combination of $\mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_{\beta'}$, with $\mathcal{F}(\iota(\mathbf{x})) < \mathcal{F}(\mathbf{y})$ with respect to the area filtration of $\widehat{CF}(Y_{\alpha\beta'})$. Since ι induces an isomorphism on the group level, it is easy to see that f induces an isomorphism on the group level as well. Since f is also a chain map, it follows that it induces an isomorphism of chain complexes. □

Let $\theta_i\colon A_i \longrightarrow A_{i+3}$ be the quasi-isomorphisms defined as in the above lemma; e.g. θ_{3i+1} is the chain map $\widehat{CF}(Y_{\alpha\gamma^{(i)}}) \longrightarrow \widehat{CF}(Y_{\alpha\gamma^{(i+1)}})$ obtained by product with the canonical generator $\widehat{\Theta}_{\gamma^{(i)}\gamma^{(i+1)}}$.

We claim that

$$f_3 \circ H_1 + H_2 \circ f_1\colon A_1 \longrightarrow A_4$$

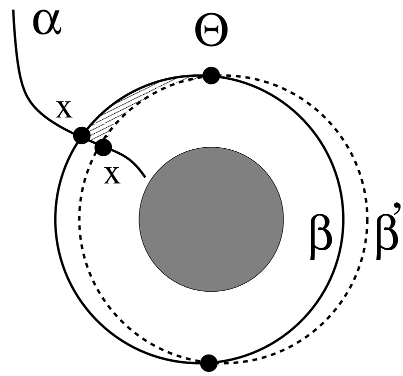


FIGURE 6. Small triangles. In the proof of Lemma 2.17, we let β_i' be a nearby isotopic translate of β_i, arranged so that the two curves meet transversally in two points. The top-dimensional generator of $\mathbb{T}_\beta \cap \mathbb{T}_{\beta'}$ is represented by the product of intersection points $\Theta_1 \times ... \times \Theta_g = \widehat{\Theta}$. Any intersection point $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ has a nearest intersection point $\mathbf{x}' \in \mathbb{T}_\alpha \cap \mathbb{T}_{\beta'}$; and there is a canonical homology class of smallest triangle $\psi \in \pi_2(\mathbf{x}, \widehat{\Theta}, \mathbf{x}')$ which supports a unique holomorphic representative. We have illustrated here an annular region in Σ (i.e. delete the shaded circle from the picture) which is a neighborhood of β_i, though we have dropped the subscripts. Θ, x, and x' represent the corresponding factor of $\widehat{\Theta}$, $\mathbf{x}$, and $\mathbf{x}'$ respectively, and the hatched region illustrates part of the region of the canonical smallest triangle.

is chain homotopic to θ_1. More precisely, counting pseudo-holomorphic pentagons with edges on $\mathbb{T}_\alpha$, $\mathbb{T}_\gamma$, $\mathbb{T}_\delta$ $\mathbb{T}_\beta$, $\mathbb{T}_\gamma^{(1)}$ can be used to give a homotopy to prove a generalized associativity law analogous to Theorem 2.7; i.e. looking at ends of one-dimensional moduli spaces of pseudo-holomorphic pentagons, we get a null-homotopy of the sum of composite maps:

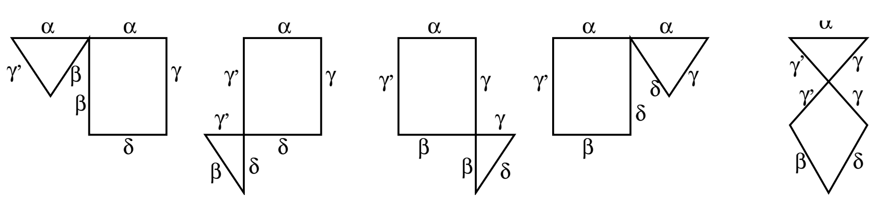


FIGURE 7. Degenerate pentagons. We have illustrated here a schematic diagram for the degenerations of pseudo-holomorphic pentagons. We have dropped the five additional degenerations, where a Whitney disk bubbles off the vertex of any pentagon.

$$
\begin{aligned}
(14) \qquad & \widehat{f}_{\alpha\beta,\gamma^{(1)}}(\widehat{h}_{\alpha\gamma\delta\beta}(\xi\otimes\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta})\otimes\widehat{\Theta}_{\beta\gamma^{(1)}}) \\
& + \quad \widehat{h}_{\alpha\gamma\delta\gamma^{(1)}}(\xi\otimes\widehat{\Theta}_{\gamma\delta}\otimes\widehat{f}_{\delta\beta\gamma^{(1)}}(\widehat{\Theta}_{\delta\beta}\otimes\widehat{\Theta}_{\beta\gamma^{(1)}})) \\
& + \quad \widehat{h}_{\alpha\gamma\beta\gamma^{(1)}}(\xi\otimes\widehat{f}_{\gamma\delta\beta}(\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta})\otimes\widehat{\Theta}_{\beta\gamma^{(1)}}) \\
& + \quad \widehat{h}_{\alpha\delta\beta\gamma^{(1)}}(\widehat{f}_{\alpha\gamma\delta}(\xi\otimes\widehat{\Theta}_{\gamma\delta})\otimes\widehat{\Theta}_{\delta\beta}\otimes\widehat{\Theta}_{\beta\gamma^{(1)}}) \\
& + \quad \widehat{f}_{\alpha\gamma\gamma^{(1)}}(\xi\otimes\widehat{h}_{\gamma\delta\beta\gamma^{(1)}}(\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta}\otimes\widehat{\Theta}_{\beta\gamma^{(1)}})).
\end{aligned}
$$

This sum is more graphically illustrated in Figure 7. Two of these terms vanish, since

$$\widehat{f}_{\delta\beta\gamma^{(1)}}(\widehat{\Theta}_{\delta\beta}\otimes\widehat{\Theta}_{\beta\gamma^{(1)}}) = 0 = \widehat{f}_{\gamma\delta\beta}(\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta}).$$

The first and fourth terms are identified with $f_3\circ H_1 + H_2\circ f_1$. To see that the final term is identified with θ_1, it suffices to show that

$$\widehat{h}_{\gamma\delta\beta\gamma^{(1)}}(\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta}\otimes\widehat{\Theta}_{\beta\gamma^{(1)}}) = \widehat{\Theta}_{\gamma\gamma^{(1)}}. \tag{15}$$

This latter equality follows from a direct inspection of the Heegaard diagram for the quadruple $(\Sigma,\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\beta},\boldsymbol{\gamma}^{(1)},z)$. (i.e. the count of pseudo-holomorphic quadrilaterals), as illustrated in Figures 8 and 9.

In Figure 8, we consider the special case where the genus $g=1$. In the picture, and in the following discussion, $\gamma_1^{(1)}$ is denoted γ_1'. The four corners of the shaded quadrilateral are the canonical generators $\widehat{\Theta}_{\gamma_1,\delta_1}$, $\widehat{\Theta}_{\delta_1,\beta_1}$, $\widehat{\Theta}_{\beta_1,\gamma_1'}$, and $\widehat{\Theta}_{\gamma_1',\gamma_1}$ (read

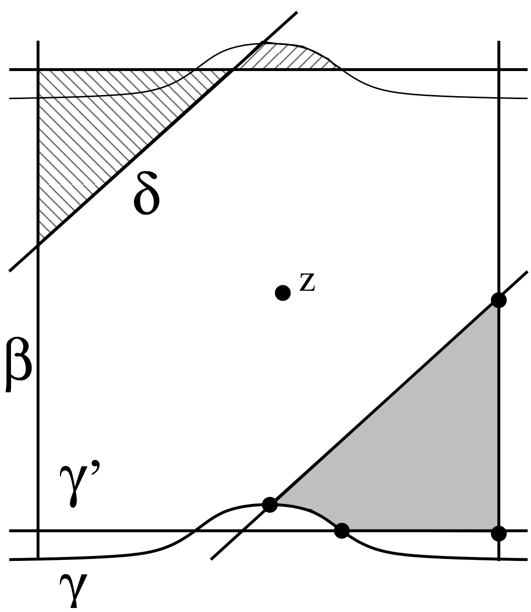


FIGURE 8. A holomorphic quadrilateral. The shaded quadrilateral has a unique holomorphic representative (by the Riemann mapping theorem), while the one indicated with the hatching does not, as it has both positive and negative local multiplicities, as indicated by the two directions in the hatching.

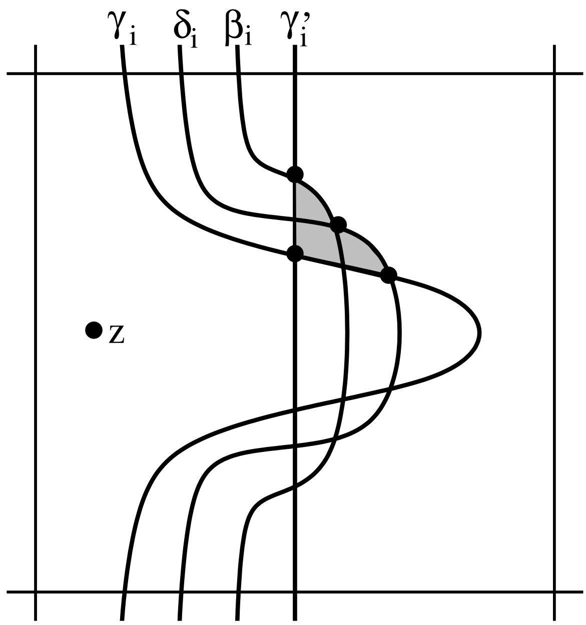


Figure 9. Other factors of the holomorphic quadrilateral. We have illustrated here a Heegaard quadruple (in a genus one surface) whose four boundary components are $S^2 \times S^1$. In the homology class indicated by the shaded quadrilateral $\varphi_i \in \pi_2(\widehat{\Theta}_{\gamma\delta}, \widehat{\Theta}_{\delta\beta}, \widehat{\Theta}_{\beta\gamma'}, \widehat{\Theta}_{\gamma'\gamma})$, there is a moduli space of pseudo-holomorphic quadrilaterals which is clearly one-dimensional, parameterized by a cut at the vertex where γ_i and δ_i meet. We take the connected sum of $g-1$ copies of this picture (at the reference point z) with the picture illustrated in Figure 8 to obtain the general case of the quadrilateral considered in the proof of Theorem 1.12.

in clockwise order). Indeed, it is straightforward to see (by passing to the universal cover), that the shaded quadrilateral represents the only homology class φ_1 of Whitney quadrilaterals connecting these four points with $n_z(\varphi_1) = 0$ and all of whose local multiplicities are non-negative. By the Riemann mapping theorem, now, this homology class φ_1 has a unique holomorphic representative u_1. (By contrast, we have also pictured here another Whitney quadrilateral with hatchings, whose local multiplicities are all 0, +1, and −1; +1 at the region where the hatchings go in one direction and −1 where they go in the other.)

For the general case ($g > 1$), we take the connected sum of the case illustrated in Figure 8 with $g-1$ copies of the torus illustrated in Figure 9. In this picture, we have illustrated the four curves γ_i, δ_i, β_i, γ_i' for $i > 1$, which are Hamiltonian translates of one another. Now, there is a homology class of quadrilateral $\varphi_i \in \pi_2(\widehat{\Theta}_{\gamma_i\delta_i}, \widehat{\Theta}_{\delta_i\beta_i}, \widehat{\Theta}_{\beta_i\gamma_i'}, \widehat{\Theta}_{\gamma_i'\gamma_i})$, and a forgetful map $\mathcal{M}(\varphi) \longrightarrow \mathcal{M}(\square)$ which remembers only the conformal class of the domain (where here $\mathcal{M}(\square)$ denotes the moduli space of rectangles, cf. Exercise 2.6). Both moduli spaces are one-dimensional (the first moduli space is parameterized by the length of the cut into the region, while the second is parameterized by the ratio of the length to the width, after the quadrilateral is uniformized to a rectangle, as in Exercise 2.6). By Gromov's compactness theorem, the forgetful map is proper; and it is easy to see that it has degree one, and hence for some generic conformal class of quadrilateral, there is an odd number of pseudo-holomorphic quadrilaterals appearing in this family whose domain has the specified conformal class. Then the holomorphic quadrilaterals in $\pi_2(\widehat{\Theta}_{\gamma\delta}, \widehat{\Theta}_{\delta\beta}, \widehat{\Theta}_{\beta\gamma'}, \widehat{\Theta}_{\gamma'\gamma})$ are easily seen to be those quadrilaterals of the form $u_1 \times ... \times u_g \in \varphi_1 \times ... \times \varphi_g$ where u_1 is the pseudo-holomorphic representative of the homology class φ_1 described in the previous paragraph, and u_i for $i > 1$ are pseudo-holomorphic representatives for φ_i whose domain supports the same conformal class. This proves Equation (15) which, in turn, yields Hypothesis (2) of Lemma 2.13. The surgery exact triangle for $\widehat{HF}$ (calculated with coefficients in $\mathbb{Z}/2\mathbb{Z}$) stated in Theorem 1.12 now follows directly from Lemma 2.13.

2.5. The case of HF^+. We outline here the modification necessary to adapt the above discussion to the case of HF^+ rather than $\widehat{HF}$.

First we define a map

$$f^+_{\alpha\beta\gamma} \colon CF^+(Y_{\alpha\beta}) \otimes_{\mathbb{F}[U]} CF^-(Y_{\beta\gamma}) \longrightarrow CF^+(Y_{\alpha\gamma})$$

by extending the following map to be U-equivariant:

$$f^+_{\alpha\beta\gamma}(U^{-i}\mathbf{x} \otimes \mathbf{y}) = \sum_{\mathbf{w} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta} \sum_{\{\psi \in \pi_2(\mathbf{x},\mathbf{y},\mathbf{w}) \big| 0 = \mu(\psi)\}} \#(\mathcal{M}(\psi)) U^{n_z(\psi)-i} \cdot \mathbf{w}. \tag{16}$$

The fact that f^+ determines a chain map follows from a suitable adaptation of the proof of Proposition 2.5, together with the additivity of n_z under juxtapositions. Finiteness of the sum is a consequence of the admissibility condition: given any integer i, there are only finitely many $\psi \in \pi_2(\mathbf{x}, \mathbf{y}, \mathbf{w})$ with $\mathcal{D}(\psi) \geq 0$ and $n_z(\psi) \leq i$.

To define the maps appearing in the exact sequence, we use the fact that $HF^-(\#^{g-1}(S^2 \times S^1)) \cong \Lambda^* H^1(\#^{g-1}(S^2 \times S^1)) \otimes \mathbb{F}[U]$. Again, we take top-dimensional generators $\Theta_{\beta\gamma}$, $\Theta_{\gamma\delta}$ and $\Theta_{\delta\beta}$ for these groups. Now, define

$$f^+(\xi) = f^+_{\alpha\beta\gamma}(\xi \otimes \Theta_{\beta\gamma}), \qquad f^+_0(\xi) = f^+_{\alpha\beta\gamma}(\xi \otimes \Theta_{\gamma\delta}), \qquad f^+_1(\xi) = f^+_{\alpha\beta\gamma}(\xi \otimes \Theta_{\delta\beta}).$$

It is not difficult to see that these maps are consistent with the earlier maps defined on $\widehat{CF}$, in the sense that the following diagram commutes:

$$\begin{tikzcd}
0 \arrow[r] & \widehat{CF}(Y) \arrow[r] \arrow[d, "\widehat{f}"'] & CF^+(Y) \arrow[r, "U"] \arrow[d, "f^+"'] & CF^+(Y) \arrow[r] \arrow[d, "f^+"] & 0 \\
0 \arrow[r] & \widehat{CF}(Y_0) \arrow[r] & CF^+(Y_0) \arrow[r, "U"] & CF^+(Y_0) \arrow[r] & 0
\end{tikzcd}$$

As before, we have an associativity law, according to which

$$f^+_{\alpha\gamma\delta}(f^+_{\alpha\beta\gamma}(\xi \otimes \Theta_{\beta\gamma})) \simeq f^+_{\alpha\beta\delta}(\xi \otimes f^-_{\beta\gamma\delta}(\Theta_{\beta\gamma} \otimes \Theta_{\gamma\delta})), \tag{17}$$

where here $f^-_{\beta\gamma\delta}$ is also obtained by counting holomorphic triangles; e.g.

$$f^-_{\alpha\beta\gamma}(\mathbf{x} \otimes \mathbf{y}) = \sum_{\mathbf{w} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta} \sum_{\{\psi \in \pi_2(\mathbf{x},\mathbf{y},\mathbf{w}) \big| \mu(\psi) = 0\}} \#(\mathcal{M}(\psi)) \cdot U^{n_z(\psi)} \cdot \mathbf{w}. \tag{18}$$

However, unlike the case of f^+, there is no longer an *a priori* finiteness statement for the number of terms on the right-hand-side (even in the presence of weak admissibility). One way of coping with this issue is to consider yet another variant of Heegaard Floer homology $CF^{--}(Y_{\beta\delta})$, where we take our coefficient ring to be formal power series in U, $\mathbb{F}[[U]]$. The map f^+ defined in Equation (16) readily extends to a map

$$CF^+(Y_{\alpha\beta}) \otimes CF^{--}(Y_{\beta\gamma}) \longrightarrow CF^+(Y_{\alpha\gamma}),$$

and now the map $f^-_{\alpha\beta\gamma}$ as defined in Equation (18) gives a well-defined map

$$CF^{--}(Y_{\alpha\beta}) \otimes CF^{--}(Y_{\beta\gamma}) \longrightarrow CF^{--}(Y_{\alpha\gamma})$$

(since the sum is no longer required to be finite). In this setting the desired homotopy associativity stated in Equation (10) holds.

In view of the above remarks, in order to verify that $F^+_{W_0} \circ F^+_W = 0$, we must prove that

$$F^-_{\beta\gamma\delta}(\Theta_{\beta\gamma} \otimes \Theta_{\gamma\delta}) = 0,$$

which in turn hinges on a generalization of Proposition 2.10. In turn, this generalization relies on a "stretching the neck" argument familiar from gauge theory and symplectic geometry. In the context of symplectic geometry, this means that in order to analyze holomorphic curves in a symplectic manifold, it is sometimes useful to degenerate the almost-complex structure, so that the space becomes singular, and the holomorphic curves localize into strata which are easier to understand. Such an argument has already appeared in the proof of stabilization invariance of HF^+ (cf. [**44**]). We cannot treat this discussion in any detail here, but rather refer the interested reader to Section 10 of [**41**], cf. also Section 6 of [**40**]. For other arguments of this type in symplectic geometry, see [**25**], [**30**]:

PROPOSITION 2.18. *The homology classes* $\psi \in \pi_2(\widehat{\Theta}_{\beta\gamma}, \widehat{\Theta}_{\gamma\delta}, \widehat{\Theta}_{\delta\beta})$ *with* $\#(\mathcal{M}(\psi)) \neq 0$ *and* $\mu(\psi) = 0$. *are of the form* $\{\Psi^{\pm}_k\}_{k=1}^{\infty}$, *where* $n_z(\Psi^{\pm}_k) = \frac{k(k-1)}{2}$.

Sketch of Proof. Of course, if we were dealing with the genus one case, then this is a consequence of Exercise 2.9. For the general case, however, we need to stretch the neck. Specifically, suppose that ψ is a homology class with the property that $\#\mathcal{M}(\psi) = 1$. This in particular means that for any choice of conformal structure on Σ, there is at least one representative for ψ. Take conformal structures on Σ which converge to the nodal Riemann surface consisting of a torus E_1 (which contains β_1, γ_1, and δ_1) meeting a Riemann surface Σ_0 (which contains the remaining curves) at a point p. In this sequence of conformal structures, one can think of $\Sigma = E_1 \# \Sigma_0$ as developing an ever-longer connected sum neck. The sequence of holomorphic representatives for ψ converges to a union of a holomorphic triangle in $E_1 \times \mathrm{Sym}^{g-1}(\Sigma)$ with spheres in $\mathrm{Sym}^g(\Sigma_0)$. According to Exercise 2.9, the projection of the holomorphic triangle into E_1 must be one of the $\{\psi^{\pm}_k\}_{k=1}^{\infty}$. Moreover, the projection onto the other factor is constrained by dimension considerations to be a product of triangles as pictured in Figure 5. These requirements uniquely determine the homology class of ψ: indeed, the possible homology classes are in one-to-one correspondence with the homology classes of $\{\psi^{\pm}_k\}_{k=1}^{\infty}$ appearing in the genus one surface E_1, and $n_z(\psi)$ coincides with n_z for the corresponding triangle in the genus one surface. Conversely, a gluing argument shows that for each homology class arising in this way, the number of holomorphic representatives (counted with sign) agrees with the number of holomorphic representatives for the corresponding $\psi_k^{\pm}$. □

To complete the argument, we must prove that

$$f^+_{\alpha\beta\beta'}(\cdot \otimes h^{--}_{\beta\gamma\delta\beta'}(\Theta_{\beta\gamma}\otimes\Theta_{\gamma\delta}\otimes\Theta_{\delta\beta'}))\colon CF^+(Y_{\alpha\beta}) \longrightarrow CF^+(Y_{\alpha\beta'})$$

induces an isomorphism in homology. To this end, it suffices to observe that the restriction of the above map to $\widehat{CF}(Y_{\alpha\beta})$ coincides with the map

$$\widehat{f}_{\alpha\beta\beta'}(\cdot \otimes \widehat{h}_{\beta\gamma\delta\beta'}(\widehat{\Theta}_{\beta\gamma}\otimes\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta'})),$$

which we have already proved induces an isomorphism from $\widehat{HF}(Y_{\alpha\beta})$ to $\widehat{HF}(Y_{\alpha\beta'})$ (cf. Exercise 1.4 part (3)).

Exercise 2.19. Suppose that Σ has genus one, and consider the curves β, γ, δ, β'. Show that in this case

$$H^{--}_{\beta\gamma\delta\beta'}(\Theta_{\beta\gamma}\otimes\Theta_{\gamma\delta}\otimes\Theta_{\delta\beta'}) \equiv \left(\sum_{k=0}^{\infty} U^{\frac{k(k+1)}{2}}\right)\Theta_{\beta\beta'} \pmod 2 \tag{19}$$

Hint: Generalizing the picture from Figure 8, show that for each k there are $2k+1$ homology classes of rectangles φ with $\mathcal{D}(\varphi)\geq 0$ and $n_z(\varphi)=\frac{k(k+1)}{2}$.

Note that this Equation (19) also holds in the case where $g(\Sigma)>1$, by another neck-stretching argument.

2.6. Other variations. There are several other variants of the long exact sequence for surgeries. One which requires the minimum additional machinery to state is the following integer surgeries exact sequence, which we will use later.

Theorem 2.20. *(Theorem 9.19 of* [**40**]*) Consider a knot $K\subset Y$ where Y is a three-manifold with $H_1(Y;\mathbb{Z})=0$, and give K its canonical Seifert framing λ. Let Y_n denote the three-manifold obtained by surgery along $K\subset Y$ with framing $n\cdot\mu+\lambda$. There are affine isomorphisms $\mathbb{Z}\cong \mathrm{Spin}^c(Y_0)$ and $\mathbb{Z}/p\mathbb{Z}\cong\mathrm{Spin}^c(Y_p)$ (cf. Exercise 1.8) so that for each $i\in\mathbb{Z}/p\mathbb{Z}$, there are exact sequences*

$$\ldots \longrightarrow \widehat{HF}(Y) \xrightarrow{\widehat{F}_{;i}} \bigoplus_{j\equiv i \pmod p}\widehat{HF}(Y_0,j) \xrightarrow{\widehat{F}_{0;i}} \widehat{HF}(Y_p,i) \xrightarrow{\widehat{F}_{p;i}} \ldots \tag{20}$$

and

$$\ldots \rightarrow HF^+(Y) \xrightarrow{F^+_{;i}} \bigoplus_{j\equiv i \pmod p} HF^+(Y_0,j) \xrightarrow{F^+_{0;i}} HF^+(Y_p,i) \xrightarrow{F^+_{p;i}} \ldots \tag{21}$$

The proof is a slight variation of the proof of Theorem 1.12.

2.7. References and remarks. Proposition 2.5, Theorem 2.7, the null-homotopy of Equation (14), and indeed the fact that $\partial^2=0$ are all special cases of a generalized associativity law satisfied by counting pseudo-holomorphic m-gons, compare [**32**], [**5**], [**19**].

The above proof of Theorem 1.12 can be found in [**38**] (for the case of $\widehat{HF}$); a different proof is given in [**40**]. In fact, in [**38**], the exact triangle from Theorem 1.12 is generalized to address the following question: suppose we have a framed link L in Y with n components, and we know the Floer homology groups of the 2^n three-manifolds which are obtained by performing 0 or 1 surgery on each of the components on the link; then what can be said about the Floer homology of Y? Of course, when $n = 1$, we have a long exact sequence relating these three groups. In the general case, there is a spectral sequence whose E_2 term consists of the direct sum $\widehat{HF}$ of all of these 2^n different three-manifolds, and whose E^∞ term calculates $\widehat{HF}(Y)$. The proof involves a generalized associativity law which is gotten by counting pseudo-holomorphic m-gons.

3. Maps from cobordisms

Recall that the maps appearing in the exact triangle are defined by counting pseudo-holomorphic triangles. These maps respect the $\mathbb{Z}/2\mathbb{Z}$ grading of Floer homology, but in general they do not respect the splitting of the groups according to Spin^c structures, or the relative $\mathbb{Z}$ gradings (in the case where the three-manifolds are rational homology spheres). However, by decomposing the maps according to (suitable equivalence classes of) homology classes of triangles, we obtain a decomposition of the maps as a sum of components which preserve these extra structures. To explain this properly, it is useful to digress to the four-dimensional interpretation of these maps.

Let W be a compact, connected, smooth, four-manifold with two boundary components, which we write as $\partial W = -Y_1 \cup Y_2$ (where here Y_1 and Y_2 are a pair of closed, oriented three-manifolds). Such a four-manifold is called a *cobordism* from Y_1 to Y_2, and we write it sometimes as $W\colon Y_1 \longrightarrow Y_2$.

Let $W\colon Y_1 \longrightarrow Y_2$ be a cobordism equipped with a Spin^c structure $\mathfrak{s} \in \mathrm{Spin}^c(W)$; then there are induced maps on Heegaard Floer homology

$$F^\circ_{W,\mathfrak{s}}\colon HF^\circ(Y_1, \mathfrak{s}|_{Y_1}) \longrightarrow HF^\circ(Y_2, \mathfrak{s}|_{Y_2}),$$

where here HF° denotes any of the variants of Heegaard Floer homology $\widehat{HF}$, HF^-, HF^∞, or HF^+, which we take throughout to be calculated with coefficients in the field $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$. The maps $F^\circ_{W,\mathfrak{s}}$ depend only on W (as a smooth four-manifold) and the Spin^c structure $\mathfrak{s} \in \mathrm{Spin}^c(W)$.

For $\widehat{HF}$ this map is non-trivial for only finitely many $\mathfrak{s} \in \mathrm{Spin}^c(W)$, and hence we can form a map

$$\widehat{F}_W\colon \widehat{HF}(Y_1) \longrightarrow \widehat{HF}(Y_2)$$

defined by

$$\widehat{F}_W = \sum_{\mathfrak{s}\in\mathrm{Spin}^c(W)} \widehat{F}_{W,\mathfrak{s}}.$$

The same construction can be made using HF^+; in this case, although there might be infinitely many $\mathfrak{s} \in \mathrm{Spin}^c(W)$ for which $F^+_{W,\mathfrak{s}}$ is non-trivial, it is still the case that for a fixed $\xi \in HF^+(Y_1)$, there are only finitely many $\mathfrak{s}$ with the property that $F^+_{W,\mathfrak{s}}(\xi)$ is non-zero. Thus, we can define

$$F^+_W\colon HF^+(Y_1) \longrightarrow HF^+(Y_2)$$

by the possibly infinite sum

$$F^+_W = \sum_{\mathfrak{s}\in\mathrm{Spin}^c(W)} F^+_{W,\mathfrak{s}}.$$

These maps are functorial under composition of cobordisms. Specifically, if $W_1\colon Y_1 \longrightarrow Y_2$ and $W_2\colon Y_2 \longrightarrow Y_3$ are two cobordisms, we can form their composition $W_1\#_{Y_2}W_2\colon Y_1 \longrightarrow Y_3$. Functoriality states that

$$\widehat{F}_{W_1\#_{Y_2}W_2} = \widehat{F}_{W_2} \circ \widehat{F}_{W_1} \quad \text{and} \quad F^+_{W_1\#_{Y_2}W_2} = F^+_{W_2} \circ F^+_{W_1}$$

These formulas can be decomposed according to Spin^c structures: assume that $b_1(Y_2) = 0$; then for Spin^c structures $\mathfrak{s}_1 \in \mathrm{Spin}^c(W_1)$, $\mathfrak{s}_2 \in \mathrm{Spin}^c(W_2)$ which agree over Y_2, we have that

$$\widehat{F}_{W_1\#_{Y_2}W_2,\mathfrak{s}_1\#\mathfrak{s}_2} = \widehat{F}_{W_2,\mathfrak{s}_2} \circ \widehat{F}_{W_1,\mathfrak{s}_1} \quad \text{and} \quad F^+_{W_1\#_{Y_2}W_2,\mathfrak{s}_1\#\mathfrak{s}_2} = F^+_{W_2,\mathfrak{s}_2} \circ F^+_{W_1,\mathfrak{s}_1},$$

where here $\mathfrak{s}_1\#\mathfrak{s}_2$ denotes the unique Spin^c structure over $W_1\#_{Y_2}W_2$ whose restriction to W_i is $\mathfrak{s}_i$ (for $i = 1, 2$).

In the case where $b_1(Y_2) > 0$, a Spin^c structure over $W_1\#_{Y_2}W_2$ is no longer necessarily determined by its restrictions to the W_i. Rather, if we consider the Poincaré dual M of the image of the map induced by inclusion $H_2(Y_2) \longrightarrow H_2(W)$, this requirement chooses an M-orbit in $\mathrm{Spin}^c(W_1\#_{Y_2}W_2)$. Now, the left-hand-sides of the equations are replaced by the sum of maps on $W_1\#_{Y_2}W_2$ induced by all the Spin^c structures in this M-orbit. For example, in the case of $\widehat{HF}$, we have

$$\sum_{\{\mathfrak{s}\in\mathrm{Spin}^c(W_1\#_{Y_2}W_2)\,\big|\,\mathfrak{s}|_{W_i}=\mathfrak{s}_i\}} \widehat{F}_{W,\mathfrak{s}} = \widehat{F}_{W_2,\mathfrak{s}_2} \circ \widehat{F}_{W_1,\mathfrak{s}_1}.$$

We will sketch the construction of $F^\circ_{W,\mathfrak{s}}$ in Subsection 3.2, but there are cases of this construction which we have seen already. Suppose that $K \subset Y$ is a knot with framing λ. Then, if $W\colon Y \longrightarrow Y_\lambda(K)$ is the cobordism obtained by attaching a two-handle with framing λ to $[0,1]\times Y$, then the induced maps $\widehat{F}_W$ and F^+_W are the maps constructed in Section 2 which appear in the exact sequence for Theorem 1.12.

Suppose that Y_1 and Y_2 are rational homology three-spheres. Then the Heegaard Floer homology groups of Y_1 and Y_2 can be given a relative $\mathbb{Z}$-grading, cf. Equation (5). In general, the map $\widehat{F}_W$ need not be homogeneous with respect to this relative grading. However, the terms $\widehat{F}_{W,\mathfrak{s}}$ are homogeneous. We can give a much stronger version of this result, after introducing some notions.

Suppose that M is a compact, oriented four-manifold with the property that $H^2(\partial M;\mathbb{Q}) = 0$. Then, there is an intersection form

$$Q_M\colon H^2(M;\mathbb{Q}) \otimes H^2(M;\mathbb{Q}) \longrightarrow \mathbb{Q}$$

defined by

$$Q(\xi\otimes\eta) = \langle \xi\cup\eta, [M]\rangle,$$

where $[M]$ is the fundamental cycle of M. To make sense of the evaluation, implicitly use an identification $H^2(M,\partial M;\mathbb{Q}) \cong H^2(M;\mathbb{Q})$ which exists thanks to the hypothesis that $H^2(\partial M;\mathbb{Q}) = 0$. Let $\sigma(M)$ denote the signature of this intersection form. Sometimes, we write ξ^2 for $Q(\xi,\xi)$. Observe that ξ^2 need not be integral, even if $\xi \in H^2(M;\mathbb{Z})$; however if $\xi \in H^2(M;\mathbb{Z})$ satisfies $n\xi|_{\partial M} = 0$, then $n\cdot\xi^2 \in \mathbb{Z}$.

Exercise 3.1. Let W be the four-manifold which is the unit disk bundle over a two-sphere with Euler number n. There is an isomorphism

$$\phi\colon \mathbb{Z} \longrightarrow H^2(W;\mathbb{Z}).$$

Find $\phi(i)^2$ for $i \in \mathbb{Z}$.

EXERCISE 3.2. If Y is a rational homology three-sphere, then there is a $\mathbb{Q}/\mathbb{Z}$-valued linking

$$q\colon H_1(Y;\mathbb{Z})\otimes H_1(Y;\mathbb{Z}) \longrightarrow \mathbb{Q}/\mathbb{Z}$$

on $H_1(Y;\mathbb{Z})$ defined as follows. Given $\alpha,\beta\in H_1(Y;\mathbb{Z})$, there is some n with the property that $n\beta=0$ in $H_1(Y;\mathbb{Z})$, and hence $n\beta=\partial F$ for some oriented two-manifold $F\subset Y$. Let $q(\alpha,\beta)=\#(\alpha\cap F)/n$. Show that this is a symmetric bilinear form, which is independent of the choice of F. If $Y=\partial W$, and $\alpha,\beta\in H_2(W,Y;\mathbb{Z})$ then show that

$$Q(\mathrm{PD}[\alpha]\otimes \mathrm{PD}[\beta]) \equiv q(\partial\alpha\otimes\partial\beta) \pmod{\mathbb{Z}}.$$

THEOREM 3.3. *(Theorem 7.1 of* [**34**]*) If Y is a rational homology three-sphere, then there is a unique $\mathbb{Q}$-lift of the relative $\mathbb{Z}$ grading on $HF^+(Y,\mathfrak{t})$, which satisfies the following properties:*

- *$\widehat{HF}(S^3)\cong\mathbb{F}$ is supported in degree zero*
- *the inclusion map $\widehat{CF}(Y,\mathfrak{t})\longrightarrow CF^+(Y,\mathfrak{t})$ is degree-preserving*
- *if ξ is a homogeneous element in $CF^+(Y,\mathfrak{t})$, then*

$$\mathrm{gr} f^+_{W,\mathfrak{s}}(\xi)-\mathrm{gr}(\xi)=\frac{c_1(\mathfrak{s})^2-2\chi(W)-3\sigma(W)}{4}, \tag{22}$$

where here $f^+_{W,\mathfrak{s}}$ is a chain map inducing $F^+_{W,\mathfrak{s}}$ on homology.

Actually, verifying the existence of this $\mathbb{Q}$-lift is rather more elementary than proving that $F^+_{W,\mathfrak{s}}$ is a topological invariant of the cobordism: Equation (22) uses only the grading of $f^+_{W,\mathfrak{s}}$, not the count of holomorphic disks.

The $\mathbb{Q}$-grading from Theorem 3.3 allows us to define a numerical invariant for rational homology three-spheres from its Heegaard Floer homology.

DEFINITION 3.4. Let Y be a rational homology three-sphere equipped with a Spin^c structure $\mathfrak{t}$. Then its *correction term* $d(Y,\mathfrak{t})$ is the minimal $\mathbb{Q}$-degree of any homogeneous element in $HF^+(Y,\mathfrak{t})$ coming from $HF^\infty(Y,\mathfrak{t})$.

The above correction term is analogous to the invariant defined in gauge theory by Kim Frøyshov, cf. [**17**].

3.1. Whitney triangles and four-manifolds. Let $(\Sigma,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},z)$ be a pointed Heegaard triple. We can form the identification space

$$X_{\alpha,\beta,\gamma}=\frac{(\Delta\times\Sigma)\coprod(e_\alpha\times U_\alpha)\coprod(e_\beta\times U_\beta)\coprod(e_\gamma\times U_\gamma)}{(e_\alpha\times\Sigma)\sim(e_\alpha\times\partial U_\alpha),(e_\beta\times\Sigma)\sim(e_\beta\times\partial U_\beta),(e_\gamma\times\Sigma)\sim(e_\gamma\times\partial U_\gamma)}.$$

Over the vertices of Δ this space has corners, which can be naturally smoothed out to obtain a smooth, oriented, four-dimensional cobordism between the three-manifolds $Y_{\alpha\beta}$, $Y_{\beta\gamma}$, and $Y_{\alpha\gamma}$ as claimed. More precisely,

$$\partial X_{\alpha,\beta,\gamma}=-Y_{\alpha\beta}-Y_{\beta\gamma}+Y_{\alpha\gamma},$$

with the obvious orientation.

The group of periodic domains for $(\Sigma,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},z)$ (cf. Definition 2.2) has a natural interpretation in terms of the homology of $X_{\alpha,\beta,\gamma}$.

EXERCISE 3.5. Show that $\mathcal{P}\cong H_2(X_{\alpha\beta\gamma};\mathbb{Z})$. Consider the quotient group $\mathcal{Q}$ of $\mathcal{P}$ by the subgroup of elements which can be written as sums of doubly-periodic domains for $Y_{\alpha\beta}$, $Y_{\alpha\gamma}$, and $Y_{\beta\gamma}$. Show that this quotient group is isomorphic to $H^2(X_{\alpha\beta\gamma};\mathbb{Z})$.

Let $\psi \in \pi_2(\mathbf{x},\mathbf{y},\mathbf{w})$ and $\psi' \in \pi_2(\mathbf{x}',\mathbf{y}',\mathbf{w}')$ where $\mathbf{x},\mathbf{x}'$, $\mathbf{y},\mathbf{y}'$ and $\mathbf{w},\mathbf{w}'$ are equivalent. We can define a difference $\delta(\psi,\psi') \in H^2(X_{\alpha\beta\gamma})$ which corresponds to $\mathcal{D}(\psi)+\mathcal{D}(\phi_1)+\mathcal{D}(\phi_2)+\mathcal{D}(\phi_3)-\mathcal{D}(\psi')$ in $\mathcal{Q}$.

We say that two homology classes ψ,ψ' are Spin^c-*equivalent* if this difference $\delta(\psi,\psi')$ vanishes. The maps corresponding to counting holomorphic triangles, cf. Equation (9) clearly split into sums of maps which are indexed by Spin^c equivalence classes of triangles.

Example 3.6. Consider the Heegaard triple in the torus obtained by three straight curves β, γ, δ as in Exercise 2.9. Observe that the triangles $\{\psi_k^{\pm}\}_{k=1}^{\infty}$ represent distinct Spin^c-equivalence classes. Moreover, the results of that exercise can be interpreted as saying that the map

$$F^{--}_{\beta\gamma\delta,[\psi_k^{\pm}]}\colon HF^{--}(S^3)\otimes_{\mathbb{F}[U]} HF^{--}(S^3) \longrightarrow HF^{--}(S^3)$$

represents the map $\mathbb{F}[[U]] \longrightarrow \mathbb{F}[[U]]$ given by multiplication by $U^{\frac{k(k-1)}{2}}$.

Recall that a pointed Heegaard diagram for Y, $(\Sigma,\boldsymbol{\alpha},\boldsymbol{\beta},z)$ gives rise to a map from equivalence classes of intersection points of $\mathbb{T}_\alpha\cap\mathbb{T}_\beta$ to Spin^c structures over Y. In a similar, but somewhat more involved manner, there is a map from Spin^c-equivalence classes of Whitney triangles to Spin^c structures over $X_{\alpha\beta\gamma}$. Moreover, there is a $\psi \in \pi_2(\mathbf{x},\mathbf{y},\mathbf{w})$ for $\mathbf{x}\in\mathbb{T}_\alpha\cap\mathbb{T}_\beta$, $\mathbf{y}\in\mathbb{T}_\beta\cap\mathbb{T}_\gamma$, $\mathbf{w}\in\mathbb{T}_\alpha\cap\mathbb{T}_\gamma$ if and only if there is $\mathfrak{s}\in\mathrm{Spin}^c(X_{\alpha\beta\gamma})$ such that $\mathfrak{s}|_{Y_{\alpha\beta}}=\mathfrak{s}_z(\mathbf{x})$, $\mathfrak{s}|_{Y_{\beta\gamma}}=\mathfrak{s}_z(\mathbf{y})$, and $\mathfrak{s}|_{Y_{\alpha\gamma}}=\mathfrak{s}_z(\mathbf{w})$. We leave the reader to consult Section 8 of [**41**] for details.

In Example 3.6 above, the four-manifold $X_{\beta\gamma\delta}$ is diffeomorphic to $\overline{\mathbb{CP}}^2$ (i.e. $\mathbb{CP}^2$ given the orientation for which its intersection form is negative definite) with three four-balls removed. The triangle $\psi_k^{\pm}$ represents the Spin^c structure over $\overline{\mathbb{CP}}^2$ whose first Chern class evaluates as $\pm(2k-1)$ on a fixed generator for $H_2(\overline{\mathbb{CP}}^2;\mathbb{Z})$.

3.2. Construction of the cobordism invariant. Let $W\colon Y_1 \longrightarrow Y_2$ be a cobordism. The induced map

$$\widehat{F}_W\colon \widehat{HF}(Y_1)\longrightarrow \widehat{HF}(Y_2)$$

is defined using a decomposition of W into handles. Specifically, W can be expressed as a union of one-, two-, and three-handles.

Suppose that W consists entirely of one-handles. Then $Y_2 \cong Y_1\#^{\ell}(S^2\times S^1)$, and a Künneth principle for connected sums ensures that

$$\widehat{HF}(Y_2)\cong \widehat{HF}(Y_1)\otimes \Lambda^* H^1(\#^{\ell}(S^2\times S^1)).$$

Letting $\widehat{\Theta}\in\Lambda^*H^1(\#^{\ell}(S^2\times S^1))$ be the a generator of the top-dimensional element of the exterior algebra, the map $\widehat{F}_W$ is defined to be the map $\xi\mapsto\xi\otimes\widehat{\Theta}$ under the above identification.

Suppose that W consists entirely of three-handles. Then, $Y_1\cong Y_2\times\#^{\ell}(S^2\times S^1)$. In this case, there is a corresponding map $\widehat{F}_W\colon\widehat{HF}(Y_1)\longrightarrow\widehat{HF}(Y_2)$ which is induced by projection onto the bottom-dimensional element of the exterior algebra $H^1(\#^{\ell}(S^2\times S^1))$ under the identification $\widehat{HF}(Y_1)\cong\widehat{HF}(Y_2)\otimes\Lambda^*H^1(\#^{\ell}(S^2\times S^1))$.

The more interesting case is when $W\colon Y_1\longrightarrow Y_2$ consists of two-handles. In this case, W can be expressed as surgery on an ℓ-component link $L\subset Y_1$. In this case, $\widehat{F}_W$ can be obtained as follows.

Consider a Heegaard decomposition of $Y_1 = U_\alpha \cup_\Sigma U_\beta$ with the property that $L = L_{i=1}^{\ell}$ is supported entirely inside U_β in a special way: the L_i is dual to the i^{th} attaching disk for U_β (i.e. it is unknotted, disjoint from all but one attaching disk, which it meets transversally in a single intersection point). Let $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z)$ be a corresponding Heegaard diagram for Y_1. The framings of the components of L_i provide an alternate set of attaching circles γ_i. For all $i > \ell$, we let γ_i be an isotopic copy of β_i. In this way, we obtain a Heegaard triple $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, z)$, where $Y_{\alpha\beta} \cong Y_1$, $Y_{\beta\gamma} \cong \#^{g-\ell}(S^2 \times S^1)$, and $Y_{\alpha\gamma} \cong Y_2$. The map

$$\widehat{F}_W \colon \widehat{HF}(Y_1) \longrightarrow \widehat{HF}(Y_2)$$

is defined now by

$$\widehat{F}_W(\xi) = \widehat{F}_{\alpha\beta\gamma}(\xi \otimes \widehat{\Theta}_{\beta\gamma}),$$

where as usual $\widehat{\Theta}_{\beta\gamma}$ represents a top-dimensional homology class for $\widehat{HF}(Y_{\beta\gamma})$.

Of course, when the number of link components $\ell = 1$, the map $\widehat{F}_W$ coincides with the construction of the map appearing in an exact sequence which contains $\widehat{HF}(Y_1)$ and $\widehat{HF}(Y_2)$.

In the general case where W has handles of all three types, we decompose $W = W_1 \cup W_2 \cup W_3$ where W_i consists of i-handles, and define $\widehat{F}_W$ to be the composite of $\widehat{F}_{W_1}$, $\widehat{F}_{W_2}$, $\widehat{F}_{W_3}$ defined as above.

The verification that the above procedure gives rise to a topological invariant of smooth four-manifolds is lengthy: one must show that it is independent of the decomposition of W into handles; and in the case where W consists of two-handles, that it is independent of the particular choice of Heegaard triple. In particular, one shows that the map (on homology) is invariant under handleslides between various handles and stabilizations. Typically, one interprets such a move as a move on the Heegaard diagram. The key technical point used frequently in these arguments is the associativity law, and some model calcuations. We refer the reader to [**34**] for details (see esp. Section 4 of [**34**]).

The decomposition of $\widehat{F}_W$ according to Spin^c structures proceeds as follows. If W consists entirely of one- or three-handles, then this decomposition is canonical: if $W \colon Y_1 \longrightarrow Y_2$ is a union of one- resp. three-handles then each Spin^c structure over Y_1 resp. Y_2 has a unique extension to a Spin^c structure over W. In the case where W consists of two-handles, the decomposition is represented by the decomposition of $\widehat{F}_{\alpha\beta\gamma}$ into the maps induced by the various Spin^c-equivalence classes of triangles over $X_{\alpha\beta\gamma}$. To identify these with Spin^c structures over W, observe that, after filling in the $Y_{\beta\gamma}$ boundary of $X_{\alpha\beta\gamma}$ by $\#^{g-\ell}(B^3 \times S^1)$, we obtain a four-manifold which is diffeomorphic to W, and hence $\mathrm{Spin}^c(X_{\alpha,\beta,\gamma}) \cong \mathrm{Spin}^c(W)$.

Maps $F^-_{W,\mathfrak{s}}$ $F^\infty_{W,\mathfrak{s}}$, and $F^+_{W,\mathfrak{s}}$ can be defined analogously. Indeed, these maps can all be thought of as induced from an $\mathbb{F}[U]$-equivariant chain map from $CF^-(Y_1, \mathfrak{s}|_{Y_1})$ $\longrightarrow CF^-(Y_2, \mathfrak{s}|_{Y_2})$, and as such, they respect the fundamental exact sequences relating $\widehat{HF}$, HF^-, HF^∞, and HF^+ (cf. Equation (2)).

Example 3.7. The results of Example 3.6 can be interpreted as follows: let W be the cobordism obtained by deleting two four-balls from $\overline{\mathbb{CP}}^2$ (equivalently, this is the cobordism obtained by attaching a two-handle to $S^3 \times [0,1]$ along the unknot with framing -1). Then, for the Spin^c structure $\mathfrak{s}$ whose first Chern class is $\pm(2k-1)$ times a generator of $H^2(W;\mathbb{Z})$, the induced map $F^-_{W,\mathfrak{s}}$ is multiplication by $U^{k(k-1)/2}$. Thus, if $c_1(\mathfrak{s})$ is a generator of $H^2(W;\mathbb{Z})$, then the map

$$\widehat{F}_{W,\mathfrak{s}}\colon \widehat{HF}(S^3) \longrightarrow \widehat{HF}(S^3)$$

is an isomorphism

3.3. Absolute gradings. Let Y be a rational homology three-sphere. The $\mathbb{Q}$-lift of the relative $\mathbb{Z}$ grading on $\widehat{HF}(Y)$ is defined as follows. For any three-manifold Y, there is a cobordism $W\colon S^3 \longrightarrow Y$ consisting entirely of two-handles. As indicated above, this gives a Heegaard triple $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, z)$ with $Y_{\alpha\beta} \cong S^3$, $Y_{\beta\gamma} \cong \#^m(S^2 \times S^1)$, and $Y_{\alpha\gamma} \cong Y$. Indeed, there exists a triangle $\psi \in \pi_2(\widehat{\Theta}_{\alpha\beta}, \widehat{\Theta}_{\beta\gamma}, \mathbf{x})$, where here $\widehat{\Theta}_{\alpha\beta}$, $\Theta_{\beta\gamma}$ are generators representing the canonical (top-dimensional) homology classes of S^3 and $\#^m(S^2 \times S^1)$. We then define

$$\mathrm{gr}(\mathbf{x}) = -\mu(\psi) + 2n_z(\psi) + \frac{c_1(\mathfrak{s})^2 - 2\chi(W) - 3\sigma(W)}{4}.$$

The verification that this is well-defined can be found in Theorem 7.1 of [**34**].

Exercise 3.8. Consider $X = \#^n\overline{\mathbb{CP}}^2$. Let $\mathfrak{s}$ be the Spin^c structure with

$$c_1(\mathfrak{s}) = E_1 + ... + E_n,$$

where $E_i \in H_2(\overline{\mathbb{CP}}^2;\mathbb{Z}) \cong \mathbb{Z}$ is a generator. Show that X can be decomposed along $L(n,1)$ as a union $X_1 \#_{L(n,1)} X_2$ in such a way that X_2 is composed of a single zero- and two-handle, and $c_1(\mathfrak{s})|_{X_1} = 0$. Deduce from the composition law that there is a Spin^c structure $\mathfrak{s}$ over $L(n,1)$ with $c_1(\mathfrak{s}) = 0$ and $d(L(n,1),\mathfrak{s}) = \frac{n-1}{4}$. *Hint*: Let $W_1\colon S^3 \longrightarrow L(n,1)$, $W_2\colon L(n,1) \longrightarrow S^3$ denote X_1 and X_2 with two four-balls removed, so that $W = W_1 \#_{L(n,1)} W_2$ is X with two four-balls removed. According to Example 3.7, $\widehat{F}_W\colon \widehat{HF}(S^3) \cong \mathbb{F} \longrightarrow \widehat{HF}(S^3) \cong \mathbb{F}$ is an isomorphism, and hence so is $\widehat{F}_{W_1}\colon \widehat{HF}(S^3) \longrightarrow \widehat{HF}(L(m,1), \mathfrak{s}|_{L(m,1)})$.

3.4. Construction of the closed four-manifold invariant. If X is a four-manifold, let $b_2^+(X)$ denote the maximal dimension of any subspace of $H^2(X;\mathbb{Z})$ on which the intersection form is positive-definite. Let X be a closed, smooth four-manifold with $b_2^+(X) > 1$. Then, the maps associated to cobordisms can be used to construct a smooth invariant for X analogous to the Seiberg-Witten invariant for closed manifolds. Its construction uses the following basic fact about the map induced by cobordisms:

Proposition 3.9. *If $W\colon Y_1 \longrightarrow Y_2$ is a four-manifold with $b_2^+(X) > 0$, then $F^\infty_{W,\mathfrak{s}} \equiv 0$.*

The proof can be found in Lemma 8.2 of [**34**]

Deleting two four-balls from X, we obtain a cobordism $W\colon S^3 \longrightarrow S^3$. When $b_2^+(X) > 1$, we can always find a separating hypersurface $N \subset W$ which decomposes W as a union of two cobordisms $W = W_1 \#_N W_2$ with $b_2^+(W_i) > 0$ and the image of $H_2(N;\mathbb{Z})$ in $H_2(W;\mathbb{Z})$ is trivial (so that each Spin^c structure over X is uniquely determined by its restrictions to W_1 and W_2). Such a separating hypersurface N is called an *admissible cut*.

Fix $\mathfrak{s} \in \mathrm{Spin}^c(X)$, and let $\mathfrak{s}_1$ and $\mathfrak{s}_2$ denote its restrictions to W_1 and W_2 respectively. In view of Proposition 3.9, the image of the map

$$F^-_{W_1,\mathfrak{s}_1}\colon HF^-(S^3) \longrightarrow HF^-(N, \mathfrak{s}|_N)$$

is contained in the kernel of the natural map $\ell_*\colon HF^-(N,\mathfrak{s}|_N) \longrightarrow HF^\infty(N,\mathfrak{s}|_N)$ (cf. Equation (1)). Another application of the same proposition shows that the map

$$F^+_{W_2,\mathfrak{s}_2}\colon HF^+(N,\mathfrak{s}|_N) \longrightarrow HF^+(S^3).$$

induces a well-defined map on the cokernel of $q_*\colon HF^\infty(N,\mathfrak{s}|_N) \longrightarrow HF^+(N,\mathfrak{s}|_N)$. Using the canonical identification between the kernel of ℓ_* and the cokernel of q_* (following from exactness in Equation (1)), we can compose the two maps to obtain a map

$$\Phi_{X,\mathfrak{s}}\colon \mathbb{F}[U] \cong HF^-(S^3) \longrightarrow \mathcal{T}^+ \cong HF^+(S^3).$$

By U-invariance, we can view $\Phi_{X,\mathfrak{s}}$ as a function from $\mathbb{F}[U]$ to $\mathbb{F}$. In fact, by Equation (22), $\Phi_{X,\mathfrak{s}}$ is a homogeneous function from $\mathbb{F}[U] \longrightarrow \mathbb{F}$, with degree given by

$$\deg(X,\mathfrak{s}) = \frac{c_1(\mathfrak{s})^2 - 2\chi(X) - 3\sigma(W)}{4};$$

i.e. $\Phi_{X,\mathfrak{s}}(U^i) = 0$ if $2i \neq \deg(X,\mathfrak{s})$. Thus, $\Phi_{X,\mathfrak{s}}$ is determined by the element $\Phi_{X,\mathfrak{s}}(U^{\deg(X,\mathfrak{s})/2}) \in \mathbb{F}$ and, of course, the degree $\deg(X,\mathfrak{s})$. (Indeed, with more work, one can lift this to an integer, uniquely determined up to sign.) The following is proved in Section 8 of [**34**]:

Theorem 3.10. *Let X be a smooth four-manifold with $b_2^+(X) > 1$. Then the function $\Phi_{X,\mathfrak{s}}$ depends on the diffeomorphism type of X and the choice of $\mathfrak{s} \in$* $\mathrm{Spin}^c(X)$.

In particular, $\Phi_{X,\mathfrak{s}}$ is independent of the choice of admissible cut used in its definition.

3.5. Properties of the closed four-manifold invariant. The following is a combination of the functoriality of W under cobordisms and the definition of $\Phi_{X,\mathfrak{s}}$:

Proposition 3.11. *Let X be a closed, smooth four-manifold which is separated (smoothly) by a three-manifold Y as $X = X_1\#_Y X_2$ with $b_2^+(X_i) > 0$, and choose $\mathfrak{s}_i \in \mathrm{Spin}^c(X_i)$ whose restriction to Y is some fixed* Spin^c *structure $\mathfrak{t} \in \mathrm{Spin}^c(Y)$. Then, if*

$$\sum_{\mathfrak{s}\in\mathrm{Spin}^c(X)\big|\mathfrak{s}|_{X_i}=\mathfrak{s}_i} \Phi_{X,\mathfrak{s}} \neq 0$$

then $HF^+(Y,\mathfrak{t}) \neq 0$. In fact, in this case the natural map $HF^\infty(Y,\mathfrak{t}) \longrightarrow HF^+(Y,\mathfrak{t})$ has non-trivial cokernel.

In particular, it follows that if X is the connected sum of two four-manifolds, each of which has $b_2^+ > 0$, then $\Phi_{X,\mathfrak{s}} \equiv 0$. This is interesting when combined with the following non-vanishing theorem:

Theorem 3.12. *(Theorem 1.1 of [**42**]; compare also Taubes [**52**]) Let (M,ω) be a symplectic four-manifold with $b_2^+(M) > 1$, and let k represent the canonical* Spin^c *structure; then $\Phi_{M,k} \neq 0$.*

The above theorem is proved by first constructing a Lefschetz pencil [**6**], and using a naturally induced handle decomposition on a suitable blow-up of M.

3.6. References and remarks. The cobordism invariant, absolute gradings on Heegaard Floer homology, and the closed four-manifold invariant are all defined in [**34**]. Further applications of the absolute grading, and also the correction term $d(Y, \mathfrak{t})$ are given in [**36**].

The closed four-manifold invariant $\Phi_{X,\mathfrak{s}}$ is analogous to the Seiberg-Witten invariant for $(X, \mathfrak{s})$, compare [**53**]. That invariant, too, vanishes for connected sums (of four-manifolds with $b_2^+ > 0$) [**53**], and is non-trivial for symplectic manifolds, according to a theorem of Taubes [**50**], [**51**]. In fact, it is conjectured that $\Phi_{X,\mathfrak{s}}$ agrees with the Seiberg-Witten invariant for four-manifolds. (Note also that Donaldson's theory behaves similarly: Donaldson's invariants for such connected sums vanish, and for Kähler surfaces, they are known not to vanish, cf. [**8**].) Example (3.7) corresponds to the "blow-up" formula for Seiberg-Witten invariants, cf. [**14**]. The correction term $d(Y, \mathfrak{s})$ is analogous to Frøyshov's invariants [**17**], cf. also [**18**] for the corresponding invariants using Donaldson's theory.

4. Dehn surgery characterization of the unknot

Suppose that $K \subset S^3$. For each rational number r, we can construct a new three-manifold $S^3_r(K)$ by Dehn filling. Not every three-manfiold can be obtained as Dehn surgery on a single knot, but for those which are, it is a natural question to ask how much of the Dehn surgery description the three-manifold remembers. There are also many examples of three-manifolds which are obtained as surgery descriptions in more than one way. For example, $+5$ surgery on the right-handed trefoil is the lens space obtained as -5 surgery on the unknot. Note that for this example, the surgery coefficients are opposite in sign.

Exercise 4.1. Consider the three-manifold Y obtained as surgery on the Borromean rings with surgery coefficients $+1$, $+1$, and -1. By blowing down the two circles with coefficient $+1$ we obtain a description of Y as -1 surgery on a knot K_1. By blowing down two circles with coefficients $+1$ and -1, we obtain a description of Y as $+1$ surgery on K_2. What are K_1 and K_2? What is Y?

More interesting examples were described by Lickorish [**31**], who gives two distinct knots K_1 and K_2 with the property that $S^3_{-1}(K_1) \cong S^3_{-1}(K_2)$. His examples are constructed from a two-component link $L_1 \cup L_2$, each of whose components is individually unknotted, and hence K_1 is the knot induced from L_1 in $S^3_{-1}(L_2) \cong S^3$, while K_2 is the knot induced from L_2 in $S^3_{-1}(L_1) \cong S^3$, cf. Figure 4.

For suitably simple three-manifolds, though, the phenomenon illustrated above does not occur. Specifically, our aim here is to sketch the proof of the following conjecture of Gordon [**23**], first proved by the authors in collaboration with Peter Kronheimer and Tomasz Mrowka [**29**], using Floer homology for Seiberg-Witten monopoles constructed by Kronheimer and Mrowka (cf. [**26**]).

Theorem 4.2. *(Kronheimer-Mrowka-Ozsváth-Szabó* [**29**]*) Let $\mathcal{U}$ denote the unknot in S^3, and let K be any knot. If there is an orientation-preserving diffeomorphism $S^3_r(K) \cong S^3_r(\mathcal{U})$ for some rational number r, then $K = \mathcal{U}$.*

Of course, $S^3_{p/q}(\mathcal{U})$ is the lens space $L(p, q)$. (The reader should be warned that this orientation convention on the lens space is opposite to the one adopted by some other authors.) This result has the following immediate application, where one can discard orientations:

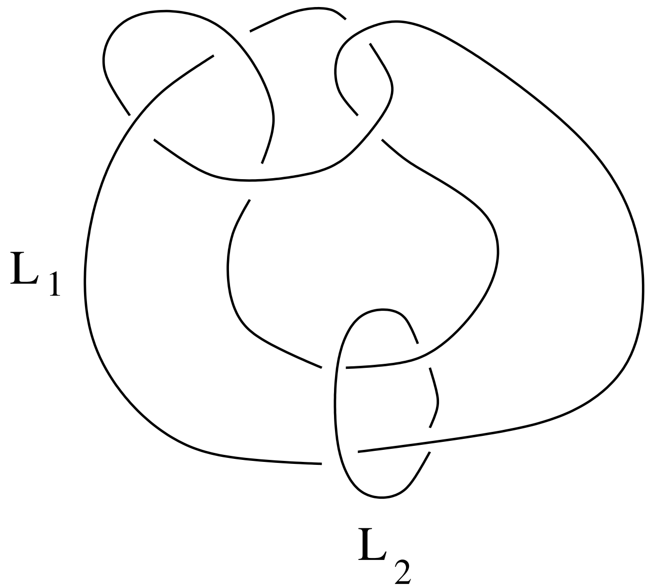


FIGURE 10. A two-component link. Each component is unknot -ted; blowing down either one or the other component gives a pair of distinct knots K_1 and K_2 in S^3 with $S^3_{-1}(K_1) \cong S^3_{-1}(K_2)$.

COROLLARY 4.3. *If K is a knot with the property that some surgery on K is the real projective three-space $\mathbb{RP}^3$, then K is the unknot.*

Many cases of Theorem 4.2 had been known previously. The case where $r = 0$ was the "Property R" conjecture proved by Gabai [**21**]; the case where r is non-integral follows from the cyclic surgery theorem of Culler, Gordon, Luecke, and Shalen [**4**], the case where $r = \pm 1$ is a theorem of Gordon and Luecke [**24**].

We outline here the proof for integral $r \neq 0$, using Heegaard Floer homology, though a re-proof of the result for all rational r can be given by adapting the arguments from [**29**]. The Heegaard Floer homology proof is strictly logically independent of the proof using monopole Floer homology, though the two proofs are formally quite analogous. Moreover, to keep the discussion simple, we prove only that $g(K) \leq 1$. To exclude the possibility that $g(K) = 1$, we require a little of the theory beyond what has been explained so far: either a discussion of "twisted coefficients" or an extra discussion of knot Floer homology (compare [**35**]). Or, alternatively, one could appeal to an earlier result of Goda and Teragaito [**22**].

The proof can be subdivided into two components: first, one proves that $HF^+(S^3_r(K)) \cong HF^+(S^3_r(\mathcal{U}))$ implies a corresponding isomorphism $HF^+(S^3_0(K)) \cong HF^+(S^3_0(\mathcal{U}))$. In the second component, one shows that the Heegaard Floer homology of $S^3_0(K)$ distinguishes any non-trivial knot from the unknot. The first component follows from a suitably enhanced application of the long exact sequence for surgeries. The second component rests on fundamental work by a large number of researchers, including the construction of taut foliations by Gabai [**20**], [**21**], Eliashberg and Thurston [**10**], Eliashberg [**9**] and Etnyre [**11**], and Donaldson [**6**].

We describe these two components in more detail in the following two subsections.

4.1. The first component: $HF^+(S^3_p(K)) \cong HF^+(S^3_p(\mathcal{U})) \Rightarrow HF^+(S^3_0(K)) \cong HF^+(S^3_0(\mathcal{U}))$

In view of our earlier remarks, it will suffice to prove that

$$HF^+(S^3_p(K)) \cong HF^+(S^3_p(\mathcal{U})) \Rightarrow HF^+(S^3_0(K), i) = 0$$

for all $i \neq 0$. In fact, for simplicity, we always work with Heegaard Floer homology with coefficients in some field $\mathbb{F}$ (which the reader can take to be $\mathbb{Z}/2\mathbb{Z}$), although since the field is generic, the results hold over $\mathbb{Z}$, as well.

The proof hinges on the following application of the exact triangle, combined with absolute gradings. In the following statement (cf. Equation (23)), we fix an identification $\mathbb{Z}/p\mathbb{Z} \cong \mathrm{Spin}^c(L(p,1))$, made explicit later.

Theorem 4.4. *(Theorem 7.2 of* [**36**]*) Suppose that $K \subset S^3$ is a knot in S^3 with the property that some integral $p > 0$ surgery on K gives the L-space Y; then there is a map $\sigma\colon \mathbb{Z}/p\mathbb{Z} \longrightarrow \mathrm{Spin}^c(Y)$ with the property that for each $i \neq 0$, with $|i| \leq p/2$,*

$$HF^+(S^3_0(K), i) \cong \mathbb{F}[U]/U^{\ell_i}$$

where

$$2\ell_i = -d(Y, \sigma(i)) + d(L(p,1), i), \tag{23}$$

while $HF^+(S^3_0(K), i) = 0$ for $|i| > p/2$. In particular, each $\ell_i \geq 0$.

The proof of this result uses the integer surgeries long exact sequence, Theorem 2.20, with the understanding that the map appearing there, $F^+_{p;i}\colon HF^+(Y_p, i) \longrightarrow HF^+(Y)$, is the sum of maps induced by the two-handle cobordism $W_p(K)\colon Y_p(K) \longrightarrow Y$, where we sum over all Spin^c structures whose restriction to $Y_p(K)$ corresponds to $i \in \mathbb{Z}/p\mathbb{Z}$. In fact, given $i \in \mathbb{Z}/p\mathbb{Z}$, the set of Spin^c structures over $W_p(K)$ whose restriction to $Y_p(K)$ corresponds to i is the set of Spin^c structures $\mathfrak{s} \in \mathrm{Spin}^c(W_p(K))$, for which

$$c_1(\mathfrak{s}) \equiv 2i + p \pmod{2p}, \tag{24}$$

under an isomorphism $H^2(W_p(K);\mathbb{Z}) \cong \mathbb{Z}$. Indeed, Equation (24) can also be viewed as determining the map $\mathbb{Z}/p\mathbb{Z} \longrightarrow \mathrm{Spin}^c(S^3_p(K))$ (up to an irrelevant overall sign – irrelevant due to the conjugation symmetry of the groups in question) arising in Theorem 2.20: $\mathfrak{s} \in \mathrm{Spin}^c(W_p(K);\mathbb{Z})$ is uniquely determined by its first Chern class, and in turn its equivalence class modulo $2p$ uniquely determines its restriction to $S^3_p(K)$.

Exercise 4.5. Show that if $\mathfrak{t} \in \mathrm{Spin}^c(S^3_p(K))$ is a Spin^c structure which corresponds to $i = 0$ under Equation (24), then $c_1(\mathfrak{t}) = 0$. (Note also that when p is odd, there is only one Spin^c structure over $S^3_p(K)$ with trivial first Chern class; when p is even, there are two.)

It will also be useful to have the following:

Exercise 4.6. Let $\mathcal{T}^+ \cong \mathbb{F}[U, U^{-1}]/U \cdot \mathbb{F}[U]$. Given any formal power series in U, $\sum_{i=0}^{\infty} a_i \cdot U^i$, there is a corresponding endomorphism of $\mathcal{T}^+$, defined by

$$\xi \mapsto \sum_{i=0}^{\infty} a_i U^i \cdot \xi.$$

Show that in fact every endomorphism of $\mathcal{T}^+$ can be described in this manner. In particular, every non-trivial endomorphism of $\mathcal{T}^+$ is surjective, with kernel isomorphic to $\mathbb{F}[U]/U^\ell$, where $\ell = \min\{i \big| a_i \neq 0\}$.

LEMMA 4.7. *Let $K \subset S^3$ be a knot. Then for all $i \neq 0$, $HF^+(S^3_0(K), i)$ is a finite-dimensional vector space (over $\mathbb{F}$).*

PROOF. Clearly, there are only finitely many integers i for which $HF^+(S^3_0(K), i) \neq 0$. It follows that for sufficiently large N, we can arrange that there is some $i \in \mathbb{Z}/N\mathbb{Z}$ with the property that $\bigoplus_{j \equiv i \pmod N} HF^+(S^3_0, j) = 0$. According to Theorem 2.20, this forces $F^+_{W_N(K)} \colon HF^+(S^3_N(K), j) \longrightarrow HF^+(S^3)$ to be an isomorphism. It follows from this that $F^\infty_{W_N(K),\mathfrak{s}} \colon HF^\infty(S^3_N(K), \mathfrak{s}|_{S^3_N(K)}) \longrightarrow HF^\infty(S^3)$ is an isomorphism for some choice of $\mathfrak{s}$.

We would like to conclude that it holds for all $\mathfrak{s} \in \mathrm{Spin}^c(W_N(K))$. To this end, recall first that for a rational homology three-sphere Y such as $S^3_N(K)$, the group $HF^\infty(Y, \mathfrak{s})$ is indpendent of the choice of $\mathfrak{s}$. This follows readily from the definition of the differential: for two different choices of reference point z_1 and z_2 and a fixed $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$, the $\mathbf{y}$ component of $\partial \mathbf{x}$ differs only by multiplication by some power of U. It is easy to see that by changing basis for $CF^\infty(Y, \mathfrak{s})$ (multiplying each generator $\mathbf{x}$ for $CF^\infty(Y, \mathfrak{s})$ by $U^{m_{\mathbf{x}}}$), we get an isomorphism between the chain complex defining $CF^\infty(Y, \mathfrak{s}_1)$ and $CF^\infty(Y, \mathfrak{s}_2)$ (where here $\mathfrak{s}_i = \mathfrak{s}_{z_i}(\mathbf{x})$).

Modifying this argument, we can also see that the induced map

$$F^\infty_{W_N(K),\mathfrak{s}} \colon HF^\infty(S^3_N(K), i) \longrightarrow HF^\infty(S^3),$$

where $\mathfrak{s}$ is any Spin^c structure over $W_N(K)$ whose restriction to $S^3_N(K)$ corresponds to i, depends on $\mathfrak{s}$ only up to an overall multiplication by some U-power. Again, this is seen from the definition of $F^\infty_{W_N(K),\mathfrak{s}}$ as a count of holomorphic triangles in a Heegaard triple representing some fixed Spin^c equivalence class, and then moving the reference point. Moreover, the precise dependence of the U-power on the choice of $\mathfrak{s}$ is determined by $c_1(\mathfrak{s})^2$, according to Equation (22) (which determines the grading of the image of any element).

In view of Equation (24), it is easy to see that for all $i \not\equiv 0 \pmod p$, for all $\mathfrak{s} \in \mathrm{Spin}^c(W_p(K))$ whose restriction to $S^3_p(K)$ corresponds to i, the lengths $c_1(\mathfrak{s})^2$ are all distinct. Thus, the homomorphism

$$\mathcal{T}^+ \cong \mathrm{Im}\left(HF^\infty(S^3_p(K), i) \subset HF^+(S^3_p(K), i)\right) \longrightarrow \mathcal{T}^+ \cong HF^+(S^3)$$

gotten by restricting $F^+_{W_N(K)}$ is non-trivial, and in particular, according to Exercise 4.6, it follows that $HF^+(S^3_0(K), i)$ is a finite-dimensional vector space. $\square$

Proof of Theorem 4.4. We use the integral surgeries long exact sequence, Theorem 2.20.

As a preliminary step, we argue that the only integer $j \equiv 0 \pmod p$ with $HF^+(S^3_0(K), j) \neq 0$ is $j = 0$. This follows easily from the exact sequence in the form of Equation (20). (We leave the details to the reader; it is a straightforward adaptation of the proof of Corollary 1.16, together with the observation that $\widehat{HF}(S^3_0(K), j) \neq 0$ if and only if $HF^+(S^3_0(K), j) \neq 0$, cf. Exercise 1.4.)

Next, we consider $j \not\equiv 0 \pmod p$. If the map

$$F^+_{W_p(K)}|_{HF^+(S^3_p(K),i)} \colon HF^+(S^3_p(K), i) \longrightarrow HF^+(S^3)$$

were trivial, the long exact sequence would would force $HF^+(S^3_0(K),j)$ to be infinitely generated (as an $\mathbb{F}$-vector space) for some $j\neq 0$, contradicting Lemma 4.7. Thus Exercise 4.6, together with the long exact sequence, gives us that

$$\bigoplus_{j\equiv i \pmod{\mathbb{Z}}} HF^+(S^3_0(K),j)\cong \mathbb{F}[U]/U^\ell \tag{25}$$

for some $\ell\geq 0$. In particular, for each $i\in\mathbb{Z}/p\mathbb{Z}$, there is at most one $j\equiv i \pmod{p}$ with $HF^+(S^3_0(K),j)\neq 0$. Next, we argue that in fact if $HF^+(S^3_0(K),m)\neq 0$, then $|m|\leq p/2$ as follows. If it were not the case, then since $2m\geq p$, $S^3_{2m}(K)$ would also be an L-space (cf. Exercise 1.14); but now both $m\equiv -m \pmod{2m}$ and $HF^+(S^3_0(K),m)\neq 0$ and $HF^+(S^3_0(K),-m)\neq 0$, violating the principle just established.

It remains now to show that the power of U, ℓ, appearing in Equation (25) is the quantity ℓ_i given by Equation (23).

Let $c(p,i)$ be the maximal value of

$$\frac{c_1(\mathfrak{s})^2+1}{4}$$

for any $\mathfrak{s}\in\mathrm{Spin}^c(W_p(K))$ with $\mathfrak{s}|_{S^3_p(K)}$ corresponding to $i\in\mathbb{Z}/p\mathbb{Z}$, and let $\mathfrak{s}_0\in\mathrm{Spin}^c(W_p(K))$ be the Spin^c structure with given restriction to $S^3_p(K)$ which achieves this maximal value. Note that $c(p,i)$ is independent of the choice of $K\subset S^3$. The element of $HF^\infty(Y_p(K),i)$ of degree $-c(p,i)$ is mapped by $F^\infty_{W_p(K);\mathfrak{s}_0}$ to the generator of $HF^+_0(S^3)$, and hence its image in $HF^+(S^3_p(K),i)$ is mapped to the generator of $HF^+_0(S^3)$, in view of the diagram:

$$\begin{array}{ccc}
HF^\infty_{-c(p,i)}(S^3_p(K),i) & \xrightarrow{F^\infty_{W_p(K),\mathfrak{s}_0}} & HF^\infty_0(S^3)\\
\Big\downarrow & & \cong\Big\downarrow\\
HF^+_{-c(p,i)}(S^3_p(K),i) & \xrightarrow{F^+_{W_p(K),\mathfrak{s}_0}} & HF^+_0(S^3).
\end{array}$$

(where here the subscripts on Heegaard Floer groups denote the summands with specified $\mathbb{Q}$ grading). But since $\mathfrak{s}_0$ is the unique Spin^c structure which maximizes $c_1(\mathfrak{s})^2$ among all $\mathfrak{s}\in\mathrm{Spin}^c(W_p(K))$ with given restriction to $S^3_p(K)$, it follows that $F^+_{W_p(K)}$ carries $HF^+_{-c(p,i)}(S^3_p(K))$ isomorphically to $HF^+_0(S^3)$.

Moreover, it also follows from this formula that all elements in $HF^+(S^3_p(K),i)$ of degree less than $-c(p,i)$ are mapped to zero, and the set of such elements form a vector space of rank

$$-1-\left(\frac{c(p,i)+d(S^3_p(K),i)}{2}\right).$$

We can conclude now that the kernel of $F^+_{W_p(K)}\colon HF^+(S^3_p(K),i)\longrightarrow HF^+(S^3)$ is isomorphic to $\mathbb{F}[U]/U^\ell$, with $2\ell=-c(p,i)-d(S^3_p(K),i)$. By comparing with the unknot $\mathcal{U}$, and recalling that $HF^+(S^3_0(\mathcal{U}),i)=HF^+(S^2\times S^1,i)=0$ for all $i\neq 0$, we conclude that $-c(p,i)=d(L(p,1),i)$. $\square$

Exercise 4.8. Using the above proof (and Equation (24)) calculate $d(L(p,1),i)$ for all $i\neq 0$. As a test, when p is even, you should find that $d(L(p,1),p/2)=-\frac{1}{4}$.

Corollary 4.9. *If $HF^+(S^3_p(K)) \cong HF^+(S^3_p(\mathcal{U}))$ as $\mathbb{Q}$-graded Abelian groups, then $HF^+(S^3_0(K),i)=0$ for all $i\neq 0$.*

Proof. The expression $S^3_p(\mathcal{U})\cong L(p,1)$ gives an affine identification $\mathbb{Z}/p\mathbb{Z}\cong \mathrm{Spin}^c(L(p,1))$ (determined by Equation (24)), and hence the affine identification $\mathbb{Z}/p\mathbb{Z}\cong \mathrm{Spin}^c(L(p,1))$ induced from the expression of $S^3_p(K)\cong L(p,1)$ can be viewed as a permutation $\sigma\colon \mathbb{Z}/p\mathbb{Z}\longrightarrow \mathbb{Z}/p\mathbb{Z}$. According to Theorem 4.4, this permutation σ has the property that for all $i\neq 0$, $-d(L(p,1),\sigma(i))+d(L(p,1),i)\geq 0$. Moreover, in the case where p is odd (cf. Exercise 4.5), σ fixes 0, inducing a permutation on the remaining $\{d(L(p,1),i)\}_{i=1}^{p-1}$. It follows readily that $d(L(p,1),\sigma(i)) = d(L(p,1),i)$ for all i. From Theorem 4.4, it follows that $HF^+(S^3_0(K),i)=0$ for all $i\neq 0$. In the case where p is even, write $p=2n$, and observe that σ either fixes 0 or it permutes 0 and n (Exercise 4.5); if it fixes both, the previous argument applies. We must rule out the possibility that $\sigma(n)=0$. Observe, however, that $d(L(2n,1),0)=\frac{2n-1}{4}$ according to Exercise 3.8, while $d(L(2n,1),n)=-\frac{1}{4}$ according to Exercise 4.8, so in this case, it would not be possible for $-d(L(2n,1),\sigma(n))+d(L(2n,1),n)\geq 0$, as required by Theorem 4.4. □

4.2. The second component: $HF^+(S^3_0(K))\cong HF^+(S^3_0(\mathcal{U}))\Rightarrow K=\mathcal{U}$

Again, we set slightly more modest goals in this article, sketching the proof that $HF^+(S^3_0(K),i)=0$ for all $i\neq 0$ implies that $g(K)\leq 1$.

We rely on the following fundamental result of Gabai. For our purposes, an oriented foliation $\mathcal{F}$ of an oriented three-manifold Y is *taut* if there is a closed two-form ω_0 over Y whose restriction to the tangent space to $\mathcal{F}$ is always non-degenerate.

Theorem 4.10. *(Gabai [**21**]) If K is a knot with Seifert genus $g(K)>1$, then there is a smooth taut foliation over $S^3_0(K)$ whose first Chern class is $2g-2$ times a generator of $H^2(S^3_0(K);\mathbb{Z})$.*

Gabai's taut foliation can be interpreted as an infinitesimal symplectic structure, according to the following result:

Theorem 4.11. *(Eliashberg-Thurston [**10**]) Let Y be a three-manifold which admits a taut foliation $\mathcal{F}$, and ω_0 be a two-form positive on the leaves. Then there is a symplectic two-form ω over $[-1,1]\times Y$ which is convex at the boundary, and whose restriction to $\{0\}\times Y$ agrees with ω_0.*

We use here the usual notion of convexity from symplectic geometry (see for example [**49**] or [**12**]). This in turn can be extended to a symplectic structure over a closed manifold according to the following convex filling result:

Theorem 4.12. *(Eliashberg [**9**] and Etnyre [**11**]) If (X,ω) is a symplectic manifold with convex boundary, then there is a closed symplectic four-manifold $(\widetilde{X},\widetilde{\omega})$ which contains (X,ω) as a submanifold.*

There is considerable flexibility in constructing $\widetilde{X}$; in particular, it is technically useful to note that one can always arrange that $b_2^+(\widetilde{X})>1$.

In sum, the above three theorems say the following: if $K\subset S^3$ is a knot with Seifert genus $g(K)>1$, then there is a closed symplectic four-manifold (M,ω) which is divided in two by $S^3_0(K)$, in such a way that $c_1(k)|_{S^3_0(K)}\neq 0$, where here k is the canonical Spin^c structure of the symplectic form specified by ω, and hence $c_1(k)$ restricts to $2g-2$ times a generator of $H^2(S^3_0(K);\mathbb{Z})$.

Proof of Theorem 4.2. As explained in the discussion preceding the statement, it suffices to consider the case where the Seifert genus g of K is greater than one (according to [**22**]), r integral (according to [**4**]) and $|r| > 1$ (according to [**24**]). After reflecting K if necessary (cf. Exercise 1.9), we can assume that $r > 1$. Since $g(K) > 1$, as explained in the above discussion (combining Theorems 4.10, 4.11, and 4.12) we obtain a symplectic four-manifold (M, ω) which is divided in two by $S^3_0(K)$ in such a way that $c_1(k)|_{S^3_0(K)} \neq 0$. According to Theorem 3.12, $\Phi_{M,k} \neq 0$, and hence, according to Proposition 3.11, $HF^+(S^3_0(K), g-1) \neq 0$. (Note that Proposition 3.11 requires the non-vanishing of a sum of invariants associated to $\{k + n\mathrm{PD}[\Sigma]\}_{n\in\mathbb{Z}}$; but since each has distinct $c_1(\mathfrak{s})^2$ and hence $\deg(X, \mathfrak{s})$, these terms are linearly independent.) But this now contradicts the conclusion of Theorem 4.4. $\square$

4.3. Comparison with Seiberg-Witten theory. The original proof of Theorem 4.2 was obtained using the monopole Floer homology for Seiberg-Witten monopoles, cf. [**29**]. The basic components of the proof are analogous: an exact triangle argument reduces the problem to showing that the monopole Floer homologies of $S^3_0(K))$ and $S^3_0(\mathcal{U})$ coincide, and a second component proves that if this holds, then $K = \mathcal{U}$. This second component had been established by Kronheimer and Mrowka [**27**] and [**28**], shortly after the discovery of the Seiberg-Witten equations. More specifically, combining Gabai's foliation with the Eliashberg-Thurston filling, one obtains a symplectic structure on $[-1, 1] \times S^3_0(K)$ with convex boundary. For four-manifolds with symplectically convex boundary, Kronheimer and Mrowka construct an invariant analogous to the invariant for closed symplectic manifolds. Using the symplectic form as a perturbation for the Seiberg-Witten equations (as Taubes did in the case of closed four-manifolds, cf. [**50**]), Kronheimer and Mrowka show that their invariant for $[-1, 1] \times S^3_0(K)$ is non-trivial. It follows that the Seiberg-Witten monopoloe Floer homology of $S^3_0(K)$ is non-trivial.

4.4. Further remarks. Theorem 4.4, including an analysis of the case where $i = 0$, was first proved in Theorem 7.2 [**36**]. This result can be used to give bounds on genera of knots admitting lens space surgeries. Further bounds on the genera of these knots have been obtained by Rasmussen [**45**].

See also [**43**] for a generalization of Theorem 4.4 to the case of knots which admit rational L-space surgeries.

References

[1] P. Braam and S. K. Donaldson. Floer's work on instanton homology, knots, and surgery. In H. Hofer, C. H. Taubes, A. Weinstein, and E. Zehnder, editors, *The Floer Memorial Volume*, number 133 in Progress in Mathematics, pages 195–256. Birkhäuser, 1995.

[2] G. Burde and H. Zieschang. *Knots*. Number 5 in de Gruyter Studies in Mathematics. Walter de Gruyter & Co., 1985.

[3] D. Calegari and N. M. Dunfield. Laminations and groups of homeomorphisms of the circle. *Invent. Math.*, 152(1):149–204, 2003.

[4] M. Culler, C. McA. Gordon, J. Luecke, and P. B. Shalen. Dehn surgery on knots. *Ann. of Math*, 125(2):237–300, 1987.

[5] V. de Silva. *Products in the symplectic Floer homology of Lagrangian intersections*. PhD thesis, Oxford University, 1999.

[6] S. K. Donaldson. Lefschetz pencils on symplectic manifolds. *J. Differential Geom.*, 53(2):205–236, 1999.

[7] S. K. Donaldson. *Floer homology groups in Yang-Mills theory*, volume 147 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2002. With the assistance of M. Furuta and D. Kotschick.

[8] S. K. Donaldson and P. B. Kronheimer. *The Geometry of Four-Manifolds*. Oxford Mathematical Monographs. Oxford University Press, 1990.

[9] Y. Eliashberg. A few remarks about symplectic filling. *Geom. Topol.*, 8:277–293 (electronic), 2004.

[10] Y. M. Eliashberg and W. P. Thurston. *Confoliations*. Number 13 in University Lecture Series. American Mathematical Society, 1998.

[11] J. B. Etnyre. On symplectic fillings. *Algebr. Geom. Topol.*, 4:73–80, 2004.

[12] J. B. Etnyre. Lectures on open book decompositions and contact structures. In *Floer homology, gauge theory, and low-dimensional topology*, 2006.

[13] R. Fintushel and R. J. Stern. Instanton homology of Seifert fibered homology three spheres. *Proc. of the London Math. Soc.*, 61:109–137, 1990.

[14] R. Fintushel and R. J. Stern. Immersed spheres in 4-manifolds and the immersed Thom conjecture. *Turkish J. Math*, 19(2):145–157, 1995.

[15] A. Floer. An instanton-invariant for 3-manifolds. *Comm. Math. Phys.*, 119:215–240, 1988.

[16] A. Floer. Instanton homology and Dehn surgery. In H. Hofer, C. H. Taubes, A. Weinstein, and E. Zehnder, editors, *The Floer Memorial Volume*, number 133 in Progress in Mathematics, pages 77–97. Birkhäuser, 1995.

[17] K. A. Frøyshov. The Seiberg-Witten equations and four-manifolds with boundary. *Math. Res. Lett*, 3:373–390, 1996.

[18] K. A. Frøyshov. Equivariant aspects of Yang-Mills Floer theory. *Topology*, 41(3):525–552, 2002.

[19] K. Fukaya, Y-G. Oh, K. Ono, and H. Ohta. *Lagrangian intersection Floer theory—anomaly and obstruction*. Kyoto University, 2000.

[20] D. Gabai. Foliations and the topology of 3-manifolds. *J. Differential Geom.*, 18(3):445–503, 1983.

[21] D. Gabai. Foliations and the topology of 3-manifolds III. *J. Differential Geom.*, 26(3):479–536, 1987.

[22] H. Goda and M. Teragaito. Dehn surgeries on knots which yield lens spaces and genera of knots. *Math. Proc. Cambridge Philos. Soc.*, 129(3):501–515, 2000.

[23] C. McA. Gordon. *Some aspects of classical knot theory*, pages pp. 1–60. Number 685 in Lecture Notes in Math. Springer-Verlag, 1978.

[24] C. McA. Gordon and J. Luecke. Knots are determined by their complements. *J. Amer. Math. Soc.*, 2(2):371–415, 1989.

[25] E-N. Ionel and T. H. Parker. Relative Gromov-Witten invariants. *Ann. of Math. (2)*, 157(1):45–96, 2003.

[26] P. B. Kronheimer and T. S. Mrowka. Floer homology for Seiberg-Witten Monopoles. Preprint.

[27] P. B. Kronheimer and T. S. Mrowka. Monopoles and contact structures. *Invent. Math.*, 130(2):209–255, 1997.

[28] P. B. Kronheimer and T. S. Mrowka. Scalar curvature and the Thurston norm. *Math. Res. Lett.*, 4(6):931–937, 1997.

[29] P. B. Kronheimer, T. S. Mrowka, P. S. Ozsváth, and Z. Szabó. Monopoles and lens space surgeries. math.GT/0310164.

[30] A-M. Li and Y. Ruan. Symplectic surgery and Gromov-Witten invariants of Calabi-Yau 3-folds. *Invent. Math.*, 145(1):151–218, 2001.

[31] W. B. R. Lickorish. *An introduction to knot theory*, volume 175 of *Graduate Texts in Mathematics*. Springer-Verlag, 1997.

[32] D. McDuff and D. Salamon. *J-holomorphic curves and quantum cohomology*. Number 6 in University Lecture Series. American Mathematical Society, 1994.

[33] J. W. Milnor. *Topology from the differentiable viewpoint*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Based on notes by David W. Weaver, Revised reprint of the 1965 original.

[34] P. S. Ozsváth and Z. Szabó. Holomorphic triangles and invariants for smooth four-manifolds. math.SG/0110169.

[35] P. S. Ozsváth and Z. Szabó. On knot Floer homology and lens space surgeries. math.GT/0303017, to appear in *Topology*.

[36] P. S. Ozsváth and Z. Szabó. Absolutely graded Floer homologies and intersection forms for four-manifolds with boundary. *Advances in Mathematics*, 173(2):179–261, 2003.

[37] P. S. Ozsváth and Z. Szabó. On the Floer homology of plumbed three-manifolds. *Geometry and Topology*, 7:185–224, 2003.
[38] P. S. Ozsváth and Z. Szabó. On the Heegaard Floer homology of branched double-covers. math.GT/0309170, 2003.
[39] P. S. Ozsváth and Z. Szabó. Holomorphic disks and genus bounds. *Geom. Topol.*, 8:311–334 (electronic), 2004.
[40] P. S. Ozsváth and Z. Szabó. Holomorphic disks and three-manifold invariants: Properties and applications. *Ann. of Math.*, 159(3):1159–1245, 2004.
[41] P. S. Ozsváth and Z. Szabó. Holomorphic disks and topological invariants for closed three-manifolds. *Ann. of Math.*, 159(3):1027–1158, 2004.
[42] P. S. Ozsváth and Z. Szabó. Holomorphic triangle invariants and the topology of symplectic four-manifolds. *Duke Math. J.*, 121(1):1–34, 2004.
[43] P. S. Ozsváth and Z. Szabó. Knot Floer homology and rational surgeries. In preparation, 2005.
[44] P. S. Ozsváth and Z. Szabó. An introduction to Heegaard Floer homology. In *Floer homology, gauge theory, and low-dimensional topology*, 2006.
[45] J. Rasmussen. Lens space surgeries and a conjecture of Goda and Teragaito. *Geometry and Topology*, 8:1013–1031, 2004.
[46] J. Robbin and D. Salamon. The Maslov index for paths. *Topology*, 32(4):827–844, 1993.
[47] R. Roberts, J. Shareshian, and M. Stein. Infinitely many hyperbolic 3-manifolds which contain no Reebless foliation. *J. Amer. Math. Soc.*, 16(3):639–679 (electronic), 2003.
[48] P. Seidel. A long exact sequence for symplectic Floer cohomology. *Topology*, 42(5):1003–1063, 2003.
[49] A. I. Stipsicz. Contact surgery and Heegaard Floer theory. In *Floer homology, gauge theory, and low-dimensional topology*, 2006.
[50] C. H. Taubes. The Seiberg-Witten invariants and symplectic forms. *Math. Research Letters*, 1(6):809–822, 1994.
[51] C. H. Taubes. More constraints on symplectic forms from Seiberg-Witten invariants. *Math. Research Letters*, 2(1):9–13, 1995.
[52] C. H. Taubes. *Metrics, connections and gluing theorems.* Number 89 in CBMS Regional Conference Series in Mathematics. AMS, 1996.
[53] E. Witten. Monopoles and four-manifolds. *Math. Research Letters*, 1:769–796, 1994.

Department of Mathematics, Columbia University, New York 10027
E-mail address: `petero@math.columbia.edu`

Department of Mathematics, Princeton University, New Jersey 08544
E-mail address: `szabo@math.princeton.edu`

Clay Mathematics Proceedings
Volume 5, 2006

Circle Valued Morse Theory for Knots and Links

Hiroshi Goda

ABSTRACT. We apply a circle valued Morse map to the complements of knots and links in the 3-sphere, and observe their topology including the (twisted) Alexander polynomial, Novikov homology, and two types of Reidemeister torsion.

1. Introduction

Let M be a smooth manifold. Traditional Morse theory deals with a real-valued function $f : M \to \mathbb{R}$. This function corresponds to a handle decomposition of M via Morse's lemma giving Morse's inequality. It describes the relationship between the number of critical points of f and the Betti number and the torsion number of M. The critical points of a Morse function f generate the Morse-Smale complex $C^{MS}(f)$ over $\mathbb{Z}$, using the gradient flow to define the differentials. It is easy to see Morse's inequality from the isomorphism $H_*(C^{MS}(f)) \cong H_*(M)$.

The more recent Morse theory of a circle valued Morse map $f : M \to S^1$ is more complicated, but shares many features of the real valued theory. As in the case of a real valued Morse theory, we have an inequality, which is called the Morse-Novikov inequality, and then the critical points of a circle valued Morse map f generate the Novikov complex $C^{\mathrm{nov}}(f)$ over the Novikov ring $\mathbb{Z}((z))$ of formal power series with integer coefficients, using the gradient flow of the real valued Morse function $\overline{f} : \overline{M} \to \mathbb{R}$ on the infinite cyclic cover to define the differentials. The Novikov homology is the $\mathbb{Z}((z))$-coefficient homology of $\overline{M}$. This theory was started by Novikov [**31**]. See [**34**] for a survey of these topics.

Recently, there are some works on the circle valued Morse theory for the complement of a knot or link in the 3-sphere S^3. We focus on it in this paper, and give a survey.

We define a circle valued Morse map with some conditions on the complement of a link L in S^3, and the *Morse-Novikov number* $\mathcal{MN}(L)$, i.e., the minimal possible number of critical points, roughly speaking. In particular, a link L is fibred if and only if $\mathcal{MN}(L) = 0$. We observe some properties of $\mathcal{MN}(L)$ in Section 2. There is a handle decomposition which corresponds to this Morse map, which is called

1991 *Mathematics Subject Classification.* 57M25, 57M50.

The author is supported in part by the Grant-in-Aid for Scientific Research (No 15740031), The Ministry of Education, Culture, Sports, Science and Technology, Japan.

Heegaard splitting for sutured manifolds. We introduce this notion in Section 3, and give some concrete examples. Furthermore we consider the behavior of $\mathcal{MN}(L)$.

By using Alexander ideals (polynomials), we describe a Morse-type inequality, originally due to Pajitnov, Rudolph and Weber in [**33**]. In Section 4, we review their theorem and see some examples. Note that this theorem may be regarded as an extension of the results of Neuwirth [**30**] and Stallings [**38**]. In Section 5, we generalize the theorem of Neuwirth and Stallings using the twisted Alexander invariant which was defined by Wada [**39**]. This observation leads to the definition of *twisted Novikov homology*. We present the definition in Section 6, and show an estimate which generalizes the theorem of Pajitnov, Rudolph and Weber.

Hutchings and Lee showed the relationship between Reidemeister torsion of an ordinary complex and that of a Novikov complex in [**17**], [**18**]. In Section 7, we observe this result through some calculations of both torsions for some 3-manifolds. Mark's work [**26**] plays an important role here.

Terminology and notation. Throughout this paper, we work in the C^∞-category. Thus, the functions, maps, curves, etc. are assumed to be of class C^∞.

Let L be an oriented link in S^3, and let $C_L = S^3 - L$. Further, let $E(L) = S^3 - \operatorname{Int} N(L)$ be its exterior, where $N(L)$ is a regular neighborhood of L in S^3.

A *Seifert surface* is an oriented compact 2-submanifold of S^3 with no closed component. The boundary $L = \partial\bar{R}$ of a Seifert surface $\bar{R}$ is an oriented link; $\bar{R}$ is called a *Seifert surface for* L. The intersection of $\bar{R}$ with $E(L)$, $R = \bar{R} \cap E(L)$, is also called a Seifert surface for L.

2. Circle valued Morse map for knots and links

In this section, we review some definitions and the basic properties of circle valued Morse maps for knots and links.

A Morse map $f : C_L \to S^1$ is said to be *regular* if each component L_i of L has a neighborhood framed as $S^1 \times D^2$ such that $L_i = S^1 \times \{0\}$ and the restriction $f| : S^1 \times (D^2 - \{0\}) \to S^1$ is given by $(x, y) \to y/|y|$. Let $m_p(f)$ be the number of critical points of f of index p. We say that a Morse map $f : C_L \to S^1$ is *minimal* if it is regular and for each p, $m_p(f)$ is minimal possible among all regular maps homotopic to f. Suppose f is minimal. We call $\mathcal{MN}(L) = \sum_p m_p(f)$ the *Morse-Novikov number* of L. Note that even in the case of a real valued Morse function on a manifold M, minimal Morse functions do not always exist. The problem is that, in general, $m_p(f)$ cannot be minimized for all indices p simultaneously. However, in [**33**], Pajitnov, Rudolph and Weber show that in the case where $M = C_L$ a minimal Morse map exists with some nice properties.

Definition 2.1. A regular Morse map $f : C_L \to S^1$ is said to be *moderate* if

(i) $m_0(f) = m_3(f) = 0$,
(ii) all critical values corresponding to critical points of the same index coincide,
(iii) $f^{-1}(x)$ is a connected Seifert surface for any regular value $x \in S^1$.

Theorem 2.2 ([**33**]). *Every link has a minimal Morse map which is moderate.*

From this theorem we have:

Corollary 2.3. (1) Let f be a moderate map; then $m_1(f) = m_2(f)$.

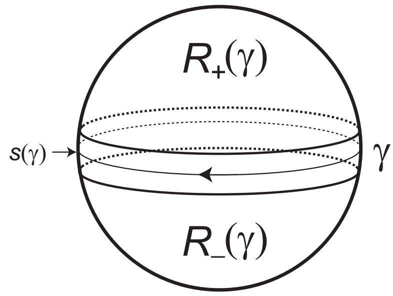


FIGURE 1. The trivial sutured manifold

(2) Let f be a regular Morse map realizing $\mathcal{MN}(L)$; then we may suppose $\mathcal{MN}(L) = m_1(f) + m_2(f)$.

3. Heegaard splitting for sutured manifolds

3.1. Definition. The concept of sutured manifold was defined by Gabai [**7**]. It is a very useful tool in studying knots and links. Here we present an application.

First, we define a sutured manifold in our setting.

DEFINITION 3.1 (sutured manifold). A sutured manifold (M, γ) is a compact oriented 3-manifold M together with a subset $\gamma \subset \partial M$ which is a union of finitely many mutually disjoint annuli. For each component of γ, a *suture*, that is, an oriented core circle, is fixed, and $s(\gamma)$ denotes the set of sutures. Every component of $R(\gamma) = \partial M - \text{Int}\,\gamma$ is oriented so that the orientations on $R(\gamma)$ are coherent with respect to $s(\gamma)$, i.e., the orientation of each component of $\partial R(\gamma)$, which is induced by that of $R(\gamma)$, is parallel to the orientation of the corresponding component of $s(\gamma)$. Let $R_+(\gamma)$ (resp. $R_-(\gamma)$) denotes the union of those components of $R(\gamma)$ whose normal vectors point out of (resp. into) M. In the case that (M, γ) is homeomorphic to $(F \times [0, 1], \partial F \times [0, 1])$ where F is a compact oriented 2-manifold, (M, γ) is called a *product sutured manifold.*

Let L be an oriented link in S^3, and $\bar{R}$ a Seifert surface of L. Set $R = \bar{R} \cap E(L)$ ($E(L) = \text{cl}(S^3 - N(L))$), and $(P, \delta) = (N(R, E(L)), N(\partial R, \partial E(L)))$. We call (P, δ) a *product sutured manifold for* R. Let $(M, \gamma) = (\text{cl}(E(L) - P), \text{cl}(\partial E(L) - \delta))$ with $R_\pm(\gamma) = R_\mp(\delta)$. We call (M, γ) a *complementary sutured manifold* for R. In this paper, we call this a *sutured manifold* for short.

Let (V, γ) be a sutured manifold such that V is a 3-ball and γ is an annulus embedded in ∂V. Then, we call (V, γ) the *trivial sutured manifold*. See Figure 1.

EXAMPLE 3.2. Let K be the trivial knot. Then K has a Seifert surface D that is a disk. The product sutured manifold for D is the trivial sutured manifold. Further, the (complementary) sutured manifold for D is also the trivial sutured manifold.

EXAMPLE 3.3. The left hand side figure in Figure 2 is the trefoil knot K, and the middle is a Seifert surface R of K. The (complementary) sutured manifold for R is homeomorphic to the manifold the right hand side of the figure. (Note that the 'outside' of the genus 2 surface is the complementary sutured manifold.)

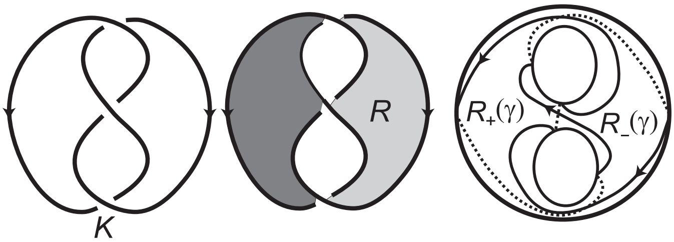


FIGURE 2

In [**2**], the notion of compression body was introduced. It is a generalization of a handlebody, and important in defining a Heegaard splitting for 3-manifolds with boundaries.

DEFINITION 3.4 (compression body). A *compression body* W is a cobordism rel ∂ between surfaces $\partial_+ W$ and $\partial_- W$ such that $W \cong \partial_+ W \times I \cup$ 2-handles $\cup$ 3-handles and $\partial_- W$ has no 2-sphere components. We can see that if $\partial_- W \neq \emptyset$ and W is connected, W is obtained from $\partial_- W \times I$ by attaching a number of 1-handles along the disks on $\partial_- W \times \{1\}$, where $\partial_- W$ corresponds to $\partial_- W \times \{0\}$.
We denote by $h(W)$ the number of these 1-handles.

DEFINITION 3.5. (W, W') is a Heegaard splitting for (M, γ) if

(i) W, W' are connected compression bodies,
(ii) $W \cup W' = M$,
(iii) $W \cap W' = \partial_+ W = \partial_+ W', \partial_- W = R_+(\gamma)$, and $\partial_- W' = R_-(\gamma)$.

Set $h(R) = \min\{h(W)(= h(W')) \mid (W, W')$ is a Heegaard splitting for the sutured manifold of $R\}$. We call $h(R)$ the *handle number* of R. The handle number is an invariant of a Seifert surface. A link L is *fibred* if L has a Seifert surface R such that $h(R) = 0$, i.e., the sutured manifold for R is a product sutured manifold. We call this Seifert surface a *fibre surface*, that is, R is a fibre surface if and only if $h(R) = 0$. It is known that a fibre surface of a fibred link L is a minimal genus Seifert surface of L.

EXERCISE 3.6. Confirm that the trivial knot is a fibred knot.

The disk is the fibre surface, and the sutured manifold is a 3-ball which decomposes into two copies of (disk)$\times[0, 1]$.

Suppose a Morse map f is moderate; then, as in case of real valued Morse theory, we observe that there is a correspondence between f and a Heegaard splitting for the sutured manifold for a Seifert surface. The handle number is the number of 1-handles, while the Morse-Novikov number stands for the number of 1-handles and 2-handles, i.e., 2$\times$(the number of 1-handles). Hence we have:

PROPOSITION 3.7. Let L be an oriented link in S^3; then

$$\mathcal{MN}(L) = 2 \times \min\{h(R) \mid R \text{ is a Seifert surface for } L\}$$

Let R be a Seifert surface for a link L. We define the $\mathcal{MN}(R) = 2 \times h(R)$, and call it the *Morse-Novikov number for a Seifert surface* R. Thus $\mathcal{MN}(L) = 0$ if and only if L is fibred.

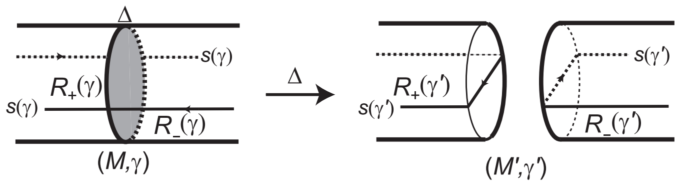


FIGURE 3. Product decomposition

3.2. Detecting fibred links. Gabai gave a useful method to detect fibred links in [**8**]. We review it in this subsection. A key method is a 'product decomposition'.

DEFINITION 3.8. (product decomposition) Let (M, γ) be a sutured manifold. A *product disk* $\Delta(\subset M)$ is a properly embedded disk such that $\partial\Delta$ intersects $s(\gamma)$ transversely in two points. We obtain a new sutured manifold (M', γ') from (M, γ) by cutting along Δ and connecting $s(\gamma)$ naturally. See Figure 3. This decomposition

$$(M, \gamma) \xrightarrow{\Delta} (M', \gamma')$$

is called a *product decomposition*.

In [**8**], the next theorem is proved:

THEOREM 3.9 ([**8**]). *Let L be a link and R a Seifert surface of L. Then L is a fibre surface with fibre surface R if and only if there exists a sequence of product decompositions:*

$$(M, \gamma) \xrightarrow{\Delta_1} (M_1, \gamma_1) \xrightarrow{\Delta_2} \cdots \xrightarrow{\Delta_n} (M_n, \gamma_n)$$

such that (M, γ) is the sutured manifold for R and (M_n, γ_n) is a union of trivial sutured manifolds.

PROOF. (only if part) From the definition, L is a fibred link with fibre surface R if and only if $(S^3 - \text{Int } N(L)) - \text{Int } N(R) \cong R \times [0, 1]$. Let $\Delta_i = \alpha_i \times [0, 1]$, where $\alpha_1, \ldots, \alpha_n$ is a set of mutually disjoint properly embedded arcs in R such that $R - \sum_{i=1}^{n} \text{Int } N(\alpha_i) = D^2$: a disk. It follows that there is the sequence of product decompositions:

$$(R \times [0, 1], \partial R \times [0, 1]) \cong (M, \gamma) \xrightarrow{\Delta_1} (M_1, \gamma_1) \xrightarrow{\Delta_2} \cdots \xrightarrow{\Delta_n} (M_n, \gamma_n)$$

such that (M_n, γ_n) is the trivial sutured manifold.

(if part) Let $(M_{i-1}, \gamma_{i-1}) \xrightarrow{\Delta_i} (M_i, \gamma_i)$ be a product decomposition, and suppose (M_i, γ_i) is a product sutured manifold. By the definition of a product decomposition, (M_{i-1}, γ_{i-1}) inherits the product property from (M_i, γ_i) under the converse operation of the product decomposition. Since (M_n, γ_n) is a product sutured manifold, then (M, γ) becomes a product sutured manifold by induction. □

EXAMPLE 3.10. Let R be the Hopf band as illustrated in Figure 4. Then the sutured manifold for R, (M, γ), consists of the solid torus and 2 component annuli as in Figure 4. By the product decomposition $(M, \gamma) \xrightarrow{\Delta_1} (M_1, \gamma_1)$, we have the trivial sutured manifold. Thus the Hopf band R is the fibre surface.

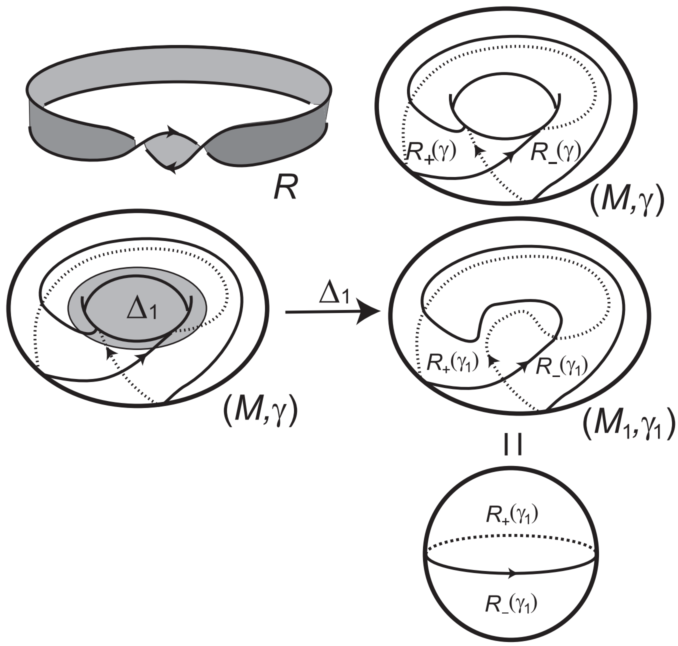


FIGURE 4

EXAMPLE 3.11. Let K be the trefoil knot and R a Seifert surface illustrated in Figure 2. We can see that K is fibred with fibre surface R as follows. Let (M, γ) be the sutured manifold for R and Δ_1 the product disk for (M, γ) as in Figure 5. After the product decomposition along Δ_1, we have the sutured manifold (M_1, γ_1) in Figure 5. Similarly, we denote by Δ_2 the product disk for (M_1, γ_1) in Figure 5, and then we obtain the trivial sutured manifold (M_2, γ_2) by the product decomposition along Δ_2. Thus the trefoil knot K is a fibred knot with fibre surface R.

EXERCISE 3.12. Find a fibred knot and fibre surface for the prime knots of $\leq$ 7 crossings in the list of Rolfsen [**36**].

3.3. Some calculations of Morse-Novikov number. We present some examples of calculations of Morse-Novikov numbers in this subsection.

EXAMPLE 3.13. Let L be the trivial link with 2 components. Suppose that the annulus R is the Seifert surface of L. The sutured manifold for R, say (M, γ), is the solid torus $S^1 \times D^2$ with 2 sutures, each of which bounds a meridian disk. Let α be an arc properly embedded in M with $\partial\alpha \subset R_+(\gamma)$ as illustrated in Figure 6. It is easy to see that the regular neighborhood of $R_+(\gamma) \cup \alpha$ in M, say W, is homeomorphic to $R_+(\gamma) \times [0, 1] \cup$ (a 1-handle), namely, it is a compression body. On the other hand, we can see that $\mathrm{cl}(M - W) = W'$ is homeomorphic to $R_-(\gamma) \times [0, 1] \cup$ (a 1-handle). Thus (W, W') is a Heegaard splitting for (M, γ). Hence, we have $\mathcal{MN}(R) \leq 2$ $(h(R) \leq 1)$.

EXERCISE 3.14. Let L be the μ component trivial link. Then L spans the μ punctured sphere R as a Seifert surface. Show that $\mathcal{MN}(R) \leq 2(\mu - 1)$.

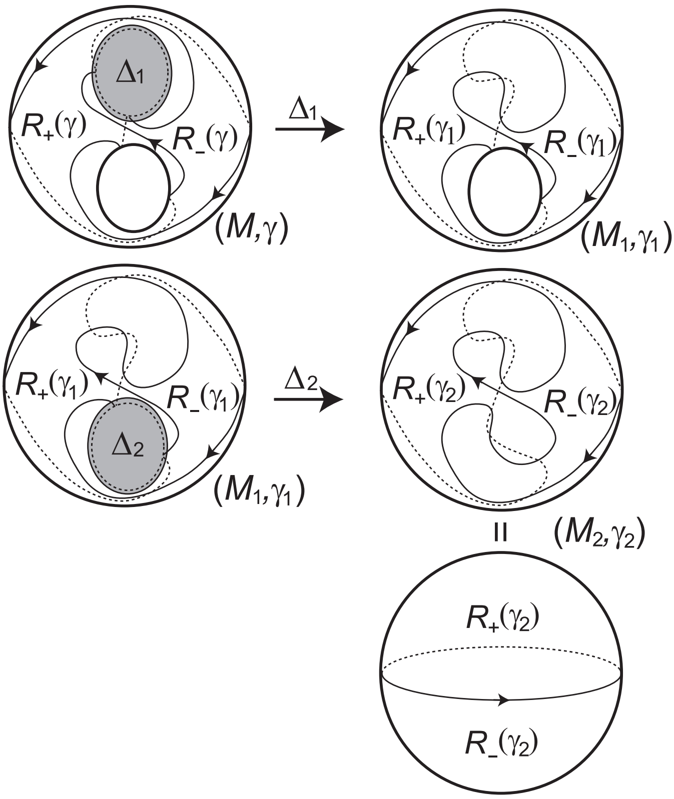


FIGURE 5

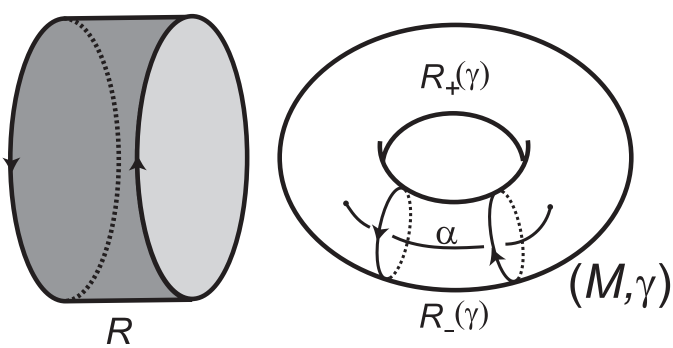


FIGURE 6

Indeed, we have $\mathcal{MN}(L) = \mathcal{MN}(R) = 2(\mu - 1)$ by Example 4.9.

Let W be a compression body with $\partial(\partial_- W) \neq \emptyset$. If we set $\gamma = \partial(\partial_- W) \times [0,1]$ and regard $\partial_+ W$ (resp. $\partial_- W$) as $R_+(\gamma)$ (resp. $R_-(\gamma)$), then W can be regarded as a sutured manifold. We denote by (W, γ) this sutured manifold.

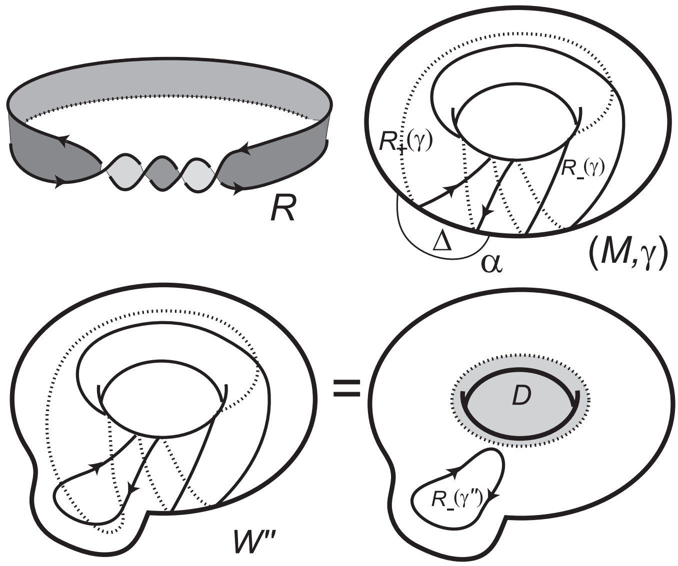


FIGURE 7

LEMMA 3.15. *Let* $(W, \gamma) \xrightarrow{\Delta} (W', \gamma')$ *be a product decomposition. Then* W *is a compression body if and only if* W' *is a compression body. Moreover,* $h(W) = h(W')$.

EXAMPLE 3.16. Let R be an unknotted annulus with 2-full twists. Then the sutured manifold for R, say (M, γ), is formed as in Figure 7. Note that M is the 'outside' of the torus indicated in the figure. Let α be an arc properly embedded in M with $\partial\alpha \subset R_+(\gamma)$ as illustrated in Figure 7. The regular neighborhood of $R_+(\gamma) \cup \alpha$ in M, say W, is a compression body with $h(W) = 1$. In what follows, we will see that the complement, i.e., $\mathrm{cl}(M - W) = W'$, is also a compression body so that (W, W') is a Heegaard splitting for (M, γ). Let Δ be a disk properly embedded in W' as in Figure 7. We may regard W' as a sutured manifold (W', γ'), so Δ is a product disk in the sutured manifold W'. Consider the product decomposition $(W'\gamma') \xrightarrow{\Delta} (W'', \gamma'')$. The sutured manifold (W'', γ'') is a compression body such that W'' is homeomorphic to (a disk)$\times[0, 1] \cup$ (a 1-handle). In the figure D shows the cocore of the 1-handle. By Lemma 3.15, we have that W' is a compression body with $h(W') = 1$. Therefore, (W, W') is a Heegaard splitting for (M, γ). Thus we have $\mathcal{MN}(R) \leq 2$.

Actually, $\mathcal{MN}(R) = 2$. See Example 4.10.

EXERCISE 3.17. Let R be an unknotted annulus with n-full twists ($n \geq 3$). Observe that $\mathcal{MN}(R) \leq 2$.

EXAMPLE 3.18. Let L be the pretzel link of type $(4, 4, 4)$ with the orientation given in Figure 8. We denote by R the Seifert surface of L as illustrated in the figure. As in the previous examples, let α_1 and α_2 be arcs properly embedded in M with $\partial\alpha_1, \partial\alpha_2 \subset R_+(\gamma)$, where (M, γ) is the sutured manifold for R. Let $W =$ the regular neighborhood of $R_+(\gamma) \cup \alpha_1 \cup \alpha_2$ and $W' =\mathrm{cl}(M - W)$. Then, by the

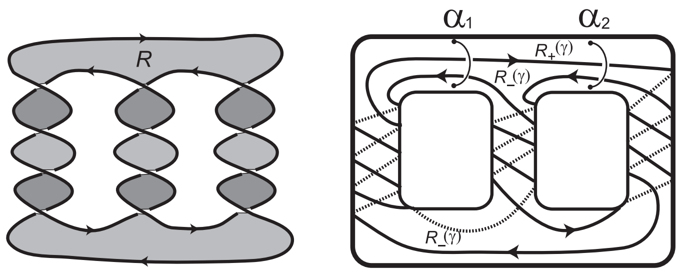


FIGURE 8

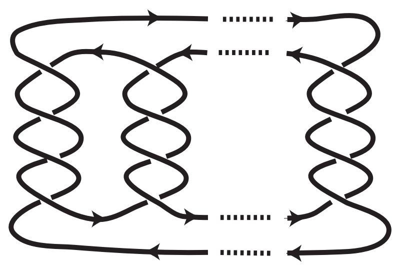


FIGURE 9

same argument as in the previous example, we can show that (W, W') is a Heegaard splitting for (M, γ) with $h(W) = h(W') = 2$. Hence we have $\mathcal{MN}(R) \leq 4$.

EXERCISE 3.19. Let L be the pretzel link of type $\underbrace{(4, 4, \ldots, 4)}_{n}$ with the orientation given in Figure 9. Show that $\mathcal{MN}(R) \leq 2(n-1)$.

In Example 4.11, we shall see that $\mathcal{MN}(L) = \mathcal{MN}(R) = 2(n-1)$.

EXAMPLE 3.20. Let K be a non-fibred prime knot of up to 10 crossings listed in [**36**]. Then, $\mathcal{MN}(K) = 2$. See [**10**].

In 2001, Hirasawa proved the next theorem, but it has not been published.

THEOREM 3.21. *Let K be a non-fibred 2-bridge knot. Then, $\mathcal{MN}(K) = 2$.*

3.4. Morse-Novikov number under a connected sum and Murasugi sum. In this subsection, we see the behavior of Morse-Novikov number of links or Seifert surfaces under connected sum and Murasugi sum. An oriented surface $\bar{R} \subset S^3$ is a $(2n)$-*Murasugi sum* of compact oriented surfaces $\bar{R}_1$ and $\bar{R}_2$ if there are 3-balls V_1 and V_2 satisfying the following property:

$$V_1 \cup V_2 = S^3, V_1 \cap V_2 = \partial V_1 = \partial V_2, \bar{R}_i \subset V_i (i = 1, 2),$$
$$\bar{R} = \bar{R}_1 \cup \bar{R}_2 \text{ and } D = \bar{R}_1 \cap \bar{R}_2 \text{ is a } 2n\text{-gon.}$$

The 2-Murasugi sum is called a *connected sum* of $\bar{R}_1$ and $\bar{R}_2$. The 4-Murasugi sum is called a *plumbing* of $\bar{R}_1$ and $\bar{R}_2$. Set $L = \partial\bar{R}, L_i = \partial\bar{R}_i, R = \bar{R} \cap E(L)$ and $R_i = \bar{R}_i \cap E(L_i)$. Then we will also say that R is a Murasugi sum of R_1 and R_2. Here we can see that $\bar{R}' = (\bar{R} - D) \cup D'$ is an oriented surface with $\partial\bar{R}' = L$ where $D' = \partial V_1 - \operatorname{Int} D$. By a tiny isotopy of S^3 keeping L fixed we can move $\bar{R}'$ so that

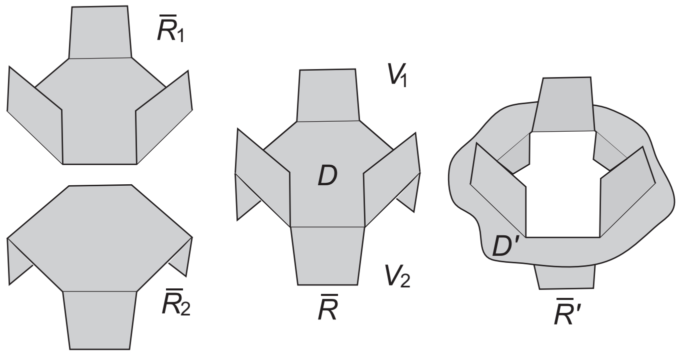


FIGURE 10. 6-Murasugi sum

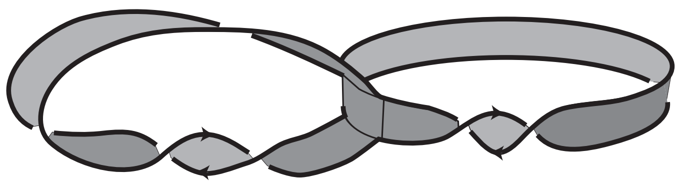


FIGURE 11

$\bar{R}' \cap \bar{R} \cap E(L) = \emptyset$. We will say that $\bar{R}'$ ($R' = \bar{R}' \cap E(L_i)(i = 1, 2)$ resp.) is the *dual* Seifert surface of $\bar{R}$ (R resp.). Note that $\bar{R}'$ (R' resp.) is also a Murasugi sum of $\bar{R}'_1$ and $\bar{R}'_2$ (R'_1 and R'_2) where $\bar{R}'_i = (\bar{R}_i - D) \cup D'$ ($R'_i = \bar{R}'_i \cap E(L_i)$ resp.) $(i = 1, 2)$.

Gabai [**6**] showed the following:

THEOREM 3.22. *Let $\bar{R}$ be a Murasugi sum of two surfaces $\bar{R}_1$ and $\bar{R}_2$. Then $L = \partial \bar{R}$ is a fibred link with fibre surface R if and only if for $i = 1, 2$, $L_i = \partial \bar{R}_i$ is a fibre link with fibre surface $\bar{R}_i$.*

EXERCISE 3.23. Show that the surface illustrated in Figure 11 is a fibre surface (cf. Example 3.10).

The connected sum can be defined without surfaces, while the $2n$-Murasugi sum ($n \geq 2$) must be defined by using surfaces. The behavior of the Morse-Novikov number under a Murasugi sum with surfaces is known; however, the behavior of Morse-Novikov number under connected sum of two knots is still not known. See the next open problem.

THEOREM 3.24 ([**9**]). *Let R be a $2n$-Murasugi sum of R_1 and R_2; then*

$$\mathcal{MN}(R_1) + \mathcal{MN}(R_2) - 2(n-1) \leq \mathcal{MN}(R) \leq \mathcal{MN}(R_1) + \mathcal{MN}(R_2).$$

In [**10**], we can find another estimate in the case of plumbing. Moreover, it was shown that there are a Seifert surface R and its dual R' such that $\mathcal{MN}(R) \neq \mathcal{MN}(R')$.

THEOREM 3.25 ([**9**]). *Let R be a $2n$-Murasugi sum of R_1 and R_2. Suppose R_1 is a fiber surface, i.e., $\mathcal{MN}(R_1) = 0$. Then $\mathcal{MN}(R) = \mathcal{MN}(R_2)$.*

COROLLARY 3.26. Let $K_1\sharp K_2$ be the knot obtained from knots K_1 and K_2 by a connected sum. Then,

$$\mathcal{MN}(K_1\sharp K_2) \leq \mathcal{MN}(K_1) + \mathcal{MN}(K_2).$$

Open problem([**33**]). Is it true that $\mathcal{MN}(K_1\sharp K_2) = \mathcal{MN}(K_1) + \mathcal{MN}(K_2)$?

4. Morse-Novikov inequality for knots and links

In this section, we introduce an inequality which is called the Morse-Novikov inequality for knots and links. This inequality and the sutured manifold theory stated in the previous section combine favorably to enable calculations in some cases.

4.1. Morse-Novikov inequality. Let L be an oriented link in S^3. Since S^3 is oriented, so is the normal circle bundle of L. For each component L_i of L, there is a unique element $\mu_i \in H_1(C_L)$ represented by any oriented fiber of the normal circle bundle of L_i. There is a unique cohomology class $\xi_L \in H^1(C_L)$ such that for each i we have $\xi_L(\mu_i) = 1$. Let $\overline{C_L} \to C_L$ be an infinite cyclic covering associated with this cohomology class.

Set $\Lambda = \mathbb{Z}[t, t^{-1}]$ and

$$\widehat{\Lambda} = \mathbb{Z}((t)) = \mathbb{Z}[[t]][t^{-1}] = \{\Sigma_{n=0}^{k} a_n t^{-n} \mid a_n \in \mathbb{Z}[[t]]\},$$
$$\text{where } \mathbb{Z}[[t]] = \{a_n(t) = \Sigma_{n=0}^{\infty} a'_n t^n \mid a'_n \in \mathbb{Z}\}.$$

That is,

$$\widehat{\Lambda} = \{\Sigma_{n=-\infty}^{\infty} a'_n t^n,\ (a'_n \in \mathbb{Z}) \text{ such that } \{n \leq 0 \mid a'_n \neq 0\} \text{ is finite}\}.$$

Note that the homology $H_*(\overline{C_L})$ is a Λ-module. We denote $H_*(\overline{C_L}) \otimes_\Lambda \widehat{\Lambda}$ by $\widehat{H}_*(L)$, and set $\widehat{b}_i(L) = \mathrm{rank}_{\widehat{\Lambda}} \widehat{H}_i(L)$, $\widehat{q}_i(L) =$ the torsion number of $\widehat{H}_i(L)$. Here, the 'torsion number' means the minimal possible number of generators of the torsion part over $\widehat{\Lambda}$.

THEOREM 4.1 ([**31**], [**33**], Morse-Novikov inequality for knots and links).

$$m_i(f) \geq \widehat{b}_i(L) + \widehat{q}_i(L) + \widehat{q}_{i-1}(L).$$

By using the Poincaré duality theorem and the fact that link complements are 3-dimensional manifolds (see [**33**] for the details), we have:

COROLLARY 4.2. $\mathcal{MN}(L) \geq 2 \times (\widehat{b}_1(L) + \widehat{q}_1(L))$.

In what follows, we study how to compute $\widehat{b}_1(L)$ and $\widehat{q}_1(L)$.

Let A be the Alexander matrix of a link L, whose size is $n \times n$. We say that Δ_s is the *s-th Alexander polynomial* of L if Δ_s is the greatest common divisor of the determinants of all $(n-s) \times (n-s)$ minors of A. For the details of the Alexander matrix and s-th Alexander (link) polynomials, see [**1**] and [**4**].

LEMMA 4.3 ([**33**]).

$$\widehat{H}_1(L) \cong \bigoplus_{s=0}^{m-1} \widehat{\Lambda}/\gamma_s\widehat{\Lambda},$$

where Δ_s is the s-th Alexander polynomial (Δ_0 is the Alexander polynomial), and $\gamma_s = \Delta_s/\Delta_{s+1}$. In particular, $\gamma_{s+1}|\gamma_s$ for every s.

Note that we use here the notations $0/0 = 0$ and $\Delta_m = 1$. In order to prove this lemma, we have to check 2 points. One of them is the fact that $\widehat{\Lambda}$ is a principal ideal domain, so that $\widehat{H}_*(L)$ admits a decomposition into a direct sum of cyclic modules. Another is the fact that $\mathrm{GCD}_\Lambda(a,b) = \mathrm{GCD}_{\widehat{\Lambda}}(a,b)$ for $a, b \in \Lambda$.

A polynomial $a_m t^m + \cdots + a_1 t + a_0 \in \mathbb{F}[t]$ is called *monic* if the coefficient a_m is one.

THEOREM 4.4 ([**33**]). (1) $\widehat{b}_1(L) =$ *the number of polynomials* Δ_s *that are equal to* 0.
(2) $\widehat{q}_1(L) =$ *the number of* γ_s *that are nonzero and nonmonic.*

REMARK 4.5. If L is a knot, then $\widehat{b}_1(L) = 0$ ([**25**]). Therefore, the monic nature of s-th Alexander polynomials is crucial in estimating the Morse-Novikov number of a knot. The most recent version of Kodama's soft 'knot' [**24**] can calculate s-th Alexander polynomials (elementary ideals).

The following was shown by several knot theorists, see [**30**], [**35**] and [**38**].

COROLLARY 4.6. Suppose that L is fibred; then the Alexander polynomial of L is monic.

We shall generalize this corollary in Section 5.

EXERCISE 4.7. Check the Alexander polynomial for knots up to 7 crossings in the list of [**36**], and compare Exercise 3.12.

EXERCISE 4.8. Determine the Morse-Novikov number for twist knots. (See [**1**] for the definition of the twist knot. Note that these knots have genus one Seifert surfaces.)

4.2. Examples. Here we present some examples.

EXAMPLE 4.9 ([**33**]). Let L be the trivial link with μ components. Obviously, $\widehat{b}_1(L) = \widehat{b}_2(L) = \mu - 1$. By Corollary 4.2, we have $\mathcal{MN}(L) \geq 2(\mu - 1)$. Together with Example 3.13 and Exercise 3.14, we have $\mathcal{MN}(L) = 2(\mu - 1)$.

EXAMPLE 4.10. Let $\bar{R}$ be an unknotted annulus with n-full twists ($n \geq 1$). See Figure 7 for the case of $n = 2$. Set $L = \partial \bar{R}$. Then $\widehat{H}_1(L) = \widehat{\Lambda}/n(1-t)\widehat{\Lambda}$. Hence, $\mathcal{MN}(L) \geq 0$ if $n = 1$, and $\mathcal{MN}(L) \geq 2$ if $n \geq 2$. Together with Examples 3.10, 3.16 and Exercise 3.17, we have $\mathcal{MN}(L) = 0$ if $n = 1$, and $\mathcal{MN}(L) = 2$ if $n \geq 2$.

EXAMPLE 4.11. Let L be the pretzel link of type $\underbrace{(4, 4, \ldots, 4)}_{n}$ that is oriented as in Figure 9. Then we have:

$$\begin{aligned}
\Delta_0(L) &= n \cdot 2^{n-1} \cdot (1-t)^{n-1}, & \gamma_0 &= n \cdot 2 \cdot (1-t), \\
\Delta_1(L) &= 2^{n-2} \cdot (1-t)^{n-2}, & \gamma_1 &= 2 \cdot (1-t), \\
\Delta_2(L) &= 2^{n-3} \cdot (1-t)^{n-3}, & \gamma_2 &= 2 \cdot (1-t), \\
&\vdots & &\vdots \\
\Delta_{n-2}(L) &= 2 \cdot (1-t), & \gamma_{n-2} &= 2 \cdot (1-t). \\
\Delta_{n-1}(L) &= 1.
\end{aligned}$$

Thus we have $\mathcal{MN}(L) \geq 2(n-1)$. Together with Exercise 3.19, we have $\mathcal{MN}(L) = 2(n-1)$.

Let K be a prime knot of up to 10 crossings listed in Rolfsen's book [**36**]. In [**19**], Kanenobu has checked that K is fibred if and only if its Alexander polynomial $\Delta_0(K)$ is monic. Then, together with Example 3.20, we have:

EXAMPLE 4.12. Equality in the Morse-Novikov inequality holds for prime knots of up to 10 crossings.

Note that Nakanishi calculated the s-th Alexander polynomials for prime knots of up to 10 crossings, see [**20**].

5. Twisted Alexander invariants and fibred knots

In this section, we generalize Corollary 4.6 using the twisted Alexander invariant, and present some examples. Twisted Alexander invariants were defined by several people. Here we use Wada's notation [**39**]. Let us start with the definition.

5.1. Definition of Twisted Alexander invariant. In [**39**], Wada defined the twisted Alexander invariant for finitely presentable groups. Here we focus on a knot group. Let K be a knot in the 3-sphere and $G(K)$ the fundamental group of the exterior $E = S^3 -$ Int $N(K)$ of K, i.e., the knot group of K. We denote by $F_u = \langle x_1, x_2, \ldots, x_u \rangle$ a free group of rank u and $T = \langle t \rangle$ an infinite cyclic group. We choose and fix a Wirtinger presentation:

$$P(G(K)) = \langle\ x_1, x_2, \ldots, x_u \mid r_1, r_2, \ldots, r_{u-1}\ \rangle$$

of $G(K)$ and let

$$\phi :\ F_u \longrightarrow G(K)$$

be the associated surjective homomorphism of the free group F_u to the knot group $G(K)$. This ϕ induces a ring homomorphism

$$\widetilde{\phi} :\ \mathbb{Z}[F_u] \longrightarrow \mathbb{Z}[G(K)].$$

On the other hand, the canonical abelianization

$$\alpha :\ G(K) \longrightarrow H_1(E;\mathbb{Z}) \cong T$$

is given by

$$\alpha(x_1) = \alpha(x_2) = \cdots = \alpha(x_u) = t,$$

and the homomorphism α induces a ring homomorphism of the integral group ring

$$\widetilde{\alpha} :\ \mathbb{Z}[G(K)] \longrightarrow \mathbb{Z}[t^{\pm 1}].$$

Let

$$\rho :\ G(K) \longrightarrow GL(n, \mathbb{F})$$

be a representation, where $\mathbb{F}$ is a field. The corresponding ring homomorphism of the integral ring $\mathbb{Z}[G(K)]$ to the matrix algebra $M(n, \mathbb{F})$ of degree n over $\mathbb{F}$ is denoted by

$$\widetilde{\rho} :\ \mathbb{Z}[G(K)] \longrightarrow M(n, \mathbb{F}).$$

We denote by Φ the composed mapping of the ring homomorphism $\widetilde{\phi}$ and the tensor product homomorphism

$$\widetilde{\rho} \otimes \widetilde{\alpha} : \mathbb{Z}[G(K)] \longrightarrow M(n, \mathbb{F}[t^{\pm 1}]).$$

That is,

$$\Phi = (\widetilde{\rho} \otimes \widetilde{\alpha}) \circ \widetilde{\phi} : \mathbb{Z}[F_u] \longrightarrow M(n, \mathbb{F}[t^{\pm 1}]).$$

Let us consider the $(u-1) \times u$ matrix $M_{\rho\otimes\alpha}$ whose (i,j)th component is the $n \times n$ matrix

$$\Phi\Big(\frac{\partial x_i}{\partial x_j}\Big) \in M(n, \mathbb{F}[t^{\pm 1}]),$$

where $\dfrac{\partial}{\partial x}$ denotes the free differential calculus [**4**]. This matrix $M_{\rho\otimes\alpha}$ is called the *Alexander matrix of the presentation $P(G(K))$ associated to the representation ρ.* We note that the classical Alexander matrix is $M_{\mathbf{1}\otimes\alpha}$, where $\mathbf{1}$ is the 1-dimensional trivial representation of $G(K)$. For $1 \le {}^\forall j \le u$, we denote by $M^j_{\rho\otimes\alpha}$ the $(u-1) \times (u-1)$ matrix obtained from $M_{\rho\otimes\alpha}$ by removing the j-th column. Finally, we may regard $M^j_{\rho\otimes\alpha}$ as a $(u-1)n \times (u-1)n$ matrix with coefficients in $\mathbb{F}[t^{\pm 1}]$.

The following two lemmas are the foundation for the definition of the twisted Alexander invariant.

Lemma 5.1. $\det \Phi(x_j - 1) \neq 0$ *for* $1 \le {}^\forall j \le u$.

Lemma 5.2. $\det M^j_{\rho\otimes\alpha} \det \Phi(x_{j'} - 1) = \pm \det M^{j'}_{\rho\otimes\alpha} \det \Phi(x_j - 1)$, *for* $1 \le {}^\forall j < {}^\forall j' \le u$.

Definition 5.3. We may define the *twisted Alexander invariant* of a knot K associated to the representation ρ to be the rational expression

$$\Delta_{K,\rho}(t) = \frac{\det M^j_{\rho\otimes\alpha}}{\det \Phi(x_j - 1)}.$$

Remark 5.4. Let $\Delta_K(t)$ be the Alexander polynomial of a knot K, and let $\rho_0 = \mathbf{1} : G(K) \to \mathbb{R} - \{0\}$ be the 1-dimensional trivial representation. Then we have:

$$\Delta_{K,\mathbf{1}}(t) = \frac{\Delta_K(t)}{t-1}.$$

The right hand side is the Reidemeister torsion of the knot complement of K ([**28**]). We will see this observation in Section 7 again. More generally, the twisted Alexander invariant also may be regarded as a Reidemeister torsion [**22**].

For the next theorem, see [**39**] or [**22**].

Theorem 5.5. *The twisted Alexander invariant $\Delta_{K,\rho}(t)$ is well-defined up to a factor $\pm t^{nk}$ $(k \in \mathbb{Z})$ if n is odd and up to only t^{nk} if n is even.*

The following example is given by Wada [**39**].

In order to calculate the (twisted) Alexander polynomial, it is convenient to deal with relations rather than relators. A relation $v = w$ $(v, w \in F_u)$ corresponds to the relator vw^{-1}. From $d(vw^{-1}) = dv - (vw^{-1})dw$, we obtain:

$$\Phi\Big(\frac{\partial}{\partial x_j}(vw^{-1})\Big) = \Phi\Big(\frac{\partial}{\partial x_j}(v - w)\Big), \ \ (j = 1, \dots s).$$

This shows that we may use $r = v - w$ instead of $r = vw^{-1}$ for the computation of the Alexander matrix.

EXAMPLE 5.6. Let K be the trefoil knot (see Figure 2). The knot group $G(K)$ has a representation

$$G(K) = \langle\ x, y \mid xyx = yxy\ \rangle.$$

Let us write $r = xyx - yxy$. The free derivatives of r are

$$\frac{\partial r}{\partial x} = 1 - y + xy,$$

and

$$\frac{\partial r}{\partial y} = -1 + x - yx.$$

First, we consider the trivial 1-dimensional representation over $\mathbb{Z}$: $\rho_0 : G(K) \to GL(n, \mathbb{Z})$, namely, $\rho_0(x) = \rho_0(y) = 1$. Then we have:

$$\Big(\Phi\Big(\frac{\partial r}{\partial x}\Big), \Phi\Big(\frac{\partial r}{\partial y}\Big)\Big) = (1 - t + t^2, -1 + t - t^2)$$

and

$$\Phi(x - 1) = \Phi(y - 1) = t - 1.$$

Thus the twisted Alexander invariant of $G(K)$ associated to ρ_0 is

$$\Delta_{K,\rho_0}(t) = \frac{t^2 - t + 1}{t - 1}.$$

Compare Remark 5.4.

Next, let us consider the 2-dimensional representation

$$\rho : G(K) \to GL(2, \mathbb{Z}[s^{\pm 1}])$$

of $G(K)$ over the Laurent polynomial ring $\mathbb{Z}[s^{\pm 1}]$, known as the reduced Burau representation of the braid group B_3. It is given by

$$\rho(x) = \begin{pmatrix} -s & 1 \\ 0 & 1 \end{pmatrix}, \text{ and } \rho(y) = \begin{pmatrix} 1 & 0 \\ s & -s \end{pmatrix}$$

Then we have:

$$\det \Phi\Big(\frac{\partial r}{\partial x}\Big) = \det \begin{pmatrix} 1 - t & -st^2 \\ -st + st^2 & 1 + st - st^2 \end{pmatrix} = (1 - t)(1 + st)(1 - st^2)$$

and

$$\det \Phi(y - 1) = \det \begin{pmatrix} t - 1 & 0 \\ st & -st - 1 \end{pmatrix} = (1 - t)(1 + st)$$

Hence, the twisted Alexander invariant of $G(K)$ associated to ρ is

$$\Delta_{K,\rho}(t) = \frac{(1 - t)(1 + st)(1 - st^2)}{(1 - t)(1 + st)} = 1 - st^2.$$

In general, it is not easy to calculate the twisted Alexander invariant. However, Kodama ([**24**]) wrote software to calculate it.

5.2. Twisted Alexander invariant and fibred knots. In this subsection, we give a necessary condition for a knot K in S^3 to be fibred. This is a 'twisted' version of Corollary 4.6.

THEOREM 5.7 ([**11**]). *For a fibred knot K in S^3 and a unimodular representation $\rho : G(K) \to SL(2n, \mathbb{F})$, the twisted Alexander invariant $\Delta_{K,\rho}(t)$ is expressed as a rational function of monic polynomials.*

For a proof of this theorem, see [**11**]. Note that Cha ([**3**]) treated the same problem.

The trefoil knot K is a fibred knot (Example 3.11). In Example 5.6, the representation $\rho(\cdot)$ of $G(K)$ is in $SL(2, \mathbb{Z})$ in the case of $s = -1$, i.e., this case satisfies the assumption of the theorem. Thus, the twisted Alexander invariant $\Delta_{K,\rho}(t)$ is expressed as a rational function of monic polynomials.

EXAMPLE 5.8. Let K be the figure eight knot (Figure 12). The fundamental group of the exterior has a presentation

$$G(K) = \langle\ x, y \mid zxz^{-1}y^{-1}\ \rangle,$$

where $z = x^{-1}yxy^{-1}x^{-1}$. Let $\rho : G(K) \to SL(2, \mathbb{C})$ be a noncommutative representation defined by

$$\rho(x) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \rho(y) = \begin{pmatrix} 1 & 0 \\ -\omega & 1 \end{pmatrix},$$

where ω is a complex number satisfying $\omega^2 + \omega + 1 = 0$. Applying free differential calculus to the relation $r : zx = yz$ by using the observation just before Example 5.6, we obtain

$$\frac{\partial r}{\partial x} = -x^{-1} + x^{-1}y + yx^{-1} - yx^{-1}y + yx^{-1}yxy^{-1}x^{-1}.$$

Thus we have the matrix

$$M^2_{\rho\otimes\alpha} = \left(\Phi\left(\frac{\partial r}{\partial x}\right)\right) = \begin{pmatrix} -(\omega+1)t + \omega + 2 - t^{-1} & t + \omega - 2 + t^{-1} \\ (\omega - 1)t - \omega + 1 & -(\omega+1)t + 3 - t^{-1} \end{pmatrix}.$$

Then the numerator of $\Delta_{K,\rho}$ is given by

$$\begin{aligned} \det M^2_{\rho\otimes\alpha} &= t^{-2}(t^4 - 6t^3 + \omega^4 t^2 + \omega^2 t^2 + 11t^2 - 6t + 1) \\ &= t^{-2}(t-1)^2(t^2 - 4t + 1). \end{aligned}$$

On the other hand, the denominator of $\Delta_{K,\rho}$ is given by

$$\begin{aligned} \det \Phi(y - 1) &= \det(t\rho(y) - I) \\ &= \det \begin{pmatrix} t - 1 & 0 \\ -\omega t & t - 1 \end{pmatrix} \\ &= (t-1)^2. \end{aligned}$$

Thus we have:

$$\Delta_{K,\rho}(t) = \frac{t^{-2}(t-1)^2(t^2 - 4t + 1)}{(t-1)^2} \doteq t^2 - 4t + 1.$$

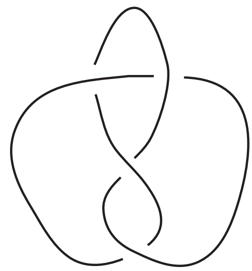


FIGURE 12. The figure eight knot

EXAMPLE 5.9. Let K be the knot illustrated in Figure 13. The normalized Alexander polynomial of K is equal to the monic polynomial $t^4 - t^3 + t^2 - t + 1$. Moreover, the genus of K is equal to 2. The knot group $G(K)$ has a presentation with seven generators $x_1, \dots, x_7$ and six relations:

$$
\begin{aligned}
&r_1 : x_2x_1 = x_3x_2x_1x_2x_1^{-1}x_2^{-1},\\
&r_2 : x_6x_5x_6^{-1} = x_4x_3x_1^{-1}x_3x_1^{-1}x_3x_1x_3^{-1}x_1x_3^{-1}x_1x_3^{-1}x_4^{-1},\\
&r_3 : x_6x_7x_6^{-1} = x_4x_3x_1^{-1}x_3x_1^{-1}x_3x_1x_3^{-1}x_1x_3^{-1}x_4^{-1},\\
&r_4 : x_5x_6x_5^{-1} = x_7x_2x_7^{-1},\\
&r_5 : x_2x_6x_2^{-1} = x_3x_2x_1x_2x_1^{-1}x_2^{-1}x_3^{-1}x_7x_3x_2x_1x_2^{-1}x_1^{-1}x_2^{-1}x_3^{-1},\\
&r_6 : x_5x_4x_5^{-1}x_7 = x_7x_3x_2x_1x_2x_1^{-1}x_2^{-1}x_3^{-1}.
\end{aligned}
$$

Let $\mathbb{F}_5$ be the finite field of characteristic 5, and $\rho : G(K) \to SL(2, \mathbb{F}_5)$ a noncommutative representation over $\mathbb{F}_5$ defined as follows:

$$
\rho(x_1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \rho(x_2) = \begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix}, \quad \rho(x_3) = \begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix}, \quad \rho(x_4) = \begin{pmatrix} 2 & 1 \\ 4 & 0 \end{pmatrix},
$$

$$
\rho(x_5) = \begin{pmatrix} 2 & 4 \\ 1 & 0 \end{pmatrix}, \quad \rho(x_6) = \begin{pmatrix} 3 & 1 \\ 1 & 4 \end{pmatrix} \quad \text{and} \quad \rho(x_7) = \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix}.
$$

By the same method as in previous examples, we have:

$$
\begin{aligned}
\Delta_{K,\rho}(t) &= \frac{\det M^7_{\rho\otimes\alpha}}{\det\Phi(x_7 - 1)}\\
&= \frac{t^{12}(3t^4 + 4t^3 + t^2 + 4t + 3)}{t^2 + 3t + 1}\\
&\doteq 3t^2 + 3.
\end{aligned}
$$

Hence this knot K is not fibred.

5.3. Some remarks and open problems. As we saw in the previous subsections, the twisted Alexander invariant sometimes becomes a polynomial, and sometimes not. In [**23**], this problem is discussed. To be precise, the next theorem was shown:

THEOREM 5.10 ([**23**]). *Let $\rho : G(K) \to SL(2, \mathbb{F})$ be a nonabelian representation of a knot group $G(K)$. Then the twisted Alexander invariant $\Delta_{K,\rho}(t)$ becomes a polynomial.*

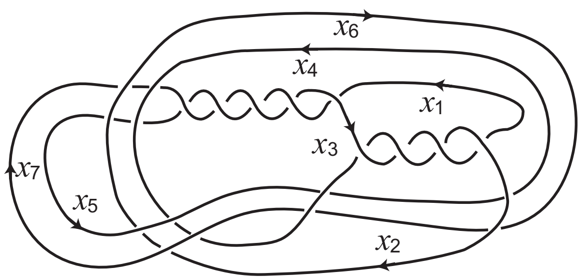


FIGURE 13

We have seen in Theorem 5.7 that the twisted Alexander invariant $\Delta_{K,\rho}(t)$ of a fibred knot associated to a representation $\rho : G(K) \to SL(2, \mathbb{F})$ is expressed as a rational function of monic polynomials. Then, Theorem 5.10 induces the following:

THEOREM 5.11. *Let $\rho : G(K) \to SL(2, \mathbb{F})$ be a nonabelian representation of a fibred knot with genus g. Then the twisted Alexander invariant $\Delta_{K,\rho}(t)$ is a monic polynomial of degree $4g - 2$.*

EXAMPLE 5.12. In [**23**], the following representation $\rho : G(K) \to SL(2, \mathbb{F}_7)$ was found, where K is the knot in Figure 13 and $\mathbb{F}_7$ is the finite field of characteristic 7.

$$\rho(x_1) = \begin{pmatrix} 3 & 3 \\ 3 & 1 \end{pmatrix}, \quad \rho(x_2) = \begin{pmatrix} 5 & 1 \\ 1 & 6 \end{pmatrix}, \quad \rho(x_3) = \begin{pmatrix} 0 & 1 \\ 6 & 4 \end{pmatrix}, \quad \rho(x_4) = \begin{pmatrix} 6 & 4 \\ 2 & 5 \end{pmatrix},$$

$$\rho(x_5) = \begin{pmatrix} 6 & 6 \\ 6 & 5 \end{pmatrix}, \quad \rho(x_6) = \begin{pmatrix} 6 & 1 \\ 1 & 5 \end{pmatrix} \quad \text{and} \quad \rho(x_7) = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}.$$

Using this representation, we have:

$$\begin{aligned} \Delta_{K,\rho}(t) &= \frac{t^6 + 2t^5 + 4t^4 + 2t^3 + 4t^2 + 2t + 1}{t^2 + 3t + 1} \\ &= t^4 + 6t^3 + 6t^2 + 6t + 1. \end{aligned}$$

Recall that the genus of K is equal to 2. So, the twisted Alexander polynomial must have the degree 6 $(= 4 \cdot 2 - 2)$ if K is fibred. Thus the knot K is again not fibred.

We close this section by presenting problems which arise naturally.

Open problem. Can we generalize Theorem 5.10 ? That is, let $\rho : G(K) \to SL(2n, \mathbb{F})$ $(n \geq 2)$ be a nonabelian representation of a knot group $G(K)$. Does the twisted Alexander invariant $\Delta_{K,\rho}(t)$ become a polynomial ?

Open problem ([**14**]). Does every non-fibred knot K have a unimodular representation $\rho : G(K) \to SL(2n, \mathbb{F})$ so that $\Delta_{K,\rho}(t)$ is a rational function of nonmonic polynomials ?

6. Twisted Novikov homology and Morse-Novikov inequality

In this section, we present some results in [**15**]. We introduce the notion 'twisted' Novikov homology, which is a module over the ring $\mathbb{Z}((t))$ associated to a representation of the fundamental group. It allows us to keep track of the non-abelian homological algebra associated to the group ring of the fundamental group of the space considered. Then we generalize the Morse-Novikov inequality for knots and links (Theorem 4.1), that is, Theorem 6.1 gives a lower bound for the Morse-Novikov number of $\mathcal{MN}(L)$ of a link L in terms of the twisted Novikov homology. See [**32**] for the precise relationship between the twisted Alexander invariant and twisted Novikov homology.

6.1. Twisted Novikov homology. Recall that

$$\Lambda = \mathbb{Z}[t, t^{-1}], \quad \widehat{\Lambda} = \mathbb{Z}((t)) = \mathbb{Z}[[t]][t^{-1}].$$

The ring Λ is isomorphic to the group ring $\mathbb{Z}[\mathbb{Z}]$, via the isomorphism sending $t \in \Lambda$ to the element $-1 \in \mathbb{Z}$. The ring $\widehat{\Lambda}$ is then identified with *the Novikov completion* of $\mathbb{Z}[\mathbb{Z}]$.

Let X be a CW complex. Set $G = \pi_1 X$, and let $\xi : G \to \mathbb{Z}$ be a homomorphism. Let $\rho : G \to GL(n, \mathbb{Z})$ be a map such that $\rho(g_1 g_2) = \rho(g_2)\rho(g_1)$ for every $g_1, g_2 \in G$. Such map will be called a *right representation* of G. The homomorphism ξ extends to a ring homomorphism $\mathbb{Z}[G] \to \Lambda$, which will be denoted by the same symbol ξ. The tensor product $\rho \otimes \xi$ (where ξ is considered as a representation $G \to GL(1, \Lambda)$) induces a right representation $\rho_\xi : G \to GL(n, \Lambda)$. The composition of this right representation with the natural inclusion $\Lambda \hookrightarrow \widehat{\Lambda}$ gives a right representation $\widehat{\rho}_\xi :$ $G \to GL(n, \widehat{\Lambda})$. We define a chain complex

$$\widehat{C}_*(\widetilde{X}; \xi, \rho) = \widehat{\Lambda}^n \otimes_{\widehat{\rho}_\xi} C_*(\widetilde{X}).$$

Here $\widetilde{X}$ is the universal cover of X, $C_*(\widetilde{X})$ is a module over $\mathbb{Z}[G]$, and $\widehat{\Lambda}^n$ is a right $\mathbb{Z}G$-module via the right representation $\widehat{\rho}_\xi$. Then this is a chain complex of free left modules over $\widehat{\Lambda}$, and the same is true for its homology. The modules

$$\widehat{H}_*(X; \xi, \rho) = H_*(\widehat{C}_*(\widetilde{X}; \xi, \rho)),$$

are called *twisted Novikov homology*. When these modules are finitely generated (this is the case for example for any X homotopy equivalent to a finite CW complex), we set, as in Section 4,

$$\widehat{b}_i(X; \xi, \rho) = \operatorname{rank}_{\widehat{\Lambda}}(\widehat{H}_i(X; \xi, \rho)),$$

$$\widehat{q}_i(X; \xi, \rho) = \text{ torsion number of } (\widehat{H}_i(X; \xi, \rho)) \text{ over } \widehat{\Lambda}.$$

Here the 'torsion number' stands for the minimal possible number of generators of the torsion part over $\widehat{\Lambda}$.

By the same argument as Lemma 4.3, the numbers $\widehat{b}_i(X; \xi, \rho)$ and $\widehat{q}_i(X; \xi, \rho)$ can be recovered from the canonical decomposition of $\widehat{H}_i(X; \xi, \rho)$ into a direct sum of cyclic modules. That is, let

$$\widehat{H}_i(X; \xi, \rho) = \widehat{\Lambda}^{\alpha_i} \oplus \big(\oplus_{j=1}^{\beta_i} \widehat{\Lambda}/\lambda_j^{(i)} \widehat{\Lambda} \big)$$

where $\lambda_j^{(i)}$ are non-zero non-invertible elements of $\widehat{\Lambda}$ and $\lambda_{j+1}^{(i)} | \lambda_j^{(i)}$ $\forall j$. (Such a decomposition exists since $\widehat{\Lambda}$ is a principal ideal domain.) Then $\alpha_i = \widehat{b}_i(X; \xi, \rho)$ and $\beta_i = \widehat{q}_i(X; \xi, \rho)$. It is not difficult to show that we can always choose $\lambda_j^{(i)} \in \Lambda$ $\forall i$, $\forall j$.

When ρ_0 is the trivial 1-dimensional representation, we obtain the usual Novikov homology, which can also be calculated from the infinite cyclic covering $\overline{X}$ associated to ξ, namely

$$\widehat{H}_*(X; \xi, \rho_0) = \widehat{\Lambda} \otimes_\Lambda H_*(\overline{X}), \qquad \text{for } \rho_0 = \mathbf{1} : G \to GL(1, \mathbb{Z}).$$

In general, we may define the twisted Novikov homology using a commutative ring R, $Q = R[t, t^{-1}]$, and $\widehat{Q} = R((t)) = R[[t]][t^{-1}]$. In particular, we may use a field instead of $\mathbb{Z}$. However, the numbers $\widehat{q}_i(X;\xi,\rho)$ vanish if R is a field. This is crucial when we treat knots and links in the 3-sphere.

6.2. Twisted version of Morse-Novikov inequality for knots and links. As we defined in the previous sections, let L be an oriented link in the 3-sphere, and put $C_L = S^3 - L$. Note that there is a unique element $\xi \in H^1(C_L, \mathbb{Z})$ such that for every positively oriented meridian μ_i of a component of L, we have $\xi(\mu_i) = 1$. Let $\rho : G(L) \to GL(n, \mathbb{Z})$ be a representation. We identify the cohomology class ξ with the corresponding homomorphism $G(L) \to \mathbb{Z}$. Since the cohomology class ξ is determined by the orientation of L, we omit it, and then we shall denote $\widehat{H}_*(C_L;\xi,\rho)$ by $\widehat{H}_*(L,\rho)$. The numbers $\widehat{b}_i(C_L;\xi,\rho)$ and $\widehat{q}_i(C_L;\xi,\rho)$ will be denoted by $\widehat{b}_i(L,\rho)$ and $\widehat{q}_i(L,\rho)$. Then, we have:

THEOREM 6.1 ([**15**]). *Let $f : C_L \to S^1$ be any regular map. Then*

$$m_i(f) \geq \frac{1}{n}\Big(\widehat{b}_i(L,\rho) + \widehat{q}_i(L,\rho) + \widehat{q}_{i-1}(L,\rho)\Big)$$

for every i.

See Section 2 for the definition of 'regular map'.

PROPOSITION 6.2. The following equations hold :

$$\widehat{b}_i(L,\rho) = \widehat{q}_i(L,\rho) = \widehat{q}_2(L,\rho) = 0 \qquad \text{for } i = 0,\ i \geq 3,$$
$$\widehat{b}_1(L,\rho) = \widehat{b}_2(L,\rho).$$

From these results, we have:

$$m_1(f) \geq \frac{1}{n}\Big(\widehat{b}_1(L,\rho) + \widehat{q}_1(L,\rho)\Big);$$
$$m_2(f) \geq \frac{1}{n}\Big(\widehat{b}_1(L,\rho) + \widehat{q}_1(L,\rho)\Big).$$

Thus we have the following, which is a twisted version of Corollary 4.2.

COROLLARY 6.3.

$$\mathcal{MN}(L) \geq \frac{2}{n} \times (\widehat{b}_1(L,\rho) + \widehat{q}_1(L,\rho)).$$

In [**15**], the Kinoshita-Terasaka knot, the Conway knot and their connected sum are discussed as examples. Note that the Alexander polynomials of these knots are equal to 1.

7. On a calculation of Hutchings-Lee type invariant

In the last part of my talk in the conference, I gave a progress report on a joint work with Hiroshi Matsuda [**12**].

For any closed oriented Riemannian manifold X with $\chi(X) = 0$ and $b_1(X) > 0$, Hutchings and Lee investigated the Morse-theoretic Reidemeister torsion T_{Morse} associated to a circle valued Morse map $f : X \to S^1$, the topological Reidemeister torsion T_{top} and a zeta function ζ in [**17**] and [**18**].

More precisely, set, as in the previous sections, $\Lambda = \mathbb{Z}[t, t^{-1}]$, and $\widehat{\Lambda} = \mathbb{Z}((t))$. Let C_*^{nov} denote the free $\widehat{\Lambda}$-module chain complex generated by the set of critical points of index $*$. If $C_*^{\text{nov}} \otimes \mathbb{Q}((t))$ is acyclic, the corresponding Reidemeister torsion is called the *Morse-theoretic Reidemeister torsion* T_{Morse}. Let $\overline{X}$ be the infinite cyclic $\mathbb{Z}$-cover of X induced by f; then $C_*^{\text{cell}}(\overline{X})$ is the cellular chain complex of $\overline{X}$ as a module over Λ. The Novikov theorem says that

$$H_*(C_*^{\text{nov}}) \cong H_*(C_*^{\text{cell}}(\overline{X}) \otimes_\Lambda \widehat{\Lambda}).$$

If $C_*^{\text{cell}}(\overline{X}) \otimes \mathbb{Q}(t)$ is acyclic, then the corresponding Reidemeister torsion T_{top} is a homotopy invariant of f, i.e., it depends only on the cohomology class $[d\phi] \in H^1(X, \mathbb{Z})$. Let $\varphi^n \colon \Sigma \to \Sigma$ be the return maps of f; then the zeta function is defined to be

$$\zeta = \exp\Big(\sum_{n=1}^{\infty} \text{Fix}(\varphi^n) t^n / n\Big)$$

where $\text{Fix}(\varphi^n)$ counts the fixed points of φ^n with sign, and the sign is given by the sign of $\det(1 - d\varphi^n)$. Under these notations, Hutchings and Lee showed that

$$T_{\text{Morse}} \cdot \zeta = \iota(T_{\text{top}})$$

in $\mathbb{Q}((t))$, up to sign and multiplication by powers of t. Here ι is the inclusion map $\iota : \mathbb{Q}(t) \hookrightarrow \mathbb{Q}((t))$. If $f : X \to S^1$ has no critical point, we define $T_{\text{Morse}} = 1$.

After Hutchings and Lee's work, Mark ([**26**]) revealed the 'structure' of $T_{\text{Morse}} \cdot \zeta$ by making use of a topological quantum field theory ([**5**]), which makes the calculations explicit. The purpose of this section is to give a rough idea of doing the concrete calculations in the case of knot or link complements using Heegaard splitting for sutured manifolds stated in Section 3.

7.1. A monodromy of a fibred link via its Heegaard diagram. For simplicity, we study the knot case. We can treat the link case similarly. Let K be a fibred knot in the 3-sphere S^3. Then K has a Seifert surface R whose sutured manifold (M, γ) is a product sutured manifold. If we glue $R_+(\gamma)$ and $R_-(\gamma)$ by the corresponding homeomorphism, we obtain a manifold that is homeomorphic to the exterior of K in S^3. This homeomorphism is called the *monodromy* h of K. The monodromy h induces the transformation matrix $H_i : H_i(R_+(\gamma)) \to H_i(R_-(\gamma))$. We call H_i the *monodromy matrix* of the fibred knot K. Concretely, let $a_1, a_2, \ldots, a_{2g}$ be generators of $H_1(R)$, where g is the genus of R. Push them off along the normal vector of R, and put them on $R_+(\gamma)$ and $R_-(\gamma)$. Then we may see that they are generators on $R_+(\gamma)$ and $R_-(\gamma)$. We denote the generators on $R_+(\gamma)$ (resp. $R_-(\gamma)$) by $a_1^+, a_2^+, \ldots, a_{2g}^+$ (resp. $a_1^-, a_2^-, \ldots, a_{2g}^-$). Then we have:

$$\begin{pmatrix} a_1^- \\ a_2^- \\ \cdot \\ \cdot \\ \cdot \\ a_{2g}^- \end{pmatrix} = H_1 \begin{pmatrix} a_1^+ \\ a_2^+ \\ \cdot \\ \cdot \\ \cdot \\ a_{2g}^+ \end{pmatrix}$$

For example, let K be the trefoil knot and R the Seifert surface as shown in Figure 2. Set a_1 and a_2 as generators of R illustrated in Figure 14. Then we can observe that

$$\begin{pmatrix} a_1^- \\ a_2^- \end{pmatrix} = \begin{pmatrix} a_1^+ + a_2^+ \\ -a_1^+ \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} a_1^+ \\ a_2^+ \end{pmatrix}$$

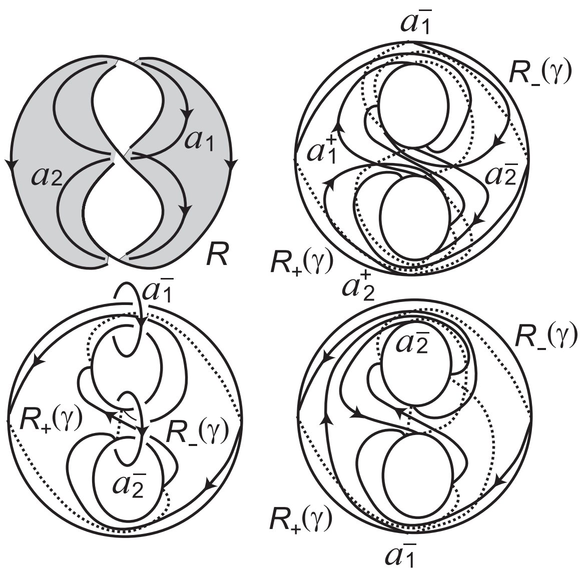


FIGURE 14

Thus we have

$$H_1 = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}.$$

7.2. Alexander polynomial and Lefschetz zeta function. In this subsection, we review a relationship between Alexander polynomials and Lefschetz zeta functions. Let F be a manifold, and $g : F \to F$: a continuous map. We define the *zeta function*

$$\zeta_g(t) = \exp \sum_{n=1}^{\infty} \frac{\Lambda(g^n)}{n} t^n,$$

$$\text{where } \Lambda(g) = \sum_{i=0}^{\dim F} (-1)^i \ \mathrm{Trace}(g_{*,i} : H_i(F, \mathbb{Q}) \to H_i(F, \mathbb{Q})).$$

This zeta function has several expressions, see [**37**] for example.

Now we focus on a knot complement. Let K be a fibred knot in the 3-sphere, and we denote by h the monodromy of K. Then the following equation is known (see [**28**] and [**29**]):

THEOREM 7.1.

$$\zeta_h(t) = \frac{\Delta_K(t)}{1-t}.$$

Compare Remark 5.4. Let us observe an example.

EXAMPLE 7.2. Let K be the trefoil knot as illustrated in Figure 14. Then, as seen in the previous subsection, the monodromy matrix is

$$H_1 = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}, \ H_1^2 = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}, \ H_1^3 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$

$$H_1^4 = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix},\ H_1^5 = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix},\ H_1^6 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Since $\mathrm{Trace}(H_0 : H_0(R) \to H_0(R)) = 1$ and $H_2 = 0$, we have:

$$\begin{aligned}
\zeta_h(t) &= \exp\{(1-1)t + (1+1)\frac{t^2}{2} + (1+2)\frac{t^3}{3} \\
&\qquad + (1+1)\frac{t^4}{4} + (1-1)\frac{t^5}{5} + (1-2)\frac{t^6}{6} + \cdots\} \\
&= \exp\left(\log \frac{1-t+t^2}{1-t}\right) \\
&= \frac{\Delta_K(t)}{1-t}.
\end{aligned}$$

Remark 7.3. Let $K(0)$ be the 3-manifold obtained by 0-Dehn surgery on a fibred knot K, and h the monodromy of $K(0)$ induced from the fibre structure of K. Then we have the following:

$$\zeta_h(t) = \frac{\Delta_K(t)}{(1-t)^2}.$$

Exercise 7.4. (1) Confirm this equation in case of the trefoil knot.
(2) Calculate $\zeta_h(t)$ for the figure eight knot (Figure 12).

We note that it is known that the topological Reidemeister torsion of $K(0)$ is equal to $\Delta_K(t)/(1-t)^2$ even if K is non-fibred.

7.3. A monodromy matrix via Heegaard diagram. In this subsection, we will consider a more general situation, namely, the non-fibred case. For simplicity, we describe the knot case here. We can treat the link case similarly (cf. the next subsection). Let K be a knot in the 3-sphere and R a Seifert surface in $E(K)$. Let (M,γ) be the sutured manifold for R. We denote by (W_1, W_2) a Heegaard splitting for (M,γ). Note that $\partial_+ W_1 = \partial_+ W_2$ is the Heegaard surface of this splitting. We glue $R_+(\gamma)$ and $R_-(\gamma)$ in $E(K)$ so that $R = R_+(\gamma) = R_-(\gamma)$ and $W_1 \cup W_2 = E(K)$. Further, once we cut $E(K)$ along $\partial_+ W_1 = \partial_+ W_2$, we may suppose that $E(K)$ is restored by gluing $\partial_+ W_1$ and $\partial_+ W_2$ using a homeomorphism h. This homeomorphism h is called the *monodromy* of K. Set $N = h(W_1) = h(W_2)$ (see Definition 3.4). Then, we may denote the generators on $\partial_+ W_1$ (resp. $\partial_+ W_2$) by $a_1^1, a_2^1, \ldots, a_{2g}^1, m_1^1, \ell_1^1, \ldots, m_N^1, \ell_N^1$ (resp. $a_1^2, a_2^2, \ldots, a_{2g}^2, m_1^2, \ell_1^2, \ldots, m_N^2, \ell_N^2$). Here, a_j^i comes from R as in case of a fibred knot, and m_k^i and ℓ_k^i are derived from the 'attaching 1-handles' of W_i, namely, m_k^i is a cocore of the 'attaching 1-handle' of W_i and ℓ_k^i is a 'longitude' corresponding to m_k^i. As in case of a fibred knot, the monodromy h induces the transformation matrix $H_i : H_i(\partial_+ W_1) \to H_i(\partial_+ W_2)$. In particular, we may describe:

$$\begin{pmatrix} a_1^2 \\ a_2^2 \\ \cdot \\ \cdot \\ \cdot \\ a_{2g}^2 \\ m_1^1 \\ \ell_1^1 \\ \cdot \\ \cdot \\ \cdot \\ m_N^2 \\ \ell_N^2 \end{pmatrix} = H_1 \begin{pmatrix} a_1^1 \\ a_2^1 \\ \cdot \\ \cdot \\ \cdot \\ a_{2g}^1 \\ m_1^1 \\ \ell_1^1 \\ \cdot \\ \cdot \\ \cdot \\ m_N^1 \\ \ell_N^1 \end{pmatrix}$$

For $n \geq 1$, we define:

$$\begin{pmatrix} h^n(a_1^1) \\ h^n(a_2^1) \\ \cdot \\ \cdot \\ \cdot \\ h^n(a_{2g}^1) \\ h^n(m_1^1) \\ h^n(\ell_1^1) \\ \cdot \\ \cdot \\ \cdot \\ h^n(m_N^1) \\ h^n(\ell_N^1) \end{pmatrix} = H_1^n \begin{pmatrix} a_1^1 \\ a_2^1 \\ \cdot \\ \cdot \\ \cdot \\ a_{2g}^1 \\ m_1^1 \\ \ell_1^1 \\ \cdot \\ \cdot \\ \cdot \\ m_N^1 \\ \ell_N^1 \end{pmatrix}$$

Similar to the case of a fibred knot, we define:

$$\zeta_h(t) = \exp \sum_{n=1}^{\infty} \frac{\Lambda(h^n)}{n} t^n,$$

$$\text{where } \Lambda(h^n) = \text{ Trace } H_0^n - \text{ Trace } H_1^n = 1 - \text{ Trace } H_1^n.$$

REMARK 7.5. To be precise, the equation $T_{\mathrm{Morse}} \cdot \zeta = \iota(T_{\mathrm{top}})$ has been proved for a closed manifold, so we should treat $K(0)$ instead of K and multiply the correction term $1/(1-t)$, which corresponds to using $\Lambda(h^n) = \text{ Trace } H_0^n - \text{ Trace } H_1^n + \text{ Trace } H_2^n = 2 - \text{ Trace } H_1^n$.

Next, we would like to define $\tau_h(t)$ to count 'flow lines' from critical points of index 2 to those of index 1 in the infinite cyclic covering $\overline{E(K)}$ of $E(K)$. The intersection points of $m_j^2 \cap m_i^1$ correspond 1 to 1 to flow lines from critical points of index 2 to those of index 1, which does not intersect the Seifert surface R. Then, the algebraic intersection number of m_j^2 and m_i^1 is equal to the algebraic number of such flow lines. Let us represent $h^n(m_j^1)$ by generators of $H_1(\partial_+ W_1)$, and let β_{ij}^n be the coefficient of ℓ_i^1. Thus we denote by β_{ij}^n the $(2g+2j-1) \times (2g+2i)$th-component of H_1^n.

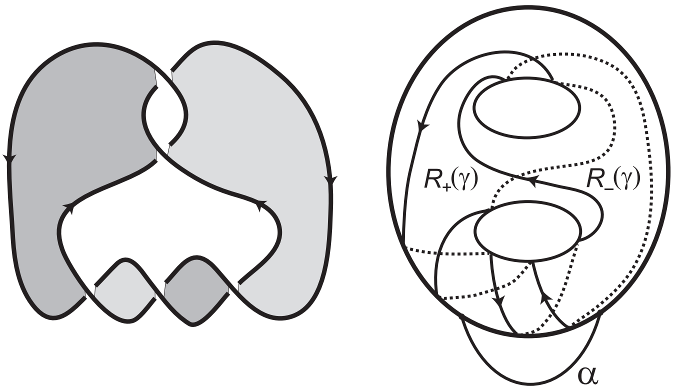


FIGURE 15

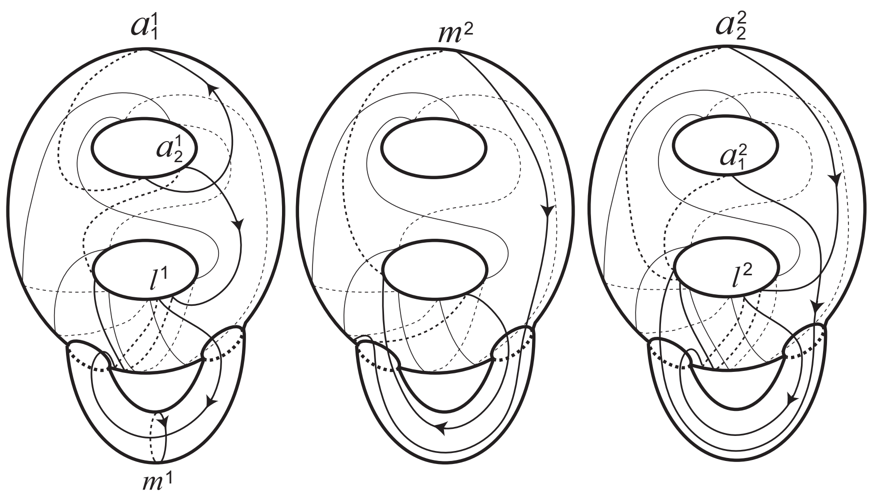


FIGURE 16

We define

$$\beta_{ij} = \sum_{k=1}^{\infty} (\beta_{ij}^k \cdot t^{k-1}), \quad \text{and} \quad \tau_h(t) = \det(\beta_{ij}).$$

If $E(K)$ has no critical point, i.e., K is a fibred knot, $\tau_h(t)$ is defined to be 1.

7.4. Some calculations. Let K be the knot 5_2 in the list in [**36**]. This is a twist knot, and has a genus one Seifert surface R. Moreover, we can observe $\mathcal{MN}(R) = 2$ by the same method as in the previous sections. In fact, the Alexander polynomial of this knot is $2-3t+2t^2$ and both the regular neighborhood of $R_-(\gamma)\cup\alpha$ and the complement in M are compression bodies, say W_1 and W_2, see Figure 15. Here (M, γ) is the sutured manifold of R (cf. Exercise 4.8).

By the same method as in the fibred case, we have $a_1^2 = a_2^1 + \ell^1$, $a_2^2 = -a_1^1 + a_2^1$. See Figure 16. Further, we can find a disk that is a cocore of an attaching 1-handle of W_2, whose boundary circle is m^2 in the middle figure of Figure 16. Then we have:

$$\begin{cases} a_1^2 = a_2^1 + \ell^1 \\ a_2^2 = -a_1^1 + a_2^1 \\ m^2 = -a_1^1 + a_2^1 + m^1 + 2\ell^1 \\ \ell^2 = \ell^1 \end{cases} \qquad \text{i.e.,} \qquad H = \begin{pmatrix} 0 & 1 & 0 & 1 \\ -1 & 1 & 0 & 0 \\ -1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

By direct calculations, we can obtain:

$$H^{6n+1} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ -1 & 1 & 0 & 0 \\ -1 & 1 & 1 & 6n+2 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad H^{6n+2} = \begin{pmatrix} -1 & 1 & 0 & 1 \\ -1 & 0 & 0 & -1 \\ -2 & 1 & 1 & 6n+3 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$H^{6n+3} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & -2 \\ -2 & 0 & 1 & 6n+3 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad H^{6n+4} = \begin{pmatrix} 0 & -1 & 0 & -1 \\ 1 & -1 & 0 & -2 \\ -1 & -1 & 1 & 6n+3 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$H^{6n+5} = \begin{pmatrix} 1 & -1 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 6n+4 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad H^{6n+6} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 6n+6 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Therefore we have:

$$\begin{aligned} \zeta_h(t) &= \exp\Big(\sum_{n=0}^{\infty}\Big\{(1-3)\frac{t^{6n+1}}{6n+1} + (1-1)\frac{t^{6n+2}}{6n+2} + (1-0)\frac{t^{6n+3}}{6n+3} \\ &\qquad + (1-1)\frac{t^{6n+4}}{6n+4} + (1-3)\frac{t^{6n+5}}{6n+5} + (1-4)\frac{t^{6n+6}}{6n+6}\Big\}\Big) \\ &= \exp\Big(\log(1-2t+2t^2-t^3)\Big) \\ &= (1-t+t^2)(1-t). \end{aligned}$$

$$\begin{aligned} \tau_h(t) &= \sum_{n=0}^{\infty}\Big\{(6n+2)t^{6n} + (6n+3)t^{6n+1} + (6n+3)t^{6n+2} \\ &\qquad + (6n+3)t^{6n+3} + (6n+4)t^{6n+4} + (6n+6)t^{6n+5}\Big\} \\ &= \frac{2-3t+2t^2}{(1-t+t^2)(1-t)^2} \end{aligned}$$

Thus we can see:

$$\tau_h(t)\cdot\zeta_h(t) = \frac{2-3t+2t^2}{1-t} = \frac{\Delta_K(t)}{(1-t)}.$$

Let $\bar{R}$ be an unknotted annulus with -2-full twists and $L = \partial\bar{R}$ the 2-component link in S^3. By the same argument as in the case of the 5_2 knot (cf. Figure 17), we have:

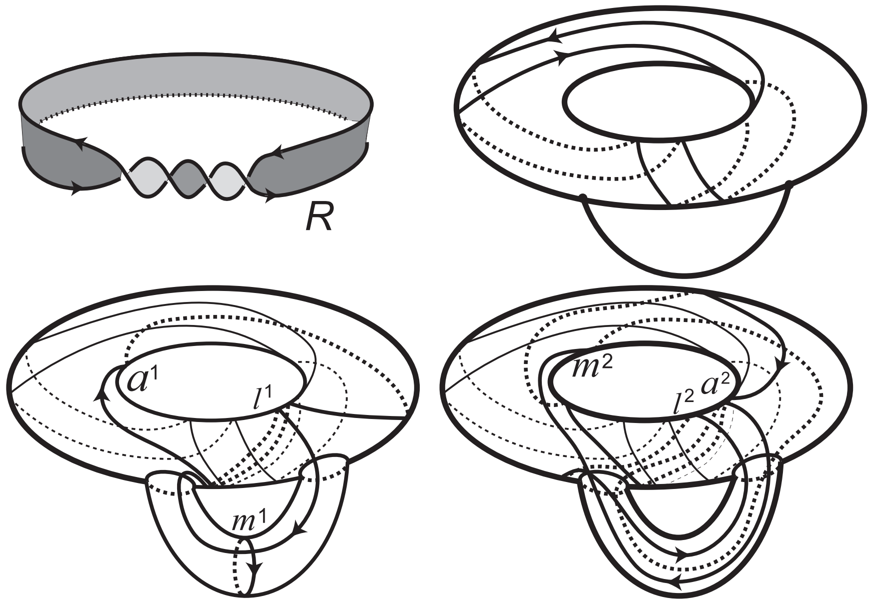


FIGURE 17

$$\begin{pmatrix} a^2 \\ m^2 \\ \ell^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 2 \\ 0 & -1 & -1 \end{pmatrix} \begin{pmatrix} a^1 \\ m^1 \\ \ell^1 \end{pmatrix}$$

Then we can observe:

$$\tau_h(t) \cdot \zeta_h(t) = \left(\frac{2}{1+t^2} \right) \cdot (1+t^2) = 2.$$

ACKNOWLEDGEMENT. I would like to thank the organizers of the conference and the institute for giving me this opportunity. I also thank Dr. Teruhisa Kadokami, Dr. Teruaki Kitano, Dr. Hiroshi Matsuda, Dr. Takayuki Morifuji, and Dr. Akio Noguchi for useful conversations.

References

[1] G. Burde and H. Zieschang, Knots, De Gruyter Studies in Mathematics, 5. Berlin - New York: Walter de Gruyter. XII, 399 p. DM 49.95 (1985).

[2] A, Casson and C. McA. Gordon, *Reducing Heegaard splittings*, Topology Appl. 27 (1987), no. 3, 275–283.

[3] J. C. Cha, *Fibred knots and twisted Alexander invariants*, Trans. Amer. Math. Soc. 355 (2003), no. 10, 4187–4200.

[4] R.H. Crowell and R.H. Fox, Introduction to knot theory, Reprint of the 1963 original. Graduate Texts in Mathematics, No. 57. Springer-Verlag, New York-Heidelberg, 1977.

[5] S. K. Donaldson, *Topological field theories and formulae of Casson and Meng-Taubes*, Proceedings of the Kirbyfest (Berkeley, CA, 1998), 87–102 (electronic), Geom. Topol. Monogr., 2, Geom. Topol. Publ., Coventry, 1999.

[6] D. Gabai, *The Murasugi sum is a natural geometric operation*, Low-dimensional topology (San Francisco, Calif., 1981), 131–143, Contemp. Math., 20, Amer. Math. Soc., Providence, RI, 1983.

[7] D. Gabai, *Foliations and the topology of* 3*-manifolds*, J. Differential Geom. 18 (1983), 445-503.

[8] D. Gabai, *Detecting fibred links in* S^3, Comment. Math. Helv. 61 (1986), 519-555.

[9] H. Goda, *Heegaard splitting for sutured manifolds and Murasugi sum*, Osaka J. Math. 29 (1992), 21-40.

[10] H. Goda, *On handle number of Seifert surfaces in* S^3, Osaka J. Math. 30 (1993), 63-80.

[11] H. Goda, T. Kitano and T. Morifuji, *Reidemeister torsion, twisted Alexander polynomial and fibered knots*, preprint (2002), to appear in Comment. Math. Helv.

[12] H. Goda and H. Matsuda, in preparation.

[13] H. Goda, H. Matsuda and T. Morifuji, *Knot Floer homology of* $(1,1)$*-knots*, to appear in Geom. Dedicata.

[14] H. Goda and T. Morifuji, *Twisted Alexander polynomial for* $SL(2,C)$*-representations and fibered knots*, C. R. Math. Acad. Sci. Soc. R. Can. 25, (2003), 97-101.

[15] H. Goda and A. Pajitnov, *Twisted Novikov homology and Circle-valued Morse theory for knots and links*, preprint (2003), to appear in Osaka J. Math.

[16] M. Hirasawa and L. Rudolph, *Constructions of Morse maps for knots and links, and upper bounds on the Morse-Novikov number*, preprint (2003).

[17] M. Hutchings and Y-J. Lee, *Circle-valued Morse theory, Reidemeister torsion, and Seiberg-Witten invariants of 3-manifolds*, Topology, 38 (1999), 861-888.

[18] M. Hutchings and Y-J. Lee, *Circle-valued Morse theory and Reidemeister torsion*, Geom. Topol. 3 (1999), 369–396.

[19] T. Kanenobu, *The augmentation subgroup of a pretzel link*, Math. Sem. Notes Kobe Univ. 7 (1979), no. 2, 363–384.

[20] A. Kawauchi (ed.), A survey of knot theory, Birkhauser Verlag, Basel, 1996.

[21] P. Kirk and C. Livingston, *Twisted Alexander invariants, Reidemeister torsion, and Casson-Gordon invariants*, Topology 38 (1999), 635–661.

[22] T. Kitano, *Twisted Alexander polynomial and Reidemeister torsion*, Pacific J. Math. 174 (1996), 431–442.

[23] T. Kitano and T. Morifuji, *Divisibility of twisted Alexander polynomial and fibered knots*, preprint (2004).

[24] K. Kodama, Knot program: http://www.math.kobe-u.ac.jp/ kodama/knot.html

[25] A. Lazarev, *The Novikov homology in knot theory*, Math. Notes 51 (1992), no. 3-4, 259–262.

[26] T. Mark, *Torsion, TQFT, and Seiberg-Witten invariants of 3-manifolds*, Geom. Topol. 6 (2002), 27-58.

[27] G. Meng and C. Taubes, $\underline{\mathrm{SW}} =$ *Milnor torsion*, Math. Res. Lett. 3 (1996), 661–674.

[28] J. Milnor, *A duality theorem for Reidemeister torsion*, Ann. Math. 76 (1962), 137–147.

[29] J. Milnor, *Infinite cyclic coverings*, 1968 Conference on the Topology of Manifolds (Michigan State Univ., E. Lansing, Mich., 1967) pp. 115–133 Prindle, Weber & Schmidt, Boston, Mass.

[30] L. Neuwirth, Knot Groups, Annals of Mathematics Studies, No. 56, Princeton University Press, Princeton, N.J.(1965).

[31] S. P. Novikov, *Multivalued functions and functionals, An analogue of the Morse theory*, Soviet Math. Dokl. 24 (1981), 222-226.

[32] A. Pajitnov, *Novikov homology, Twisted Alexander polynomials and the Thurston cones*, preprint (2004).

[33] A. Pajitnov, L. Rudolph and C. Weber, *Morse-Novikov number for knots and links*, St. Petersburg Math. J. 13 (2002), 417-426.

[34] A. Ranicki, *Circle valued Morse theory and Novikov homology*, Lecture Notes at the Summer School on High-dimensional manifold Topology, Trieste, (2001).

[35] E. Rapaport, *On the commutator subgroup of a knot group*, Ann. of Math. (2) 71 (1960), 157–162

[36] D. Rolfsen, Knots and links. Mathematics Lecture Series, No. 7. Publish or Perish, Inc., Berkeley, Calif., 1976.

[37] D. Salamon, *Seiberg-Witten invariants of mapping tori, symplectic fixed points, and Lefschetz numbers*, Proceedings of 6th Gokova Geometry-Topology Conference. Turkish J. Math. 23 (1999), no. 1, 117–143.

[38] J. Stallings, *On fibering certain* 3*-manifolds*, 1962 Topology of 3-manifolds and related topics (Proc. The Univ. of Georgia Institute, 1961), 95–100.

[39] M. Wada, *Twisted Alexander polynomial for finitely presentable groups*, Topology 33 (1994), 241–256.

Department of Mathematics, Tokyo University of Agriculture and Technology, Naka, Koganei, Tokyo 184-8588, Japan

E-mail address: goda@cc.tuat.ac.jp

Floer Homologies and Contact Structures

Clay Mathematics Proceedings
Volume 5, 2006

Lectures on Open Book Decompositions and Contact Structures

John B. Etnyre

Abstract. This article provides a brief introduction to open book decompositions of 3-manifolds and sketches the proof of Giroux's correspondence between these open books and oriented contact structures on closed 3-manifolds. We then discuss applications of this correspondence to symplectic fillings. This circle of ideas has been essential to recent progress in contact geometry and applications of Heegaard Floer homology and gauge theory to low-dimensional topology.

1. Introduction

The main goal of this survey is to discuss the proof and examine some consequences of the following fundamental theorem of Giroux.

Theorem 1.1 (Giroux 2000, [**21**]). *Let M be a closed oriented 3-manifold. Then there is a one-to-one correspondence between*

$$\{\textit{oriented contact structures on } M \textit{ up to isotopy}\}$$

and

$$\{\textit{open book decompositions of } M \textit{ up to positive stabilization}\}.$$

This theorem plays a pivotal role in studying cobordisms of contact structures and understanding filling properties of contact structures, see [**2, 6, 13, 14, 16, 19**]. This better understanding of fillings leads to various topological applications of contact geometry. Specifically, the much studied property P for knots was established by P. Kronheimer and T. Mrowka in [**28**]. A non-trivial knot has property P if non-trivial surgery on it never gives a homotopy sphere. In addition P. Ozsváth and Z. Szabó in [**34**] gave an alternate proof of a characterization of the unknot via surgery which was originally established in [**29**]. This characterization says that the unknot is the only knot on which p-surgery yields $-L(p,1)$. Moreover, in [**34**] it is shown that the Thurston norm is determined by Heegaard Floer Homology.

Ideally the reader should be familiar with low-dimensional topology at the level of, say [**36**]. In particular, we will assume familiarity with Dehn surgery, mapping tori and basic algebraic topology. At various points we also discuss branch coverings, Heegaard splittings and other notions; however, the reader unfamiliar with these

1991 *Mathematics Subject Classification.* Primary 53D35; Secondary 57R17.

notions should be able to skim these parts of the paper without missing much, if any, of the main line of the arguments. Since diffeomorphisms of surfaces play a central role in much of the paper and specific conventions are important we have included an Appendix discussing basic facts about this. We also assume the reader has some familiarity with contact geometry. Having read [**15**] should be sufficient background for this paper. In order to accommodate the reader with little background in contact geometry we have included brief discussions, scattered throughout the paper, of all the necessary facts. Other good introductions to contact geometry are [**1, 20**], though a basic understanding of convex surfaces is also useful but is not covered in these sources.

In the next three sections we give a thorough sketch of the proof of Theorem 1.1. In Section 2 we define open book decompositions of 3-manifolds, discuss their existence and various constructions. The following two sections discuss how to get a contact structure from an open book and an open book from a contact structure, respectively. Finally in Section 5 we will consider various applications of Theorem 1.1. While we prove various things about open books and contact structures our main goal is to prove the following theorem which is the basis for most of the above mentioned applications of contact geometry to topology.

Theorem (Eliashberg 2004 [**6**]; Etnyre 2004 [**14**]). *If (X, ω) is a symplectic filling of (M, ξ) then there is a closed symplectic manifold (W, ω') and a symplectic embedding $(X, \omega) \to (W, \omega')$.*

Acknowledgments: I am grateful to David Alexandre Ellwood, Peter Ozsváth, András Stipsicz, Zoltan Szabó, the Clay Mathematics Institute and the Alfréd Rényi Institute of Mathematics for organizing the excellent summer school on "Floer Homology, Gauge Theory, and Low Dimensional Topology" and for giving me an opportunity to give the lectures on which these notes are based. I also thank Emmanuel Giroux who gave a beautiful series of lectures at Stanford University in 2000 where I was first exposed to the strong relation between open books and contact structures. I am also grateful to Danny Calegari, Noah Goodman, Gordana Matić, András Némethi and Burak Ozbagci for many illuminating conversations. Finally I thank Paolo Lisca, Stephan Schoenenberger and the referee for valuable comments on the first draft of this paper. This work was supported in part by NSF CAREER Grant (DMS–0239600) and FRG-0244663.

2. Open book decompositions of 3-manifolds

Throughout this section (and these notes)

M is always a closed oriented 3-manifold.

We also mention that when inducing an orientation on the boundary of a manifold we use the "outward normal first" convention. That is, given an oriented manifold N then $v_1, \ldots, v_{n-1}$ is an oriented basis for ∂N if $\nu, v_1, \ldots, v_{n-1}$ is an oriented basis for N.

Definition 2.1. An *open book decomposition of* M is a pair (B, π) where

(1) B is an oriented link in M called the *binding* of the open book and
(2) $\pi : M \setminus B \to S^1$ is a fibration of the complement of B such that $\pi^{-1}(\theta)$ is the interior of a compact surface $\Sigma_\theta \subset M$ and $\partial\Sigma_\theta = B$ for all $\theta \in S^1$. The surface $\Sigma = \Sigma_\theta$, for any θ, is called the *page* of the open book.

One should note that it is important to include the projection in the data for an open book, since B does not determine the open book, as the following example shows.

EXAMPLE 2.2. Let $M = S^1 \times S^2$ and $B = S^1 \times \{N, S\}$, where $N, S \in S^2$. There are many ways to fiber $M \setminus B = S^1 \times S^1 \times (0, 1)$. In particular if γ_n is an embedded curve on T^2 in the homology class $(1, n)$, then $M \setminus B$ can be fibered by annuli parallel to $\gamma_n \times (0, 1)$. There are diffeomorphisms of $S^1 \times S^2$ that relate all of these fibrations but the fibrations coming from γ_0 and γ_1 are not isotopic. There are examples of fibrations that are not even diffeomorphic.

DEFINITION 2.3. An *abstract open book* is a pair (Σ, ϕ) where

(1) Σ is an oriented compact surface with boundary and
(2) $\phi : \Sigma \to \Sigma$ is a diffeomorphism such that ϕ is the identity in a neighborhood of $\partial\Sigma$. The map ϕ is called the *monodromy.*

We begin by observing that given an abstract open book (Σ, ϕ) we get a 3-manifold M_ϕ as follows:

$$M_\phi = \Sigma_\phi \cup_\psi \left(\coprod_{|\partial\Sigma|} S^1 \times D^2 \right),$$

where $|\partial\Sigma|$ denotes the number of boundary components of Σ and Σ_ϕ is the mapping torus of ϕ. By this we mean

$$\Sigma \times [0, 1]/ \sim,$$

where $\sim$ is the equivalence relation $(\phi(x), 0) \sim (x, 1)$ for all $x \in \Sigma$. Finally, $\cup_\psi$ means that the diffeomorphism ψ is used to identify the boundaries of the two manifolds. For each boundary component l of Σ the map $\psi : \partial(S^1 \times D^2) \to l \times S^1 \subset \Sigma_\phi$ is defined to be the unique (up to isotopy) diffeomorphism that takes $S^1 \times \{p\}$ to l where $p \in \partial D^2$ and $\{q\} \times \partial D^2$ to $(\{q'\} \times [0, 1]/ \sim) = S^1$, where $q \in S^1$ and $q' \in \partial\Sigma$. We denote the cores of the solid tori in the definition of M_ϕ by B_ϕ.

Two abstract open book decompositions (Σ_1, ϕ_1) and (Σ_2, ϕ_2) are called *equivalent* if there is a diffeomrophism $h : \Sigma_1 \to \Sigma_2$ such that $h \circ \phi_1 = \phi_2 \circ h$.

LEMMA 2.4. *We have the following basic facts about open books and abstract open books:*

(1) *An open book decomposition (B, π) of M gives an abstract open book (Σ_π, ϕ_π) such that $(M_{\phi_\pi}, B_{\phi_\pi})$ is diffeomorphic to (M, B).*
(2) *An abstract open book determines M_ϕ and an open book (B_ϕ, π_ϕ) up to diffeomorphism.*
(3) *Equivalent open books give diffeomorphic 3-manifolds.*

EXERCISE 2.5. Prove this lemma.

REMARK 2.6. Clearly the two notions of open book decomposition are closely related. The basic difference is that when discussing open books (non-abstract) we can discuss the binding and pages up to *isotopy* in M, whereas when discussing abstract open books we can only discuss them up to *diffeomorphism.* Thus when discussing Giroux's Theorem 1.1 we need to use (non-abstract) open books; however, it is still quite useful to consider abstract open books and we will frequently not make much of a distinction between them.

EXAMPLE 2.7. Let S^3 be the unit sphere in $\mathbb{C}^2$, and $(z_1, z_2) = (r_1 e^{i\theta_1}, r_2 e^{i\theta_2})$ be coordinates on $\mathbb{C}^2$.

(1) Let $U = \{z_1 = 0\} = \{r_1 = 0\} \subset S^3$. Thus U is a unit S^1 sitting in S^3. It is easy to see that U is an unknotted S^1 in S^3. The complement of U fibers:
$$\pi_U : S^3 \setminus U \to S^1 : (z_1, z_2) \mapsto \frac{z_1}{|z_1|}.$$
In polar coordinates this map is just $\pi_U(r_1 e^{i\theta_1}, r_2 e^{i\theta_2}) = \theta_1$. This fibration is related to the well known fact that S^3 is the union of two solid tori. Pictorially we see this fibration in Figure 1.

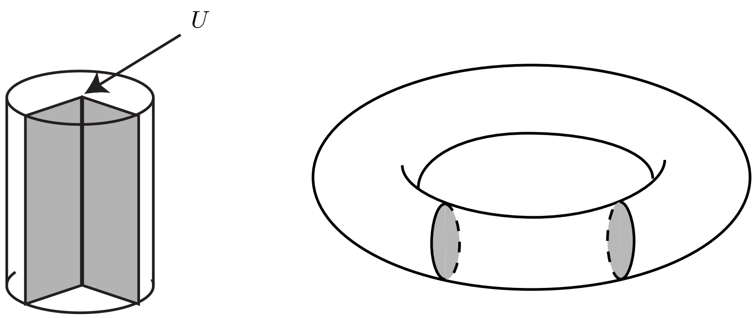


FIGURE 1. S^3 broken into two solid tori (to get the one on the left identify top and bottom of the cylinder). The union of the shaded annuli and disks give two pages in the open book.

(2) Let $H^+ = \{(z_1, z_2) \in S^3 : z_1 z_2 = 0\}$ and $H^- = \{(z_1, z_2) \in S^3 : z_1 \overline{z_2} = 0\}$.

EXERCISE 2.8. Show H^+ is the positive Hopf link and H^- is the negative Hopf link. See Figure 2. (Recall H^+ gets an orientation as the boundary of a complex hypersurface in $\mathbb{C}^2$, and H^- may be similarly oriented.)

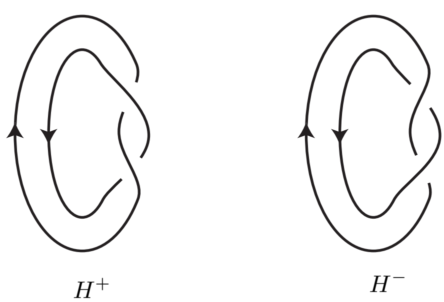


FIGURE 2. The two Hopf links.

We have the fibrations
$$\pi_+ : S^3 \setminus H^+ \to S^1 : (z_1, z_2) \mapsto \frac{z_1 z_2}{|z_1 z_2|}, \text{ and}$$

$$\pi_- : S^3 \setminus H^- \to S^1 : (z_1, z_2) \mapsto \frac{z_1\overline{z_2}}{|z_1\overline{z_2}|}.$$

In polar coordinates these maps are just $\pi_\pm(r_1e^{i\theta_1}, r_2e^{i\theta_2}) = \theta_1 \pm \theta_2$.

Exercise 2.9. Picture these fibrations.

(3) More generally, let $f : \mathbb{C}^2 \to \mathbb{C}$ be a polynomial that vanishes at $(0,0)$ and has no critical points inside S^3 except possibly $(0,0)$. Then $B = f^{-1}(0) \cap S^3$ gives an open book of S^3 with fibration

$$\pi_f : S^3 \setminus B \to S^1 : (z_1, z_2) \mapsto \frac{f(z_1, z_2)}{|f(z_1, z_2)|}.$$

This is called the Milnor fibration of the hypersurface singularity $(0,0)$ (note that $(0,0)$ does not have to be a singularity, but if it is not then B is always the unknot). See [**32**].

Exercise 2.10. Suppose Σ is a surface of genus g with n boundary components and ϕ is the identity map on Σ. Show $M_\phi = \#_{2g+n-1} S^1 \times S^2$.
HINT: If a is a properly embedded arc in Σ then $a \times [0,1]$ is an annulus in the mapping torus Σ_ϕ that can be capped off into a sphere using two disks in the neighborhood of the binding.

Theorem 2.11 (Alexander 1920, [**4**]). *Every closed oriented 3-manifold has an open book decomposition.*

We will sketch three proofs of this theorem.

First Sketch of Proof. We first need two facts

Fact (Alexander 1920, [**4**]). *Every closed oriented 3-manifold M is a branched cover of S^3 with branched set some link L_M.*

Fact (Alexander 1923, [**3**]). *Every link L in S^3 can be braided about the unknot.*

When we say L can be *braided about the unknot* we mean that if $S^1 \times D^2 = S^3 \setminus U$ then we can isotop L so that $L \subset S^1 \times D^2$ and L is transverse to $\{p\} \times D^2$ for all $p \in S^1$.

Now given M and $L_M \subset S^3$ as in the first fact we can braid L_M about the unknot U. Let $P : M \to S^3$ be the branch covering map. Set $B = P^{-1}(U) \subset M$. We claim that B is the binding of an open book. The fibering of the complement of B is simply $\pi = \pi_U \circ P$, where π_U is the fibering of the complement of U in S^3.

Exercise 2.12. Prove this last assertion and try to picture the fibration.

□

Before we continue with our two other proofs let's have some fun with branched covers.

Exercise 2.13. Use the branched covering idea in the previous proof to find various open books of S^3.
HINT: Any cyclic branched cover of S^3 over the unknot is S^3. Consider Figure 3. See also [**23, 36**].

Second Sketch of Proof. This proof comes from Rolfsen's book [**36**] and relies on the following fact.

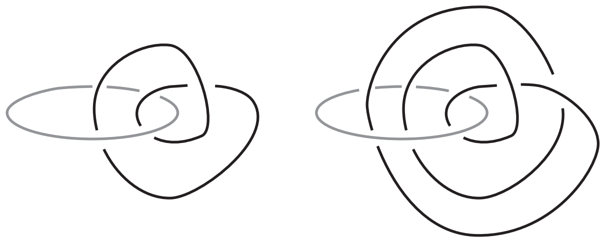


FIGURE 3. Here are two links each of which is a link of unknots. If we do a cyclic branched cover of S^3 over the grey component and lift the black component to the cover it will become the binding of an open book decomposition of S^3.

FACT (Lickorish 1962, [**30**]; Wallace 1960, [**40**]). *Every closed oriented 3-manifold may be obtained by ± 1 surgery on a link L_M of unknots. Moreover, there is an unknot U such that L_M is braided about U and each component of L_M can be assumed to link U trivially one time. See Figure 4.*

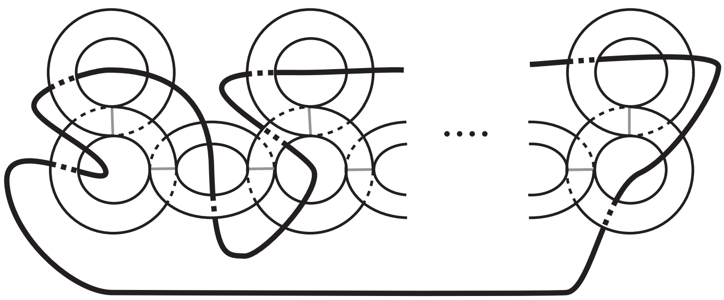


FIGURE 4. All the unknots in the link L_M can be isotoped to be on the annuli depicted here. The heavy black line is the unknot U.

Now (U, π_U) is an open book for S^3. Let N be a small tubular neighborhood of L_M. Each component N_C of N corresponds to a component C of L_M and we can assume that N_C intersects the fibers of the fibration π_U in meridional disks. So the complement of $U \cup N$ fibers so that each ∂N_C is fibered by meridional circles. To perform ± 1 surgery on L_M we remove each of the N_C's and glue it back sending the boundary of the meridional disk to a $(1, \pm 1)$ curve on the appropriate boundary component of $\overline{S^3 \setminus N}$. After the surgery we have M and inside M we have the union of surgery tori N', the components of which we denote N'_C and the cores of which we denote C'. We denote the union of the cores by L'. Inside M we also have the "unknot" U (of course U may not be an unknot any more, for example it could represent non-trivial homology in M). Since $M \setminus (U \cup N') = S^3 \setminus (U \cup N)$ we have a fibration of $M \setminus (U \cup N')$ and it is easy to see that the fibration induces on $\partial N'_C$ a fibration by $(1, \pm 1)$ curves. We can fiber $N'_C \setminus C'$ by annuli so that the induced fibration on $\partial N'_C$ is by $(1, \pm 1)$ curves. Thus we may extend the fibration of $M \setminus (U \cup N')$ to a fibration of $M \setminus (U \cup L')$, hence inducing an open book of M.

EXERCISE 2.14. Convince yourself of these last statements.

REMARK 2.15. We have produced an open book for M with planar pages!

□

THIRD SKETCH OF PROOF. This proof is due to Harer. We need

FACT (Harer 1979, [**25**]). *An oriented compact 4-manifold has an achiral Lefschetz fibration with non-closed leaves over a disk if and only if it admits a handle decomposition with only 0-, 1-, and 2-handles.*

An achiral Lefschetz fibration of a 4-manifold X over a surface S is simply a map $\pi : X \to S$ such that the differential $d\pi$ is onto for all but a finite number of points $p_1, \dots p_k \in \mathrm{int}(X)$, where there are complex coordinate charts U_i of p_i and V_i of $\pi(p_i)$ such that $\pi_{U_i}(z_1, z_2) = z_1^2 + z_2^2$. Note the definition implies that π restricted to $X \setminus \pi^{-1}(\pi(\{p_1 \dots, p_k\}))$ is a locally trivial fibration. We denote a generic fiber by Σ_π.

FACT (Lickorish 1962, [**30**]; Wallace 1960, [**40**]). *Every closed oriented 3-manifold is the boundary of a 4-manifold built with only 0- and 2-handles.*

Given a 3-manifold M we use this fact to find a 4-manifold X with $\partial X = M$ and X built with only 0- and 2-handles. Then the previous fact gives us an achiral Lefschetz fibration $\pi : X \to D^2$. Set $B = \partial\pi^{-1}(x)$ for a non-critical value $x \in \mathrm{int}(D^2)$. We claim that B is the binding of an open book decomposition for M and the fibration of the complement is the restriction of π to $M \setminus B$. □

DEFINITION 2.16. Given two abstract open books $(\Sigma_i, \phi_i), i = 0, 1$, let c_i be an arc properly embedded in Σ_i and R_i a rectangular neighborhood of c_i, $R_i = c_i \times [-1, 1]$. The *Murasugi sum* of (Σ_0, ϕ_0) and (Σ_1, ϕ_1) is the open book $(\Sigma_0, \phi_0) * (\Sigma_1, \phi_1)$ with page

$$\Sigma_0 * \Sigma_1 = \Sigma_0 \cup_{R_1 = R_2} \Sigma_1,$$

where R_0 and R_1 are identified so that $c_i \times \{-1, 1\} = (\partial c_{i+1}) \times [-1, 1]$, and the monodromy is $\phi_0 \circ \phi_1$.

THEOREM 2.17 (Gabai 1983, [**18**]).

$$M_{(\Sigma_0,\phi_0)} \# M_{(\Sigma_1,\phi_1)} \textit{ is diffeomorphic to } M_{(\Sigma_0,\phi_0)*(\Sigma_1,\phi_1)}.$$

SKETCH OF PROOF. The proof is essentially contained in Figure 5. The idea is that $B_0 = R_0 \times [\frac{1}{2}, 1]$ is a 3-ball in $M_{(\Sigma_1,\phi_1)}$ and similarly for $B_1 = R_1 \times [0, \frac{1}{2}]$ in $M_{(\Sigma_0,\phi_0)}$. Now $(\Sigma_0 * \Sigma_1) \times [0, 1]$ can be formed as shown in Figure 5.

Think about forming the mapping cylinder of ϕ_0 by gluing $\Sigma_0 \times \{0\}$ to $\Sigma_0 \times \{1\}$ using the identity and then cutting the resulting $\Sigma_0 \times S^1$ along $\Sigma_0 \times \{\frac{1}{4}\}$ and regluing using ϕ_0. Similarly think about the mapping cylinder for ϕ_1 as $\Sigma_1 \times S^1$ reglued along $\Sigma_1 \times \{\frac{3}{4}\}$ and the mapping cylinder for $\phi_0 \circ \phi_1$ as $(\Sigma_0 * \Sigma_1) \times S^1$ reglued by ϕ_0 along $(\Sigma_0 * \Sigma_1) \times \{\frac{1}{4}\}$ and by ϕ_1 along $(\Sigma_0 * \Sigma_1) \times \{\frac{3}{4}\}$. Thus we see how to fit all the mapping cylinders together nicely.

EXERCISE 2.18. Think about how the binding fits in and complete the proof.

□

DEFINITION 2.19. A *positive (negative) stabilization* of an abstract open book (Σ, ϕ) is the open book

(1) with page $\Sigma' = \Sigma \cup$ 1-handle and

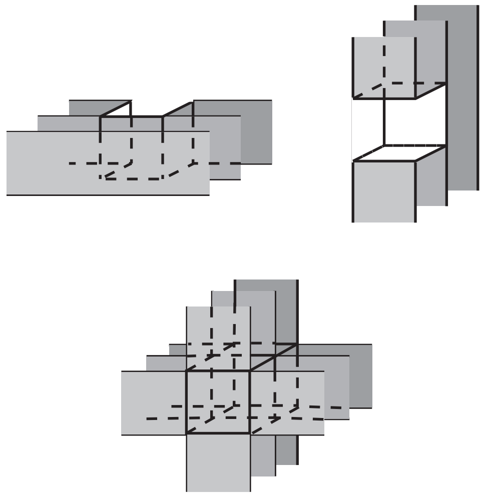


FIGURE 5. At the top left is a piece of $\Sigma_0 \times [0,1]$ near c_0 with B_0 cut out. The lightest shaded part is $\Sigma_0 \times \{0\}$ the medium shaded part is $\Sigma_0 \times \{\frac{1}{2}\}$ and the darkest shaded part is $\Sigma_0 \times \{1\}$. The top right is a similar picture for Σ_1. The bottom picture is $(\Sigma_0 * \Sigma_1) \times [0,1]$.

(2) monodromy $\phi' = \phi \circ \tau_c$ where τ_c is a right- (left-)handed Dehn twist along a curve c in Σ' that intersects the co-core of the 1-handle exactly one time.

We denote this stabilization by $S_{(a,\pm)}(\Sigma, \phi)$ where $a = c \cap \Sigma$ and $\pm$ refers to the positivity or negativity of the stabilization. (We omit the a if it is unimportant in a given context.)

EXERCISE 2.20. Show

$$S_\pm(\Sigma, \phi) = (\Sigma, \phi) * (H^\pm, \pi_\pm)$$

where $H^\pm$ is the positive/negative Hopf link and $\pi_\pm$ is the corresponding fibration of its complement.

From this exercise and Theorem 2.17 we immediately have:

COROLLARY 2.21.

$$M_{(S_\pm(\Sigma,\phi))} = M_{(\Sigma,\phi)}.$$

EXERCISE 2.22. Show how to do a Murasugi sum ambiently. That is show how to perform a Murasugi sum for open book decompositions (not abstract open books!). Of course one of the open books must be an open book for S^3.
HINT: See Figure 6.

EXERCISE 2.23. Use Murasugi sums to show the right- and left-handed trefoil knots and the figure eight knot all give open book decompositions for S^3.

EXERCISE 2.24. Use Murasugi sums to show all torus links give open book decompositions of S^3.

EXERCISE 2.25. Show that every 3-manifold has an open book decomposition with connected binding.

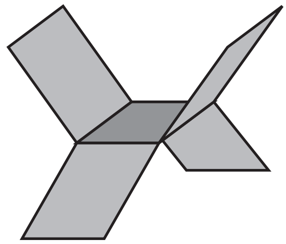


FIGURE 6. Two ambient surfaces being summed together.

EXERCISE 2.26. Use the previous exercise to prove a theorem of Bing: A closed oriented 3-manifold is S^3 if and only if every simple closed curve in M is contained in a 3-ball.
HINT: The only surface bundle over S^1 that is an orientable manifold but not irreducible is $S^1 \times S^2$.

3. From open books to contact structures

DEFINITION 3.1. An *(oriented) contact structure* ξ on M is an oriented plane field $\xi \subset TM$ for which there is a 1-form α such that $\xi = \ker \alpha$ and $\alpha \wedge d\alpha > 0$. (Recall: M is oriented.)

REMARK 3.2. What we have really defined is a *positive* contact structure, but since this is all we will talk about we will stick to this definition.

EXAMPLE 3.3. (1) On $\mathbb{R}^3$ we have the standard contact structure $\xi_{std} = \ker(dz + r^2 d\theta)$. See Figure 7.

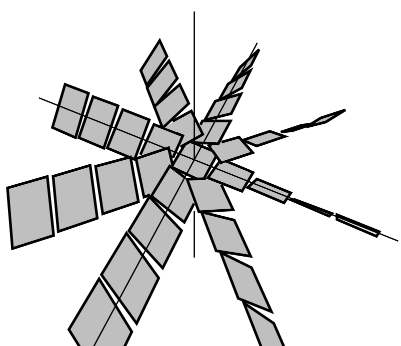


FIGURE 7. The standard contact structure on $\mathbb{R}^3$. (Picture by Stephan Schönenberger.)

(2) On S^3, thought of as the unit sphere in $\mathbb{C}^2$, we have ξ_{std} the set of complex tangents. That is $\xi_{std} = TS^3 \cap i(TS^3)$. We can also describe this plane field as $\xi_{std} = \ker(r_1^2 d\theta_1 + r_2^2 d\theta_2)$, where we are using coordinates $(r_1 e^{i\theta_1}, r_2 e^{i\theta_2})$ on $\mathbb{C}^2$.

We will need the following facts about contact structures. Most of these facts are proven or discussed in [**1, 15, 20**].

- All 3-manifolds admit a contact structure. See Theorem 3.13 below.
- Locally all contact structures look the same. This is called Darboux's Theorem and means that if p_i is a point in the contact manifold $(M_i, \xi_i), i = 0, 1$, then there is a neighborhood U_i of p_i and a diffeomorphism $f : U_0 \to U_1$ such that $f_*(\xi_0) = \xi_1$. Such a diffeomorphism is called a *contactomorphism.*
- Given two contact manifolds $(M_i, \xi_i), i = 0, 1$, we can form their *contact connected sum* $(M_0 \# M_1, \xi_0 \# \xi_1)$ as follows: there are balls $B_i \subset M_i$ and an orientation reversing diffeomorphism $f : \partial(\overline{M_0 \setminus B_0}) \to \partial(\overline{M_1 \setminus B_1})$ such that $M_0 \# M_1$ is formed by gluing $M_0 \setminus B_0$ and $M_1 \setminus B_1$ together using f and
$$\xi_0|_{M_0 \setminus B_0} \cup \xi_1|_{M_1 \setminus B_1}$$
extends to a well defined contact structure on $M_0 \# M_1$. See [**5**].
- Given a 1-parameter family of contact structures $\xi_t, t \in [0, 1]$, there is a 1-parameter family of diffeomorphisms $\phi_t : M \to M$ such that $(\phi_t)_*(\xi_0) = \xi_t$. This is called Gray's Theorem.

Two contact structures ξ_0 and ξ_1 are called isotopic if there is a 1-parameter family of contact structures connecting them.

Definition 3.4. A contact structure ξ on M is *supported* by an open book decomposition (B, π) of M if ξ can be isotoped through contact structures so that there is a contact 1-form α for ξ such that

(1) $d\alpha$ is a positive area form on each page Σ_θ of the open book and
(2) $\alpha > 0$ on B (Recall: B and the pages are oriented.)

Lemma 3.5. *The following statements are equivalent*

(1) *The contact manifold (M, ξ) is supported by the open book (B, π).*
(2) *(B, π) is an open book for M and ξ can be isotoped to be arbitrarily close (as oriented plane fields), on compact subsets of the pages, to the tangent planes to the pages of the open book in such a way that after some point in the isotopy the contact planes are transverse to B and transverse to the pages of the open book in a fixed neighborhood of B.*
(3) *(B, π) is an open book for M and there is a Reeb vector field X for a contact structure isotopic to ξ such that X is (positively) tangent to B and (positively) transverse to the pages of π.*

The condition in part (2) of the lemma involving transversality to the pages is to prevent excess twisting near the binding and may be dispensed with for tight contact structures.

Recall that a vector field X is a Reeb vector field for ξ if it is transverse to ξ and its flow preserves ξ. This is equivalent to saying there is a contact form α for ξ such that $\alpha(X) = 1$ and $\iota_X d\alpha = 0$. The equivalence of (1) and (2) is supposed to give some intuition about what it means for a contact structure to be supported by an open book. We do not actually use (2) anywhere in this paper, but it is interesting to know that "supported" can be defined this way. Similarly condition (3) should be illuminating if you have studied Reeb vector fields in the past.

Proof. We begin with the equivalence of (1) and (2). Suppose (M, ξ) is supported by (B, π). So $M \setminus B$ fibers over S^1. Let $d\theta$ be the coordinate on S^1. We also use $d\theta$ to denote the pullback of $d\theta$ to $M \setminus B$, that is, for $\pi^* d\theta$. Near each component for the binding we can choose coordinates $(\psi, (r, \theta))$ on $N = S^1 \times D^2$ in such a way that $d\theta$ in these coordinates and $\pi^* d\theta$ agree. Choosing the neighborhood N small enough we can assume that $\alpha(\frac{\partial}{\partial \psi}) > 0$ (since α is positive on B). Choose an increasing non-negative function $f : [0, \epsilon] \to \mathbb{R}$ that equals r^2 near 0 and 1 near ϵ, where ϵ is chosen so that $\{(\psi, (r, \theta)) | r < \epsilon\} \subset N$. Now consider the 1-form $\alpha_R = \alpha + Rf(r) d\theta$, where R is any large constant. (Here we of course mean that outside the region $\{(\psi, (r, \theta)) | r < \epsilon\}$ we just take f to be 1.) Note that α_R is a contact 1-form for all $R > 0$. Indeed

$$\alpha_R \wedge d\alpha_R = \alpha \wedge d\alpha + Rf d\theta \wedge d\alpha + Rf' \alpha \wedge dr \wedge d\theta.$$

The first term on the right is clearly positive since α is a contact form. The second term is also positive since $d\alpha$ is a volume form for the pages, $d\theta$ vanishes on the pages and is positive on the oriented normals to the pages. Finally the last term is non-negative since $dr \wedge d\theta$ vanishes on $\frac{\partial}{\partial \psi}$ while $\alpha(\frac{\partial}{\partial \psi}) > 0$. As $R \to \infty$ we have a 1-parameter family of contact structures $\xi_R = \ker \alpha_R$ that starts at $\xi = \xi_0$ and converges to the pages of the open book away from the binding while staying transverse to the binding (and the pages near the binding).

Now for the converse we assume (2). Let ξ_s be a family of plane fields isotopic to ξ that converge to a singular plane field tangent to the pages of the open book (and singular along the binding) as $s \to \infty$. Let α_s be contact forms for the ξ_s. We clearly have that $\alpha_s > 0$ on B.

Thinking of $M \setminus B$ as the mapping torus Σ_ϕ we can use coordinates $(x, \theta) \in \Sigma \times [0, 1]$ (we use θ for the coordinate on $[0, 1]$ since on the mapping torus Σ_ϕ this is the pullback of θ on S^1 under the fibration) and write

$$\alpha_s = \beta_s(\theta) + u_s(\theta)\, d\theta,$$

where $\beta_s(\theta)$ is a 1-form on Σ and $u_s(\theta)$ is a function on Σ for each s and θ. Let N be a tubular neighborhood of B on which ξ_s is transverse to the pages of the open book and let N' be a tubular neighborhood of B contained in N. We can choose N' so that $\overline{M \setminus N'}$ is a mapping torus $\Sigma'_{\phi'}$ where $\Sigma' \subset \Sigma$ is Σ minus a collar neighborhood of $\partial \Sigma$ and $\phi' = \phi|_{\Sigma'}$ is the identity near $\partial \Sigma'$. For s large enough $u_s(\theta) > 0$ on $\overline{M \setminus N'}$ for all θ, since α_s converges uniformly to some positive multiple of $d\theta$ on $\overline{M \setminus N'}$ as $s \to \infty$. Thus, for large s, on $\overline{M \setminus N'} = \Sigma'_{\phi'}$ we can divide α_s by $u_s(\theta)$ and get a new family of contact forms

$$\alpha'_s = \beta'_s(\theta) + d\theta.$$

We now claim that $d\alpha'_s|_{\text{page}} = d\beta'_s$ is a positive volume form on Σ'. To see this note that

$$\alpha'_s \wedge d\alpha'_s = d\theta \wedge (d\beta'_s(\theta) - \beta'_s(\theta) \wedge \frac{\partial \beta'_s}{\partial \theta}(\theta)).$$

So clearly

$$d\beta'_s(\theta) - \beta'_s(\theta) \wedge \frac{\partial \beta'_s}{\partial \theta}(\theta) > 0. \tag{1}$$

To see that $d\beta'_s(\theta) > 0$ for s large enough, we note that the second term in this equation vanishes to higher order than the first as s goes to infinity. From this one can easily conclude that $d\beta'_s(\theta) \geq 0$ for s large enough.

EXERCISE 3.6. Verify this last statement.
HINT: Assume, with out loss of generality, the 1-forms α_s are analytic and are analytic in s.

By adding a small multiple of a 1-form, similar to the one constructed on the mapping torus in the proof of Theorem 3.13 below, we easily see that for a fixed s, large enough, we can assume $d\beta'_s(\theta) > 0$.

EXERCISE 3.7. Show how to add this 1-form to α'_s preserving all the properties of α'_s in N' but still having $d\alpha'_s > 0$ on Σ'_θ.
HINT: Make sure the 1-form and its derivative are very small and use a cutoff function that is C^1-small too.

We may now assume that $d\alpha'_s$ is a volume form on the pages of $\Sigma'_{\phi'}$. Denote α'_s by α. We are left to verify α can be modified to have the desired properties in N.

EXERCISE 3.8. Show that we may assume each component of N is diffeomorphic to $S^1 \times D^2$ with coordinates $(\psi, (r, \theta))$ such that the pages of the open book go to constant θ annuli in $S^1 \times D^2$ and the contact structure $\ker \alpha$ on M maps to $\ker(d\psi + f(r)\, d\theta)$, for some function $f(r)$.
HINT: This is more than a standard neighborhood of a transverse curve. Think about the foliation on the pages of the open book near the binding and on the constant θ annuli.

Under this identification α maps to some contact form $\alpha' = h\,(d\psi + f(r)\, d\theta)$ near the boundary of $S^1 \times D^2$, where h is function on this neighborhood. By scaling α if necessary we may assume that $h > 1$ where it is defined.

EXERCISE 3.9. Show that $d\alpha'$ is a volume form on the (parts of the) constant θ annuli (where it is defined) if and only if $h_r > 0$.

Since that we know $d\alpha$ is a volume form on the pages of the mapping torus, $h_r > 0$ where it is defined. Moreover we can extend it to all of $S^1 \times D^2$ so that it is equal to 1 on $r = 0$ and so that $h_r > 0$ everywhere. Thus the contact form equal to α off of N' and equal to $h(d\psi + f(r)\, d\theta)$ on each component of N is a globally defined contact form for ξ_s and satisfies conditions (1) of the lemma.

EXERCISE 3.10. Try to show that the condition in part (2) of the lemma involving transversality to the pages of the open book near B is unnecessary if the contact structure is tight.
HINT: There is a unique universally tight contact structure on a solid torus with a fixed non-singular characteristic foliation on the boundary that is transverse to the meridional circles.

We now establish the equivalence of (1) and (3). Assume (3) and let X be the vector field discussed in (3). Since X is positively tangent to the binding we have $\alpha > 0$ on oriented tangent vectors to B. Moreover, since X is positively transverse to the pages of the open book we have $d\alpha = \iota_X \alpha \wedge d\alpha > 0$ on the pages. Thus (M, ξ) is supported by (B, π). Conversely assume (1) is true and let α be the contact form implicated in the definition of supporting open book. Let X be the Reeb vector field associated to α. It is clear that X is positively transverse to the pages of the open book since $d\alpha$ is a volume form on the pages. Thus we are left to check that X is positively tangent to B. To this end consider coordinates $(\psi, (r, \theta))$ on a neighborhood of a component of B such that constant θ's give the pages of the open book in the neighborhood. Switching (r, θ) coordinates to Cartesian coordinates (x, y) we can write $X = f\frac{\partial}{\partial \psi} + g\frac{\partial}{\partial x} + h\frac{\partial}{\partial y}$, where f, g, h are functions. We need to see that g and h are zero when $(x, y) = (0, 0)$. This is clear, for if say $g > 0$ at some point $(c, (0, 0))$ then it will be positive in some neighborhood of this point in particular at $(c, (0, \pm\epsilon))$ for sufficiently small ϵ. But at $(c, (0, \epsilon))$ the $\frac{\partial}{\partial x}$ component of X must be negative, not positive, in order to be positively transverse to the pages. Thus g and h are indeed zero along the binding. $\square$

EXAMPLE 3.11. Let (U, π_U) be the open book for S^3, where U is the unknot and

$$\pi_U : S^3 \setminus U \to S^1 : (r_1, \theta_1, r_2, \theta_2) \mapsto \theta_1.$$

(Recall that we are thinking of S^3 as the unit sphere in $\mathbb{C}^2$.) This open book supports the standard contact structure $\xi_{std} = \ker(r_1^2 d\theta_1 + r_2^2 d\theta_2)$. To see this notice that for fixed θ_1 the page $\pi_U^{-1}(\theta_1)$ is parameterized by

$$f(r, \theta) = (\sqrt{1 - r^2}, \theta_1, r, \theta).$$

Thus $d(f^*(r_1^2 d\theta_1 + r_2^2 d\theta_2)) = 2r\, dr \wedge d\theta$ which is the volume form on the disk. Moreover the positively oriented tangent to U is $\frac{\partial}{\partial \theta_2}$ and $\alpha(\frac{\partial}{\partial \theta_2}) > 0$

EXERCISE 3.12. Show that (H^+, π_+) also supports ξ_{std} but that (H^-, π_-) does not.

THEOREM 3.13 (Thurston-Winkelnkemper 1975, [**38**]). *Every open book decomposition* (Σ, ϕ) *supports a contact structure* ξ_ϕ *on* M_ϕ.

PROOF. Recall

$$M_\phi = \Sigma_\phi \cup_\psi \left(\coprod_{|\partial\Sigma|} S^1 \times D^2 \right),$$

where Σ_ϕ is the mapping torus of ϕ. We first construct a contact structure on Σ_ϕ. To this end we consider the set

$$\begin{aligned} S = \{\text{1-forms } \lambda : &(1)\ \lambda = (1 + s)d\theta \text{ near } \partial\Sigma \text{ and} \\ &(2)\ d\lambda \text{ is a volume form on } \Sigma\} \end{aligned}$$

where near each boundary component of Σ we use coordinates $(s, \theta) \in [0, 1] \times S^1$.

EXERCISE 3.14. Show this set is convex.

To show this set is non-empty let λ_1 be any 1-form on Σ that has the right form near the boundary. Note that

$$\int_\Sigma d\lambda_1 = \int_{\partial\Sigma} \lambda_1 = 2\pi|\partial\Sigma|.$$

Let ω be any volume form on Σ whose integral over Σ is $2\pi|\partial\Sigma|$ and near the boundary of Σ equals $ds \wedge d\theta$. We clearly have

$$\int_\Sigma \left(\omega - d\lambda_1\right) = 0$$

and $\omega - d\lambda_1 = 0$ near the boundary. Thus the de Rham theorem says we can find a 1-form β vanishing near the boundary such that $d\beta = \omega - d\lambda_1$. One may check $\lambda = \lambda_1 + \beta$ is a form in S.

Now given $\lambda \in S$ note that $\phi^*\lambda$ is also in S. Consider the 1-form

$$\lambda_{(t,x)} = t\lambda_x + (1-t)(\phi^*\lambda)_x$$

on $\Sigma \times [0,1]$ where $(x,t) \in \Sigma \times [0,1]$ and set

$$\alpha_K = \lambda_{(t,x)} + K dt.$$

EXERCISE 3.15. Show that for sufficiently large K this form is a contact form.

It is clear that this form descends to a contact form on the mapping torus Σ_ϕ. We now want to extend this form over the solid tori neighborhood of the binding. To this end consider the map ψ that glues the solid tori to the mapping torus. In coordinates $(\varphi, (r, \vartheta))$ on $S^1 \times D^2$ where D^2 is the unit disk in the $\mathbb{R}^2$ with polar coordinates we have

$$\psi(\varphi, r, \vartheta) = (r - 1 + \epsilon, -\varphi, \vartheta).$$

This is a map defined near the boundary of $S^1 \times D^2$. Pulling back the contact form α_K using this maps gives

$$\alpha_\psi = K\, d\vartheta - (r + \epsilon)\, d\varphi.$$

We need to extend this over all of $S^1 \times D^2$. We will extend using a form of the form

$$f(r)\, d\varphi + g(r)\, d\vartheta.$$

EXERCISE 3.16. Show this form is a contact form if and only if $f(r)g'(r) - f'(r)g(r) > 0$. Said another way, that

$$\begin{pmatrix} f(r) \\ g(r) \end{pmatrix}, \begin{pmatrix} f'(r) \\ g'(r) \end{pmatrix}$$

is an oriented basis for $\mathbb{R}^2$ for all r.

Near the boundary α_ψ is defined with $f(r) = -(r + \epsilon)$ and $g(r) = K$. Near the core of $S^1 \times D^2$ we would like $f(r) = 1$ and $g(r) = r^2$.

EXERCISE 3.17. Show that $f(r)$ and $g(r)$ can be chosen to extend α_ψ across the solid torus.
HINT: Consider the parameterized curve $(f(r), g(r))$. This curve is defined for r near 0 and 1; can we extend it over all of $[0,1]$ so that the position and tangent vector are never collinear?

$\square$

PROPOSITION 3.18 (Giroux 2000, [**21**]). *Two contact structures supported by the same open book are isotopic.*

PROOF. Let α_0 and α_1 be the contact forms for ξ_0 and ξ_1, two contact structures that are supported by (B, π). In the proof of Lemma 3.5 we constructed a contact form $\alpha_R = \alpha + Rf(r)d\theta$ from α. (See the proof of the lemma for the definitions of the various terms.) In a similar fashion we can construct α_{0R} and α_{1R} from α_0 and α_1. These are all contact forms for all $R \geq 0$. Now consider

$$\alpha_s = s\alpha_{1R} + (1-s)\alpha_{0R}.$$

EXERCISE 3.19. For large R verify that α_s is a contact form for all $0 \leq s \leq 1$.
HINT: There are three regions to consider when verifying that α_s is a contact form. The region near the binding where $f(t) = r^2$, the region where f is not 1 and the region where f is 1. Referring back to the proof of Lemma 3.5 should help if you are having difficulty when considering any of these regions.

Thus we have an isotopy from α_0 to α_1. $\square$

We now know that for each open book (B, π) there is a unique contact structure supported by (B, π). We denote this contact structure by $\xi_{(B,\pi)}$. If we are concerned with abstract open books we denote the contact structure by $\xi_{(\Sigma,\phi)}$.

THEOREM 3.20. *We have*

$$\xi_{(\Sigma_0,\phi_0)} \# \xi_{(\Sigma_1,\phi_1)} = \xi_{(\Sigma_0,\phi_0)*(\Sigma_1,\phi_1)}$$

$\square$

This theorem follows immediately from the proof of Theorem 2.17 concerning the effect of Murasugi sums on the 3–manifold. The theorem seems to have been known is some form or another for some time now but the first reference in the literature is in Torisu's paper [**39**].

EXERCISE 3.21. Go back through the proof of Theorem 2.17 and verify that the contact structures are also connect summed.
HINT: If you have trouble see the proof of Theorem 4.6.

COROLLARY 3.22 (Giroux 2000, [**21**]). *Let a be any arc in Σ, then*

$$M_{S_{(\pm,a)}(\Sigma,\phi)} \text{ is diffeomorphic to } M_{(\Sigma,\phi)}$$

and

$$\xi_{S_{(+,a)}(\Sigma,\phi)} \text{ is isotopic to } \xi_{(\Sigma,\phi)}$$

(where the corresponding manifolds are identified using the first diffeomorphism).

REMARK 3.23. The contact structure $\xi_{S_{(-,a)}(\Sigma,\phi)}$ is not isotopic to $\xi_{(\Sigma,\phi)}$! One can show that these contact structures are not even homotopic as plane fields.

PROOF. The first statement was proven in the previous section. For the second statement recall if (H^+, π_+) is the open book for the positive Hopf link then $\xi_{(H^+,\pi_+)}$ is the standard contact structure on S^3. Thus

$$\xi_{S_{(+,a)}(\Sigma,\phi)} = \xi_{(\Sigma,\phi)*(H^+,\pi_+)} = \xi_{(\Sigma,\phi)} \# \xi_{(H^+,\pi_+)} = \xi_{(\Sigma,\phi)}$$

where all the equal signs mean isotopic. The last equality follows from the following exercise.

EXERCISE 3.24. Show the contact manifold (S^3, ξ_{std}) is the union of two standard (Darboux) balls.

$\square$

EXERCISE 3.25. Show $\xi|_{M\setminus B}$ is tight. (If you do not know the definition of *tight* then see the beginning of the next section.)
HINT: Try to show the contact structure pulled back to the cover $\Sigma \times \mathbb{R}$ of $M \setminus B$ is tight. This will be much easier after you know something about convex surfaces and, in particular, Giroux's tightness criterion which is discussed at the beginning of the next section.

Note that we now have a well defined map

$$\Psi : \mathcal{O} \to \mathcal{C} \tag{2}$$

where

$$\mathcal{C} = \{\text{oriented contact structures on } M \text{ up to isotopy}\}$$

and

$$\mathcal{O} = \{\text{open book decompositions of } M \text{ up to positive stabilization}\}.$$

In the next section we will show that Ψ is onto and one-to-one.

4. From contact structures to open books

We begin this section by recalling a few more basic facts and definitions from contact geometry. Again for more details see [**1, 15, 20**]. This is not meant to be an introduction to contact geometry, but simply to remind the reader of some important facts or to allow the reader with little background in contact geometry to follow some of the arguments below. First some facts about Legendrian curves.

- A curve γ in (M, ξ) is Legendrian if it is always tangent to ξ.
- Any curve may be C^0 approximated by a Legendrian curve.
- If $\gamma \subset \Sigma$ is a simple closed Legendrian curve in Σ then $tw(\gamma, \Sigma)$ is the twisting of ξ along γ relative to Σ. Said another way, both ξ and Σ give γ a framing (that is, a trivialization of its normal bundle) by taking a vector field normal to γ and tangent to ξ or Σ, respectively; then $tw(\gamma, \Sigma)$ measures how many times the vector field corresponding to ξ rotates as γ is traversed measured with respect to the vector field corresponding to Σ.

We now turn to surfaces in contact 3-manifolds and the fundamental dichotomy in 3-dimensional contact geometry: tight vs. overtwisted.

- If Σ is a surface in (M, ξ) then if at each point $x \in \Sigma$ we consider $l_x = \xi_x \cap T_x\Sigma$ we get a singular line field on Σ. This (actually any) line field can be integrated to give a singular foliation on Σ. This singular foliation is called the *characteristic foliation* and is denoted Σ_ξ.
- The contact structure ξ is called *overtwisted* if there is an embedded disk D such that D_ξ contains a closed leaf. Such a disk is called an overtwisted disk. If there are no overtwisted disks in ξ then the contact structure is called *tight*.
- The standard contact structures on S^3 and $\mathbb{R}^3$ are tight.
- If ξ is a tight contact structure and Σ is a surface with Legendrian boundary then we have the weak-Bennequin inequality

$$tw(\partial\Sigma, \Sigma) \leq \chi(\Sigma).$$

We now begin a brief discussion of convex surfaces. These have proved to be an invaluable tool in studying 3-dimensional contact manifolds.

- A surface Σ in (M, ξ) is called *convex* if there is a vector field v transverse to Σ whose flow preserves ξ. A vector field whose flow preserves ξ is called a *contact vector field.*
- Any closed surface is C^∞ close to a convex surface. If Σ has Legendrian boundary such that $tw(\gamma, \Sigma) \leq 0$ for all components γ of $\partial\Sigma$ then after a C^0 perturbation of Σ near the boundary (but fixing the boundary) Σ will be C^∞ close to a convex surface.

- Let Σ be convex, with v a transverse contact vector field. The set
$$\Gamma_\Sigma = \{x \in \Sigma | v(x) \in \xi_x\}$$
is a multi-curve on Σ and is called the *dividing set.*
- Let $\mathcal{F}$ be a singular foliation on Σ and let Γ be a multi-curve on Σ. The multi-curve Γ is said to *divide* $\mathcal{F}$ if
  1. $\Sigma \setminus \Gamma = \Sigma_+ \coprod \Sigma_-$
  2. Γ is transverse to $\mathcal{F}$ and
  3. there is a vector field X and a volume form ω on Σ so that
     (a) X is tangent to $\mathcal{F}$ at non-singular points and $X = 0$ at the singular points of $\mathcal{F}$ (we summarize this by saying X *directs* $\mathcal{F}$)
     (b) the flow of X expands (contracts) ω on Σ_+ (Σ_-) and
     (c) X points out of Σ_+.
- If Σ is convex then Γ_Σ divides Σ_ξ.
- On any compact subset of Σ_+ we can isotop ξ to be arbitrarily close to $T\Sigma_+$ while keeping it transverse to Γ_Σ.
- If Σ is convex and $\mathcal{F}$ is any other foliation divided by Γ_Σ then there is a C^0 small isotopy, through convex surfaces, of Σ to Σ' so that $\Sigma'_\xi = \mathcal{F}$.
- If γ is a properly embedded arc or a closed curve on Σ, a convex surface, and all components of $\Sigma \setminus \gamma$ contain some component of $\Gamma_\Sigma \setminus \gamma$ then Σ may be isotoped through convex surfaces so that γ is Legendrian. This is called *Legendrian realization.*
- If Σ_1 and Σ_2 are convex, $\partial\Sigma_1 = \partial\Sigma_2$ is Legendrian and the surfaces meet transversely, then the dividing curves interlace as shown in Figure 8 and we can round the corner to get a single smooth convex surface with dividing curves shown in Figure 8.

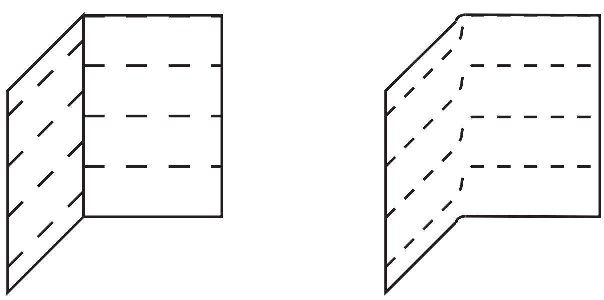


FIGURE 8. Two convex surfaces intersecting along their boundary, left, and the result of rounding their corners, right.

- If γ is a Legendrian simple closed curve on a convex surface Σ then $tw(\gamma, \Sigma) = -\frac{1}{2}(\gamma \cap \Gamma_\Sigma)$.
- If Σ is a convex surface then a small neighborhood of Σ is tight if and only if $\Sigma \neq S^2$ and no component of Γ_Σ is a contractible circle or $\Sigma = S^2$ and Γ_Σ is connected. This is called Giroux's tightness criterion.

Now that we know about convex surfaces we can discuss a fourth way to say that a contact structure is supported by an open book decomposition.

Lemma 4.1. *The contact structure ξ on M is supported by the open book decomposition (B, π) if and only if for every two pages of the open book that form a smooth surface Σ' the contact structure can be isotoped so that Σ' is convex with dividing set $B \subset \Sigma'$ and ξ is tight when restricted to the components of $M \setminus \Sigma'$.*

This criterion for an open book to be supported by a contact structure is one of the easiest to check and is quite useful in practice. This lemma is essentially due to Torisu [**39**].

Proof. Assume ξ is supported by (B, π). Let V_0 and V_1 be the closures of the complement of Σ' in M.

Exercise 4.2. Show V_i is a handlebody.
HINT: Each V_i is diffeomorphic to $\Sigma \times [0, 1]$ where Σ is a surface with boundary (*i.e.* the page).

It is not hard to show that Σ' is convex. Indeed Σ' is the union of two pages Σ_1 and Σ_2. Each Σ_i has a transverse contact vector field v_i. Along $\partial\Sigma_i$ the v_i point in opposite directions.

Exercise 4.3. Show v_1 and $-v_2$ can be altered in a neighborhood of $\partial\Sigma_1 = \partial\Sigma_2 = B$ so that they give a contact vector field v on Σ' so that B is the dividing set.
HINT: If you have trouble see [**22**].

In Exercise 3.25 you checked that ξ restricted to $M \setminus B$ is tight. It is easy to contact isotop B to be disjoint from V_i so ξ restricted to V_i is tight.

The other implication immediately follows from the next lemma. □

Lemma 4.4 (Torisu 2000, [**39**]). *Given an open book decomposition (B, π) of M there is a unique contact structure ξ that makes Σ' (the smooth union of two pages) convex with dividing set B and that is tight when restricted to each component of $M \setminus \Sigma'$.*

Sketch of Proof. Let $\Sigma \subset \Sigma'$ be a page of the open book. Let $\alpha_1, \ldots, \alpha_n$ be a collection of disjoint properly embedded arcs in Σ that cut Σ into a 2-disk. Since each component of $M \setminus \Sigma'$ is a handlebody $V_i = \Sigma \times [0, 1]$ we can consider the disks $D_j = \alpha_j \times [0, 1]$. These disks cut V_i into a 3-ball. We can Legendrian realize ∂D_j on Σ' and make all the disks D_j convex. Now cutting V_i along these disks and rounding corners we get a tight contact structure on the 3-ball. Eliashberg [**8**] has shown that there is a unique tight contact structure on the 3-ball with fixed characteristic foliation. From this it follows that there is a unique tight contact structure on V_i with any fixed characteristic foliation divided by B. Finally this implies there is at most one contact structure on M satisfying the conditions in the lemma. The existence of one contact structure satisfying these conditions is given by Theorem 3.13.

Exercise 4.5. Fill in the details of this proof.
HINT: If you have trouble read the section on convex surfaces in [**15**] or see [**39**].

□

We are now ready to show that the map Ψ from open books to contact structures is onto.

THEOREM 4.6 (Giroux 2000, [**21**]). *Every oriented contact structure on a closed oriented 3–manifold is supported by an open book decomposition.*

PROOF. We begin the proof with a definition.

DEFINITION 4.7. A *contact cell decomposition* of a contact 3–manifold (M, ξ) is a finite CW-decomposition of M such that

(1) the 1-skeleton is a Legendrian graph,
(2) each 2-cell D satisfies $tw(\partial D, D) = -1$, and
(3) ξ is tight when restricted to each 3-cell.

LEMMA 4.8. *Every closed contact 3-manifold* (M, ξ) *has a contact cell decomposition.*

PROOF. Cover M by a finite number of Darboux balls (this is clearly possible since M is compact). Note that since Darboux balls are by definition contactomorphic to a ball in the standard contact structure on $\mathbb{R}^3$ we know ξ restricted to the Darboux balls is tight. Now take any finite CW-decomposition of M such that each 3-cell sits in some Darboux ball. Isotop the 1-skeleton to be Legendrian (this can be done preserving the fact that 3-cells sit in Darboux balls). Note that we have a CW-decomposition satisfying all but condition (2) of contact cell decomposition. To achieve this condition consider a 2-cell D. By the weak-Bennequin inequality we have $tw(\partial D, D) \leq -1$. Thus we can perturb each 2-cell to be convex (care must be taken at the boundary of the 2-cells). Since Γ_D contains no simple closed curves and $tw(\partial D, D) = -\frac{1}{2}(\Gamma_D \cap \partial D)$ we know that there are $\frac{1}{2}(\Gamma_D \cap \partial D)$ components to Γ_D. If $tw(\partial D, D) \neq -1$ there is more than one component to Γ_D and we can thus use Legendrian realization to realize arcs separating all the components of Γ_D by Legendrian arcs. If we add these arcs to the 1-skeleton and subdivide the 2-skeleton then condition (2) of the definition is also satisfied. □

Suppose we have a contact cell decomposition of (M, ξ). Denote its 1-skeleton by G. Given the (or any) Legendrian graph G the *ribbon of* G is a compact surface $R = R_G$ satisfying

(1) R retracts onto G,
(2) $T_pR = \xi_p$ for all $p \in G$,
(3) $T_pR \neq \xi_p$ for all $p \in R \setminus G$.

Clearly any Legendrian graph has a ribbon. Let $B = \partial R$ and note that B is a transverse link.

CLAIM. *B is the binding of an open book decomposition of M that supports ξ.*

Clearly this claim finishes the proof. □

PROOF OF CLAIM. Since B is a transverse link there is a contactomorphism from each component of a neighborhood $N(B)$ of B to an ϵ-neighborhood of the z-axis in $(\mathbb{R}^3, \ker(dz + r^2\, d\theta))/\sim$ where $(r, \theta, z) \sim (r, \theta, z+1)$. Let $X(B) = \overline{M \setminus N(B)}$ be the complement of $N(B)$ and $R_X = R \cap X(B)$. We can choose a neighborhood $N(R) = R_X \times [-\delta, \delta]$ of R_X in $X(B)$ such that $\partial R_X \times \{pt\}$ thought of as sitting in $N(B)$ is a line with constant θ value. Clearly $N(R)$ is an R_X bundle over $[-\delta, \delta]$. Set $X(R) = X(B) \setminus N(R)$. See Figure 9.

We first show that $X(R)$ is diffeomorphic to $R_X \times [0, 1]$ and that $X(B)$ is formed by identifying $R_X \times \{0\}$ $(R_X \times \{1\})$ in $X(R)$ and $R_X \times \{\delta\}$ $(R_X \times \{-\delta\})$ in $N(R)$.

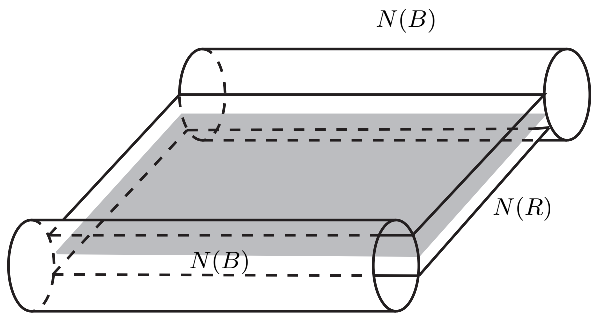


FIGURE 9. The neighborhoods $N(B)$ and $N(R)$. The grey part is R_X.

Clearly this implies that $X(B)$, the complement of a neighborhood of B, is fibered and the fibration can be extended over $N(B)\setminus B$ so that the boundary of the fibers is B. Note that $\partial X(R) = A \cup F$ where $A = (\partial X(R)) \cap N(B)$ is a disjoint union of annuli (one for each component of $N(B)$) that are naturally fibered by circles of constant θ value in $N(B)$. The subsurface F is defined to be the closure of the complement of A in $\partial X(R)$. Note that we can write $F = F^- \cup F^+$, where $F^\pm$ is identified with $R_X \times \{\pm\delta\}$ in $N(R)$.

EXERCISE 4.9. Show that $\partial X(R)$ is a convex surface with dividing set $\Gamma_{\partial X(R)}$ equal to the union of the cores of A and such that $F^\pm \subset (\partial X(R))_\pm$. (Note, $\partial X(R)$ is only piecewise smooth, but if we rounded the edges it would be convex.)

REMARK 4.10. Throughout this part of the proof we will be discussing manifolds whose boundaries have corners. We do not want to smooth the corners. However, sometimes to understand the annuli A better we will think about rounding the corners, but once we have understood A sufficiently we actually will not round the corners.

Let $D_1, \ldots, D_k$ be the two cells in the contact cell decomposition of (M, ξ). Recall that ∂D_i is Legendrian and has twisting number -1. Thus since R twists with the contact structure along the 1-skeleton G we can assume that B intersects D_i exactly twice for all i. Let $D_i' = D_i \cap X(R)$.

EXERCISE 4.11. Show that it can be arranged that the D_i''s intersect the region $A \subset \partial X(R)$ in exactly two properly embedded arcs, and each arc runs from one boundary component of A to another.

EXERCISE 4.12. Show that the interior of $X(R)$ cut along all the D_i''s is homeomorphic to M minus its 2-skeleton. That is, $X(R)$ cut along the D_i''s is a union of balls.

Using the Legendrian realization principle we can assume $\partial D_i'$ is Legendrian. (Again as in the remark above we only Legendrian realize $\partial D_i'$ to see what happens to the dividing curves when we cut $X(R)$ along these disks. Once we have seen this we don't actually do the Legendrian realization.) Let's consider what happens to $X(R)$ when we cut along D_1'; denote the resulting manifold by X_1. Note: in ∂X_1 there are two copies of D_1'. Let A_1 be A sitting in ∂X_1 union the two copies of D_1. See Figure 10.

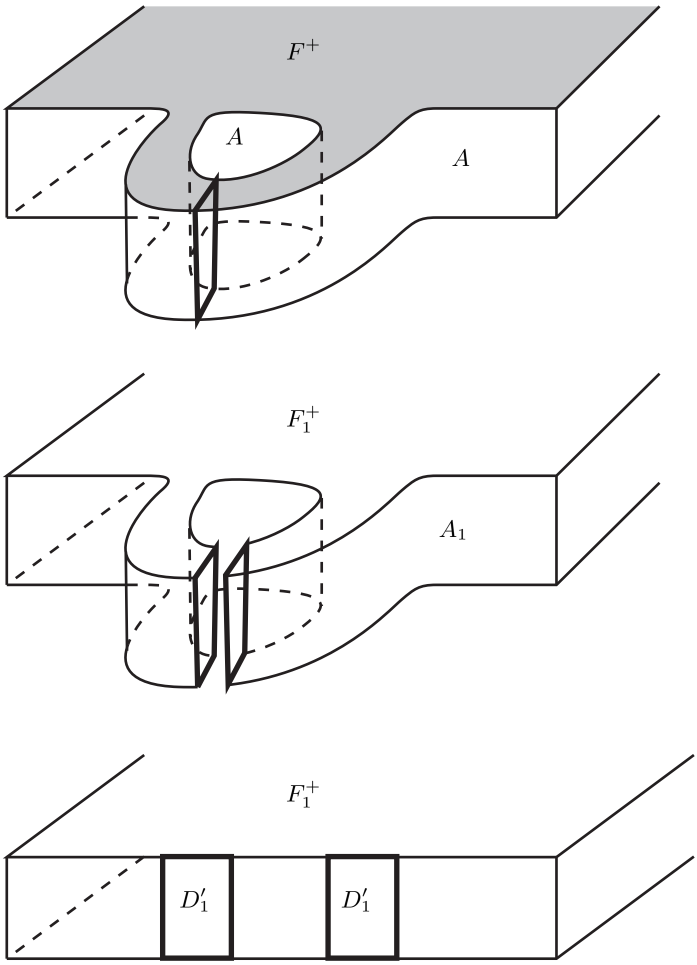


FIGURE 10. Top picture is $X(R)$ with the boundary of D_1' drawn darkly. The middle picture is X_1 right after the cut and the bottom picture is X_1 after isotoping a little.

EXERCISE 4.13. Show that the components of A_1 are annuli and that they have a natural fibration by S^1 that is naturally related to the fibration on A.

Note that $\partial X_1 \setminus A_1$ naturally breaks into two surfaces F_1^+ and F_1^-, where $F_1^{\pm}$ is obtained from $F^{\pm}$ by cutting along a properly embedded arc.

EXERCISE 4.14. Show ∂X_1 is convex (once the corners are rounded) and its dividing set is the union of the cores of A_1 and $F_1^{\pm} \subset (\partial X_1)_{\pm}$.

If we continue to cut along the D_i''s we eventually get to X_k once we have cut along all the disks. From above we know X_k is a disjoint union of balls (all contained in 3-cells of our contact cell decomposition). Moreover, on ∂X_k we have that A_k a union of annuli whose cores give the dividing curves for ∂X_k. By the definition of contact cell decomposition we know that the contact structure when restricted to each component of X_k is tight. Thus we know A_k has exactly one component on the boundary of each component of X_k. Thus each component of X_k is a ball B^3 with an annulus S that has a natural fibration by circles. Clearly B^3 has a natural fibration by D^2's that extends the fibration of S by circles. That is, $B^3 = D^2 \times [0,1]$ with $(\partial D^2) \times [0,1] = S$.

Exercise 4.15. Show that as we glue X_k together along the two components of D'_k in ∂X_k to get X_{k-1} we can glue the fibration of X_k by D^2's together to get a fibration of X_{k-1} by surfaces that extend the fibration of A_{k-1} by circles.

Thus, continuing in this fashion we get back to $X(R)$ and see that it is fibered by surfaces that extend the fibration of A by circles. This clearly implies that $X(R) = F^- \times [0,1] = R_X \times [0,1]$ and the surfaces $R_X \times \{0\}$ and $R_X \times \{1\}$ are glued to the boundary of $N(R)$ as required above. Hence we have shown that $X(B)$ is fibered over the circle by surfaces diffeomorphic to R and that the fibers all have boundary B. That is, we have demonstrated that B is the binding of an open book.

We now must show that this open book supports the contact structure ξ. Looking back through the proof it is not hard to believe that one may isotop the contact planes to be arbitrarily close to the pages of the open book, but it seems a little difficult to prove this directly. We will show the open book is compatible with the contact structure by showing that there is a Reeb vector field that is tangent to the binding and transverse to the pages. Recall that the neighborhood $N(B)$ of the binding is contactomorphic to an ϵ-neighborhood of the z-axis in $(\mathbb{R}^3, \ker(dz + r^2\, d\theta))/\sim$ where $(r,\theta,z) \sim (r,\theta,z+1)$. Moreover, we can assume the pages intersect this neighborhood as the constant θ annuli.

Exercise 4.16. In the explicit model for $N(B)$ find a Reeb vector field that is tangent to the binding and positively transverse to the pages of the open book in the neighborhood. Also make sure the boundary of $N(B)$ is preserved by the flow of the Reeb field.

We can think of the Reeb fields just constructed as giving a contact field in the neighborhood of the boundary of R_X (recall this is the ribbon of the Legendrian 1-skeleton G intersected with the complement of $N(B)$).

Exercise 4.17. Show that this contact vector field defined in a neighborhood of ∂R_X can be extended to a contact vector field v over the rest of R_X so that it is transverse to R_X and there are no dividing curves. (This is OK since R_X is not a closed surface.) Note that since there are no dividing curves v is also transverse to ξ.

Use v to create the neighborhood $N(R)$ of R_X. Since v is never tangent to the contact planes along R_X we can assume that this is the case in all of $N(R)$.

Exercise 4.18. Show that a contact vector field which is never tangent to the contact planes is a Reeb vector field.

Thus we have a Reeb vector field defined on $N(B) \cup N(R)$ that has the desired properties.

We now need to extend the Reeb vector field v over $X(R)$. From the construction of v we can assume we have v defined near the boundary of $X(R)$ and as a vector field defined there it satisfies the following:

(1) v is tangent to $A \subset \partial X(R)$.

(2) There is a neighborhood $N(A)$ of A in $X(R)$ such that $\partial N(A) = A \cup A^+ \cup A^- \cup A'$ where $A^\pm = N(A) \cap F^\pm$ and A' is a parallel copy of A on the interior of $X(R)$. v is defined in $N(A)$, tangent to $A \cup A'$, $\pm v$ points transversely out of $N(A)$ at $A^\pm$ and v is transverse to the pages of the open book intersected with $N(A)$. Moreover the flow of v will take A^- to A^+.

(3) $\pm v$ points transversely out of $X(R)$ along $F^\pm$.

We now want to construct a model situation into which we can embed $X(R)$. To this end let $\Sigma = R_X \cup A' \cup -R_X$, where $A' = (\partial R_X) \times [0,1]$ and the pieces are glued together so that Σ is diffeomorphic to the double of R_X. On Σ let $\mathcal{F}$ be the singular foliation $(R_X)_\xi$ on each of R_X and $-R_X$ and extend this foliation across A' so that it is non singular there and the leaves of the foliation run from one boundary component to another. Let Γ be the union of the cores of the annuli that make up A'. It is easy to see that $\mathcal{F}$ is divided by Γ. Given this one can create a vertically invariant contact structure ξ' on $\Sigma \times \mathbb{R}$ such that $(\Sigma \times \{t\})_{\xi'} = \mathcal{F}$ and the dividing set on $\Sigma \times \{t\}$ is Γ, for all $t \in \mathbb{R}$. (See [**22**].) Note that $\frac{\partial}{\partial t}$ restricted to $R_X \times \mathbb{R}$ is a Reeb vector field since it is a contact vector field and positively transverse to ξ' in this region. Pick a diffeomorphism $f : F^- \to (R_X \times \{0\} \subset \Sigma \times \{0\})$ that sends $(R_X)_\xi$ to $\mathcal{F}$. (Recall $F^+ \cup F^- = (\partial X(R)) \setminus A$.)

Exercise 4.19. Show that the flow of v on R_X and $\frac{\partial}{\partial t}$ on $\Sigma \times \mathbb{R}$ allow you to extend f to a contact embedding of $N(A)$ into $\Sigma \times \mathbb{R}$.

Thus we can use the flow of v and $\frac{\partial}{\partial t}$ to extend f to a contact embedding of a neighborhood of $\partial X(R)$ in $X(R)$ into $\Sigma \times \mathbb{R}$.

Exercise 4.20. Make sure you understand how to get the embedding near F^+.
HINT: From the previous exercise we have a neighborhood of the boundary of F^+ embedded into $R_X \times \{t_0\}$, for some t_0. Show that there is an obvious way to extend this to an embedding of all of F^+ to $R_X \times \{t_0\}$.

Of course this extension of f, which we also call f, takes the Reeb field v to the Reeb field $\frac{\partial}{\partial t}$. We can clearly extend f to an embedding, but not necessarily a contact embedding, of all of $X(R)$ into $\Sigma \times \mathbb{R}$. The following exercises allow us to isotop f, relative to a neighborhood of the boundary, to a contact embedding and thus we may extend v to all of $X(R)$ by $\frac{\partial}{\partial t}$. This gives us a Reeb vector field on M which demonstrates that the open book supports ξ. $\square$

Exercise 4.21. Let H be a handlebody and $D_1, \ldots, D_g$ be properly embedded disks that cut H into a 3-ball. Given any singular foliation $\mathcal{F}$ on the boundary of H that is divided by Γ for which $\partial D_i \cap \Gamma = 2$, for all i, then there is at most one tight contact structure on H, up to isotopy, that induces $\mathcal{F}$ on ∂H.
HINT: This is a simple exercise in convex surface theory. See [**15**].

Exercise 4.22. Show the contact structure ξ' on $\Sigma \times \mathbb{R}$ is tight. (This is easy using Giroux's tightness criterion.) Also show the contact structure ξ restricted to $X(R)$ is tight.
HINT: The second part is not so easy. The idea is that if you can cut up a handlebody by disks, as in the previous exercise, and the 3-ball you end up with has a tight contact structure on it, then the original contact structure on the handlebody is tight. See [**26**].

We have the following immediate useful corollaries.

Corollary 4.23. *If L is a Legendrian link in (M, ξ) then there is an open book decomposition supporting ξ such that L sits on a page of the open book and the framing given by the page and by ξ agree.*

Proof. Simply include the Legendrian link L in the 1-skeleton of the contact cell decomposition. ☐

Example 4.24. Figure 11 illustrates Corollary 4.23 for two knots in S^3 with its standard contact structure.

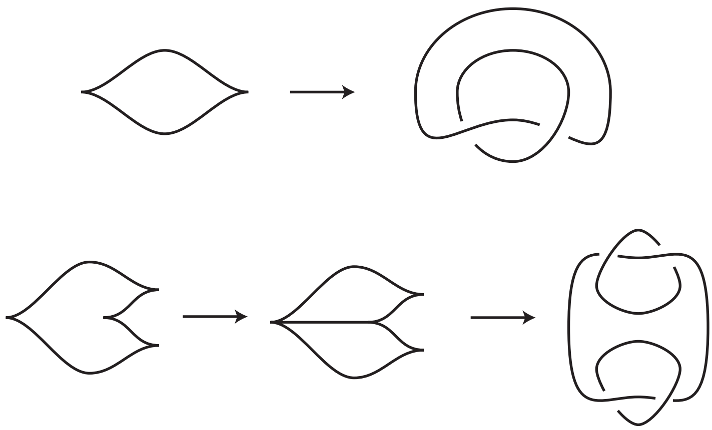


Figure 11. On the top left we have a Legendrian unknot that is the 1-skeleton of a contact cell decomposition of S^3. The resulting open book is shown on the upper right. On the bottom left we start with a Legendrian unknot, moving to the right we add a Legendrian arc to get the 1-skeleton of a contact cell decomposition. The bottom right shows the resulting open book.

Using positive stabilizations we can see the following.

Corollary 4.25. *Any contact manifold is supported by an open book with connected binding.*

Corollary 4.26 (Contact Bing). *A contact manifold (M, ξ) is the standard tight contact structure on S^3 if and only if every simple closed curve is contained in a Darboux ball.*

Exercise 4.27. Prove this last corollary.
HINT: There is a unique tight contact structure on B^3 inducing a fixed characteristic foliation on the boundary [**8**].

Theorem 4.28 (Giroux 2002, [**21**]). *Two open books supporting the same contact manifold (M, ξ) are related by positive stabilizations.*

To prove this theorem we need the following lemma.

Lemma 4.29. *Any open book supporting (M, ξ), after possibly positively stabilizing, comes from a contact cell decomposition.*

Proof. Let Σ be a page of the open book and G be the core of Σ. That is, G is a graph embedded in Σ onto which Σ retracts. We can Legendrian realize G.

Remark 4.30. The Legendrian realization principle is for curves, or graphs, on a closed convex surface or a convex surface with Legendrian boundary. The pages of an open book are convex but their boundary is transverse to the contact structure so we cannot apply the Legendrian realization principle as it is usually stated. Nonetheless since we can keep the characteristic foliation near the boundary fixed while trying to realize a simple closed curve or graph, we can still realize it. But recall the curve or graph must be non-isolating. In this context this means that all components of the complement of the curve in the surface should contain a boundary component. To see this review Giroux's proof of realization.

Note that Σ is the ribbon of G. Let N be a neighborhood of Σ such that $\partial\Sigma \subset \partial N$. Let α_i be a collection of properly embedded arcs on Σ that cut Σ into a disk. Let $\widetilde{A}_i$ be $\alpha_i \times [0,1]$ in $M \setminus \Sigma = \Sigma \times [0,1]$ and $A_i' = \widetilde{A}_i \cap \overline{(M \setminus N)}$. Note that A_i' intersects $\partial\Sigma$ on ∂N exactly twice. Thus if we extend the A_i's into N so their boundaries lie on G then the twisting of Σ, and hence ξ, along ∂A_i with respect to A_i is $-1, 0$ or 1. If all the twisting is -1 then we have a contact cell decomposition (recall that the contact structure restricted to the complement of Σ is tight). Thus we just need to see how to reduce the twisting of ξ along ∂A_i.

Suppose ∂A_i has twisting 0. Positively stabilize Σ as shown in Figure 12. Note

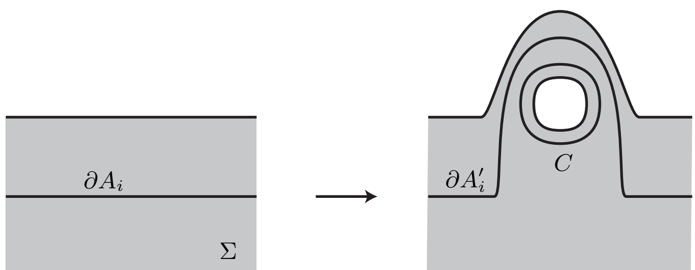


Figure 12. On the left is part of Σ and ∂A_i near $\partial\Sigma$. On the right is the stabilized Σ, C and $\partial A_i'$. (This picture is abstract. If we were drawing it ambiently the added 1-handle would have a full left-handed twist in it.)

the curve C shown in the picture can be assumed to be Legendrian and bounds a disk D in M. Now isotop G across D to get a new Legendrian graph with all the A_j's unchanged except that A_i is replace with the disk A_i' obtained from A_i by isotoping across D. We also add C to G and add D to the 2-skeleton.

Exercise 4.31. Show that the twisting of ξ along $\partial A_i'$ is one less than the twisting along ∂A_i.

Clearly the twisting of ξ along D is -1. Thus we can reduce the twisting of ξ along ∂A_i as needed and after sufficiently many positive stabilizations we have an open book that comes from a contact cell decomposition. $\square$

Proof of Theorem 4.28. Given two open books (B, π) and (B', π') supporting (M, ξ) we can assume they both come from contact cell decompositions by using Lemma 4.29. Now, given two contact cell decompositions one can show that they are related by a sequence of the following:

(1) A subdivision of a 2-cell by a Legendrian arc intersecting the dividing set one time.
(2) Add a 1-cell c' and a 2-cell D so that $\partial D = c' \cup c$ where c is part of the original 1-skeleton and $tw(\partial D, D) = -1$.
(3) Add a 2-cell D whose boundary is already in the 1-skeleton and $tw(\partial D, D) = -1$.

Thus the theorem follows from the following exercises.

Exercise 4.32. Show that (3) does not change the open book associated to the cell decomposition.

Exercise 4.33. Show that (1) and (2) positively stabilize the open book associated to the cell decomposition.
HINT: Show that an arc a is added to the 1-skeleton and a disk D to the 2-skeleton so that ∂D is part of the old 1-skeleton union a, Γ_D is a single arc and $a \cap \Gamma_D$ is one point. Show that adding such an arc to the one skeleton is equivalent to positively stabilizing the open book.

□

5. Symplectic cobordisms and caps.

A contact manifold (M, ξ) is called *(weakly) symplectically fillable* if there is a compact symplectic 4–manifold (X, ω) such that $\partial X = M$ and $\omega|_\xi \neq 0$. Many applications of contact geometry to topology (see [**28, 34**] and the discussion in the introduction) rely on the following theorem.

Theorem 5.1 (Eliashberg 2004 [**6**]; Etnyre 2004 [**14**]). *If (X, ω) is a symplectic filling of (M, ξ) then there is a closed symplectic manifold (W, ω') and a symplectic embedding $(X, \omega) \to (W, \omega')$.*

Partial results aimed towards this theorem were obtained by many people. Specifically, Lisca and Matić established this result for Stein fillable manifolds in [**31**] and later work of Akbulut and Ozbagci [**2**] coupled with work of Plamenevskaya [**35**] provided an alternate proof in this case (for unfamiliar terminology see the next paragraph). For strongly fillable manifolds this was proven by Gay in [**19**] and follows trivially from Theorem 1.3 in [**16**]. The full version of this theorem also follows fairly easily from [**37**].

In the process of proving Theorem 5.1 we will need to take a few detours. The first concerns various types of symplectic fillings and the second concerns Legendrian/contact surgery. These two detours occupy the next two subsections. We return to the proof of Theorem 5.1 in Subsection 5.3. In the final section we discuss the relation between open book decompositions and overtwistedness.

5.1. Symplectic fillings. A contact manifold (M, ξ) is said to be *strongly symplectically filled* by the symplectic manifold (X, ω) if X is compact, $\partial X = M$ and there is a vector field v transversely pointing out of X along M such that the flow of v dilates ω (that is to say the Lie derivative of ω along v is a positive multiple of ω). The symplectic manifold (X, ω) is said to have convex boundary if there is a contact structure ξ on ∂X that is strongly filled by (X, ω). We say that

(X, ω) is a *strong concave filling* if (X, ω) and v are as above except that v points into X. Note that given a symplectic manifold (X, ω) with a dilating vector field v transverse to its boundary then $\iota_v\omega$ is a contact form on ∂X. If v points out of X then the contact form gives an oriented contact structure on ∂X and if v points into X then it gives an oriented contact structure on $-\partial X$. (Recall Remark 3.2)

Given a contact manifold (M, ξ) let α be a contact form for ξ, then consider $W = M \times \mathbb{R}$ and set $\omega_W = d(e^t\alpha)$, where t is the coordinate on $\mathbb{R}$. It is easy to see that ω_W is a symplectic form on W and the vector field $v = \frac{\partial}{\partial t}$ is a dilating vector field for ω_W. The symplectic manifold (W, ω_W) is called the *symplectization* of (M, ξ).

Exercise 5.2. Given any other contact form α' for ξ (note that this implies that $\alpha' = g\alpha$ for some function $g : M \to (0, \infty)$) show there is some function f such that $\alpha' = F^*(\iota_v\omega_W)$ where $F : M \to W : x \mapsto (x, f(x))$.

It can be shown that if (X, ω) is a strong symplectic filling (strong concave filling) of (M, ξ) then there is a neighborhood N of M in X, a function f, a one-sided neighborhood N_f of the graph of f in W with N_W lying below (above) the graph and a symplectomorphism $\psi : N_W \to N$. See Figure 13. Thus we have a

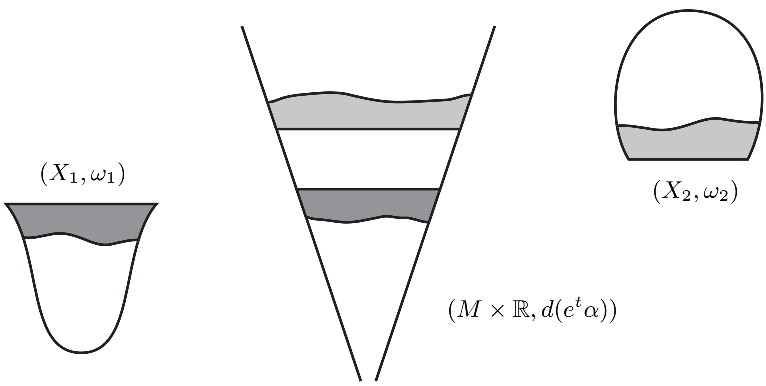


Figure 13. The symplectization of (M, ξ), middle, and a symplectic manifold with convex, left, and concave, right, boundary.

model for a neighborhood of a contact manifold in a strong symplectic filling.

Exercise 5.3. If (X_1, ω_1) is a strong symplectic filling of (M, ξ) and (X_2, ω_2) is a strong concave filling of (M, ξ) then show $X = X_1 \cup X_2$ has a symplectic structure ω such that $\omega|_{X_1} = \omega_1$ and $\omega|_{X_2 \setminus N} = c\omega_2$ where N is a neighborhood of ∂X_2 in X_2 and $c > 0$ is a constant.
HINT: Look at Figure 13.

Thus we can use strong symplectic fillings to glue symplectic manifolds together. This is not, in general, possible with a weak symplectic filling.

Recall that a Stein manifold is a triple (X, J, ψ) where J is a complex structure on X and $\omega_\psi(v, w) = -d(d\psi \circ J)(v, w)$ is non-degenerate. A contact manifold (M, ξ) is called *Stein fillable* (or *holomorphically fillable*) if there is a Stein manifold (X, J, ψ) such that ψ is bounded from below, M is a non-critical level of ψ and $-(d\psi \circ J)$ is a contact form for ξ. It is customary to think of X as $\psi^{-1}((-\infty, c])$

where $M = \psi^{-1}(c)$. Thus we can think of X as a compact manifold (Stein manifolds themselves are never compact).

In [**11, 41**] it was shown how to attach a 1-handle to the boundary of a symplectic manifold with convex boundary and extend the symplectic structure over the 1-handle so as to get a new symplectic manifold with convex boundary. They also showed the same could be done when a 2-handle is attached along a Legendrian knot with framing one less than the contact framing. In fact we have the following characterization of Stein manifolds.

THEOREM 5.4 (Eliashberg 1990, [**11**]). *A 4–manifold X is Stein if and only if X has a handle decomposition with only 0-handles, 1-handles and 2-handles attached along Legendrian knots with framing one less than the contact framing.* □

Summarizing the relations between various notions of filling and tightness we have

$$\text{Tight} \quad \supset \quad \text{Weakly Fillable} \quad \supset \quad \text{Strongly Fillable} \quad \supset \quad \text{Stein Fillable.}$$

The first two inclusions are strict, see [**16**] and [**7**] respectively. It is unknown whether or not the last inclusion is strict. We have the following useful fact.

THEOREM 5.5 (Eliashberg 1991, [**9**]; Ohta and Ono 1999, [**33**]). *If M is a rational homology sphere then any weak filling of (M, ξ) can be deformed into a strong filling.* □

5.2. Contact surgery. Let L be a Legendrian knot in a contact 3–manifold (M, ξ). It is well known (see [**15, 20**]) that L has a neighborhood N_L that is contactomorphic to a neighborhood of the x-axis in

$$(\mathbb{R}^3, \ker(dz - y\,dx))/\sim,$$

where $\sim$ identifies (x, y, z) with $(x+1, y, z)$. With respect to these coordinates on N_L we can remove N_L from M and topologically glue it back with a ± 1-twist (that is, we are doing Dehn surgery along L with framing the contact framing ± 1). Call the resulting manifold $M_{(L,\pm 1)}$. There is a unique way, up to isotopy, to extend $\xi|_{M\setminus N_L}$ to a contact structure $\xi_{(L,\pm 1)}$ over all of $M_{(L,\pm 1)}$ so that $\xi_{(L,\pm 1)}|_{N_L}$ is tight (see [**27**]). The contact manifold $(M, \xi)_{(L,\pm 1)} = (M_{(L,\pm 1)}, \xi_{(L,\pm 1)})$ is said to be obtained from (M, ξ) by ± 1-*contact surgery* along L. It is customary to refer to -1-contact surgery along L as *Legendrian surgery* along L.

QUESTION 1. *Is $\xi_{(L,-1)}$ tight if ξ is tight?*

If the original contact manifold (M, ξ) is not closed then it is known that the answer is sometimes NO, see [**26**]. But there is no known such example on a closed manifold. It is known, by a combination of Theorems 5.4 and 5.5, that Legendrian surgery (but not $+1$-contact surgery!) preserves any type of symplectic fillability. (Similarly, $+1$-contact surgery preserves non-fillability.) We have the following result along those lines.

THEOREM 5.6 (Eliashberg 1990, [**11**]; Weinstein 1991, [**41**]). *Given a contact 3–manifold (M, ξ) let $(W = M \times [0, 1], \omega = d(e^t\alpha))$ be a piece of the symplectization of (M, ξ) discussed in the last section. Let L be a Legendrian knot sitting in (M, ξ) thought of as $M \times \{1\}$. Let W' be obtained from W by attaching a 2-handle to W along $L \subset M \times \{1\}$ with framing one less than the contact framing. Then ω may be extended over W' so that the upper boundary is still convex and the induced contact* *manifold is* $(M_{(L,-1)}, \xi_{(L,-1)})$. *Moreover, if the 2-handle was added to a Stein filling (respectively weak filling, strong filling) of* (M, ξ) *then the resulting manifold would be a Stein filling (respectively weak filling, strong filling) of* $(M_{(L,-1)}, \xi_{(L,-1)})$. $\square$

We now want to see how contact surgery relates to open book decompositions. The main result along these lines is the following.

THEOREM 5.7. *Let* (Σ, ϕ) *be an open book supporting the contact manifold* (M, ξ). *If* L *is a Legendrian knot on the page of the open book then*

$$(M, \xi)_{(L,\pm 1)} = (M_{(\Sigma, \phi\circ D_L^{\mp})}, \xi_{(\Sigma, \phi\circ D_L^{\mp})}).$$

PROOF. We begin by ignoring the contact structures and just concentrating on the manifold. We have a simple closed curve L on the page Σ of the open book. Recall $M \setminus \text{nbhd } B$ is the mapping cylinder Σ_ϕ. We will think of L as sitting on $\Sigma \times \{\frac{1}{2}\}$ in $\Sigma \times [0, 1]$, then by moding out by the identification $(\phi(x), 0) \sim (x, 1)$ we will have L on a page in $\Sigma_\phi \subset M$.

EXERCISE 5.8. Show that cutting Σ_ϕ open along $\Sigma \times \{\frac{1}{2}\}$ and regluing using $D_L^{\pm}$ will give you a manifold diffeomorphic to $\Sigma_{\phi\circ D_L^{\pm}}$.

Let N_Σ be a closed tubular neighborhood of L in $\Sigma \times \{\frac{1}{2}\}$. Then a neighborhood N of L in M is given by $N_\Sigma \times [\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon]$. We can assume the support of $D_L^{\mp}$ is in N_Σ. Thus if N' is neighborhood of L in $\Sigma_{\phi\circ D_L^{\mp}}$ corresponding to N in Σ_ϕ then

$$\Sigma_\phi \setminus N = \Sigma_{\phi\circ D_L^{\mp}} \setminus N'.$$

So clearly $M_{(\Sigma,\phi)} \setminus N$ is diffeomorphic to $M_{(\Sigma,\phi\circ D_L^{\mp})} \setminus N'$. Said another way $M_{(\Sigma,\phi\circ D_L^{\mp})}$ is obtained from $M_{(\Sigma,\phi)}$ by removing a solid torus and gluing it back in, *i.e.* by a Dehn surgery along L. We are left to see that the Dehn surgery is a ± 1 Dehn surgery with respect to the framing on L coming from the page on which it sits. See Figure 14 while reading the rest of this paragraph. Note that we get N' from N by

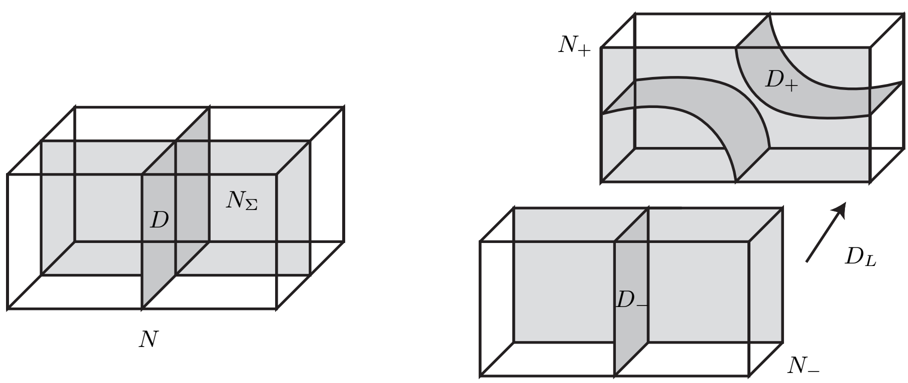


FIGURE 14. On the right is the neighborhood N with its meridional disk D. On the right is the neighborhood N'. (The right- and left-hand sides of each cube are identified to get a solid torus.)

cutting N along N_Σ and regluing using $D_L^{\mp}$. We get a meridional disk D for N by taking an arc a on $N_\Sigma \times \{\frac{1}{2} - \epsilon\}$ running from one boundary component to the other and setting $D = a\times[\frac{1}{2}-\epsilon, \frac{1}{2}+\epsilon]$. Let $N_- = N_\Sigma\times[\frac{1}{2}-\epsilon, \frac{1}{2}]$ and $N_+ = N_\Sigma\times[\frac{1}{2}, \frac{1}{2}+\epsilon]$. Thus we get N' by gluing $\Sigma\times\{\frac{1}{2}\}$ in N_+ to $\Sigma\times\{\frac{1}{2}\}$ in N_-. Set $D_- = D\cap N_-$. Then $a_- = D_-\cap\Sigma\times\{\frac{1}{2}\}$ in N_- is taken to $a_+\subset\Sigma\times\{\frac{1}{2}\}\subset N_+$. So D', the meridional disk in N', is $D_-\cup a_+\times[\frac{1}{2}-\epsilon, \frac{1}{2}]$. Thus on $\partial M_{(\Sigma,\phi)}\setminus N = \partial M_{(\Sigma,\phi\circ D_L^{\mp})}\setminus N'$ the curve $\partial D'$ is homologous to $\partial D\pm L'$, where L' is a parallel copy of L lying on ∂N. Thus $M_{(\Sigma,\phi\circ D_L^{\mp})}$ is obtained from $M_{(\Sigma,\phi)}$ by a ± 1 Dehn surgery on L.

We now must see that the contact structure one gets from ±1-contact surgery on L is the contact structure supported by the open book $(\Sigma, \phi\circ D_L^{\mp})$. To do this we consider the definition of compatible open book involving the Reeb vector field.

EXERCISE 5.9. Think about trying to show that ξ can be isotoped arbitrarily close to the pages of the open book. Intuitively this is not too hard to see (but as usual, making a rigorous proof out of this intuition is not so easy).

Since (Σ,ϕ) is compatible with ξ there is a Reeb vector field X for ξ such that X is positively transverse to the pages and tangent to the binding. Notice our neighborhood $N = N_\Sigma\times[\frac{1}{2}-\epsilon, \frac{1}{2}+\epsilon]$ is such that X is transverse to all the $N_\Sigma\times\{t\}$ and the flow of X preserves these pages. As usual we will now consider a model situation. Consider $(\mathbb{R}^3, \ker(dz-y\,dx))/\sim$, where $(x,y,z)\sim(x+1,y,z)$. It is easy to arrange that the foliation on N_Σ is the same as the foliation on $\{(x,y,z)|-\delta\leq y\leq\delta\}$. Thus we can assume the contact structure on N is contactomorphic to the contact structure on $N_m = \{(x,y,z)|-\epsilon\leq z\leq\epsilon, -\delta\leq y\leq\delta\}$. Moreover, we can assume this contactomorphism takes a Reeb vector field for ξ to $\frac{\partial}{\partial z}$, the Reeb vector field for $dz-y\,dx$. (We do this by picking an identification of N_Σ and the annulus in the xy-plane that preserves the characteristic foliation and then using the Reeb vector fields to extend this identification.) Let $\psi : \partial N_m\to\partial N$ be the diffeomorphism that agrees with the above contactomorphism everywhere except on $S_u = \{(x,y,z)|z=\epsilon, -\delta\leq y\leq\delta\}\subset\partial N_m$, where it differs by $D_L^{\pm}$. Note that gluing N_m to $M\setminus N$ using ψ will yield the manifold $M_{(L,\pm1)}$. Use ψ to pullback the characteristic foliation on $\partial(M\setminus N)$ to ∂N_m. Note that the characteristic foliation agrees with the characteristic foliation on $\partial N_m\setminus S_u$ induced by $\ker(dz-y\,dx)$.

EXERCISE 5.10. Show there is a function $f : \{(x,y)|-\delta\leq y\leq\delta\}\to\mathbb{R}$ that equals ϵ near $|y|=\delta$ such that the characteristic foliation on the graph of f agrees with the foliation on S_u that is pulled back from N by ψ. (By "agrees with" I mean that S_u and the graph of f are isotopic rel boundary so that the isotopy takes the pulled back foliation to the characteristic foliation on the graph of f.)
HINT: Figure out what the pullback foliation is first. Then experiment with perturbing the graph of the constant function.

Now let N'_m be the region bounded by $\partial N_m\setminus S_u$ union the graph of f. There is a natural way to think of ψ as a map from $\partial N'_m$ to $\partial(M\setminus N)$ that preserves the characteristic foliation. Furthermore we can extend ψ to a neighborhood of the boundary so that it preserves Reeb vector fields. The contact structures $\xi_{M\setminus N}$ and $\ker(dz-y\,dx)|_{N'_m}$ glue to give a contact structure on $M_{(L,\pm1)}$. We can also glue up the Reeb vector fields to get a Reeb vector field on $M_{(L,\pm1)}$ that is transverse to the pages of the obvious open book and tangent to the binding. □

An easy corollary of this theorem is the following.

THEOREM 5.11 (Giroux 2002, [**21**]). *A contact manifold (M,ξ) is Stein fillable if and only if there is an open book decomposition for (M,ξ) whose monodromy can be written as a composition of right-handed Dehn twists.*

PROOF. We start by assuming that there is an open book (Σ,ϕ) supporting (M,ξ) for which ϕ is a composition of right-handed Dehn twists. Let us begin by assuming that ϕ is the identity map on Σ. In Exercise 2.10 you verified that $M = \#_{2g+n-1}S^1\times S^2$, where g is the genus of Σ and n is the number of boundary components. Eliashberg has shown that $\#_{2g+n-1}S^1\times S^2$ has a unique strong symplectic filling [**10**]. This filling (W,ω) is also a Stein filling. Thus we are done if ϕ is the identity map.

Now assume $\phi = D^+_\gamma$ where γ is a simple non-separating closed curve on Σ. We can use the Legendrian realization principle to make γ a Legendrian arc on a page of the open book. (Recall that even though our convex surface does not have Legendrian boundary we can still use the Legendrian realization principle. See Remark 4.30.) (Note that we required γ to be non-separating so that we could use the Legendrian realization principle.) We know $(M_{(\Sigma,id)},\xi_{(\Sigma,id)})$ is Stein filled by (W,ω) so by Theorems 5.6 and 5.7 we can attach a 2-handle to W to get a Stein filling of $(M_{(\Sigma,D^+_\gamma)},\xi_{(\Sigma,D^+_\gamma)})$. If ϕ is a composition of more than one right-handed Dehn twist along non-separating curves in Σ we may clearly continue this process to obtain a Stein filling of $(M_{(\Sigma,\phi)},\xi_{(\Sigma,\phi)})$. The only thing left to consider is when one or more of the curves on which we Dehn twist is separating. Suppose γ is separating. If both components of the complement of γ contain parts of the boundary of the page then we can still realize γ. Thus we only have a problem when there is a subsurface Σ' of Σ such that $\partial\Sigma' = \gamma$. In this case we can use the "chain relation" (see Theorem 6.5 in the Appendix) to write D^+_γ as a composition of positive Dehn twists along non-separating curves in Σ'.

For the other implication we assume that (M,ξ) is Stein fillable by say (W,J,ψ). According to Eliashberg's Theorem 5.4, W has a handle decomposition with only 1-handles and 2-handles attached along Legendrian knots with framing one less than the contact framing. Let W' be the union of the 0- and 1-handles. Clearly $M' = \partial W' = \#_k S^1\times S^2$ and the induced contact structure ξ' is tight. So we have a Legendrian link L in M' on which we can perform Legendrian surgery to obtain (M,ξ). Now according to Corollary 4.23 there is an open book decomposition (Σ,ϕ) for (M',ξ') such that L sits on a page of the open book. At the moment ϕ might not be the composition of right-handed Dehn twists. According to a theorem of Eliashberg [**8**] and Colin [**5**] M' has a unique tight contact structure. We know there is a surface Σ' such that $M' = M_{(\Sigma',id)}$ and the supported contact structure is tight. According to Giroux's Stabilization Theorem 4.28 we can positively stabilize (Σ',id) and (Σ,ϕ) so that they become isotopic. Let (Σ'',ϕ'') be their common stabilization. Since the stabilizations were all positive ϕ'' is a composition of positive Dehn twists. Moreover L sits on a page of this open book. As in the previous paragraph of the proof, performing Legendrian surgery on L will change the open book (Σ'',ϕ'') by composing ϕ'' by right-handed Dehn twists. Thus we eventually get an open book for (M,ξ) whose monodromy consists of a composition of right-handed Dehn twists. □

5.3. Proof of Theorem 5.1. We are ready to begin the proof of Theorem 5.1.

LEMMA 5.12 (Etnyre 2004, [**14**]; *cf.* Stipsicz 2003, [**37**]). *Suppose (X,ω) is a weak filling of (M,ξ). Then there is a compact symplectic manifold (X',ω') such that (X,ω) embeds in (X',ω') and the boundary of (X',ω') is strongly convex.*

PROOF. Let (B,π) be an open book for (M,ξ). Using positive stabilizations of the open book we can assume the binding is connected. Let ϕ be the monodromy of the open book. It is well known, see Lemma 6.7 in the Appendix, that ϕ can be written

$$\phi = D_c^m \circ D_{\gamma_1}^{-1} \circ \ldots \circ D_{\gamma_n}^{-1},$$

where γ_i are non-separating curves on the interior of the page Σ and c is a curve on Σ parallel to the boundary of Σ.

We know $M \setminus B$ is the mapping cylinder Σ_ϕ and an identical argument to the one in the first paragraph of the proof of Theorem 5.7 says we may think of Σ_ϕ as

$$\coprod_{i=1}^{n} \Sigma_i / \sim,$$

where $\Sigma_i = \Sigma \times [\frac{i-1}{n}, \frac{i}{n}]$ and $\sim$ is the equivalence relation that glues $\Sigma \times \{\frac{i}{n}\}$ in Σ_i to $\Sigma \times \{\frac{i}{n}\}$ in Σ_{i+1} by $D_{\gamma_i}^{-1}$ and $\Sigma \times 1$ to $\Sigma \times 0$ by D_c^m. See Figure 15.

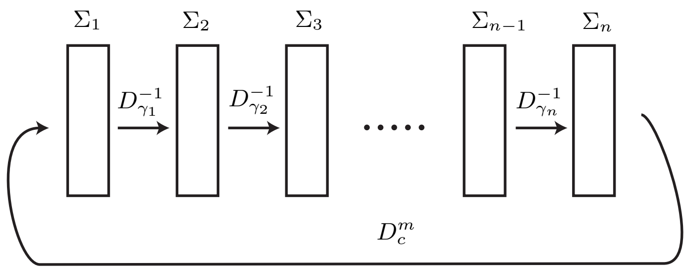


FIGURE 15. Breaking up the monodromy.

Choose a point $p \in (0, \frac{1}{n})$. We can Legendrian realize γ_1 on the surface $\Sigma \times \{p\}$ in $\Sigma_1 \subset M$. If we cut Σ_ϕ along $\Sigma \times \{p\}$ and reglue using the diffeomorphism D_{γ_1} then the new open book decomposition will have page Σ and monodromy $D_c^m \circ D_{\gamma_2}^{-1} \circ \ldots \circ D_{\gamma_n}^{-1}$. Continuing in this way we can get an open book with page Σ and monodromy D_c^m. Denote the contact manifold supported by this open book by (M',ξ'). By Theorem 5.7 we know we can get from (M,ξ) to (M',ξ') by a sequence of Legendrian surgeries. Thus by Theorem 5.6 we can add 2-handles to (X,ω) in a symplectic way to get a symplectic manifold (X'',ω'') with weakly convex boundary equal to (M',ξ').

Let $a_1,\ldots,a_{2g}$ be the curves on Σ pictured in Figure 16. We can Legendrian realize these curves on separate pages of the open book for (M',ξ') and do Legendrian surgery on them to get the contact manifold (M'',ξ''). Moreover, we can add 2-handles to (X'',ω'') to obtain the symplectic manifold (X',ω') with weakly convex boundary (M'',ξ''). The open book supporting (M'',ξ'') has page Σ and monodromy $D_{a_1} \circ \ldots \circ D_{a_{2g}} \circ D_c^m$.

EXERCISE 5.13. Show M'' is topologically obtained from S^3 by $\frac{1}{m}$ Dehn surgery on the knot in Figure 17. Thus M'' is a homology sphere.

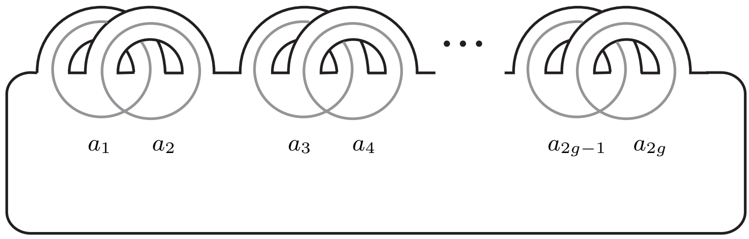


FIGURE 16. The curves $a_1, \ldots a_{2g}$ on Σ.

FIGURE 17. Topological description of M''.

Now Theorem 5.5 says for a homology sphere a weak filling can be deformed into a strong filling. Thus we may deform (X', ω') into a strong filling of (M'', ξ'') and clearly (X, ω) symplectically embeds into this manifold. $\square$

LEMMA 5.14 (Etnyre and Honda 2002, [**16**]; Gay 2002, [**19**]). *Given any contact manifold (M, ξ) there is a strong concave filling of (M, ξ).*

This Lemma was also proven by Lisca and Matić [**31**] for Stein fillable contact structures. An alternate proof in the Stein case was provided by the work of Akbulut and Ozbagci [**2**] coupled with that of Plamenevskaya [**35**]. The proof below is in the spirit of Gay's work.

PROOF. We start with the symplectic manifold $(W, \omega_W) = (M \times [0,1], d(e^t\alpha))$, where α is a contact form of ξ and t is the coordinate on $[0,1]$. It is easy to see that $M \times \{0\}$ is a strongly concave boundary component of W and $M \times \{1\}$ is a strongly convex boundary. Our strategy will be to cap off the convex boundary component so we are left with only the concave boundary component. Throughout this proof we will call the concave boundary of a symplectic manifold the *lower boundary component* and the convex boundary component the *upper boundary component*.

Let (Σ, ϕ) be an open book for (M, ξ) with a connected boundary. As in the proof of Lemma 5.12 we can add 2-handles to W to get a symplectic manifold (W', ω') with lower boundary M and upper boundary a contact manifold with open book having page Σ and monodromy D_c^m where c is a boundary parallel curve in Σ.

We want to argue that we can assume that $m = 1$. If $m < 1$ then by adding symplectic 2-handles to (W', ω') we can get to the situation where $m = 1$. (Note c is separating, but we can handle this as we did in the proof of Theorem 5.11.) Throughout the rest of the proof as we add handles to (W', ω') we still denote the resulting manifold by (W', ω'). We are left to consider the situation where $m > 1$. For this we observe that we can increase the genus of Σ as follows.

EXERCISE 5.15. Show that if we add a symplectic 1-handle to (W', ω') this has the effect on the upper boundary of W' of connect summing with the standard (unique tight) contact structure on $S^1 \times S^2$.

But we know connect summing the contact manifold can be achieved by Murasugi summing their open books. The tight contact structure on $S^1 \times S^2$ has open book with page an annulus and monodromy the identity map. Thus adding symplectic 1-handles to W' has the effect on the open book of the upper boundary component of adding a 1-handle to Σ and extending the old monodromy over this handle by the identity. So by adding 1-handles to W we can arrange that the open book for the upper boundary component has page Σ' shown in Figure 18 and mon-

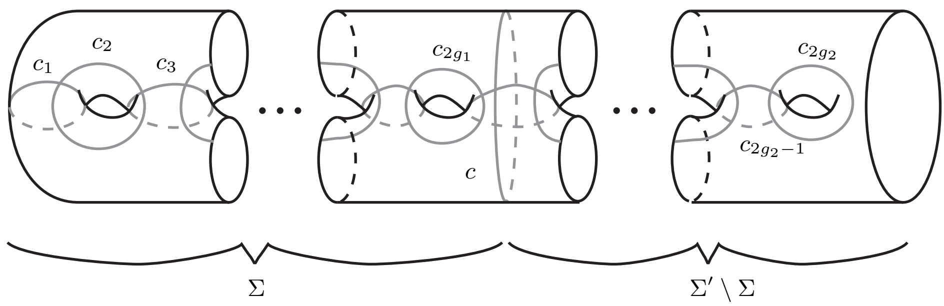


FIGURE 18. The surfaces Σ and Σ' and the curves c and $c_1, \ldots a_{2g_2}$.

odromy D_c^m. Let $c_1, \ldots, c_{2g_1}$ be the curves in Σ shown in Figure 18; g_1 is the genus of Σ. The Chain Relation (see Theorem 6.5) says $D_c^m = (D_{c_1} \circ \ldots \circ D_{c_{2g_1}})^{m(4g_1+2)}$. Now let $c_{g_1+1}, \ldots, c_{2g_2}$ be the curves shown in Figure 18; g_2 is the genus of Σ'.

EXERCISE 5.16. Show that we can assume m is such that we can choose the genus g_2 of Σ' so that $m(4g_1 + 2) = 4g_2 + 2$.
HINT: Attach symplectic 2-handles.

Thus we can attach symplectic 2-handles to (W', ω') so that the upper boundary (M', ξ') has an open book decomposition with page Σ' and monodromy $D_{c'} = (D_{c_1} \circ \ldots \circ D_{c_{2g_2}})^{4g_2+2}$, where c' is a curve on Σ' parallel to the boundary.

EXERCISE 5.17. Show M' is an S^1 bundle over Σ'' with Euler number -1, where Σ'' is Σ' with a disk capping of its boundary.

EXERCISE 5.18. Let C be the D^2 bundle over Σ'' with Euler number 1. Construct a natural symplectic structure ω_C on C.
HINT: On the circle bundle ∂C there is a connection 1-form α that is also a contact form on ∂C. Use this to construct the symplectic form on C. Note that if you think about the symplectization of a contact structure you can easily get a symplectic structure on C minus the zero section. Some care is needed to extend over the zero section.

EXERCISE 5.19. Show (C, ω_C) has a strongly concave boundary, $\partial C = -M'$, and ξ' is the induced contact structure.
HINT: The contact structure induced on ∂C is transverse to the circle fibers. If you remove a neighborhood of one of the fibers the resulting manifold is $\Sigma' \times S^1$.

This is the mapping cylinder of the identity on Σ'. Show that the contact planes can be isotoped arbitrarily close to the pages. Now consider how the neighborhood of the fiber is glued back in.

We now simply glue (C, ω_C) to the top of (W', ω') to get our concave filling of (M, ξ). $\square$

We are now ready to prove the main result of this section.

Proof of Theorem 5.1. We start with a weak symplectic filling (X, ω) of (M, ξ). Now apply Lemma 5.12 to embed (X, ω) symplectically into (X', ω') where (X', ω') has a strongly convex boundary (M', ξ'). Now use Lemma 5.14 to find a symplectic manifold (X'', ω'') that is a strong concave filling of (M', ξ'). Using an exercise from Subsection 5.1 we can glue (X', ω') and (X'', ω'') together to get a closed symplectic manifold (W, ω_W) into which (X, ω) embeds. $\square$

5.4. Sobering arcs and overtwisted contact structures. Theorem 5.11 gives a nice characterization of Stein fillable contact structures in terms of open book decompositions. It turns out there is a similar characterization of overtwisted contact structures due to Goodman [**24**]. Suppose we are given an oriented surface Σ. Given two properly embedded oriented arcs a and b on Σ with $\partial a = \partial b$ we can isotop them relative to the boundary so that the number of intersection points between the arcs is minimized. At a boundary point x of a define $\epsilon(x)$ to be $+1$ if the oriented tangent to a at x followed by the oriented tangent to b at x is an oriented basis for Σ, otherwise we set $\epsilon(x) = -1$. Let $i(a, b) = \frac{1}{2}(\epsilon(x) + \epsilon(y))$, where x and y are the boundary points of a.

Definition 5.20. Let Σ be an oriented surface and $\phi : \Sigma \to \Sigma$ a diffeomorphism that fixes the boundary. An arc b properly embedded in Σ is a *sobering arc* for the pair (Σ, ϕ) if $i(b, \phi(b)) \geq 0$ and there are no positive intersection points of b with $\phi(b)$ (after isotoping to minimize the number of intersection points).

Note the definition of sobering arc does not depend on an orientation on b.

Theorem 5.21 (Goodman 2004, [**24**]). *If (Σ, ϕ) admits a sobering arc then the corresponding contact structure $\xi_{(\Sigma,\phi)}$ is overtwisted.* $\square$

We will not prove this theorem but indicate by an example how one shows a contact structure is overtwisted if a supporting open book admits a sobering arc. Indeed, consider (A, ϕ) where $A = S^1 \times [-1, 1]$ and ϕ is a left-handed Dehn twist about $S^1 \times \{0\}$. Of course this is the open book describing the negative Hopf link in S^3. Earlier we claimed that the associated contact structure is overtwisted; we now find the overtwisted disk. (Actually we find a disk whose Legendrian boundary violates the Bennequin inequality, but from this one can easily locate an overtwisted disk.) The arc $b = \{pt\} \times [-1, 1] \subset A$ is obviously a sobering arc. Let T be the union of two pages of the open book. Clearly T is a torus that separates S^3 into two solid tori V_0 and V_1. We can think of V_0 as $A \times [0, \frac{1}{2}]$ (union part of the neighborhood of the binding if you want to be precise) and V_1 is $A \times [\frac{1}{2}, 1]$. In V_0 we can take $D_0 = b \times [0, \frac{1}{2}]$ to be the meridional disk and in V_1 we take $D_1 = b \times [\frac{1}{2}, 1]$ to be the meridional disk. If we think of T as the boundary of V_1 then ∂D_1 is simply a meridional curve, that is, a $(1, 0)$ curve. Note that ∂D_0 does not naturally sit on T. We must identify $A \times \{\frac{1}{2}\} \subset V_0$ with $A \times \{\frac{1}{2}\} \subset V_1$ using the identity map and

$A \times \{0\} \subset V_0$ with $A \times \{1\} \subset V_1$ via ϕ. Thus ∂D_1 in T is a $(-1, 1)$ curve. In particular these two meridional curves intersect once on T. We can Legendrian realize a $(0, 1)$ curve on T. Call the Legendrian curve L. Note since $(0, 1) = (1, 0) + (-1, 1)$, L bounds a disk D in S^3. With respect to the framing induced by T the twisting of L is 0.

EXERCISE 5.22. Show that the framing induced on L by T is one larger than the framing induced by D. Thus $tw(\xi, D) = 1$.

So we see that ∂D violates the Bennequin inequality and thus ξ is overtwisted.

EXERCISE 5.23. Starting with D find an overtwisted disk.

In general, in the proof of the above theorem you will not always be able to find an explicit overtwisted disk, but in a manner similar to what we did above you will always be able to construct a Legendrian knot bounding a surface that violates the Bennequin inequality.

Lastly we have the following theorem.

THEOREM 5.24 (Goodman 2004, [**24**]). *A contact structure is overtwisted if and only if it is there is an open book decomposition supporting the contact structure that admits a sobering arc.*

The if part of this theorem is the content of the previous theorem. The only if part follows from:

EXERCISE 5.25. Show that any overtwisted contact structure is supported by an open book that has been negatively stabilized. Show that this implies there is a sobering arc.
HINT: You will need to use Eliashberg's classification of overtwisted contact structures by their homotopy class of plain field [**12**]. If you are having trouble you might want to consult [**13**] or, of course, the original paper [**24**].

6. Appendix

We recall several important facts about diffeomorphisms of surfaces. First, given an embedded curve γ in an oriented surface Σ let $N = \gamma \times [0, 1]$ be a (oriented) neighborhood of the curve. We then define the *right-handed Dehn twists along* γ, denoted D_γ, to be the diffeomorphism of Σ that is the identity on $\Sigma \setminus N$ and on N is given by $(\theta, t) \mapsto (\theta + 2\pi t, t)$, where θ is the coordinate on $\gamma = S^1$ and t is the coordinate on $[0, 1]$ and we have chosen the product structure so that $\frac{\partial}{\partial \theta}, \frac{\partial}{\partial t}$ is an oriented basis for $N \subset \Sigma$. (Note that to make D_γ a diffeomorphism one needs to "smooth" it near ∂N.) A *left-handed Dehn twists about* γ is D_γ^{-1}.

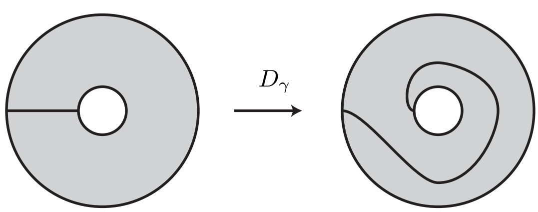


FIGURE 19. A right-handed Dehn twist.

EXERCISE 6.1. Show the following

(1) D_γ does not depend on an orientation on γ.
(2) If γ and γ' are isotopic then D_γ and $D_{\gamma'}$ are isotopic diffeomorphisms.

THEOREM 6.2 (Lickorish 1962, [**30**]). *Any diffeomorphism of a compact oriented surface can be written as a composition of Dehn twists about non-separating curves and curves parallel to the boundary of the surface.* □

There are several important relations among Dehn twists. For example:

(1) For any γ and diffeomorphism f we have $f \circ D_\gamma \circ f^{-1} = D_{f(\gamma)}$.
(2) If γ and δ are disjoint then $D_\gamma \circ D_\delta = D_\delta \circ D_\gamma$.
(3) If γ and δ intersect in one point then $D_\delta \circ D_\gamma(\delta)$ is isotopic to γ.
(4) If γ and δ intersect in one point then $D_\delta \circ D_\gamma \circ D_\delta = D_\gamma \circ D_\delta \circ D_\gamma$.

EXERCISE 6.3. Prove the above relations.

EXERCISE 6.4. Show that given two non-separating curves γ and δ on Σ there is a diffeomorphism of Σ taking γ to δ.

In many of the above applications of open books we needed the following fundamental relation called the Chain Relation.

THEOREM 6.5. *Let $\gamma_1, \ldots, \gamma_k$ be a chain of simple closed curves in Σ, that is, the curves satisfy $\gamma_i \cdot \gamma_j$ is 1 if $|i-j| = 1$ and is 0 otherwise, where $\cdot$ means geometric intersection. Let N be a neighborhood of the union of the γ_i's. If k is odd then N has two boundary components d_1 and d_2. If k is even then N has one boundary component d. We have the following relations*

$$(D_{\gamma_1} \circ \ldots \circ D_{\gamma_k})^{2k+2} = D_d, \quad \textit{if } k \textit{ is even, and}$$

$$(D_{\gamma_1} \circ \ldots \circ D_{\gamma_k})^{k+1} = D_{d_1} \circ D_{d_2} \quad \textit{if } k \textit{ is odd.}$$

□

EXERCISE 6.6. Try to prove this theorem. Note: for $k = 1$ it is trivial, for $k = 2, 3$ it is quite easy to explicitly check the relation.

An important consequence of this theorem is the following lemma.

LEMMA 6.7. *Let Σ be a surface with one boundary component, then any diffeomorphism of Σ can be written as the composition of right-handed Dehn twists about non-separating curves on the interior of Σ and arbitrary Dehn twists about a curve parallel to the boundary of Σ.*

PROOF. We find a chain of curves $\gamma_1, \ldots, \gamma_{2g}$ in Σ such that Σ is a neighborhood of their union. Thus the chain relation tells us that $(D_{\gamma_1} \circ \ldots \circ D_{\gamma_{2g}})^{4g+2} = D_d$. So clearly we can replace $D_{\gamma_i}^{-1}$ by a composition of right-handed Dehn twists and one left-handed Dehn twist about d. Now by the exercises above any left-handed Dehn twist about a separating curve can be written as right-handed Dehn twists about non-separating curves and a left-handed Dehn twist about d. □

References

[1] B. Aebischer, et. al., *Symplectic Geometry*, Progress in Math. **124**, Birkhäuser, Basel, Boston and Berlin, 1994.
[2] S. Akbulut, B. Ozbagci, *On the topology of compact Stein surfaces*, Int. Math. Res. Not. 2002, no. **15**, 769–782.
[3] J. W. Alexander, *A lemma on systems of knotted curves*, Proc. Nat. Acad. Sci. USA, **9** (1923), 93–95.
[4] J. W. Alexander, *Note on Riemann spaces*, Bull. Amer. Math. Soc., **26** (1920), 370–372.
[5] V. Colin, *Chirurgies d'indice un et isotopies de sphères dans les variétés de contact tendues*, C. R. Acad. Sci. Paris Sér. I Math. **324** (1997), 659–663.
[6] Y. Eliashberg, *A few remarks about symplectic filling*, Geom. Topol. **8** (2004) 277–293.
[7] Y. Eliashberg, *Unique holomorphically fillable contact structure on the* 3*-torus*, Internat. Math. Res. Notices **2** (1996), 77–82.
[8] Y. Eliashberg, *Contact 3-manifolds twenty years since J. Martinet's work*, Ann. Inst. Fourier **42** (1992), 165–192.
[9] Y. Eliashberg, *On symplectic manifolds with some contact properties*, J. Diff. Geom. **33** (1991), 233–238.
[10] Y. Eliashberg, *Filling by holomorphic discs and its applications*, Geometry of low-dimensional manifolds, Vol. II (Ed. Donaldson and Thomas), Cambridge, 1990.
[11] Y. Eliashberg, *Topological characterization of Stein manifolds of dimension > 2*, Int. J. of Math. **1** (1990), 29–46.
[12] Y. Eliashberg, *Classification of overtwisted contact structures on 3-manifolds*, Invent. Math. **98** (1989), 623–637.
[13] J. Etnyre, *Planar open book decompositions and contact structures*, Int. Math. Res. Not. **2004:79** (2004) 4255–4267.
[14] J. Etnyre, *On Symplectic Fillings*, Algebr. Geom. Topol. **4** (2004), 73–80.
[15] J. Etnyre, *Introductory Lectures on Contact Geometry*, Topology and geometry of manifolds (Athens, GA, 2001), 81–107, Proc. Sympos. Pure Math., **71**, Amer. Math. Soc., Providence, RI, 2003.
[16] J. Etnyre, K. Honda, *On Symplectic Cobordisms*, Math. Annalen **323** (2002), no. 1, 31-39.
[17] J. Etnyre, K. Honda, *Tight contact structures with no symplectic fillings*, Invent. Math **148** (2002), no. 3, 609-626.
[18] D. Gabai, *The Murasugi sum is a natural geometric operation*, Low-dimensional topology (San Francisco, Calif., 1981), 131–143, Contemp. Math., **20**, Amer. Math. Soc., Providence, RI, 1983.
[19] D. Gay, *Explicit concave fillings of contact three-manifolds*, Proc. Cam. Phil. Soc. **133** (2002), 431–441.
[20] H. Geiges, *Contact Geometry*, to appear in the Handbook of Differential Geometry, vol. 2.
[21] E. Giroux, *Géométrie de contact: de la dimension trois vers les dimensions supérieures*, Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002), 405–414, Higher Ed. Press, Beijing, 2002.
[22] E. Giroux, *Convexité en topologie de contact*, Comment. Math. Helv. **66** (1991), no. 4, 637–677.
[23] D. L. Goldsmith, *Symmetric fibered links*, Knots, groups, and 3-manifolds (Papers dedicated to the memory of R. H. Fox), pp. 3–23. Ann. of Math. Studies, No. **84**, Princeton Univ. Press, Princeton, N.J., 1975.
[24] N. Goodman, *Overtwisted Open Books From Sobering Arcs*, arXiv:math.GT/0407420
[25] J. Harer, *Pencils of curves on 4-manifolds*, UC Berkeley PhD thesis 1979.
[26] K. Honda, *Gluing tight contact structures*, Duke Math. J., **115** (2002), no. 3, 435–478.
[27] K. Honda, *On the classification of tight contact structures I*, Geom. Topol. **4** (2000), 309–368.
[28] P. Kronheimer, T. Mrowka, *Witten's conjecture and property P*, Geom. Topol. **8** (2004), 295–310.
[29] P. Kronheimer, T. Mrowka, P. Ozsváth, Z. Szabó. *Monopoles and lens space surgeries*, preprint (arXiv: math.GT/0310164).
[30] W. B. R. Lickorish, *A representation of orientable combinatorial* 3*-manifolds*, Ann. of Math. (2) **76** 1962 531–540.

[31] P. Lisca, G. Matić, *Tight contact structures and Seiberg-Witten invariants*, Invent. Math. **129** (1997), no. 3, 509–525.

[32] J. Milnor, *Singular points of complex hypersurfaces*, Volume **61**, Ann. of Math. Stud. Princeton Univ. Press. 1968.

[33] H. Ohta, K. Ono, *Simple singularities and topology of symplectically filling 4-manifold*, Comment. Math. Helv. **74** (1999), 575–590.

[34] P. Ozsváth, Z. Szabó, *Holomorphic disks and genus bounds*, Geom. Topol. **8** (2004), 311–334.

[35] O. Plamenevskaya, *Contact structures with distinct Heegaard-Floer invariants*, Math. Res. Lett. **11** (2004), no. 4, 547–561.

[36] D. Rolfsen, *Knots and Links*, Mathematics Lecture Series **7** (1976), Publish or Perish Inc.

[37] A. Stipsicz, *On the geography of Stein fillings of certain 3-manifolds*, Michigan Math. J. **51** (2003), no. 2, 327–337.

[38] W. P. Thurston, H. Winkelnkemper, *On the existence of contact forms*, Proc. Amer. Math. Soc. **52** (1975), 345–347.

[39] I. Torisu, *Convex contact structures and fibered links in 3-manifolds*, Int. Math. Res. Not. **2000:9**, 441–454.

[40] A. H. Wallace, *Modifications and cobounding manifolds*, Canad. J. Math. **12** 1960 503–528

[41] A. Weinstein, *Contact surgery and symplectic handlebodies*, Hokkiado Math. Journal **20** (1991), 241–251.

University of Pennsylvania, Philadelphia, PA 19104
E-mail address: etnyre@math.upenn.edu
URL: http://www.math.upenn.edu/~etnyre

Clay Mathematics Proceedings
Volume 5, 2006

Contact Surgery and Heegaard Floer Theory

András I. Stipsicz

Abstract. The fundamental theorem of Giroux — relating contact structures and open book decompositions — provides a way to study contact structures on closed 3–manifolds from a topological point of view. Contact surgery diagrams allow us to use some form of Kirby calculus in the study of contact 3–manifolds, and Heegaard Floer theory — through the Ozsváth–Szabó knot invariant of the binding of a compatible open book decomposition — gives a very sensitive contact invariant, which seems to be crucial in attacking the classification problem of tight contact structures on certain types of closed 3–manifolds. In these notes we collected the basic ideas of contact surgery and computation of contact Ozsváth–Szabó invariants. We paid special attention to explicit computations, hoping to convince the reader that the usual Heegaard Floer package, together with some simple homotopy–theoretic arguments might be used to derive exciting new results in contact topology.

1. Contact 3–manifolds

General definitions. We start our discussion by recalling basic notions of contact topology — for a more complete treatment of the topics just mentioned here, see [**7, 11**].

Let Y be a given closed, oriented, smooth 3–manifold. A 1–form α is a (positive) *contact form* if $\alpha \wedge d\alpha > 0$ (with respect to the given orientation). A 2–plane field ξ is a positive, coorientable contact structure if there is a contact 1–form $\alpha \in \Omega^1(Y)$ such that $\xi = \ker \alpha$. By fixing α (up to multiplication by smooth functions $f\colon Y \to \mathbb{R}^+$) we also fix an orientation for the 2–plane field ξ: the basis $\{v_1, v_2\} \subset \xi_p$ is positive if $\{v_1, v_2, n\}$ with normal vector n satisfying $\alpha(n) > 0$ provides an oriented basis for T_pY. In this case the contact structure is *cooriented.*

Let (X, ω) be a given compact, symplectic 4–manifold, that is, X is a smooth, compact, oriented 4–manifold with possibly non–empty boundary and ω is a closed 2–form with $\omega \wedge \omega > 0$ (with respect to the given orientation). The contact 3–manifold (Y, ξ) is *compatible* with (X, ω), or (X, ω) is a *filling* of (Y, ξ) if $\partial X = Y$ as oriented manifolds and $\omega|_\xi \neq 0$. In this case (X, ω) is also called a *weak symplectic filling* of (Y, ξ).

A symplectic filling (X, ω) is a *strong filling* of (Y, ξ) if ω is exact near $\partial X = Y$ and there is a 1-form α near ∂X with $\omega = d\alpha$, $d\alpha|_\xi \neq 0$ and $\xi = \{\alpha|_Y = 0\}$. It can

2000 *Mathematics Subject Classification.* Primary 57R17; Secondary 57M27, 53D35, 57R58.
The author would like to thank Burak Ozbagci for his help in preparing the figures.

be shown that the existence of such α is equivalent to the existence of a vector field v defined near ∂X which is transverse to the boundary and is a symplectic dilation, that is, $\mathfrak{L}_v\omega = \omega$. In particular, for a strong filling (X,ω) the symplectic structure on a collar of the boundary can be shown to have a model in a symplectic manifold (in the *symplectization* of (Y,ξ)) which depends only on the contact structure ξ; therefore strong fillings are suitable for performing symplectic surgeries. Notice that if ω is nonexact near ∂X then (X,ω) is not a strong filling. It turns out that this is the only obstruction, more precisely

LEMMA 1.1 (Eliashberg [**6**], Ohta–Ono [**24**]). *If (X,ω) is a weak filling of (Y,ξ) and ω is exact on a collar neighbourhood of ∂X then ω can be perturbed near ∂X to a symplectic form $\tilde{\omega}$ such that $(X,\tilde{\omega})$ is a strong filling of (Y,ξ).* □

Since on a rational homology 3–sphere any 2–form is exact, this implies

COROLLARY 1.2. *Suppose that Y is a rational homology 3–sphere, i.e., $b_1(Y) = 0$. If (X,ω) is a weak filling of (Y,ξ) for some contact structure ξ then ω can be perturbed to provide a strong filling $(X,\tilde{\omega})$ of (Y,ξ).* □

The compact complex manifold (X,J) with complex structure J is a *Stein filling* of (Y,ξ) if $\partial X = Y$, ξ is given as the oriented 2–plane field of complex tangencies on Y and (X,J) is a Stein domain, that is, it admits a proper function $\varphi\colon X \to [0,\infty)$ with $\partial X = \varphi^{-1}(a)$ for some regular value $a \in \mathbb{R}$ which is *plurisubharmonic*, i.e., the 2–form $\omega_\varphi = -d^{\mathbb{C}}d\varphi$ is a Kähler form with associated Kähler metric g_φ. It is not hard to see that a Stein filling is always a strong filling and a strong filling is automatically a weak filling. For more about fillings see [**6**].

EXAMPLE 1.3. It is easy to see that the 1–form $\alpha = dz + xdy$ induces a contact structure on the 3–dimensional Euclidean space $\mathbb{R}^3$. It turns out that this contact structure extends to the 3–sphere S^3. In addition, the resulting 2–plane field is isotopic to the 2–plane field of complex tangencies on S^3 when viewed as the boundary of the unit 4–ball in the complex vector space $\mathbb{C}^2$. The above structures are the *standard* contact structures on $\mathbb{R}^3$ and S^3, and we will denote them by ξ_{st}.

A knot $K \subset (Y,\xi)$ is called *Legendrian* if it is tangent to ξ, i.e., if ξ is defined by the 1–form α then $\alpha(TK) = 0$. Every knot can be smoothly isotoped to a Legendrian knot, in fact, for every knot there is a C^0–close Legendrian knot smoothly isotopic to it.

Legendrian knots in $(\mathbb{R}^3,\xi_{st})$ (and so in (S^3,ξ_{st})) can be depicted by their *front projections* to the yz–plane, since according to the equation $x = -\frac{dz}{dy}$ the slope of the tangent of the front projection determines the x–coordinate. After possibly isotoping, every Legendrian knot admits a front projection with no triple points, transverse double points and $(2,3)$-cusps instead of vertical tangencies. Conversely, any front projection having cusps instead of vertical tangencies and not admitting crossings with higher slope in front uniquely specifies a Legendrian knot. For this reason we will symbolize Legendrian knots in $(\mathbb{R}^3,\xi_{st})$ (and so in (S^3,ξ_{st})) by their front projections.

Notice that if $L \subset (Y,\xi)$ is Legendrian, it admits a canonical framing: consider the unit orthogonal of the tangent vector of L in ξ. (When choosing the particular orthogonal, we take the orientation of the 2–plane field into account.) The resulting framing is called the *contact framing* of the Legendrian knot L. If L is null-homologous in Y then it admits another framing, induced by pushing off L

along its existing Seifert surface. This latter framing is called the *Seifert framing.* When measuring the contact framing with respect to this Seifert framing we get an integer invariant of the Legendrian knot L called the *Thurston–Bennequin invariant* $\operatorname{tb}(L)$. Notice that since the Seifert framing is well–defined and independent of the chosen Seifert surface, the Thurston–Bennequin invariant depends only on the Legendrian knot L. Since knots in $\mathbb{R}^3$ and S^3 are all null–homologous, they all admit Thurston–Bennequin invariants. The computation of $\operatorname{tb}(L)$ from a front projection of L is an easy task: it is equal to

$$w(L) - \tfrac{1}{2}c(L),$$

where $w(L)$ is the *writhe* of the projection, i.e. the signed number of the double points of the projection, and $c(L)$ is the number of cusps in the projection. Since left and right cusps alternate among each other, it is easy to see that $\frac{1}{2}c(L) = c_r(L) = c_l(L)$ where $c_r(L)$ (and $c_l(L)$) stands for the right (resp. left) cusps of the projection.

EXAMPLE 1.4. Figure 1 shows the front projection of a Legendrian knot smoothly isotopic to the right–handed trefoil. The writhe of this projection is 3, and has 4 cusps, hence the Thurston–Bennequin invariant of the Legendrian knot determined by the front projection is equal to 1.

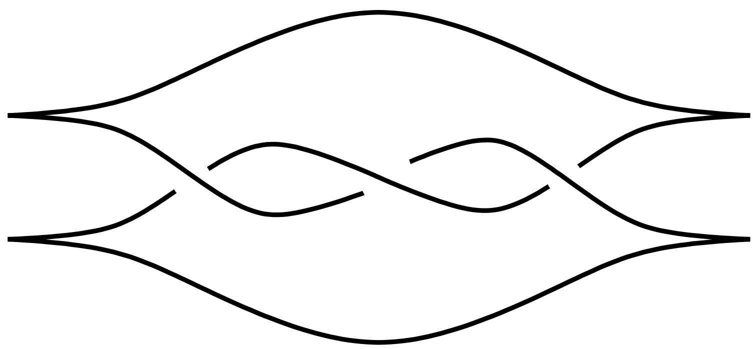


FIGURE 1. Front projection of a Legendrian trefoil knot

If $L \subset (Y, \xi)$ is null-homologous then there is another numerical invariant we can associate to it: consider a Seifert surface $\Sigma \subset (Y, \xi)$ and take the relative Euler class of ξ (as a 2–plane bundle) over Σ. For this to make sense we need to trivialize ξ over $\partial\Sigma = L$: choose the trivialization provided by the tangents of L together with their oriented normals in ξ. Note that in order to specify the tangents we need to fix an orientation on L. It is not hard to see that the resulting quantity, called the *rotation number* $\operatorname{rot}_\Sigma(L)$, will depend on the chosen Seifert surface and the orientation fixed on the knot. If $L \subset \mathbb{R}^3$ or S^3, however, the vanishing of the second homology group implies that the rotation number is independent of the chosen Seifert surface. If L is in $\mathbb{R}^3$ or in S^3, the rotation number can be computed for L given by a front projection by the formula

$$\operatorname{rot}(L) = \tfrac{1}{2}(c_d(L) - c_u(L)),$$

where $c_u(L)$ (resp. $c_d(L)$) denotes the number of up (resp. down) cusps of the projection.

Exercise 1.5. Compute the rotation number of the Legendrian trefoil given by Figure 1. Compute tb and rot for the knots given by Figures 4 and 6.

The crucial step for being able to do surgery in the contact category is to find canonical neighbourhoods. The following theorem provides such neighbourhoods for Legendrian curves. To state the theorem, consider the contact structure $\zeta_1 = \ker(\cos(2\pi\phi)dx - \sin(2\pi\phi)dy)$ on $S^1 \times \mathbb{R}^2$. (Here ϕ is the coordinate in the S^1-direction, while (x, y) are Cartesian coordinates on $\mathbb{R}^2$.)

Theorem 1.6 (Legendrian neighbourhood theorem). *If $K \subset (Y, \xi)$ is a Legendrian knot then there are neighbourhoods $U_1 \subset Y$ of K and $U_2 \subset S^1 \times D^2$ of $S^1 \times \{0\}$ such that $(U_1, \xi|_{U_1})$ and $(U_2, \zeta_1|_{U_2})$ are contactomorphic via a contactomorphism mapping K to $S^1 \times \{0\}$.* □

Overtwisted versus tight dichotomy.

Definition 1.7. A contact 3–manifold (Y, ξ) is *overtwisted* if there is an embedded 2–disk $D \subset Y$ which is tangent to ξ along its boundary. Such a disk D is called and *overtwisted disk*. If (Y, ξ) contains no overtwisted disk, we say that it is *tight*.

Theorem 1.8 (Eliashberg–Gromov). *If the contact 3–manifold (Y, ξ) is fillable then it is tight.* □

The above theorem is a major tool in proving tightness of contact structures. For a while, actually, it was unclear whether the reverse implication of the theorem is true or false, though it is now known to be false. As examples of this note show, there are many contact structures which are tight but not fillable. Regarding overtwisted contact structures we have Eliashberg's classification:

Theorem 1.9 (Eliashberg, [**4**]). *Two overtwisted contact structures on a closed 3–manifold Y are isotopic if and only if they are homotopic as oriented 2–plane fields. Moreover, for any oriented 2–plane field there is an overtwisted contact structure homotopic to it.* □

In short, the classification of overtwisted contact structures on a closed 3–manifold Y up to isotopy coincides with the classification of oriented 2–plane fields up to homotopy. This latter problem is homotopy–theoretic in nature: Fix a trivialization of TY and associate to an oriented 2–plane field ξ its unit orthogonal, providing a map $Y \to S^2$. In this way a trivialization of TY provides a bijection between homotopy types of oriented 2–plane fields and the set of homotopy classes of continuous maps from Y to S^2, $[Y, S^2]$. According to the Pontrjagin–Thom construction $[Y, S^2]$ can be identified with the framed cobordism classes of framed 1–manifolds in Y. The 1–manifold (through Poincaré duality) gives the spin^c structure induced by the oriented 2–plane field, while the framing gives an invariant in $\mathbb{Z}_d$ where d is the divisibility of the first Chern class of the 2–plane field. If $d = 0$, that is, the 2–plane field has torsion first Chern class, then this invariant (also called the *3–dimensional invariant* $d_3(\xi)$ of the 2–plane field ξ) is an element of an affine copy of $\mathbb{Z}$. An absolute $\mathbb{Q}$–lift of this invariant can be determined as follows: consider a compact almost–complex 4–manifold (X, J) such that $\partial X = Y$ and the 2–plane field of complex tangencies along ∂X is homotopic (as an oriented 2–plane field) to ξ. A homotopy–theoretic argument shows that such (X, J) always exists. Then

$$d_3(\xi) = \tfrac{1}{4}(c_1^2(X, J) - 3\sigma(X) - 2\chi(X)).$$

Notice that since $c_1(X,J)|_{\partial X} = c_1(\xi)$ is a torsion class, the square $c_1^2(X,J)$ is defined as a rational number rather than an integer. We will return to computations of 3–dimensional invariants in later sections. Tightness of a given contact structure turns out to be equivalent to a form of the adjunction inequality as given below.

THEOREM 1.10 (Eliashberg). *The inequality*

$$tb_\Sigma(L) + |\mathrm{rot}_\Sigma(L)| \leq -\chi(\Sigma)$$

is satisfied for any Legendrian knot L and surface Σ with $\partial\Sigma = L$ in (Y,ξ) if and only if the contact 3-manifold (Y,ξ) is tight. □

EXAMPLE 1.11. According to a result of Bennequin, the contact structure ξ_{st} induced by $\alpha = dz + x dy$ on $\mathbb{R}^3$ (and its extension to S^3, also denoted by ξ_{st}) is tight. On the other hand, the contact structure $\xi_1 = \ker\alpha_1$ for $\alpha_1 = \cos r dz + r\sin r d\theta$ in coordinates $(z,(r,\theta))$ on $\mathbb{R}^3$ can be easily shown to be overtwisted.

EXERCISE 1.12. Find an overtwisted disk $D \subset \mathbb{R}^3$ for the contact structure ξ_1 defined above.

The central problem of contact topology is to classify contact structures on 3–manifolds. Since overtwisted structures are classified by their homotopy type, the question reduces to understanding tight contact structures. Tight structures are much harder to find, and seem to carry important information about the geometry of the underlying 3–manifold, as is demonstrated by the successful application of contact topological arguments in the solution of several low–dimensional problems, see for example [**17, 18**], cf. also [**30**]. Great advances have been made in the recent past in classifying tight contact structures on some simple 3–manifolds, and this question is still in the focus of active research. In this note we would like to show an application of contact Ozsváth–Szabó invariants to solve the classification problem on certain classes of 3–manifolds.

2. Surgeries

Dehn surgeries. Suppose that $K \subset Y$ is a given knot in the closed 3–manifold Y. The operation of deleting a tubular neighbourhood of K and then regluing the solid torus $S^1 \times D^2$ is called *surgery* along the knot K. In order to specify the surgery uniquely, we have to determine the image of the simple closed curve $\{pt.\} \times \partial D^2$; the rest of the gluing is unique. For that matter, we need to fix a simple closed curve in the 2–torus $\partial(Y - \nu K) \cong T^2$. Since a simple closed curve in T^2 can be specified through its homology class, we only need to describe a homology class in $H_1(\partial(Y - \nu(K));\mathbb{Z}) \cong \mathbb{Z}^2$. Such a class can be represented by a pair (p,q) of relatively prime integers provided there is a basis fixed in $H_1(\partial(Y - \nu(K));\mathbb{Z})$. One basis element can be given by the boundary μ of a normal disk to K. By fixing an orientation on K, the orientation of Y equips this *meridian* μ with a canonical orientation. The other basis element, however, needs a choice. By fixing a framing of K, we get a longitude λ by pushing K off along the first basis vector of the framing. (Notice that longitudes and framings determine each other by this recipe.) Such a longitude can be chosen to be the second element of a basis in $H_1(\partial(Y - \nu(K));\mathbb{Z})$, hence the surgery can be described by the pair (p,q) satisfying that the simple closed curve $\{pt.\} \times \partial D^2$ maps to a curve homologous to $p\mu + q\lambda$. By reversing the orientation of K, both μ and λ switch sign, and therefore the ratio $\frac{p}{q}$ remains unchanged. Notice that the ratio $\frac{p}{q}$ can take its values in $\mathbb{Q} \cup \{\infty\}$.

EXAMPLE 2.1. The surgery diagram of Figure 2 provides a description of the small Seifert fibered 3–manifold $M(r_1, r_2, r_3)$.

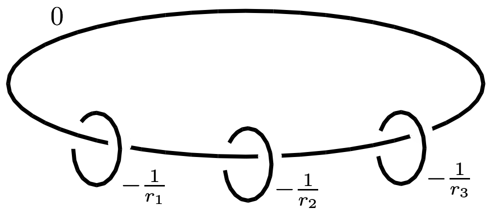


FIGURE 2. Surgery diagram for the Seifert fibered 3–manifold $M(r_1, r_2, r_3)$

EXERCISE 2.2. Verify the *slam-dunk* operation, i.e., that the two surgeries given by Figure 3 give diffeomorphic 3-manifolds. We assume that $n \in \mathbb{Z}$ and $r \in \mathbb{Q} \cup \{\infty\}$. (*Hint:* Perform surgery on K_2 first and isotop K_1 into the glued-up solid torus T. Since first we performed an integer surgery, K_1 will be isotopic to the core of T, hence when performing the second surgery we cut T out again and reglue it. Therefore it can be done by one surgery; the coefficient can be computed

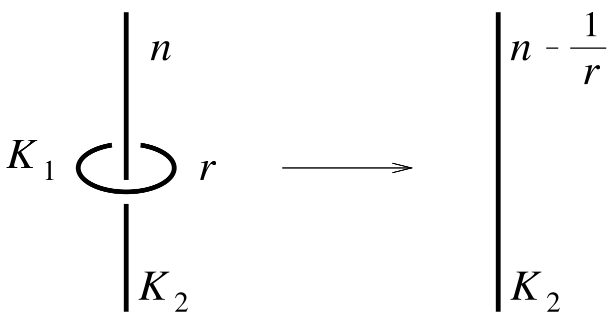


FIGURE 3. The slam-dunk operation

by first assuming $n = 0$ and then adding n extra twists. For more details see [**15**, pp 163-164].)

Contact surgery. The above surgery scheme can be extended to the contact category as follows. Suppose that $L \subset (Y, \xi)$ is a Legendrian knot in the given contact 3–manifold. Consider the contact framing on L and perform r–surgery with respect to this framing. The resulting 3–manifold is denoted by $Y_r(L)$. According to the classification of tight contact structures on solid tori [**16**], the contact structure ξ admits an extension from $Y - \nu(L)$ to $Y_r(L)$ as a tight structure on the new glued–up torus provided $r \neq 0$. (The extension might not be tight on the whole closed 3–manifold $Y_r(L)$ but it is required to be tight on the solid torus of the surgery.) Such a tight extension is not unique in general; the different extensions can be determined by the continued fraction coefficients of r. Nevertheless, the extension is unique if $r \in \mathbb{Q}$ is of the form $\frac{1}{k}$ for some integer $k \in \mathbb{Z}$. In particular, according to the above, we have

PROPOSITION 2.3. *Let* $\mathbb{L} = \mathbb{L}^+ \cup \mathbb{L}^- \subset (S^3, \xi_{st})$ *be a given Legendrian link. The result of contact* $(+1)$*–surgery along components of* $\mathbb{L}^+$ *and contact* (-1)*–surgery along components of* $\mathbb{L}^-$ *uniquely specifies a contact 3–manifold* $(Y_{\mathbb{L}}, \xi_{\mathbb{L}})$*.* □

REMARK 2.4. The reason why this surgery construction works for contact 3–manifolds is that in the neighborhood of a Legendrian knot L the contact structure can be proved to be canonical, see Theorem 1.6.

In fact, the converse of this statement also holds, namely

THEOREM 2.5 (Ding–Geiges, [**2**]). *For a given contact 3–manifold* (Y, ξ) *there exists a Legendrian link* $\mathbb{L} = \mathbb{L}^+ \cup \mathbb{L}^- \subset (S^3, \xi_{st})$ *such that* $(Y_{\mathbb{L}}, \xi_{\mathbb{L}}) = (Y, \xi)$*.*

Before turning to the proof of this theorem we recall a useful observation regarding contact surgeries.

LEMMA 2.6 ([**2**]). *Suppose that* $L \subset (Y, \xi)$ *is Legendrian and* L' *is its Legendrian push–off. If* (Y', ξ') *is given by contact* $(+1)$*–surgery on* L *and contact* (-1)*–surgery on* L' *then* Y' *is diffeomorphic to* Y *and* ξ' *is isotopic to* ξ*.* □

EXERCISE 2.7. Verify that Y' is diffeomorphic to Y.

PROOF. *of Theorem 2.5 (sketch).* Perform contact $(+1)$–surgery on $L \subset (Y, \xi)$ contained by a Darboux chart and isotopic to the Legendrian unknot with Thurston–Bennequin invariant being equal to -2; see Figure 4 for such a knot. It is not hard to see that the resulting contact structure ξ' is overtwisted, hence there is a Legendrian link in (Y, ξ') such that contact $(+1)$–surgery on it gives S^3 with some contact structure. Since contact $(+1)$–surgery can be inverted by contact (-1)–surgery, the above argument shows that a sequence of contact (-1)–surgeries on *some* contact S^3 results in (Y, ξ). In the next section we will give a surgery diagram for any contact structure on S^3. (The verification that the examples given in Section 2 comprise a complete list of contact structures on S^3 is postponed until Section 3.) □

We close this section with a statement which will be used frequently, and a more detailed explanation of the phenomenon will be discussed in the next chapter.

THEOREM 2.8 (Eliashberg, Gompf). *If* $\mathbb{L} = \mathbb{L}^-$ *then the contact 3–manifold* $(Y_{\mathbb{L}}, \xi_{\mathbb{L}})$ *is Stein fillable.* □

Examples.

Contact structures on S^3 *and* $S^1 \times S^2$. By converting the contact surgery coefficients on the diagrams of Figures 4 and 5 to smooth surgery coefficients, it is quite easy to see that these diagrams represent contact structures ξ_1 and ξ_{-1} on the 3–sphere S^3. By taking connected sums (which means simply to draw the diagrams next to each other) of n copies of Figures 4 and 5 we get sequences ξ_n and ξ_{-n} of contact structures on S^3. We define ξ_0 as the connected sum of ξ_1 and ξ_{-1}. As will be shown later, all these structures are overtwisted, and not homotopic to each other. In addition, all homotopy types of 2–plane fields are realized by one of the ξ_k ($k \in \mathbb{Z}$). Since the unique tight contact structure on S^3 can be given by the empty surgery diagram, we have the list of all contact structures on S^3 presented by surgery diagrams.

In a similar manner, we can draw diagrams providing overtwisted contact structures on $S^1 \times S^2$. (For details see [**3**].) It is a little bit more involved to find a

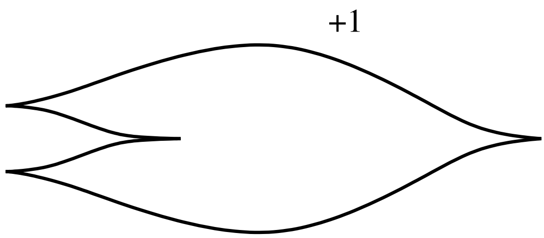


FIGURE 4. Contact structure ξ_1 on S^3

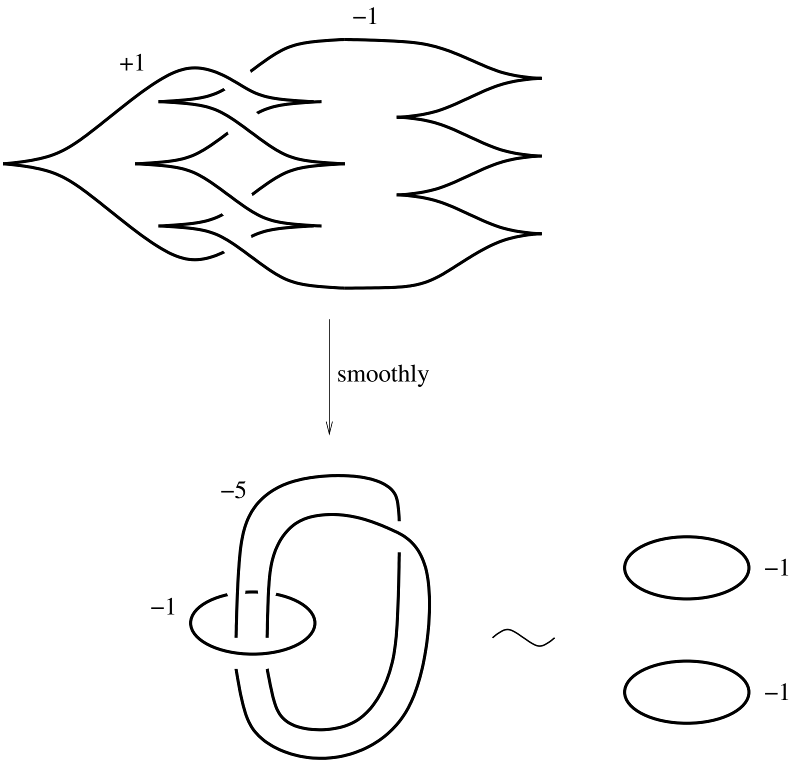


FIGURE 5. Contact structure ξ_{-1} on S^3

diagram for the unique tight contact structure on $S^1 \times S^2$. A direct argument presented in [**3**] shows that contact $(+1)$–surgery on the Legendrian unknot shown by Figure 6 is tight. We will show tightness of this structure using contact Ozsváth–Szabó invariants.

In fact, using the above ideas now it is quite simple to find a diagram for all overtwisted contact structures on a 3–manifold Y given by a smooth rational surgery diagram. Details of this algorithm are given in [**3**].

Contact structures on small Seifert fibered 3–manifolds. It is much harder to find all the *tight* structures on a given 3–manifold. In general this question is still open, but for some families of 3–manifolds we have a classification of tight structures. Here we restrict our attention to some special cases.

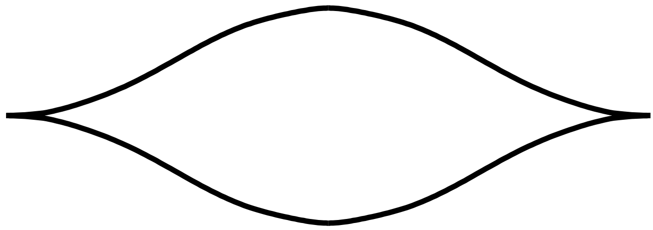


FIGURE 6. The Legendrian unknot

Let $L(p,q)$ denote the lens space we get by $(-\frac{p}{q})$–surgery along the unknot. It is not hard to see that this surgery is equivalent to a sequence of surgeries along a chain of unknots, all with integer surgery coefficients ≤ -2.

EXERCISE 2.9. Verify the above statement. (*Hint:* Apply the result of Exercise 2.2.)

Putting these unknots into Legendrian position and adding zig–zags to them we get surgery diagrams representing tight contact structures on $L(p,q)$. We can arrange that all the contact surgery coefficients are (-1), and then by appealing to Theorems 2.8 and 1.8 we conclude that the resulting structures are tight (in fact, Stein fillable). Notice that we have a freedom when putting the zig–zags on the two sides of the unknots; different choices will result in contact structures with different homotopy types. According to [**16**], these diagrams represent all tight contact structures on $L(p,q)$.

The above classification can be extended to a wider class of 3–manifolds of some *small Seifert 3–manifolds*. A Seifert fibered 3–manifold M is *small* if it fibers over S^2 with three singular fibers. Let $M = M(r_1, r_2, r_3)$ as given in Example 2.1 and define $e_0(M)$ as

$$e_0(M) = \sum_{i=1}^{3} [r_i].$$

It is not hard to see that $e_0(M)$ is an invariant of the Seifert fibered 3–manifold. (Notice that there might be many choices for r_i to present M as $M(r_1, r_2, r_3)$.) Also, if M has one or two singular fibers, then M is a lens space. For $e_0(M) \neq -1$ it is fairly easy to give surgery diagrams for tight contact structures on M, see [**12, 34**] and Exercise 2.10. In fact, for $e_0(M) \neq -2, -1$ these diagrams comprise a complete list of tight structures. In the case of $e_0(M) = -2$ there might be other structures for which we do not have the corresponding surgery diagram. If $e_0(M) = -1$, our understanding is less satisfactory. The difficulty lies in two facts: there are many small Seifert fibered spaces without any tight contact structures [**10, 21**], and even if there are tight structures, their tightness is more involved to prove since for many 3–manifolds the contact structures cannot be fillable [**20**]. We will return to these cases in Section 6.

EXERCISE 2.10. **(a)** Show that if $e_0(M) \geq 0$ then there are $r_1, r_2, r_3 \in \mathbb{Q}$ such that $M = M(r_1, r_2, r_3)$ and $r_i > 0$.
(b) Show that for any small Seifert fibered 3–manifold M with $e_0(M) = e$ there are rational numbers r_1, r_2, r_3 with $r_i \in (0,1) \cap \mathbb{Q}$ such that M can be given by the surgery diagram of Figure 7.

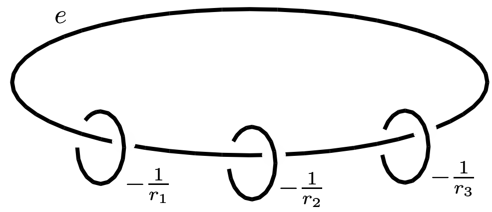


FIGURE 7. Another diagram for the small Seifert fibered 3–manifold M

(c) Verify that if $e_0(M) \leq -2$ then M can be presented by a surgery diagram of unknots along a tree with all framings ≤ -2.

3. 4–dimensional theory

Handle attachments, Kirby calculus. It is not hard to see that if the surgery coefficient of a particular knot is an integer, then the corresponding surgery can be realized by a 4–dimensional 2–handle attachment. More precisely, consider the 3–manifold Y with a knot $K \subset Y$. Fix a longitude for K and perform integer surgery along K. (Notice that the fact that a surgery coefficient is an integer does not depend on the chosen longitude, although the actual value of the surgery coefficient does.) Alternatively, consider the 4–manifold $Y \times [0,1]$ and attach a 2–handle $D^2 \times D^2$ along $\partial D^2 \times D^2$ to $Y \times \{1\}$ along K. Since $\partial D^2 \times \{0\}$ will map to K, in order to fix the gluing map we only need to specify a framing of K in Y, which (as we already remarked) is equivalent to fixing a longitude of K. Since $\partial(D^2 \times D^2) = \partial D^2 \times D^2 \cup D^2 \times \partial D^2$, after this procedure the part $\partial D^2 \times D^2$ with its image will disappear from the 3–manifold (and sinks into the 4–manifold) while $D^2 \times \partial D^2$ appears on the new boundary. Therefore the handle attachment has the same effect as surgery along K; the surgery coefficient can be determined by the chosen framing, i.e. the chosen longitude. Therefore this alternative method works if and only if the surgery coefficient is an integer. The advantage we get is that we do not only have the surgered 3–manifold, but also a 4–dimensional cobordism between the original and the resulting 3–manifolds. As we will see, such a cobordism can be very conveniently used in many specific problems. Any rational surgery diagram can be easily converted into a surgery diagram (along a possibly different link) involving only integer coefficients using the slam–dunk operation as it is given in Exercise 2.2.

Suppose that the contact 3–manifold (Y, ξ) is given as $(Y_{\mathbb{L}}, \xi_{\mathbb{L}})$ for some Legendrian link $\mathbb{L} = \mathbb{L}^+ \cup \mathbb{L}^- \subset (S^3, \xi_{st})$. Since all the surgery coefficients in such diagrams are integers, we immediately see a 4–manifold X defined by the picture: the description of integer surgeries above provides a cobordism from S^3 to Y, which can be glued to the 4–disk D^4 to get the compact 4–manifold X with $\partial X = Y$.

First we would like to find some extra structure on X. The main theorem of the subject is the following result of Eliashberg:

THEOREM 3.1 (Eliashberg, Gompf; [**5**]). *Suppose that $\mathbb{L} = \mathbb{L}^-$ in the above situation. Then the resulting 4–manifold X admits a Stein structure inducing $\xi =$*

$\xi_{\mathbb{L}}$ on $Y = \partial X$. Consequently a surgery diagram involving only contact (-1)–surgeries gives Stein fillable, hence tight contact structure. □

The idea of the proof of Eliashberg's theorem goes back to a result of Weinstein:

THEOREM 3.2 (Weinstein, [**33**]). *If (X,ω) is a weak symplectic filling of (Y,ξ) and $L \subset (Y,\xi)$ is a Legendrian knot then for the handle attachment $X' = X \cup H$ inducing contact (-1)–surgery along L the symplectic structure ω extends to a symplectic structure ω' on X'. Moreover, (X',ω') will be compatible with the resulting contact 3–manifold (Y',ξ').* □

COROLLARY 3.3. *If (Y',ξ') is given by contact (-1)–surgery on (Y,ξ) and (Y,ξ) is fillable then so is (Y',ξ').* □

REMARK 3.4. The original setup of Weinstein's handle attachment picture involves a Liouville vector field near the boundary, which exists only for strong fillings. In attaching a handle, though, this vector field is used only near the knot along which the attachment is carried out. Since the obstruction to the existence of a Liouville vector field v is cohomological, such v always exists on a neighborhood of the given knot. Therefore the general gluing scheme works for weak fillings as well.

Homotopy theory for contact surgery diagrams. The question of determining the homotopy type of a contact structure given by contact surgery diagram was first addressed by Gompf in the case when $\mathbb{L} = \mathbb{L}^-$ in [**14, 15**]. Below we sketch its extension to the general case.

First of all, note that since we use only 2–handles to build up the 4–manifold X, the result will be simply connected. Recall that spin^c structures on a simply connected 4–manifold X are in bijection with characteristic elements of $H^2(X;\mathbb{Z})$. (See the appendix of this section for more about spin^c structures.) Therefore the spin^c structure $\mathfrak{t}_\xi$ induced by the contact structure of the diagram can be specified by a characteristic cohomology class $c \in H^2(X;\mathbb{Z})$ through its corresponding spin^c structure $\mathfrak{s}$ which satisfies $\mathfrak{s}|_{\partial X} = \mathfrak{t}_\xi$. (It is easy to see that if X is simply connected then all spin^c structures from ∂X extend to X.) The definition of c is quite simple: consider the basis $\alpha_{K_1}, \ldots, \alpha_{K_n}$ of $H_2(X;\mathbb{Z})$ induced by the 2–homologies corresponding to the surgery curves $K_1, \ldots, K_n$. Then it can be shown that the class c which evaluates on α_{K_i} as $\mathrm{rot}(K_i)$ will satisfy the properties required above.

We sketch the proof (which relies on Gompf's result mentioned above) when $\mathrm{tb}(K_i) \neq 0$. For more details see [**3**]. Consider $L \subset (S^3,\xi_{st})$ and let L' denote its Legendrian push–off. By doing contact (-1)–surgery along L we get a Stein 4–manifold X_L such that $c_1(X_L)$ evaluates on the generator α_L of $H_2(X_L;\mathbb{Z})$ as $\mathrm{rot}(L)$. After performing the handle attachment corresponding to contact $(+1)$–surgery on L' we get a 4–manifold X. The complex structure on X_L will extend (for simple homotopy reasons) to a complex structure J on $X - \{P\}$, and $c = c_1(J)$ will further extend to an element $c \in H^2(X;\mathbb{Z})$. We know that c evaluates on α_L as $\mathrm{rot}(L)$; suppose that $\langle c, \alpha_{L'}\rangle = k$. Let p denote the d_3–invariant of the 2–plane field of complex tangencies along $S^3 = \partial\nu\{P\} \subset X$. Since (-1)–surgery on L and $(+1)$–surgery on L' gives (S^3,ξ_{st}) back, when computing the 3–dimensional invariant of (S^3,ξ_{st}) using X we get an equation involving k and p. By the same argument for n Legendrian push–offs of L and n Legendrian push–offs of L' this equation has the form

$$-\tfrac{1}{2} = \tfrac{1}{4}(n(k^2 - \mathrm{rot}^2(L)) - n^2\mathrm{tb}(L)(k - \mathrm{rot}(L))^2) + n(p - \tfrac{1}{2}) - \tfrac{1}{2},$$

implying that $k = \mathrm{rot}(L)$ once $\mathrm{tb}(L)$ is nonzero. The computation also shows that $p = \frac{1}{2}$, hence, provided $c_1(\xi_{\mathbb{L}})$ is torsion, the 3–dimensional invariant $d_3(\xi_{\mathbb{L}})$ can be computed as

$$d_3(\xi_{\mathbb{L}}) = \tfrac{1}{4}(c^2 - 3\sigma(X) - 2\chi(X)) + q,$$

where q denotes the number of components in $\mathbb{L}^+$.

Example 3.5. Let ξ_k $(k \in \mathbb{Z})$ denote the contact structure on S^3 defined in the previous chapter. Then, after a short calculation the above formula shows that $d_3(\xi_k) = k - \frac{1}{2}$, and hence these structures are all nonisotopic. If $k \neq 0$ then (since S^3 admits a unique tight contact structure with 3–dimensional invariant equal to $-\frac{1}{2}$) these are all overtwisted. For ξ_0 a direct argument provides overtwistedness. Similar computation applies to diagrams representing contact structures on $S^1 \times S^2$.

It is easy to see that if ξ is an oriented 2–plane field on S^3 then its 3–dimensional invariant is of the form $k - \frac{1}{2}$. This follows from the fact that for any 4–manifold X with $\partial X = S^3$ and almost–complex structure J the expression

$$d_3 + \tfrac{1}{2} = \tfrac{1}{4}(c_1^2(X,J) - 3\sigma(X) - 2\chi(X)) + \tfrac{1}{2} =$$

$$= \tfrac{1}{4}(c_1^2(X,J) - \sigma(X)) - \tfrac{\sigma(X)+\chi(X)}{2} + \tfrac{1}{2} = \tfrac{1}{4}(c_1^2(X,J) - \sigma(X)) - \tfrac{2b_2^+(X) - 2b_1(X)}{2}$$

is an integer. This is obvious since $c_1(X,J) \in H^2(X;\mathbb{Z})$ is characteristic, hence $c_1^2(X,J) \equiv \sigma(X) \pmod 8$.

In conclusion, by Eliashberg's theorem we conclude that $\{\xi_k \mid k \in \mathbb{Z}\}$ together with the standard contact structure ξ_{st} (represented by the empty diagram) comprises a complete list of contact structures (up to isotopy) on the 3–sphere S^3. Notice that this observation concludes the proof of Theorem 2.5 from the previous chapter.

Embedding theorems. Let (X,ω) be a given symplectic filling of (Y,ξ). It turned out to be very useful to study embeddings of (X,ω) into closed 4–manifolds with some additional structures. In this section we will recall the basic results and some ramifications of this theory.

Theorem 3.6 (Lisca–Matić, [**19**]). *If (X,ω) is a Stein filling of (Y,ξ) then there exists a closed minimal complex surface Z of general type such that X Kähler embeds into Z.* □

Remark 3.7. The fact that a Stein filling can be Kähler embedded into a Kähler manifold can be verified much more easily. The important point of Theorem 3.6 is that the target manifold can be chosen to be a minimal surface of general type. This property is very important in applications concerning Seiberg–Witten theory.

A nice proof (relying on the theory of Lefschetz fibrations) for the weaker statement of embedding (X,ω) into a minimal symplectic 4–manifold was found by Akbulut and Ozbagci [**1**]. A corollary of Theorem 3.6 can be used to distinguish tight contact structures:

COROLLARY 3.8. *If J_1, J_2 are two Stein structures on a fixed 4–manifold X inducing contact structures ξ_1 and ξ_2 on $Y = \partial X$ then $c_1(J_1) \neq c_1(J_2)$ implies that ξ_1 and ξ_2 are not isotopic.* □

EXERCISE 3.9. Using Corollary 3.8 find lower bounds on the number of tight contact structures on small Seifert fibered manifolds with $e_0(M) \neq -1$. (*Hint*: See [**12, 34**].)

The above embedding theorem of Lisca and Matić was extended by Eliashberg [**6**] (see also Etnyre [**8**]) to weak fillings as follows:

THEOREM 3.10. *If (X, ω) is a weak symplectic filling of (Y, ξ) then it symplectically embeds into a closed symplectic 4–manifold U.*

PROOF. Consider $\mathbb{L} = \mathbb{L}^+ \cup \mathbb{L}^- \subset (S^3, \xi_{st})$ such that $(Y_{\mathbb{L}}, \xi_{\mathbb{L}}) = (Y, \xi)$. Perform contact (-1)–surgeries along Legendrian push–offs of knots in $\mathbb{L}^+$, getting a contact 3–manifold (Y_1, ξ_1) with a weak symplectic filling (X_1, ω_1). Notice that (X, ω) symplectically embeds into (X_1, ω_1). Now consider Legendrian trefoils with tb=1 for each surgery curve linking the particular curve once (and not linking the others) and perform contact (-1)–surgeries on them. The resulting contact 3–manifold (Y_2, ξ_2) has $H_1(Y_2; \mathbb{Z}) = 0$ with a weak filling (X_2, ω_2) such that $(X, \omega) \subset (X_1, \omega_1) \subset (X_2, \omega_2)$. Since $b_1(Y_2) = 0$, the symplectic form ω_2 can be perturbed to a symplectic form $\tilde{\omega}_2$ near ∂X_2 such that $(X_2, \tilde{\omega}_2)$ becomes a strong filling of (Y_2, ξ_2). Notice that (Y_2, ξ_2) is given by a sequence of contact (-1)–surgeries, therefore it is Stein fillable with Stein filling (W, J). Apply the embedding of Theorem 3.6 to W, providing a closed complex surface Z containing W. Since both W and X_2 are strong fillings of the same contact 3–manifold, we can perform surgery along Y_2, providing a closed symplectic 4–manifold $U = (Z - W) \cup X_2$, into which (X, ω) symplectically embeds. □

This embedding theorem has far-reaching applications in low–dimensional topology: it served as the missing step in the proof of *Property P* given by Kronheimer and Mrowka [**17**], and completed a Heegaard Floer theoretic proof [**30**] of the lens space surgery theorem of Kronheimer, Mrowka, Ozsváth and Szabó [**18**]. See also [**32**].

Appendix: Spin^c structures on 3– and 4–manifolds.

DEFINITION 3.11. A *spin^c structure* on an oriented 3–manifold Y is an equivalence class of nowhere zero vector fields, where v_1 and v_2 are equivalent if they are homotopic outside a ball $B^3 \subset Y$.

Alternatively, we can consider the same equivalence relation on oriented 2–plane fields in the tangent bundle TY. Therfore a 2–plane field automatically induces a spin^c structure. Notice that for a spin^c structure $\mathfrak{t} \in Spin^c(Y)$ the class $c_1(\mathfrak{t}) \in H^2(Y; \mathbb{Z})$ is defined by taking a representative of $\mathfrak{t}$ and considering it as a complex line bundle. It is easy to see that this definition is equivalent to the conventional definition of spin^c structures: An oriented 2–plane field reduces the structure group of TY from $SO(3)$ to $U(1)$, which admits a canonical lift to $U(2) = Spin^c(3)$. Conversely, a spin^c bundle $P \to Y$, through the canonical representation of $U(2)$ on $\mathbb{C}^2$, provides a $\mathbb{C}^2$–bundle $W \to Y$ (called the bundle of spinors), and a nowhere vanishing section $\phi \in \Gamma(W)$ gives rise to a section of unit length of TY. It is not hard to verify that $H^2(Y;\mathbb{Z})$ admits a free and transitive action

$$\mathfrak{t} \mapsto a^*(\mathfrak{t})$$

($\mathfrak{t} \in Spin^c(Y)$, $a \in H^2(Y;\mathbb{Z})$) on $Spin^c(Y)$, with the property $c_1(a^*(\mathfrak{t})) = c_1(\mathfrak{t})+2a$.

The action $v \mapsto -v$ of multiplication by (-1) on vector fields induces an involution

$$\mathfrak{J} = \mathfrak{J}_Y \colon Spin^c(Y) \to Spin^c(Y).$$

A straightforward argument shows that $c_1(\mathfrak{J}\mathfrak{t}) = -c_1(\mathfrak{t})$. The fixed points of this action satisfy $c_1(\mathfrak{t}) = 0$, and since $c_1(\mathfrak{t})$ is the obstruction to reducing the structure group of the principal $Spin^c(3)$ bundle to $Spin(3) = SU(2)$, we have

Corollary 3.12. *A spin^c structure $\mathfrak{t} \in Spin^c(Y)$ is a fixed point of the involution $\mathfrak{J}$ if and only if $\mathfrak{t}$ can be given by a spin structure.* □

In a similar fashion we proceed for 4–manifolds. For a fixed 4–manifold X (with possibly nonempty boundary ∂X) let J_1, J_2 be two almost–complex structures defined on $X - \{x_1, \ldots, x_n\}$ with $\partial X \cap \{x_1, \ldots, x_n\} = \emptyset$. The two structures J_1 and J_2 are said to be homologous if there is a 1–dimensional submanifold C (with possibly nonempty boundary) containing x_i ($i = 1, \ldots, n$) such that J_1 is homotopic to J_2 on $X - C$.

Definition 3.13. An equivalence class of homologous almost–complex structures on X is called a spin^c structure on X.

The first Chern class $c_1(\mathfrak{s})$ of $\mathfrak{s} \in Spin^c(X)$ is defined as the extension of $c_1(X - \{x_1, \ldots, x_n\}, J)$ through the points where J is undefined (for an almost–complex structure J representing $\mathfrak{s}$). Since J reduces the structure group $SO(4)$ of TX to $U(2)$ outside of a finite set, and $U(2)$ admits a canonical lift to

$$Spin^c(4) = \{(A, B) \in U(2) \times U(2) \mid \det A = \det B\},$$

we can easily see that our definition above is equivalent to the traditional definition of spin^c srtructures through lifting of cocycle structures of appropriate principal bundles. For one direction we also need that a spin^c structure uniquely extends through a point, and for the converse direction we need that a section $\phi \in \Gamma(W^+)$ of the positive spinor bundle provides an isomorphism of W^- and TX away from the zero set of ϕ. Since W^- is a complex 2–plane bundle, this construction provides the necessary almost–complex structure on TX. Notice that, in fact, an oriented 2–plane bundle on $X - \{x_1, \ldots, x_n\}$ already determines an almost–complex structure by defining J as rotation on the orthogonal plane. As in the 3–dimensional case, it is also quite easy to see that $H^2(X;\mathbb{Z})$ admits a free, transitive action on $Spin^c(X)$. By considering the conjugate complex multiplication we get an involution

$$\mathfrak{J} = \mathfrak{J}_X \colon Spin^c(X) \to Spin^c(X)$$

with the property that $c_1(\mathfrak{J}\mathfrak{s}) = -c_1(\mathfrak{s})$. As in the 3–dimensional case, the spin^c structure $\mathfrak{s}$ is induced by a spin structure if and only if $\mathfrak{J}\mathfrak{s} = \mathfrak{s}$, equivalently if $c_1(\mathfrak{s}) = 0$.

If X is a manifold with boundary ∂X then the oriented 2–plane field of complex tangencies along the boundary ∂X provides a restriction map $r\colon Spin^c(X) \to Spin^c(\partial X)$.

4. Heegaard Floer theory

In this chapter we outline the basics of Heegaard Floer theory; we restrict ourselves to a short introduction, highlighting the aspects crucial for contact topological considerations. For a more detailed treatment see [**25, 26**] and the contributions [**31, 32**] in this volume.

Ozsváth–Szabó homologies of 3–manifolds. Elementary Morse theory shows that a closed, oriented 3–manifold Y admits a *Heegaard decomposition* $Y = U_1 \cup_{\Sigma_g} U_2$ into two solid genus–g handlebodies U_1 and U_2, glued together along a surface Σ_g of genus g. A solid genus–g handlebody with boundary Σ_g can be specified by g disjoint, simple closed curves $\alpha_1, \ldots, \alpha_g \subset \Sigma_g$ which are linearly independent in homology: attaching handles along α_i (together with a 3–ball) we recover the given handlebody. Therefore Y can be described by

$$(\Sigma_g, \{\alpha_i\}_{i=1}^g, \{\beta_j\}_{j=1}^g).$$

Consider the g^{th} symmetric power $Sym^g(\Sigma_g)$ and the g–dimensional tori $\mathbb{T}_\alpha = \alpha_1 \times \ldots \times \alpha_g$ and $\mathbb{T}_\beta = \beta_1 \times \ldots \times \beta_g$ in it. A symplectic structure on Σ_g gives rise to a symplectic structure on $Sym^g(\Sigma_g)$; let J be a compatible almost-complex structure. Furthermore, fix a point $z \in \Sigma_g$ distinct from all the α– and β–curves and consider the hypersurface $V_z = \{z\} \times Sym^{g-1}(\Sigma_g)$, which is disjoint from the tori $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$. For $x, y \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ let $\mathfrak{M}_{x,y}$ denote the moduli space of holomorphic maps $u\colon \Delta^2 \to Sym^g(\Sigma_g) - V_z$ from the unit disk $\Delta^2 \subset \mathbb{C}$ with the properties that $u(i) = x, u(-i) = y$ and the arc connecting i and $-i$ on $\partial\Delta^2$ is mapped into $\mathbb{T}_\alpha$ (resp. into $\mathbb{T}_\beta$) if the points on the arc have positive (resp. negative) real parts. The space $\mathfrak{M}_{x,y}$ admits an $\mathbb{R}$–action, let $\mathfrak{N}_{x,y}$ denote the result of the factorization by this action.

Consider $\widehat{CF}(Y) = \oplus_{x \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta} \mathbb{Z}_2\langle x \rangle$ and define the map $\partial\colon \widehat{CF}(Y) \to \widehat{CF}(Y)$ by the matrix element $\langle \partial x, y \rangle$ (for $x, y \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$) to be zero if $\dim \mathfrak{N}_{x,y} > 0$ and

$$\langle \partial x, y \rangle = \# \mathfrak{N}_{x,y} \qquad (mod\ 2)$$

if $\dim \mathfrak{N}_{x,y} = 0$.

Remark 4.1. For the sake of simplicity above we used $\mathbb{Z}_2$–coefficients. The theory can be set up using $\mathbb{Z}$–coefficients, in which case a coherent choice of orientations of the various moduli spaces must be made. Such a choice exists, but since we will use only mod 2 invariants, we do not deal with the details of this subtlety. We also note that when $b_1(Y) \neq 0$, not every Heegaard diagram gives rise to a well-defined theory, since for $x, y \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ there might be infinitely many homotopy types of disks connecting them, each possibly containing holomorphic representatives. By choosing *admissible* Heegaard diagrams, this case can be ruled out, and so for such decompositions the boundary operator ∂ is defined and satisfies $\partial \circ \partial = 0$. For definition and details on admissibility see [**25**].

Standard theory of Floer homologies shows that $\partial \circ \partial = 0$, hence $(\widehat{CF}(Y), \partial)$ is a chain complex. We define $\widehat{HF}(Y)$ as the homology of this chain complex.

Theorem 4.2 (Ozsváth–Szabó, [**25**]). *The Abelian group $\widehat{HF}(Y)$ is an invariant of the 3–manifold Y and is independent of the choices made throughout its definition.* □

It can be shown directly that by fixing the base point z, any intersection point $x \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ determines a spin^c structure $\mathfrak{t}_x$, and ∂x can have components only with the same induced spin^c structure. Consequently the chain complex $(\widehat{CF}(Y), \partial)$ naturally splits as a direct sum $\oplus_{\mathfrak{t}\in Spin^c(Y)}(\widehat{CF}(Y,\mathfrak{t}), \partial)$, defining a splitting as

$$\widehat{HF}(Y) = \oplus_{\mathfrak{t}\in Spin^c(Y)} \widehat{HF}(Y,\mathfrak{t}).$$

As above, it has been proved [**25**] that the group $\widehat{HF}(Y,\mathfrak{t})$ is an invariant of the spin^c 3–manifold $(Y,\mathfrak{t})$.

For $x, y \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ inducing the same spin^c structure $\mathfrak{t}$ consider the formal dimension of the moduli space $\mathfrak{M}_{x,y}$, and take this number as the difference of the gradings of x and y, denoted by $gr(x,y)$. Of course, this number might depend on the chosen component of the moduli space $\mathfrak{M}_{x,y}$, i.e., on the homotopy type of the disk connecting x and y. As application of the appropriate index theorem shows, $gr(x,y)$ is well-defined only modulo $d(\mathfrak{t})$, where $d(\mathfrak{t})$ is the divisibility of the first Chern class $c_1(\mathfrak{t}) \in H^2(Y;\mathbb{Z})$ of the spin^c structure $\mathfrak{t}$. Consequently, if $c_1(\mathfrak{t})$ is a torsion element, the above procedure provides a relative $\mathbb{Z}$–grading on $\widehat{CF}(Y)$, which descends to a relative $\mathbb{Z}$–grading on $\widehat{HF}(Y)$. By fixing the convention that the group $\widehat{HF}(S^3) = \mathbb{Z}_2$ is in degree 0, there is a lift of the above relative $\mathbb{Z}$–grading to an absolute $\mathbb{Q}$–grading (provided that $c_1(\mathfrak{t})$ is torsion).

LEMMA 4.3. *With the grading as given above, the Ozsváth–Szabó homology group $\widehat{HF}(Y,\mathfrak{t})$ with $c_1(\mathfrak{t})$ torsion splits as*

$$\widehat{HF}(Y,\mathfrak{t}) = \oplus_{d\in\mathbb{Q}} \widehat{HF}_d(Y,\mathfrak{t}).$$

The degree $d \in \mathbb{Q}$ is determined mod 1 by the spin^c structure $\mathfrak{t}$. Moreover, $\widehat{HF}_d(Y,\mathfrak{t})$ is isomorphic to $\widehat{HF}_d(Y,\mathfrak{J}\mathfrak{t})$ and to $\widehat{HF}_{-d}(-Y,\mathfrak{t})$. □

Suppose now that W is an oriented cobordism between the 3–manifolds Y_1 and Y_2. It is easy to see that W can be given as a sequence of 1–, 2– and 3–handle attachments. Since 1– and 3–handles can be attached essentially uniquely, the maps on the Ozsváth–Szabó homologies induced by those cobordisms follow a straightforward convention. If W is given by 2–handle attachments only, then W can be described by a Heegaard triple

$$(\Sigma_g, \{\alpha_i\}_{i=1}^g, \{\beta_j\}_{j=1}^g, \{\gamma_k\}_{k=1}^g),$$

and the map induced by such triple follows the same line of ideas as the definition of the homology groups, except now we count holomorphic triangles instead of holomorphic disks, see [**26**]. The case of general cobordism W now relies on the Morse theoretic argument of decomposing W into subcobordisms of the above types and composing the associated maps. In analogy with the definition of the 3–dimensional invariants, one can prove that the resulting map is independent of the choices involved; for more details see [**26**]. As in the 3–dimensional case, the maps split according to spin^c structures on the cobordisms. In the following F_W denotes the sum of the induced maps for all spin^c structures.

In addition, a spin^c cobordism $(W,\mathfrak{s})$ from $(Y_1,\mathfrak{t}_1)$ to $(Y_2,\mathfrak{t}_2)$, with $\mathfrak{t}_1,\mathfrak{t}_2$ torsion spin^c structures shifts the absolute $\mathbb{Q}$–grading by the rational number

$$\tfrac{1}{4}(c_1^2(\mathfrak{s}) - 3\sigma(W) - 2\chi(W)).$$

Basic properties. The most fundamental properties of these homology groups can be summarized as follows.

Theorem 4.4 (Adjunction inequality for homologies, [**26**]). *Suppose that $\Sigma \subset Y$ is a closed oriented surface in a closed, oriented 3–manifold Y and $\mathfrak{t} \in Spin^c(Y)$ is a given spin^c structure. The nontriviality of the Ozsváth–Szabó homology group $\widehat{HF}(Y, \mathfrak{t})$ implies that either $\Sigma = S^2$ and $\langle c_1(\mathfrak{t}), [\Sigma]\rangle = 0$, or $g(\Sigma) > 0$ and*

$$|\langle c_1(\mathfrak{t}), [\Sigma]\rangle| \leq 2g(\Sigma) - 2.$$

□

Theorem 4.5 (Adjunction inequality for maps, [**26**]). *If W is a 4-dimensional cobordism and $\Sigma \subset W$ is a closed oriented surface with positive genus in it then for $\mathfrak{s} \in Spin^c(W)$ the fact that $F_W \neq 0$ implies that $|\langle c_1(\mathfrak{s}), [\Sigma]\rangle| + [\Sigma]^2 \leq 2g(\Sigma) - 2$. In particular, if W contains a surface Σ with $[\Sigma]^2 > 2g(\Sigma) - 2 \geq 0$ then $F_{W,\mathfrak{s}} = 0$ for all spin^c structures $\mathfrak{s} \in Spin^c(W)$.* □

Finally, suppose that a 3–manifold Y and a knot $K \subset Y$ are given. Perform integer surgery along K, resulting in a 3–manifold Y_K and a cobordism X_1 from Y to Y_K. Consider a normal circle N to K and attach a 2–handle to Y_K along N with framing (-1). The resulting 3–manifold will be denoted by Y', while the cobordism is X_2. Repeat this last step, i.e., attach a 2–handle to Y' along a normal circle U of N with framing (-1). It is not hard to see that the resulting 3–manifold is diffeomorphic to Y; denote the last cobordism by X_3. The diagram below describes the situation.

$$\begin{tikzcd} Y \arrow[rr, "X_1"] & & Y_K \arrow[dl, "X_2"] \\ & Y' \arrow[ul, "X_3"] & \end{tikzcd}$$

This geometric situation induces a triangle on Ozsváth–Szabó homologies as depicted below.

$$\begin{tikzcd} \widehat{HF}(Y) \arrow[rr, "F_{X_1}"] & & \widehat{HF}(Y_K) \arrow[dl, "F_{X_2}"] \\ & \widehat{HF}(Y') \arrow[ul, "F_{X_3}"] & \end{tikzcd}$$

The central result for computing Ozsváth–Szabó homologies is the following

Theorem 4.6 (Surgery exact triangle, [**26**]). *The triangle defined above for Ozsváth–Szabó homologies is exact.* □

For an elegant proof of the exactness of the surgery triangle see [**32**] in this volume.

5. Contact invariants

The most spectacular success of Ozsváth–Szabó homologies stems from its applications to knot theory and to contact topology. In the following we will discuss the definition and basic properties of the contact invariant defined in [**28**]. Applications will be given in the next chapter.

Open book decompositions and Giroux's theorem. The definition of the contact invariant $c(Y,\xi) \in \widehat{HF}(-Y, \mathfrak{t}_\xi)$ rests on a seminal result of Giroux, providing a close connection between open book decompositions and contact structures on a given 3–manifold Y. Here we restrict ourselves to an outline of this beautiful theory; for a more complete treatment the reader is advised to turn to [**9**] in this volume.

Suppose that $L \subset Y$ is a fibered link in Y, that is, the complement $Y - L$ fibers as $f\colon Y - L \to S^1$ over the circle S^1, and the fibers of f provide Seifert surfaces for L. In this case the pair (L, f) is an *open book decomposition* of Y. The fibers of f are the *pages*, while L is the binding of the open book decomposition. The monodromy of the fibration $f\colon Y - L \to S^1$ is called the *monodromy* of the open book decomposition (L, f). A contact structure ξ on Y is said to be *compatible* with an open book decomposition (L, f) on Y if L is transverse with respect to ξ and there is a contact 1–form α defining ξ such that the 2–form $d\alpha$ is a volume form on each page. In addition, we assume that the orientation of the binding as a transverse knot coincides with its orientation as the boundary of a page.

According to a classical theorem of Thurston and Winkelnkemper, for any open book decomposition there exists a contact structure compatible with it: by slightly perturbing the tangents of the pages and extending this plane field through the binding we get the desired contact structure. Giroux proved that the converse of this statement is also true, namely for any contact structure there is an open book decomposition compatible with it. The existence of this open book can, in fact, be deduced from a surgery diagram representing the given contact structure. In addition, a simple argument shows that if two contact structures are compatible with the same open book decomposition then they are isotopic. The converse of this correspondence is more subtle and requires a definition.

Definition 5.1. Suppose that an open book decomposition is given on Y with page F and monodromy φ. Let F' denote the surface we get by adding a 1–handle to F. The open book decomposition with page F' and monodromy $\varphi \circ t_a$ is called a *positive stabilization* of (F, φ) if t_a is a right-handed Dehn twist along the simple closed curve $a \subset F'$ intersecting the cocore of the new 1–handle in a unique point.

With this definition in place, we can formulate the central result clarifying the relation between open book decompositions and compatible contact structures.

Theorem 5.2 (Giroux, [**13**]). **(a)** *For a given open book decomposition of Y there is a compatible contact structure ξ on Y. Contact structures compatible with a fixed open book decomposition are isotopic.*
(b) *For a contact structure ξ on Y there is a compatible open book decomposition of Y. Two open book decompositions compatible with a fixed contact structure admit a common positive stabilization.* □

Contact Ozsváth–Szabó invariants. The above result of Giroux shows that any invariant of the open book decomposition which is invariant under positive stabilization is a contact invariant. For simplicity let us assume that the binding of the given open book decomposition is connected. (This can always be achieved by sufficiently many positive stabilizations.) Perform 0–surgery along the binding, resulting in a fibered 3–manifold Y_B. The definition of the contact Ozsváth–Szabó invariant relies on the following lemma.

LEMMA 5.3. *Suppose that Y is a closed 3–manifold which fibers over S^1. If $\mathfrak{t}_{can}$ denotes the spin^c structure given by the tangents of the fibers of $Y \to S^1$, then $\widehat{HF}(Y, \mathfrak{t}_{can}) = \mathbb{Z}_2 \oplus \mathbb{Z}_2$ with a distinguished generator g. Moreover, $\widehat{HF}(Y, \mathfrak{t}) = 0$ for all spin^c structures distinct from $\mathfrak{t}_{can}$ satisfying $\langle c_1(\mathfrak{t}), F\rangle = \langle c_1(\mathfrak{t}_{can}), F\rangle$, where F denotes the homology class of the fiber of $Y \to S^1$.*

In specifying the distinguished element of $\widehat{HF}(Y, \mathfrak{t}_{can})$ one has to use another version of Ozsváth–Szabó homology groups (namely the group $HF^+(Y, \mathfrak{t}_{can})$) and the map connecting the two theories — we will not go into the details of the definition of g.

PROOF. *(sketch).* The argument rests on a simple application of the surgery exact triangle together with a sample computation. Let $\mathfrak{T} = \{\mathfrak{t} \in Spin^c(Y) \mid \langle c_1(\mathfrak{t}), F\rangle = \langle c_1(\mathfrak{t}_{can}), F\rangle\}$. In the first step we show that the group $\oplus_{\mathfrak{t}\in\mathfrak{T}}\widehat{HF}(Y, \mathfrak{t})$ does not depend on the monodromy of the fibration, just on the genus of the fiber. To this end, suppose that $Y_1, Y_2 \to S^1$ are two given genus–g fibrations. There is a cobordism between them which is a Lefschetz fibration over the annulus; by induction we may assume that the fibration has a unique singular fiber, that is, Y_2 is given as (-1)–surgery (with respect to the fiber framing) along the nonseparating vanishing cycle of the singular fiber. Writing down the exact triangle, it is easy to see that the third group is zero, since 0–surgery (with respect to the fiber framing) reduces the genus of the fiber, hence the adjunction inequality implies the result. The sample computation can be carried out for the result of the 0–surgery along the $(2, 2g + 1)$ torus knot. (A related exact triangle reduces this computation to simple algebra.) □

Recall that $g \in \widehat{HF}(Y_B, \mathfrak{t}_{can}) \cong \widehat{HF}(-Y_B, \mathfrak{t}_{can})$ denotes the distinguished generator of this homology group. When turning the cobordism of the above 0–surgery upside down, we get a cobordism W from $-Y_B$ to $-Y$. The contact invariant $c(Y, \xi)$ of a contact 3–manifold is defined by $F_W(g) \in \widehat{HF}(-Y)$. The proof of the fact that the resulting element is an invariant of the contact structure (and not only the open book decomposition) proceeds in two steps. First we show the following:

PROPOSITION 5.4 (Ozsváth–Szabó, [**28**] Lemma 4.4). *Suppose that an open book decomposition on Y with monodromy φ is given, and denote the element corresponding to this decomposition via the above recipe by $c(\varphi) \in \widehat{HF}(-Y)$. Then for every positive integer $h \in \mathbb{N}$ there is an element $c(\varphi, h) \in \widehat{HF}(-Y)$ such that $c(\varphi, h) = c(\varphi')$ for any monodromy φ' we get by applying $2h$ positive stabilizations to the given open book decomposition.* □

The proof of the proposition follows the same line of argument outlined in the proof of Lemma 5.3 above. Using induction, it would be sufficient to show that $c(\varphi) = c(\varphi, 1)$. This identity is, however, hard to deal with directly, since the two open book decompositions correspond to different genera, hence the fibrations we get by the 0–surgeries are not connected by any natural cobordism. Therefore one has to use more sophisticated tools in proving that the invariant $c(\varphi)$ is an invariant of the compatible contact structure, rather than only the open book decomposition.

This second step of the proof of invariance relies on the knot invariants introduced in [**29**]. Ozsváth and Szabó noticed that the knot group $\widehat{HFK}(Y, L, \mathfrak{t})$ of the binding (with an appropriate relative spin^c structure determined by the genus of the page) is cyclic, and the image of its generator in $\widehat{HF}(Y)$ can be easly seen to be invariant under simple positive stabilizations of the open book decomposition. Finally, an explicit computation with appropriate Heegaard diagrams show that this knot invariant is equal to the element $c(\varphi)$ defined above. This last step verifies that our definition provides a contact invariant.

Basic properties. Recall that the Ozsváth–Szabó homology groups split as a direct sum $\widehat{HF}(Y) = \oplus_{(\mathfrak{t},d)\in\mathfrak{P}}\widehat{HF}_d(Y,\mathfrak{t})$ where $\mathfrak{P}$ is the set of homotopy types of oriented 2–plane fields on Y and the pair $(\mathfrak{t}, d)$ stands for the spin^c structure and the 3–dimensional invariant determined by a given oriented 2–plane field. The first property of the contact invariant is that it is an element of the summand corresponding to the 2–plane field of the contact structure:

LEMMA 5.5. *For a contact 3–manifold (Y,ξ) the contact Ozsváth–Szabó invariant $c(Y,\xi)$ is an element of $\widehat{HF}_{-d_\xi}(-Y,\mathfrak{t}_\xi)$.* □

The next property provides a way for computing the invariant for contact structures given by contact surgery diagrams.

THEOREM 5.6. *Suppose that (Y_2,ξ_2) is given as contact $(+1)$–surgery along the Legendrian knot $L\subset(Y_1,\xi_1)$; the corresponding cobordism is denoted by X. Then*

$$F_{-X}(c(Y_1,\xi_1)) = c(Y_2,\xi_2).$$

PROOF. Fix an open book decomposition of Y which contains L in a page such that the contact framing and the surface framing on L coincide. A simple modification of the argument given in the proof of Proposition 5.4 now provides the argument. □

Using the definition, it can be shown that

LEMMA 5.7. *The contact invariant $c(S^3,\xi_{st})\in\widehat{HF}(S^3)=\mathbb{Z}_2$ of the standard contact 3–sphere is nonzero.*

PROOF. Consider the standard open book decomposition on S^3 with the trivial knot as binding and apply the definition. □

From the transformation rule of Theorem 5.6 the following important vanishing and nonvanishing results can be easily deduced.

PROPOSITION 5.8. *If (Y,ξ) is overtwisted then $c(Y,\xi)=0$.*

PROOF. It is not hard to see from the classification of overtwisted contact structures that there is a contact structure ξ' on Y such that (Y,ξ) is given as contact $(+1)$–surgery along the Legendrian knot of Figure 4, located in a Darboux chart of (Y,ξ'). Since the cobordism corresponding to this surgery contains a sphere of square (-1), after reversing orientation it is clear that $c(Y,\xi)=F_{-X}(c(Y,\xi'))=0$ independent of the value of $c(Y,\xi')$. □

PROPOSITION 5.9. *If (Y,ξ) is Stein fillable then $c(Y,\xi)\neq 0$.*

PROOF. Using Eliashberg's theorem it can be shown that any Stein fillable contact 3–manifold can be given by a sequence of contact (-1)–surgeries on one of the contact 3–manifolds η_k, where η_k is a contact structure on $\#_kS^1\times S^2$ $(k\geq 0)$ given by doing contact $(+1)$–surgery along the k–component Legendrian unlink. It can be shown that $c(\#_k S^1 \times S^2, \eta_k) \neq 0$ (see Exercise 5.10), and since it is given by a sequence of contact $(+1)$–surgeries on (Y, ξ), we get that for some cobordism W the equation $F_W(c(Y, \xi)) = c(\#_k S^1 \times S^2, \eta_k)$ holds, implying that $c(Y, \xi) \neq 0$. $\square$

EXERCISE 5.10. Using induction on k verify that the contact invariant $c(\#_k S^1 \times S^2, \eta_k)$ is nonzero.

6. Applications

After having discussed the basic properties of the contact Ozsváth–Szabó invariants, we present some results relying on these notions.

Surgery along knots in S^3. First we examine the problem of the existence of tight contact structures on 3–manifolds of the form $Y = S^3_r(K)$, i.e., Y can be given by a single Dehn surgery on S^3. Let us recall that the *maximal Thurston–Bennequin number* $TB(K)$ of a knot $K \subset S^3$ is defined by

$$\max\{tb(L) \mid L \text{ is smoothly isotopic to } K \text{ and Legendrian in } (S^3, \xi_{st})\}.$$

The *slice–genus* (or 4–ball genus) $g_s(K)$ of $K \subset S^3$ is by definition

$$\max\{g(F) \mid F \subset D^4, \partial F = K \subset S^3\}.$$

Using gauge theory it has been proved that $TB(K) \leq 2g_s(K) - 1$.

THEOREM 6.1. *If $TB(K) = 2g_s(K) - 1 > 0$ is satisfied for a knot K then $S^3_r(K)$ admits a positive tight contact structure for any $r \neq TB(K)$.*

Notice that if K is the (p, q) torus knot $T_{(p,q)}$, then for $p, q \geq 2$ and relative prime it has $TB(T_{(p,q)}) = pq - p - q$, which is equal to $2g_s(T_{(p,q)}) - 1$, hence those knots satisfy the assumptions of the theorem. For example, the right-handed trefoil knot T depicted in Figure 8 satisfies the assumptions. (For a more detailed and simpler proof of Theorem 6.1 for $T =$ the trefoil knot, see the contribution [**22**] in this volume.) In fact, any nontrivial algebraic knot does the same, and there are many other knots with this property. For example, if for the knots K_1, K_2 we have $TB(K_i) = 2g_s(K_i) - 1$ then the same equation holds for their connected sum $K_1 \# K_2$.

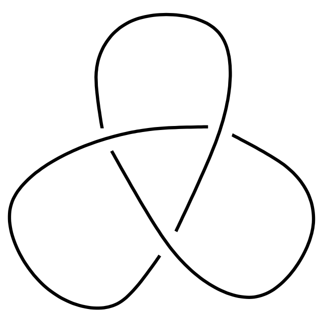


FIGURE 8. The right-handed trefoil knot

PROOF. Let L be a Legendrian knot smoothly isotopic to K with $\text{tb}(L) = TB(K)$. Let $k \geq 0$ be a fixed integer and consider $L_1, \ldots, L_k$, L_{k+1} Legendrian push–offs of L, while $C_1, \ldots, C_t$ is a chain of Legendrian unknots linked to L_{k+1}. Fix a rational number $r \neq TB(K)$.

LEMMA 6.2. *For any rational number $r \neq TB(K)$ there is an integer $k \geq 0$ and suitable stabilizations of $L_{k+1}, C_1, \ldots, C_t$, such that a sequence of contact $(+1)$–surgeries on $L_1, \ldots, L_k$ and contact (-1)–surgeries on the stabilized $L_{k+1}, C_1, \ldots, C_t$ yields a contact 3–manifold (Y, ξ) such that Y is diffeomorphic to $S^3_r(K)$.*

PROOF. *(sketch).* Let $r' = r - TB(K) \neq 0$. If $r' < 0$ then take $k = 0$ and the stabilizations are directed by the continued fraction coefficients of r'. In case $r' = \frac{p'}{q'} > 0$, take $k > 0$ satisfying $q' - kp' < 0$ and repeat the above recipe for $\frac{p'}{q'-kp'}$. For a more detailed description of the algorithm see [**3, 20**]. □

Having the above lemma at hand we have candidate contact structures on $S^3_r(K)$; in the following we will use contact Ozsváth–Szabó invariants to prove their tightness. Let (Y_k, ξ_k) denote the contact 3–manifold we get by performing only the $(+1)$–surgeries, $k \geq 0$; in particular, $(Y_0, \xi_0) = (S^3, \xi_{st})$. Since contact (-1)–surgery on a contact 3–manifold with nonvanishing contact Ozsváth–Szabó invariants is tight, $c(Y_k, \xi_k) \neq 0$ will imply that all the contact structures defined by Lemma 6.2 are tight, concluding the proof of Theorem 6.1. The fact $c(Y_k, \xi_k) \neq 0$ can be proved by induction — this is the step where the assumption $TB(K) = 2g_s(K) - 1$ is used. To start the induction, we notice that for $k = 0$ the statement clearly holds. Since (Y_{k+1}, ξ_{k+1}) is given by contact $(+1)$–surgery along a Legendrian knot in (Y_k, ξ_k), for the corresponding cobordism X_k we have

$$F_{-X_k}(c(Y_k, \xi_k)) = c(Y_{k+1}, \xi_{k+1}).$$

Therefore induction together with the injectivity of F_{-X_k} implies the result. This latter claim of injectivity can be proved by applying the surgery exact triangle for the cobordism $-X_k$.

For the sake of simplicity we sketch the proof of this last step for the case when K is the right-handed trefoil knot depicted by Figure 8 (the general case follows a very similar pattern). In this case the surgery triangle induced by the handle attachment along the $(k+1)$st Legendrian trefoil is given by Figure 9. If V_k denotes the cobordism from Y to $-Y_k$, then a simple geometric arguments shows

PROPOSITION 6.3. *The 4–manifold V_k contains a torus of self–intersection $k + 1$.* □

Now the proof of Theorem 6.1 can be easily completed for the special case when $K = T$: By the adjunction formula for cobordisms, Proposition 6.3 implies that the induced homomorphism F_{V_k} is trivial, hence the injectivity of F_{-X_k} follows from the exactness of the triangle. This concludes the proof of Theorem 6.1. □

EXERCISE 6.4. Find a genus–$g_s(L)$ surface $\Sigma \subset V_0$ with $[\Sigma]^2 = \text{tb}(L)$ for general L satisfying $\text{tb}(L) = 2g_s(L) - 1$. Let t denote $\text{tb}(L)$. For $k > 0$ find a surface of genus $\frac{1}{2}(t(t-1)k + t + 1)$ with self–intersection $t^2k + t$ in V_k. (For the solution see [**20**].)

Notice that the above result deals only with surgeries satisfying $r \neq TB(K)$. This assumption plays an important role in defining the candidate tight contact

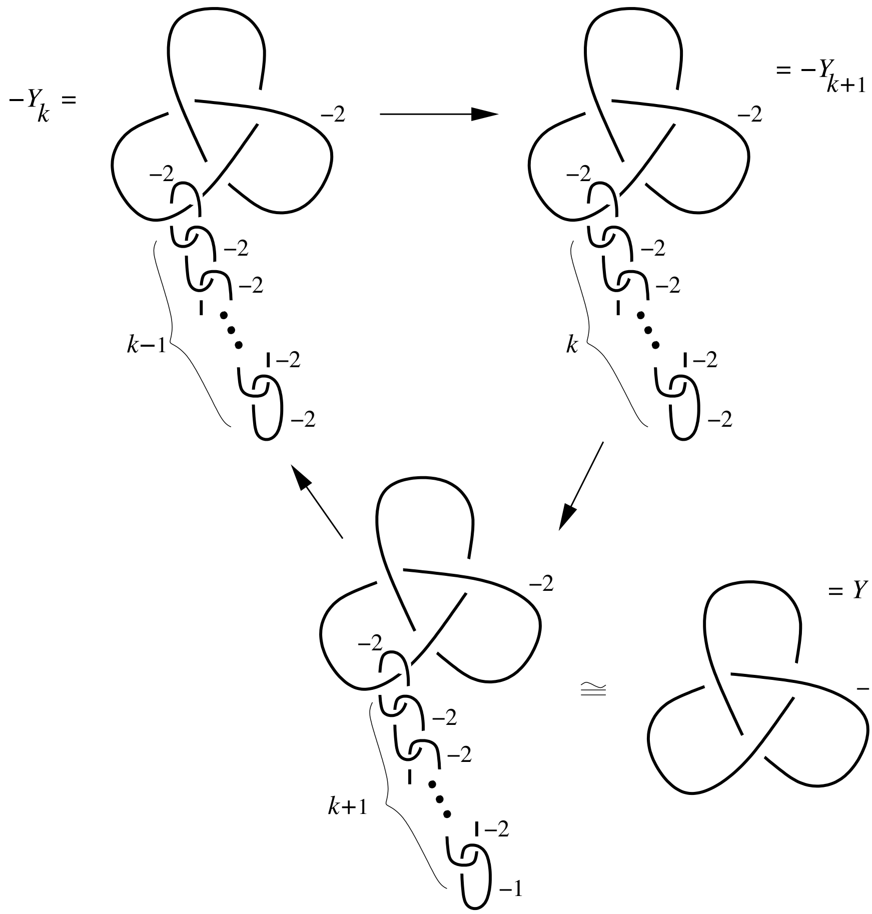


FIGURE 9. The exact triangle induced by $-X_k$

structure; for $r = TB(K)$ the previous strategy would provide an overtwisted structure. It seems to be a more subtle question to understand what happens on the 3–manifold $S^3_{TB(K)}(K)$. Below we give a partial answer to this question, concentrating on some particular families of torus knots.

A sample computation. Using convex surface theory it can be proved that the oriented 3–manifolds $S^3_{2n-1}(T_{(2,2n+1)})$ admit no positive tight contact structures once $n \geq 1$ [**21**]. It is natural to ask what happens with other knots for which Theorem 6.1 applies, when we perform the *critical* surgery with coefficient $TB(K)$. Below we show that in one sample case the corresponding 3–manifold admits a positive tight contact structure. This computation generalizes to a wider family of knots, see [**21**].

PROPOSITION 6.5. *The 3–manifold* $Y = S^3_5(T_{(3,4)})$ *admits a tight contact structure.*

The proof will obviously follow from

THEOREM 6.6. *The contact structure ξ defined by the contact surgery diagram of Figure 10 on Y is tight.*

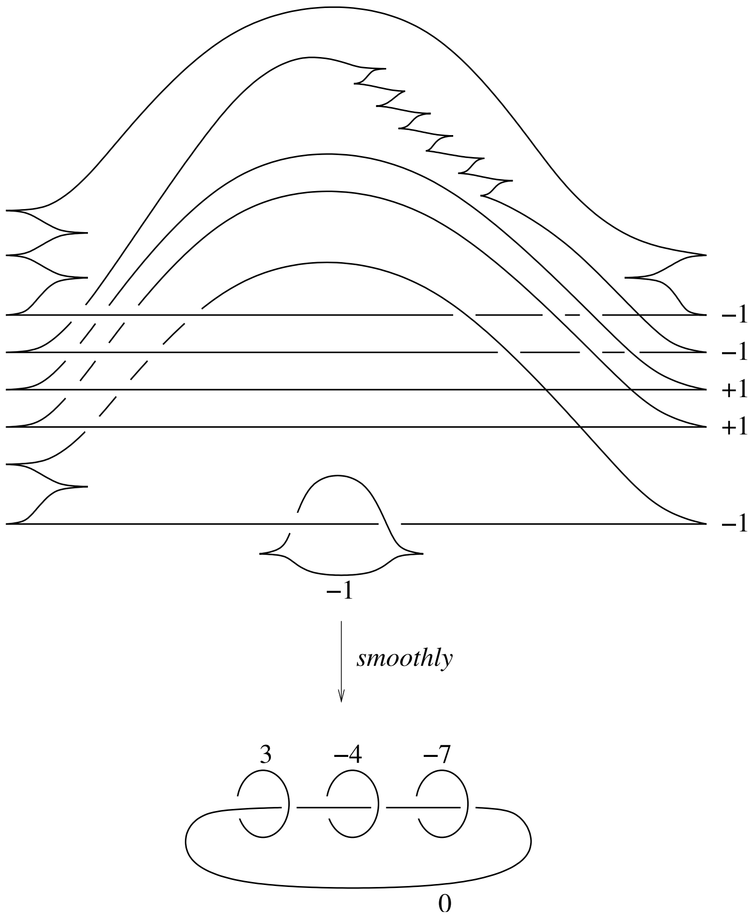


FIGURE 10. Tight contact structure on $Y = M(-\frac{1}{2}, \frac{1}{4}, \frac{1}{7}) = S^3_5(T_{(3,4)})$

PROOF. Consider the contact structure ξ' given by the diagram of Figure 10 after deleting one of the $(+1)$–framed surgery curves. The underlying 3–manifold will be denoted by S. It is easy to see that $c(S, \xi') \neq 0$, since, according to the diagram, ξ' is given by a sequence of (-1)–surgeries on the tight $S^1 \times S^2$. Let X denote the cobordism defined by the second contact $(+1)$–surgery. Our aim is to show that $c(S, \xi')$ is not in $\ker F_{-X}$, i.e.,

$$c(Y, \xi) = F_{-X}(c(S, \xi')) \neq 0.$$

In order to analyze the map F_{-X}, consider the exact triangle defined by the cobordism $-X$. Here we will follow the convention of denoting the 3–manifolds by solid surgery curves, while the cobordisms are denoted by dashed curves.

It is not hard to verify that the 3–manifolds S, Y and L are all L–spaces, that is, $\dim \widehat{HF}(S) = |H_1(S; \mathbb{Z})| = 89$, $\dim \widehat{HF}(Y) = 5$ and $\dim \widehat{HF}(L; \mathbb{Z}) = 84$. In

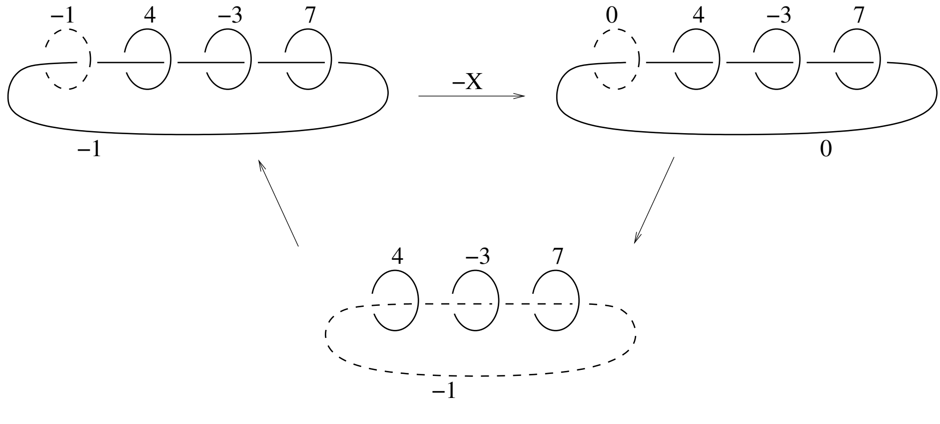


FIGURE 11. The exact triangle induced by $-X$

particular, for a given spin^c structure $\mathfrak{t}$ (on any of the above three 3–manifolds) the corresponding Ozsváth–Szabó homology group (with $\mathbb{Z}_2$–coefficients) admits a unique nontrivial element $a_{\mathfrak{t}}$, which we will denote by the spin^c structure itself. Let U denote the cobordism given by attaching a 2–handle along the dashed curve with framing 0 in Figure 11 and V the third cobordism of the same figure.

By exactness we get that $F_U = 0$, therefore F_V is injective and F_{-X} is surjective. The 3–manifold L is the connected sum of three lens spaces, more precisely $L = L(7,6)\#L(4,3)\#L(3,1)$. It admits two spin structures $\mathfrak{t}_1$ and $\mathfrak{t}_2$ and a simple geometric argument shows that exactly one of them (say $\mathfrak{t}_1$) extends to V as a spin structure and $\mathfrak{t}_2$ extends to U as a spin structure. The crucial step in the argument is the following observation:

PROPOSITION 6.7. *If $\mathfrak{t}_e$ denotes the unique spin structure on Y then for the gradings of the corresponding Ozsváth–Szabó homology elements we have*

$$gr(\mathfrak{t}_e) - gr(\mathfrak{t}_2) = \tfrac{1}{4}.$$

□

REMARK 6.8. There are several ways to prove this proposition. For example, we can directly compute the gradings of the two elements: For lens spaces the gradings are fairly easy to determine, for the small Seifert 3–manifold Y it is a little more complicated, but can be done using [**27**] or [**23**]. A less explicit, but possibly shorter argument considers the triangle induced by the cobordism $-U$ between L and $-Y$, and shows that the image of the spin structure $\mathfrak{t}_2$ has $\mathfrak{t}_e$ as nonzero component, from which it is a simple task to deduce the above proposition.

Since $\mathfrak{t}_2$ is self–conjugate under the $\mathbb{Z}_2$ action induced by conjugating the spin^c structures, its image $F_V(\mathfrak{t}_2)$ decomposes as $a+\mathfrak{J}a$ for some $a \in \widehat{HF}(-S)$. Injectivity of F_V implies that a and $\mathfrak{J}a$ are not in $\ker F_{-X}$. In a way similar to the proof of Proposition 6.7 it can be checked that $\langle F_{-X}(a), \mathfrak{t}_e\rangle = 1$, hence a has a homogeneous component a_1 with the same property. By determining the spin^c structure of a_1 and comparing it to $\mathfrak{t}_{\xi'}$ we will conclude that the spin^c structures coincide, hence $a_1 = c(S,\xi')$, implying $F_{-X}(c(S,\xi')) \neq 0$. From Proposition 6.7 the determination of the spin^c structure of a_1 is a simple task: the degree shift between $\mathfrak{t}_2$ and a_1 is at most $\frac{1}{4}(-\frac{84k^2}{89}+1)$ and $k^2>0$ since $\mathfrak{t}_2$ does not extend as a spin structure to this cobordism. Similarly the degree shift between a_1 and $\mathfrak{t}_e$ is $\frac{1}{4}(-\frac{5l^2}{89}+1)$ and $l^2>0$ since there is no spin structure on the cobordism $-X$. Since $gr(\mathfrak{t}_e)-gr(\mathfrak{t}_2)=\frac{1}{4}$, it follows that $k^2+l^2=2$, hence $k=\pm 1$, which specifies the spin^c structure of a_1 in terms of the spin structure $\mathfrak{t}_2$, which is easy to describe. Now a simple homotopy–theoretic computation verifies that $\mathfrak{t}_{a_1}=\mathfrak{t}_{\xi'}$ or $\mathfrak{t}_{a_1}=\mathfrak{J}\mathfrak{t}_{\xi'}$, concluding the proof. For details of the steps only sketched above the reader is advised to turn to [**21**]. $\square$

Fillability. As it turns out, many of the tight contact structures found by Theorem 6.1 are nonfillable. In fact

THEOREM 6.9. *The manifold $S^3_r(T_{(p,pn+1)})$ with $p,n\in\mathbb{N}$ and $r\in[p^2n-pn-1,p^2n+p-1)$ supports no fillable contact structure.*

PROOF. The proof proceeds roughly as follows. First, using the surgery exact triangle, the fact that $(pq-1)$–surgery on $T_{(p,q)}$ is a lens space, and the adjunction formula for cobordisms, one can show that the 3–manifolds encountered above are all L–spaces. According to [**30**] a weak symplectic filling of an L–space has vanishing b_2^+ invariant. Kirby calculus and some elementary algebra shows that the manifolds encountered above can be given as boundaries of positive definite 4–manifolds with intersection forms which do not embed into any diagonal definite lattice. By reversing the orientation of these 4–manifolds and gluing them to potential fillings we end up with a closed, negative definite 4–manifold with nonstandard intersection form, contradicting Donaldson's famous diagonalizability theorem. This shows that no filling of the above manifolds can exist. $\square$

By analyzing the freedom of putting stabilizations on the Legendrian knots C_i we get

COROLLARY 6.10. *For any $n\in\mathbb{N}$ there is a rational homology sphere M_n which carries at least n pairwise nonisomorphic tight contact structures, none of them fillable.* $\square$

We conclude this chapter by showing examples where the manifold carries both fillable and nonfillable tight contact structures. To this end, notice that by Theorem 6.1 contact $(+1)$–surgery on the Legendrian trefoil knot with Thruston–Bennequin invariant 1 and rotation number 0 provides a tight, nonfillable contact 3–manifold. Choosing a particular such knot (as is given by the Legendrian trefoil of Figure 12) and doing two contact $(+1)$–surgeries along the Legendrian unknots we get a contact structure which is still not fillable. Its tightness can be proved by viewing it as contact $(+1)$–surgery on the tight contact $S^1\times S^2\# S^1\times S^2$. The analysis of the induced map shows its injectivity as before, hence the contact Ozsváth–Szabó invariant of the contact structure of Figure 12 is nonzero, conluding the proof of tightness.

It is not hard to see that the above contact structure is defined on the circle bundle over the torus with Euler number equal to 2. Stein fillable structures on this 3–manifold were given by Gompf in [**14**]. From the above example appropriate contact (-1)–surgeries provide a family of tight, nonfillable structures on many Seifert fibered 3–manifolds over T^2. A small modification of the argument (by starting with the connected sum of n copies of the trefoil) extends to Seifert fibered 3–manifolds over higher genus surfaces.

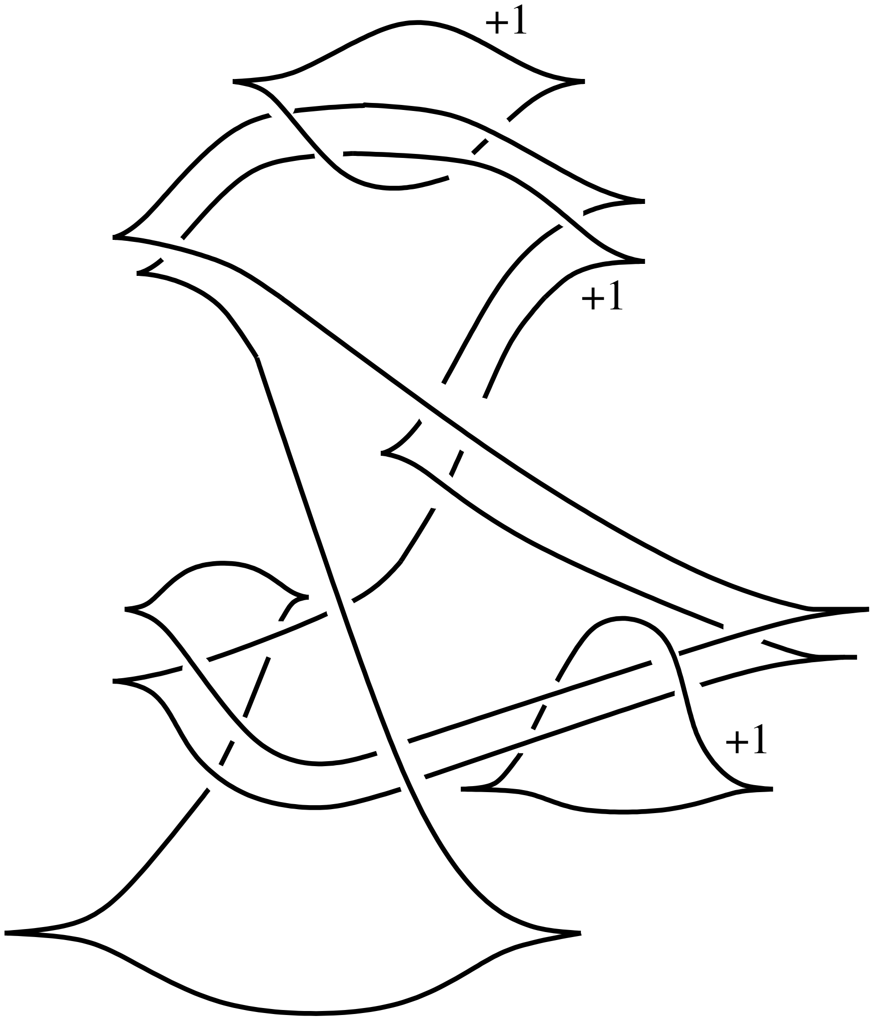


FIGURE 12. Nonfillable tight contact circle bundle

References

[1] S. Akbulut and B. Ozbagci, *On the topology of compact Stein surfaces*, Int. Math. Res. Not. **15** (2002), 769–782.
[2] F. Ding and H. Geiges, *A Legendrian surgery presentation of contact 3-manifolds*, Proc. Cambridge Philos. Soc. **136** (2004), 583–598.
[3] F. Ding, H. Geiges and A. Stipsicz, *Surgery diagrams for contact 3-manifolds*, Turkish J. Math. **28** (2004), 41–74.
[4] Y. Eliashberg, *Classification of overtwisted contact structures on 3-manifolds*, Invent. Math. **98** (1989), 623–637.
[5] Y. Eliashberg, *Topological characterization of Stein manifolds of dimension* > 2, International J. of Math. **1** (1990), 29–46.
[6] Y. Eliashberg, *A few remarks about symplectic filling*, Geom. Topol. **8** (2004), 277–293.
[7] J. Etnyre, *Introductory lectures on contact geometry*, arXiv:math.SG/0111118
[8] J. Etnyre, *On symplectic fillings*, Algebr. Geom. Topol. **4** (2004), 73–80.
[9] J. Etnyre, *Lectures on open book decompositions and contact structures*, in this volume.

[10] J. Etnyre and K. Honda, *On the non-existence of tight structures*, Ann. of Math. **153** (2001), 749–766.

[11] H. Geiges, *Contact geometry*, to appear in the Handbook of Differential Geometry, vol. 2.

[12] P. Ghiggini, P. Lisca and A. Stipsicz, *Classification of tight contact structures on small Seifert 3–manifolds with* $e_0 \geq 0$, Proc. Amer. Math. Soc., to appear, arXiv:math.SG/0406080

[13] E. Giroux, *Contact geometry: from dimension three to higher dimensions*, Proceedings of the International Congress of Mathematicians (Beijing 2002), 405–414.

[14] R. Gompf, *Handlebody construction of Stein surfaces*, Ann. of Math. **148** (1998), 619–693.

[15] R. E. Gompf and A. I. Stipsicz, *4-manifolds and Kirby calculus*, Graduate Studies in Mathematics, vol. **20**, American Math. Society, Providence 1999.

[16] K. Honda, *On the classification of tight contact structures, I.*, Geom. Topol. **4** (2000), 309–368.

[17] P. Kronheimer and T. Mrowka, *Witten's conjecture and Property P*, Geom. Topol. **8** (2004), 295–310.

[18] P. Kronheimer, T. Mrowka, P. Ozsváth and Z. Szabó, *Monopoles and lens space surgeries*, arXiv:math.GT/0310164

[19] P. Lisca and G. Matić, *Tight contact structures and Seiberg-Witten invariants*, Invent. Math. **129** (1997), 509–525.

[20] P. Lisca and A. Stipsicz, *Ozsváth–Szabó invariants and tight contact three–manifolds I.*, Geom. Topol. **8** (2004), 925–945.

[21] P. Lisca and A. Stipsicz, *Ozsváth–Szabó invariants and tight contact three–manifolds II.*, arXiv:math.SG/0404136

[22] P. Lisca and A. Stipsicz,, *Ozsváth–Szabó invariants and contact surgery*, in this volume.

[23] A. Némethi, *On the Ozsváth–Szabó invariants of negative definite plumbed 3-manifolds*,arXiv:math.GT/0310083.

[24] H. Ohta and K. Ono, *Simple singularities and topology of symplectically filling 4-manifolds*, Comment. Math. Helv. **74** (1999), 575–590.

[25] P. Ozsváth and Z. Szabó, *Holomorphic disks and topological invariants for rational homology three-spheres*, Ann. of Math. **159** (2004), 1027–1158.

[26] P. Ozsváth and Z. Szabó, *Holomorphic disks and three-manifold invariants: properties and applications*, Ann. of Math. **159** (2004), 1159–1245.

[27] P. Ozsváth and Z. Szabó, *On the Floer homology of plumbed three-manifolds*, Geom. Topol. **7** (2003), 185–224.

[28] P. Ozsváth and Z. Szabó, *Heegaard Floer homologies and contact structures*, arXiv:math.SG/0210127

[29] P. Ozsváth and Z. Szabó, *Holomorphic disks and knot invariants*, arXiv:math.GT/0209056

[30] P. Ozsváth and Z. Szabó, *Holomorphic disks and genus bounds*, Geom. Topol. **8** (2004), 311–334.

[31] P. Ozsváth and Z. Szabó, *An introduction to Heegaard Floer homology*, in this volume.

[32] P. Ozsváth and Z. Szabó, *Lectures on Heegaard Floer homology*, in this volume.

[33] A. Weinstein, *Contact surgery and symplectic handlebodies*, Hokkaido Mathematical Journal **20** (1991), 241–51.

[34] H. Wu, *Tight contact small Seifert spaces with* $e_0 \neq -2, -1, 0$, arXiv:math.GT/0402167.

Rényi Institute, Budapest and Institute for Advanced Study, Princeton
E-mail address: stipsicz@math-inst.hu and stipsicz@math.ias.edu

Clay Mathematics Proceedings
Volume 5, 2006

Ozsváth–Szabó Invariants and Contact Surgery

Paolo Lisca and András I. Stipsicz

Abstract. Let $T \subset S^3$ be a right–handed trefoil, and let $Y_r(T)$ be the closed, oriented 3–manifold obtained by performing rational r–surgery on the 3–sphere S^3 along T. In this paper we explain how to use contact surgery and the contact Ozsváth–Szabó invariants to construct positive, tight contact structures on $Y_r(T)$ for every $r \neq 1$. In particular, we give explicit constructions of positive, tight contact structures on the oriented boundaries of the positive E_6 and E_7 plumbings.

1. Introduction

We shall assume throughout the paper that every 3–manifold is connected, closed and oriented. A *contact structure* on a 3–manifold Y is a 2–dimensional distribution $\xi \subset TY$ given as the kernel of a 1–form $\alpha \in \Omega^1(Y)$ such that $\alpha \wedge d\alpha > 0$ everywhere on Y. The pair (Y, ξ) is a *contact 3–manifold.*

The *standard contact structure* ξ_{st} on $S^3 \subset \mathbb{C}^2$ is the distribution of complex tangent lines

$$\xi_{\text{st}} := TS^3 \cap i \cdot TS^3 \subset TS^3.$$

A contact 3–manifold (Y, ξ) is *overtwisted* if there exists an embedded disk $D^2 \hookrightarrow Y$ such that ξ is tangent to D^2 along its boundary ∂D^2. If there is no such disk, (Y, ξ) is *tight.*

It is known that every coorientable 2–plane field on an orientable 3–manifold is homotopic to a contact structure, so one of the central problems in present–day contact topology is:

(P) Which 3–manifolds carry tight contact structures?

The standard contact 3–sphere (S^3, ξ_{st}) is tight [**1**]. Let $T \subset S^3$ be a right–handed trefoil knot and, for every $r \in \mathbb{Q} \cup \{\infty\}$, denote by $Y_r(T)$ the oriented 3–manifold obtained by performing a rational surgery along T with coefficient r. Then, the oriented 3–manifold $Y_1(T)$ (i.e. the Poincaré homology sphere with orientation the opposite of the standard one) does not carry tight contact structures [**4**].

2000 *Mathematics Subject Classification.* AMS 2000 Classification: Primary 57R17; Secondary 57M27, 53D35, 57R58.

The authors wish to thank Peter Ozsváth and Zoltán Szabó for many useful discussions regarding their joint work. The first author was partially supported by MURST. The second author was partially supported by OTKA T034885.

Until recently, the two most important methods to deal with problem (P) were Eliashberg's *Legendrian surgery* as used, e.g. by Gompf in [**7**], and the *state traversal* method, developed by Ko Honda and based on Giroux's theory of convex surfaces. The limitations of these two methods come from the fact that Legendrian surgery can only prove tightness of Stein fillable contact structures, while the state traversal becomes too complicated in the absence of suitable incompressible surfaces. For example, both methods fail to deal with problem (P) when Y is either $Y_2(T)$ or $Y_3(T)$, because these Seifert fibered 3–manifolds do not contain vertical incompressible tori, nor do they carry symplectically fillable contact structures [**10, 11**]. As a result, for some time it was posed as an open problem whether $Y_2(T)$ or $Y_3(T)$ carried tight contact structures [**6**].

In this paper we illustrate how the contact Ozsváth–Szabó invariants [**19**] can be effectively combined with contact surgery [**2, 3**] to tackle problem (P). In particular, it follows from Theorem 1 below that $Y_2(T)$ and $Y_3(T)$ do indeed carry tight contact structures. Moreover, it follows from the proof of Theorem 1 that such contact structures can be explicitly described as in Figures 1 and 2 (see Section 2 for the explanation of the notation).

THEOREM 1. *Let $r \in \mathbb{Q} \cup \{\infty\}$, and denote by $Y_r(T)$ the closed, oriented 3–manifold obtained by performing r–surgery on the right–handed trefoil knot $T \subset S^3$. Then $Y_r(T)$ carries a tight contact structure for every $r \neq 1$.*

In proving Theorem 1 we first use contact surgery to define contact structures on $Y_r(T)$ for $r \neq 1$, and then show that the contact Ozsváth–Szabó invariants of those structures do not vanish, implying tightness. During the course of the proof we show that the contact invariants are nontrivial for infinitely many tight, not fillable contact 3–manifolds.

REMARK 2. The reader should be aware that in [**13, 14**] we prove results which are more general than the ones presented here. On the other hand, in this paper we try to keep our presentation at a more expository level by concentrating on just a few illustrative examples. In particular, the arguments given here are somewhat different from, and relatively simpler than, the ones used in [**13, 14**].

2. Contact surgery

Let (Y, ξ) be a contact 3–manifold. A knot $K \subset Y$ is *Legendrian* if K is everywhere tangent to ξ, i.e. $TK \subset \xi$. The framing of a Legendrian knot $K \subset Y$ naturally induced by ξ is called the *contact framing* of K. Given a non–zero rational number $r \in \mathbb{Q}$, one can perform *contact r–surgery* on a contact 3–manifold (Y, ξ) along a Legendrian knot $K \subset Y$ to obtain a new contact 3–manifold (Y', ξ') [**2, 3**]. Here Y' is the 3–manifold obtained by smooth r–surgery along K with respect to the contact framing, while ξ' is constructed by extending ξ from the complement of a suitable regular neighborhood of K as a tight contact structure on the glued–up solid torus. If $r \neq 0$ such an extension always exists, and for $r = \frac{1}{k}$ ($k \in \mathbb{Z}$) it is unique [**9**]. When $r = -1$, the corresponding contact surgery is usually called *Legendrian surgery* along K.

As an illustration of the contact surgery construction, consider the Legendrian link whose front projection is given by the left–hand side of Figure 1 (see e.g. [**8**, Section 11.1] for the description of Legendrian links in terms of their front projections). The coefficients next to each component of the diagram mean that one should perform contact (-1)–surgery along the Legendrian trefoil and $(+1)$–surgery along each of the Legendrian unknots. Since the contact framing of the Legendrian trefoil is $+1$ with respect to the Seifert framing while the contact framing of each Legendrian unknot is -1 (see e.g. [**8**, Section 11.1] for these calculations), converting the contact surgeries into smooth surgeries and applying some Kirby calculus gives the right–hand side of Figure 1. Therefore, the picture represents a contact structure on the oriented 3–manifold $Y_2(T)$. According to [**3**, Proposition 7], a

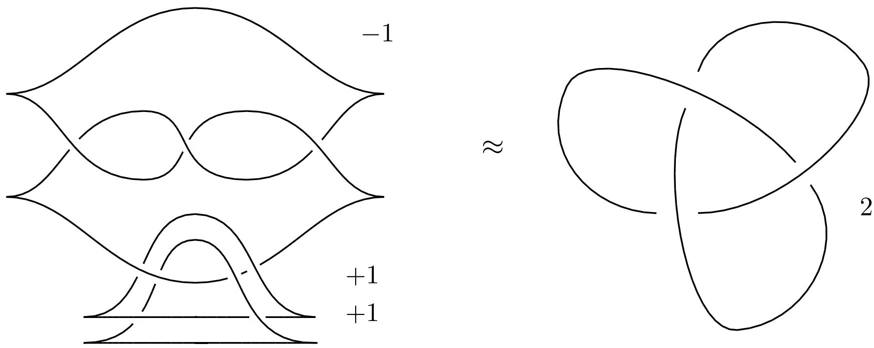


FIGURE 1. A contact structure on $Y_2(T)$

contact $r = \frac{p}{q}$–surgery ($p, q \in \mathbb{N}$) on a Legendrian knot K is equivalent to a contact $\frac{1}{k}$–surgery on K followed by a contact $\frac{p}{q-kp}$–surgery on a Legendrian pushoff of K for any integer $k \in \mathbb{N}$ such that $q - kp < 0$. Moreover, by [**3**, Proposition 3] each contact r–surgery along $K \subset (Y, \xi)$ with $r < 0$ is equivalent to a Legendrian surgery along a Legendrian link $\mathbb{L} = \cup_{i=0}^m L_i$. The set of all the Legendrian links $\mathbb{L}$ corresponding to all the possible contact r–surgeries along the Legendrian knot K is determined via a simple algorithm by K and the contact surgery coefficient r. The algorithm is the following. Since $1 - r > 1$, there is a continued fraction expansion

$$1 - r = a_0 - \cfrac{1}{a_1 - \cfrac{1}{\ddots - \cfrac{1}{a_m}}}, \quad a_0, \ldots, a_m \geq 2.$$

To obtain the first component L_0, push off K using the contact framing and stabilize it $a_0 - 2$ times. Then, push off L_0 and stabilize it $a_1 - 2$ times. Repeat the above scheme for each of the remaining pivots of the continued fraction expansion. Since there are $a_i - 1$ inequivalent ways to stabilize a Legendrian knot $a_i - 2$ times, this construction yields $\Pi_{i=0}^m (a_i - 1)$ potentially different Legendrian links.

For example, applying the algorithm just described one can check that the contact surgeries prescribed in the central picture of Figure 2 can be realized in the two ways given by the side pictures of Figure 2. Moreover, converting the coefficients into smooth surgery coefficients and applying Kirby calculus it is easy to check that the underlying 3–manifold is $Y_3(T)$.

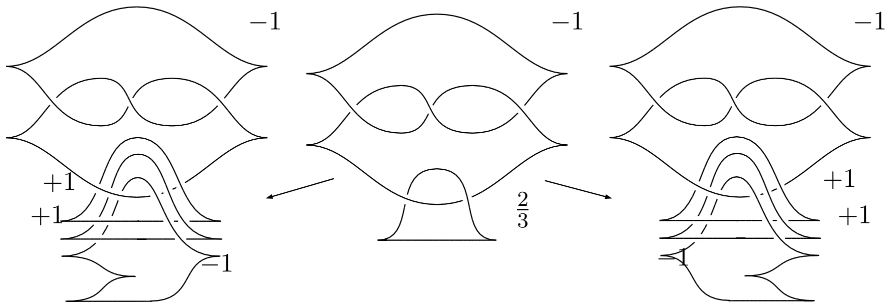


FIGURE 2. Two contact structures on $Y_3(T)$

3. Ozsváth–Szabó invariants

The *smooth Ozsváth–Szabó invariants* [**15, 16, 17**] assign to each oriented Spin^c 3–manifold $(Y, \mathbf{s})$ a finitely generated Abelian group $\widehat{HF}(Y, \mathbf{s})$, and to each oriented Spin^c cobordism $(W, \mathbf{t})$ between $(Y_1, \mathbf{s}_1)$ and $(Y_2, \mathbf{s}_2)$ a homomorphism

$$F_{W,\mathbf{t}}\colon \widehat{HF}(Y_1, \mathbf{s}_1) \to \widehat{HF}(Y_2, \mathbf{s}_2).$$

For simplicity, in the following we will use these homology theories with $\mathbb{Z}/2\mathbb{Z}$ coefficients. In this setting, $\widehat{HF}(Y, \mathbf{s})$ is a finite dimensional vector space over the field $\mathbb{Z}/2\mathbb{Z}$. Define

$$\widehat{HF}(Y) = \bigoplus_{\mathbf{s}\in \mathrm{Spin}^c(Y)} \widehat{HF}(Y, \mathbf{s}).$$

Since there are only finitely many Spin^c structures with nonvanishing invariants [**16**, Theorem 7.1], $\widehat{HF}(Y)$ is still finite dimensional.

Now we describe what is usually called the *surgery exact triangle* for the Ozsváth–Szabó homologies.

Let Y be a closed, oriented 3–manifold and let $K \subset Y$ be a framed knot with framing f. Let $Y_f(K)$ denote the 3–manifold given by surgery along $K \subset Y$ with respect to the framing f. The surgery can be viewed at the 4–manifold level as a 4–dimensional 2–handle addition. The resulting cobordism X induces a homomorphism

$$F_X := \sum_{\mathbf{t}\in \mathrm{Spin}^c(X)} F_{X,\mathbf{t}}\colon\ \widehat{HF}(Y) \to \widehat{HF}(Y_f(K))$$

obtained by summing over all Spin^c structures on X. Similarly, there is a cobordism U defined by adding a 2–handle to $Y_f(K)$ along a small normal circle N to K with framing -1 with respect to a small normal disk to K. The boundary components of U are $Y_f(K)$ and the 3–manifold $Y_{f+1}(K)$ obtained from Y by a surgery along K with framing $f+1$. As before, U induces a homomorphism

$$F_U\colon\ \widehat{HF}(Y_f(K)) \to \widehat{HF}(Y_{f+1}(K)).$$

The above construction can be repeated starting with $Y_f(K)$ and $N \subset Y_f(K)$ equipped with the framing specified above: we get U (playing the role previously played by X) and a new cobordism V starting from $Y_{f+1}(K)$, given by attaching a 4–dimensional 2–handle along a normal circle to N with framing -1 with respect to a normal disk. It is easy to check that this last operation yields Y at the 3–manifold level. The homomorphisms F_X, F_U and F_V fit into an exact triangle called the *surgery exact triangle*

$$\begin{tikzcd} \widehat{HF}(Y) \arrow[rr, "F_X"] & & \widehat{HF}(Y(K)) \arrow[dl, "F_U"] \\ & \widehat{HF}(Y'(K)) \arrow[ul, "F_V"] & \end{tikzcd} \tag{3.1}$$

The *contact Ozsváth–Szabó invariant* for a contact 3–manifold (Y, ξ) [**19**] is an element

$$c(Y, \xi) \in \widehat{HF}(-Y, \mathbf{s}_\xi)/\langle \pm 1 \rangle,$$

where $\mathbf{s}_\xi$ denotes the Spin^c structure induced by the contact structure ξ. Since in this paper we are working with $\mathbb{Z}/2\mathbb{Z}$ coefficients, the above sign ambiguity for $c(Y, \xi)$ does not occur. It is proved in [**19**] that if (Y, ξ) is overtwisted then $c(Y, \xi) = 0$, and if (Y, ξ) is Stein fillable (see e.g. [**8**, Chapter 11] for the definition) then $c(Y, \xi) \neq 0$. It follows immediately that if $c(Y, \xi) \neq 0$ then (Y, ξ) is tight, and $c(S^3, \xi_{\mathrm{st}}) \neq 0$.

In order to prove Theorem 1 we shall use the properties of $c(Y, \xi)$ given in the following theorem and corollary.

Theorem 3 ([**12**], Theorem 2.3). *Suppose that (Y', ξ') is obtained from (Y, ξ) by a contact $(+1)$–surgery. Let $-X$ be the cobordism induced by the surgery with reversed orientation. Define*

$$F_{-X} := \sum_{\mathbf{t} \in \mathrm{Spin}^c(-X)} F_{-X, \mathbf{t}}.$$

Then,

$$F_{-X}(c(Y, \xi)) = c(Y', \xi').$$

In particular, if $c(Y', \xi') \neq 0$ then (Y, ξ) is tight. □

Corollary 4 ([**12**], Corollary 2.4). *If $c(Y_1, \xi_1) \neq 0$ and (Y_2, ξ_2) is obtained from (Y_1, ξ_1) by Legendrian surgery along a Legendrian knot, then $c(Y_2, \xi_2) \neq 0$. In particular, (Y_2, ξ_2) is tight.* □

4. The proof of Theorem 1

Consider the contact structures defined by Figure 3(a) for $r' \neq 0$. Converting the picture into a smooth surgery, it is easy to check that the underlying 3–manifold is $Y_r(T)$. Observe that the contact structures are well–defined only for $r \neq 1$, because when $r = 1$ we have $r' = 0$ (in which case the corresponding contact surgery is not well–defined).

In order to prove Theorem 1 we will show that all the contact structures determined by Figure 3(a) have nonvanishing Ozsváth–Szabó invariants. As explained in Section 2, a contact r'–surgery with $r' < 0$ can be replaced by a sequence of Legendrian (i.e., contact (-1)–) surgeries. Therefore, since $c(S^3, \xi_{\mathrm{st}}) \neq 0$, by Corollary 4 the contact structures defined by Figure 3(a) have nonvanishing Ozsváth–Szabó invariants for $r' < 0$ or $r' = \infty$.

If $r' > 0$ then, as explained in Section 2, the contact structures of Figure 3(a) can be equivalently given by the diagram of Figure 3(b) for any natural number

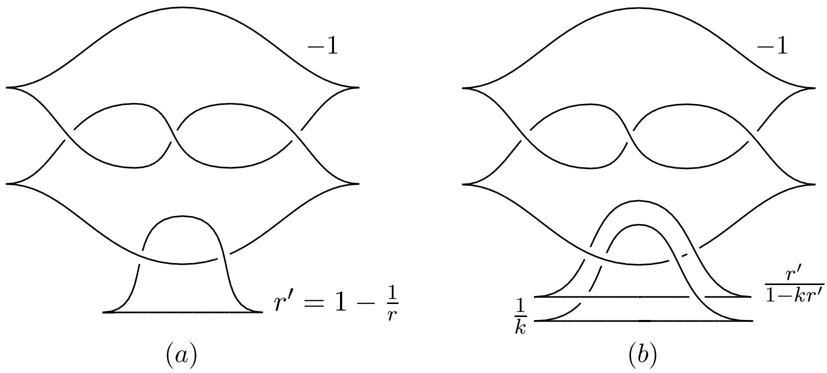


FIGURE 3. Surgery diagrams for contact structures on $Y_r(T)$

k large enough so that $1 - kr' < 0$. Therefore, since any contact $\frac{r'}{1-kr'}$–surgery in Figure 3(b) can be replaced by a sequence of Legendrian surgeries, in order to prove Theorem 1 it suffices to show that the contact 3–manifold obtained from Figure 3(a) for $r' = \frac{1}{k}$ has nonvanishing contact invariant for every $k \in \mathbb{N}$. That is exactly what we are going to do, but first we need an auxiliary result.

LEMMA 5. *The contact structure given by Figure 4 has nonvanishing Ozsváth–Szabó invariant.*

PROOF. The contact framing of the Legendrian unknot of Figure 4 is -1 with respect to the Seifert framing. Therefore, the contact 3–manifold given by Figure 4 is of the form $(S^1 \times S^2, \eta)$, and Triangle (3.1) becomes

$$\begin{tikzcd}
\widehat{HF}(S^3) \arrow[rr, "F_{-X}"] & & \widehat{HF}(S^1 \times S^2) \arrow[dl] \\
& \widehat{HF}(S^3) \arrow[ul] &
\end{tikzcd}$$

where $-X$ is the cobordism from S^3 to $S^1 \times S^2$ obtained by attaching a two–handle to S^3 along a zero–framed unknot. By Theorem 3 we have

$$F_{-X}(c(S^3, \xi_{\rm st})) = c(S^1 \times S^2, \eta).$$

By [**16**], $\widehat{HF}(S^1 \times S^2)$ is isomorphic to $(\mathbb{Z}/2\mathbb{Z})^2$, while $\widehat{HF}(S^3)$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$. Exactness of the triangle immediately implies that F_{-X} is injective. Since (S^3, ξ_{st}) is Stein fillable we have $c(S^3, \xi_{st}) \neq 0$, therefore $c(S^1 \times S^2, \eta) \neq 0$. □

Let (V_k, ξ_k) denote the contact 3–manifold obtained by choosing $r' = \frac{1}{k}$ in Figure 3(a), so that $V_k \cong Y_{\frac{k}{k-1}}$. Notice that for $r' = k = 1$ the 3–manifold $V_1 \cong Y_\infty$ is diffeomorphic to the 3–sphere S^3. By [**2**, Proposition 9], a contact $\frac{1}{k}$–surgery ($k \in \mathbb{N}$) on a Legendrian knot K can be replaced by k contact $(+1)$–surgeries on k Legendrian pushoffs of K. Therefore, the contact 3–manifold (V_k, ξ_k) can be alternatively defined by the diagram of Figure 5, which contains k contact $(+1)$–framed Legendrian unknots.

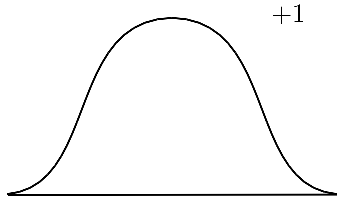


FIGURE 4. A contact structure with nonvanishing invariant

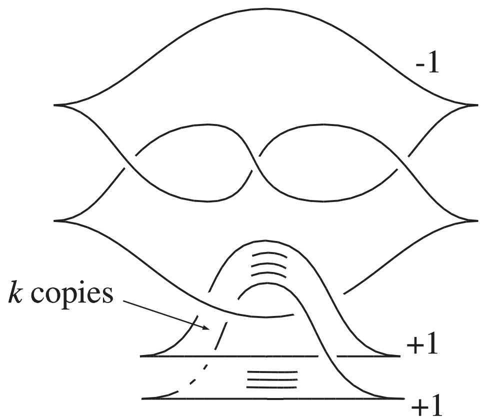


FIGURE 5. Equivalent surgery diagram for (V_k, ξ_k)

LEMMA 6. *Let $k \geq 1$ be an integer. Then, $c(V_k, \xi_k) \neq 0$.*

PROOF. Consider Figure 5 for $k = 1$, which represents (V_1, ξ_1). Clearly, (V_1, ξ_1) is obtained by performing a Legendrian surgery on the contact 3–manifold given by Figure 4. Therefore, by Lemma 5 and Corollary 4, the contact Ozsváth–Szabó invariant of (V_1, ξ_1) is nonzero. This proves the lemma for $k = 1$.

Observe that, given an exact triangle of vector spaces and linear maps

$$\begin{tikzcd} V_1 \arrow[rr, "F_3"] & & V_2 \arrow[dl, "F_1"] \\ & V_3 \arrow[ul, "F_2"] & \end{tikzcd}$$

we have

$$\dim V_i \geq |\dim V_j - \dim V_k| \tag{4.1}$$

and

$$\dim V_i \leq \dim V_j + \dim V_k \tag{4.2}$$

for $\{i, j, k\} = \{1, 2, 3\}$. Moreover, equality holds in (4.2) if and only if $F_i = 0$.

Now suppose $k \geq 1$ and $c(V_k, \xi_k) \neq 0$. Clearly, (V_{k+1}, ξ_{k+1}) is obtained from (V_k, ξ_k) by performing a contact $(+1)$–surgery. Now it is easy to check that the cobordism X_k corresponding to the surgery induces a homomorphism F_{-X_k} which fits into an exact triangle having the peculiar property that the third manifold involved in the triangle is independent of k:

$$(*) \qquad \begin{tikzcd} \widehat{HF}(-V_k) \arrow[rr, "F_{-X_k}"] & & \widehat{HF}(-V_{k+1}) \arrow[dl] \\ & \widehat{HF}(-Y_{+1}(T)) \arrow[ul] & \end{tikzcd}$$

In fact, $-Y_{+1}(T)$ is the Poincaré sphere $\Sigma(2,3,5)$, and it follows from the calculations of [**18**, Section 3.2] that $\widehat{HF}(-Y_{+1}(T)) = \mathbb{Z}/2\mathbb{Z}$. Therefore, setting $d(k) = \dim_{\mathbb{Z}/2\mathbb{Z}} \widehat{HF}(-V_k)$, Triangle $(*)$ and (4.2) imply

$$d(k+1) \leq d(k) + 1 \tag{4.3}$$

for every $k \geq 1$. Now observe that $-V_k$ can be presented by the surgery diagram of Figure 6. Let M be the 3–manifold obtained by surgery on the framed link of

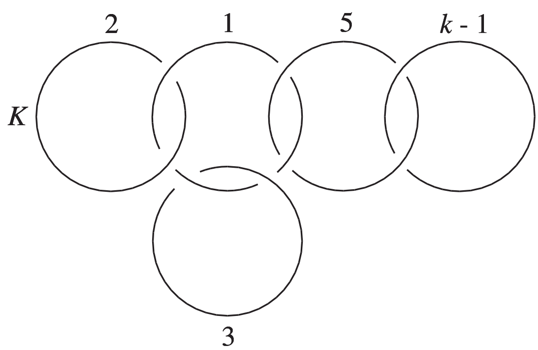


FIGURE 6. A surgery diagram for $-V_k$

Figure 6 with the 2–framed knot K deleted. It is easy to compute what the surgery exact triangle corresponding to $(M, M_1(K), M_2(K))$ looks like:

$$\begin{tikzcd} \widehat{HF}(L(7k-9,7)) \arrow[rr] & & \widehat{HF}(L(8k-9,8)) \arrow[dl] \\ & \widehat{HF}(-V_k) \arrow[ul] & \end{tikzcd}$$

Since by [**16**, Proposition 3.1] $\dim_{\mathbb{Z}/2\mathbb{Z}} \widehat{HF}(L(p,q)) = p$ for every p and q, exactness of the triangle and (4.1) imply

$$d(k) \geq k \tag{4.4}$$

for every $k \geq 1$, and since $V_1 \cong Y_\infty \cong S^3$, we have $d(1) = 1$. Therefore, by (4.3) and (4.4) we have $d(k) = k$ for every $k \geq 1$ and, in particular, equality holds in (4.3). By exactness of Triangle $(*)$ this immediately implies that F_{-X_k} is injective for every $k \geq 1$. Thus,

$$c(V_k, \xi_k) = F_{-X_{k-1}}(c(V_{k-1}, \xi_{k-1})) \neq 0$$

for every $k \geq 1$. □

Remark 7. Since $V_2 = Y_2(T)$ does not carry symplectically fillable contact structures [**10, 11**], by Lemma 6 (V_2, ξ_2) is a tight but not fillable contact 3–manifold. Moreover, since contact (+1)–surgery on a nonfillable structure produces a nonfillable structure [**2, 5**], the contact 3–manifold (V_k, ξ_k) is tight, not symplectically fillable for each $k \geq 2$.

Proof of Theorem 1. If $r' < 0$ or $r' = \infty$, any contact surgery given by Figure 3(a) can be realized by a sequence of Legendrian surgeries on (S^3, ξ_{st}), therefore by Corollary 4 the resulting contact structure has nonvanishing contact Ozsváth–Szabó invariant and hence it is tight. If $r' \neq \infty$ and $r' > 0$, choose an integer k so large that $\frac{r'}{1-kr'} < 0$. Then, each contact surgery given by Figure 3(a) is equivalent to a contact surgery given by Figure 3(b). Moreover, the resulting contact structure is obtained from (V_k, ξ_k) for some $k \in \mathbb{N}$ by a sequence of Legendrian surgeries, and therefore it is tight by Corollary 4 and Lemma 6. □

References

[1] D. Bennequin *Entrelacements et équations de Pfaff*, Astérisque **107–108** (1983), 87–161.

[2] F. Ding and H. Geiges, *Symplectic fillability of tight contact structures on torus bundles*, Alg. and Geom. Topol. **1** (2001), 153–172.

[3] F. Ding and H. Geiges, *A Legendrian surgery presentation of contact 3-manifolds*, Math. Proc. Cambridge Philos. Soc. **136** (2004), 583–598.

[4] J. Etnyre and K. Honda, *On the nonexistence of tight contact structures*, Ann. of Math. **153** (2001), 749–766.

[5] J. Etnyre and K. Honda, *Tight contact structures with no symplectic fillings*, Invent. Math. **148** (2002), 609–626.

[6] J. Etnyre and L. Ng, *Problems in Low Dimensional Contact Topology*, Geometric Topology, 2001 Georgia International Topology Conference, AMS/IP Studies in Advanced Mathematics **35** (2003), 337–357.

[7] R Gompf, *Handlebody constructions of Stein surfaces*, Ann. of Math. **148** (1998), 619–693.

[8] R. Gompf and A. Stipsicz, *4–manifolds and Kirby calculus*, Graduate Studies in Mathematics **20** AMS, 1999.

[9] K Honda, *On the classification of tight contact structures I.*, Geom. Topol. **4** (2000), 309–368.

[10] P. Lisca, *Symplectic fillings and positive scalar curvature*, Geom. Topol. **2** (1998) 103–116.

[11] P. Lisca, *On symplectic fillings of 3-manifolds*, Proceedings of the 6^{th} Gökova Geometry-Topology Conference, Turkish J. Math. **23** (1999), 151–159.

[12] P. Lisca and A. Stipsicz, *Seifert fibered contact three-manifolds via surgery*, Algebr. Geom. Topol. **4** (2004), 199–217.

[13] P. Lisca and A. Stipsicz, *Ozsváth–Szabó invariants and tight contact three-manifolds, I*, Geom. Topol. **8** (2004), 925–945.

[14] P. Lisca and A. Stipsicz, *Ozsváth–Szabó invariants and tight contact three-manifolds, II*, arXiv:math.SG/0404136.

[15] P. Ozsváth and Z. Szabó, *Holomorphic disks and topological invariants for closed three-manifolds*, Ann. of Math. (2) **159** (2004), no. 3, 1027–1158.

[16] P. Ozsváth and Z. Szabó, *Holomorphic disks and three-manifold invariants: properties and applications*, Ann. of Math. (2) **159** (2004), no. 3, 1159–1245.

[17] P Ozsváth and Z Szabó, *Holomorphic triangles and invariants of smooth* 4*–manifolds*, Duke Math. J. **121** (2004), 1–34.

[18] P. Ozsváth and Z. Szabó, *Absolutely Graded Floer homologies and intersection forms for four-manifolds with boundary*, Adv. Math. **173** (2003), 179–261.
[19] P. Ozsváth and Z. Szabó, *Heegaard Floer homologies and contact structures*, arXiv:math.SG/0210127.

Dipartimento di Matematica "L. Tonelli", Università di Pisa, Largo Bruno Pontecorvo 5, I–56127 Pisa, ITALY
E-mail address: lisca@dm.unipi.it

Rényi Institute of Mathematics, Hungarian Academy of Sciences, H-1053 Budapest, Reáltanoda utca 13–15, Hungary
E-mail address: stipsicz@math-inst.hu

Clay Mathematics Proceedings
Volume 5, 2006

Double Points of Exact Lagrangian Immersions and Legendrian Contact Homology

Tobias Ekholm

Abstract. We use contact homology to obtain lower bounds on the number of double points of self transverse exact Lagrangian immersions of closed manifolds into the product of the cotangent bundle of a manifold and $\mathbb{C}$. The inequality obtained is similar to the Morse inequalities estimating the number of critical points of a Morse function on a closed manifold in terms of its homology.

1. Introduction

Let M be a smooth manifold of dimension n. Consider the cotangent bundle $T^*M \xrightarrow{\pi} M$. The *canonical* 1*-form* θ_M on T^*M maps a tangent vector $X \in T_\alpha(T^*M)$ to $\alpha(d\pi(X))$. The *standard symplectic form* on T^*M is $\omega_M = d\theta_M$. If $(q_1, \dots, q_n)$ are local coordinates on M and $(q_1, \dots, q_n, p_1, \dots, p_n)$ are corresponding coordinates on T^*M then $\theta_M = \sum_j p_j\, dq_j$ and $\omega_M = \sum_j dp_j \wedge dq_j$.

An immersion $f\colon L \to T^*M$ of an n-dimensional manifold L is *Lagrangian* if $f^*\omega_M = 0$. This implies that the form $f^*\theta_M$ is closed. A Lagrangian immersion $f\colon L \to T^*M$ is *exact* if the form $f^*\theta_M$ is exact.

Let $f\colon L \to T^*M$ be an exact Lagrangian immersion of a connected manifold and let $h\colon L \to \mathbb{R}$ be a function such that $dh = f^*\theta_M$. Consider the map $\tilde{f} = (f, h)\colon L \to T^*M \times \mathbb{R} \approx J^1(M)$, where $J^1(M)$ is the 1-jet space of M. This map is an immersion which is everywhere tangent to the hyperplane field $\xi = \ker(dz - \theta_M)$ on $J^1(M)$, where z is a coordinate along the $\mathbb{R}$-direction in $T^*M \times \mathbb{R}$. The hyperplane field ξ is completely non-integrable: if $\alpha = dz - \theta_M$ then $\alpha \wedge (d\alpha)^n \neq 0$. Such a hyperplane field is called a *contact structure* and the 1-form α a *contact form*. In fact ξ is the *standard contact structure* on $J^1(M)$ and α the *standard contact form*. An immersion of an n-manifold into $J^1(M)$ which is everywhere tangent to ξ is called *Legendrian*. Thus, to each exact Lagrangian immersion $f\colon L \to T^*M$ corresponds a family of Legendrian immersions $\tilde{f}\colon L \to J^1(M)$, two members of which differ by a translation in the $\mathbb{R}$-direction (the choice of h is unique up

---

2000 *Mathematics Subject Classification.* 53D35 (53D12).

TE is a research fellow of the Swedish Royal Academy of Sciences sponsored by the Knut and Alice Wallenberg foundation.

to addition of constants). Moreover, for a dense open set of exact Lagrangian immersions their Legendrian lifts are embeddings.

Legendrian and Lagrangian immersions are "soft" in the sense that they obey so called h-principles, see [**14**]. For example, to determine whether or not two Lagrangian (Legendrian) immersion are regularly homotopic through Lagrangian (Legendrian) immersions is a homotopy theoretic question. In contrast to this there are also "hard" properties. For example, double points of exact Lagrangian immersions can in general not be removed even though there are no homotopy obstructions for doing so, as the following theorem of Gromov [**15**] shows.

THEOREM 1.1. *An exact Lagrangian immersion $f\colon L \to \mathbb{C}^n$ has at least one double point.*

We will present a proof of Gromov's result which uses Floer homology, see Theorem 2.8, and use similar techniques to demonstrate that the following conjecture, see [**1**], (which we state in its simplest form) holds for a certain class of exact Lagrangian submanifolds.

CONJECTURE 1.2. *Every self transverse Lagrangian immersion $f\colon L \to \mathbb{C}^n$ has at least*

$$\frac{1}{2}\dim(H_*(L;\mathbb{Z}_2))$$

double points.

The tool we use is *Legendrian contact homology*, which is part of Symplectic Field Theory, see [**5**] and also [**3**] and [**4**], and is similar to the Floer homology of Lagrangian intersections. It provides Legendrian isotopy invariants via pseudo-holomorphic curve techniques. Using a Morse-Bott argument it is straightforward to show that Conjecture 1.2 holds for any exact Lagrangian the Legendrian lift of which admits a generating function, see e.g. [**2**] or [**6**] for the definition of a generating function. In Theorem 3.5 we prove a result which implies that Conjecture 1.2 holds for exact Lagrangian immersions into $T^*(M \times \mathbb{R})$ provided their Legendrian lifts have good contact homology algebras (see Subsection 3.2 for the definition of a good algebras). This result was first proved in [**9**].

REMARK 1.3. The definitions of Floer homology and contact homology given below are streamlined in the sense that only the part of these theories needed for the proof of the double point estimates discussed above will be described. In particular, there is no mention of the grading in either of the theories. Also, for simplicity we use only $\mathbb{Z}_2$-coefficients throughout. If the Legendrian submanifolds considered in Section 3 are assumed to be spin then the $\mathbb{Z}_2$ in all double point estimates involving homology groups could be replaced by $\mathbb{Z}_p$, where p is any prime or with $\mathbb{Q}$, see [**9**].

2. Floer homology and non-injectivity of exact Lagrangian immersions

The purpose of this section is to show that the Floer homology of two compact embedded exact Lagrangian submanifolds of a cotangent bundle T^*M of some n-manifold M is well-defined.

2.1. Floer homology of Lagrangian intersections. Let L be an embedded exact Lagrangian submanifolds in a cotangent bundle T^*M and let J be an almost complex structure on T^*M compatible with ω_M. That is, ω is positive on J-complex lines and J is an ω-isomorphism. Let S be a Riemann surface with complex structure i. A map $u\colon S \to \mathbb{C}^n$ is called *J-holomorphic* is

$$du + J \circ du \circ i = 0.$$

LEMMA 2.1. *Let S be the unit disk or the Riemann sphere. The only J-holomorphic maps $u\colon S \to T^*M$ such that $u(\partial S) \subset L$ are the constant maps.*

PROOF. Note that the area of a J-holomorphic map $u\colon S \to T^*M$ agrees with its energy and satisfies

$$\mathrm{Area}(u) = \int_S u^*\omega = \int_{\partial S} u^*\theta = \int_{\partial S} dh = 0.$$

Thus any such map must be constant. □

Let L_0 and L_1 be exact Lagrangian transverse submanifolds of T^*M. Let $\mathcal{C} = \{c_1, \dots, c_m\}$ be the set of intersection points of L_1 and L_2. Let $\mathbb{Z}_2\langle\mathcal{C}\rangle$ be the vector space over $\mathbb{Z}_2$ generated by $\mathcal{C}$. We define the Floer homology differential on $\mathbb{Z}_2\langle\mathcal{C}\rangle$ by counting rigid J-holomorphic strips. More precisely, define for double points a and b the moduli space $\mathcal{M}(a;b)$ as the space of maps $u\colon \mathbb{R} \times [0,1] \to T^*M$ such that

- u is J-holomorphic, i.e. $du + J \circ du \circ i = 0$,
- $u(\mathbb{R} \times \{0\}) \subset L_0$ and $u(\mathbb{R} \times \{1\}) \subset L_1$, and
- $\lim_{\tau\to-\infty} u(\tau + it) = a$ and $\lim_{\tau\to\infty} u(\tau + it) = b$,

up to conformal reparametrization. The following lemma is proved in [**11**].

LEMMA 2.2. *For almost complex structures J in an open dense subset $\mathcal{M}(a;b)$ is a finite collection of finite dimensional manifolds with natural compactifications. In particular the 0-dimensional components of the space $\mathcal{M}(a;b)$ form a finite collection of points.*

DEFINITION 2.3. The Floer homology differential $\partial\colon Z_2\langle\mathcal{C}\rangle \to \mathbb{Z}_2\langle\mathcal{C}\rangle$ is the linear map defined on generators as

$$\partial a = \sum_{\dim \mathcal{M}(a;b)=0} |\mathcal{M}(a;b)| b,$$

where $|\mathcal{M}(a;b)|$ denotes the mod 2 number of points in the finite set $\mathcal{M}(a;b)$.

LEMMA 2.4. *The Floer homology differential is a differential, in other words, $\partial^2 = 0$.*

PROOF. To show this one applies the usual gluing argument, see [**12**]. Let a be a double point. A term contributing to $\partial^2 a$ arises through a rigid strip connecting a to b and another rigid strip connecting b to c. These strips can be glued together to a 1-parameter family of strips connecting a to c. Using Gromov compactness we find that this 1-parameter family must break. This can happen in three ways: either the strip splits off a non-constant J-holomorphic sphere or a J-holomorphic disk with boundary on L or it breaks into a rigid strip from a to b' and from b' to c. The two first cases are ruled out by Lemma 2.1. Therefore the contributions to $\partial^2 a$ cancel in pairs and the lemma holds. □

LEMMA 2.5. *The Floer homology $\ker(\partial)/\operatorname{Im}(\partial)$ is invariant under deformations of L_0 and L_1 through exact Lagrangian submanifolds.*

PROOF. See Floer [**12**]. □

2.2. Floer homology and Morse theory. We give a short description of Floer's result relating Floer homology to finite dimensional Morse theory. Let M be a smooth n-manifold and let $f\colon M \to \mathbb{R}$ be a Morse function on M. Note that the graph of the differential of f, Γ_f, parameterized by

$$m \mapsto (m, df(m)) \in T^*M, \quad m \in M,$$

is a Lagrangian submanifold of T^*M by the fact that mixed partial derivatives are equal. Consider the pair of Lagrangians Γ_f and Γ_0. Note that the intersection points of Γ_f and Γ_0 correspond exactly to critical points of f. Moreover, since f is a Morse function the intersection points are transverse.

Fix a Riemannian metric g on M. The metric g determines the Levi-Civita connection ∇ on T^*M. The connection ∇ gives a direct sum decomposition $T(T^*M) = V \oplus H$, where the vertical bundle V equals the kernel of $d\pi\colon T(T^*M) \to TM$, the differential of the projection $\pi\colon T^*M \to M$. The fiber H_α of the horizontal bundle H at $\alpha \in T^*M$ is defined as the velocity vectors of covariantly constant lifts of curves through $\pi(\alpha)$ with initial value α. The natural almost complex structure J on T^*M is required to satisfy $J(V) = H$ and is defined as follows on vertical vectors $\xi \in V_\alpha$. Translate ξ to the origin in $T^*_{\pi(\alpha)}M$. Identify this translate with a tangent vector to M and let $J(\xi)$ be the negative of its horizontal lift.

Let Φ_t be the time t Hamiltonian flow in T^*M of the function $F = f \circ \pi$, where $\pi\colon T^*M \to M$ is the natural projection and where f is the function used to define Γ_f. Define the t-dependent complex structure

$$J_t = d\Phi_{-t} \circ J \circ d\Phi_t$$

and the corresponding $\bar\partial_{J_t}$-equation

$$du + J_t \circ du \circ i = 0, \tag{2.1}$$

for $u\colon \mathbb{R} \times [0,1] \to T^*M$, where t is a coordinate in the $[0,1]$-direction. A straightforward calculation shows that if $\gamma\colon \mathbb{R} \to M$ solves the gradient equation

$$\frac{d}{d\tau}\gamma(\tau) = -\nabla f(\gamma(\tau)),$$

then $u(\tau, t) = \Phi_t(\gamma(\tau))$ solves (2.1). Moreover the following theorem guarantees that after scaling $f \to \lambda f$ with a sufficiently small $\lambda > 0$ these solutions are the only ones.

THEOREM 2.6. *For every $f\colon M \to \mathbb{R}$ of sufficiently small C^2-norm the moduli space of rigid J_t-holomorphic strips with boundary on Γ_f and Γ_0 is diffeomorphic to the moduli space of gradient trajectories of f.*

PROOF. See [**13**]. □

The usual proof of invariance of Floer homology implies that the Floer homology defined using J_t-holomorphic disks and that defined using J-holomorphic disks are the same. In particular, it follows that the Floer homology of Γ_f and Γ_0 is isomorphic to the ordinary homology of M, since this is what the Morse complex computes.

COROLLARY 2.7.

$$HF_*(\Gamma_0, \Gamma_f; \mathbb{Z}_2) = H_*(M; \mathbb{Z}_2).$$

□

2.3. Non-injectivity. Let M be any smooth manifold.

THEOREM 2.8. *An exact Lagrangian immersion $f\colon L \to T^*(M \times \mathbb{R})$ has at least one double point.*

PROOF. Assume that there exists an exact Lagrangian embedding $f\colon L \to T^*(M \times \mathbb{R})$. Then the symplectic neighborhood theorem implies that this embedding can be extended to a symplectic embedding $\phi\colon U \to T^*M$, where U is a neighborhood of the 0-section in T^*L. Let ϕ be such a map and fix $\epsilon > 0$ such that the closed 2ϵ neighborhood of the 0-section is contained in U. Let V and W be the images under ϕ of the ϵ-neighborhood and of the 2ϵ-neighborhood of the 0-section, respectively. Let $g\colon L \to \mathbb{R}$ be a Morse function. For $\lambda > 0$ small enough $\Gamma_{\lambda g} \subset U$ and using ϕ we may regard $\Gamma_{\lambda g}$ as an embedded Lagrangian submanifold of $T^*(M \times \mathbb{R})$.

We will compute the Floer homology of $L_0 = L$ and $L_1 = \Gamma_{\lambda f}$ in two ways. To this end we first show that there exists $\lambda_0 > 0$ such that, for all $\lambda < \lambda_0$, all holomorphic strips with boundary on L_0 and L_1 lie inside U. Assume this is not the case. Pick a sequence of holomorphic strips which passes through some point q in the compact region $W - V$. By Gromov compactness, see [**15**], this sequence converges to a collection of holomorphic curves with non-empty boundary on L as $\lambda \to 0$. Since this curve must contain q some component of it is non-constant and has its boundary on L. This contradicts the exactness of L by the argument in the proof of Lemma 2.1 and we conclude that such a $\lambda_0 > 0$ exists.

It follows that for $0 < \lambda < \lambda_0$, the Floer homology of L_0 and L_1 in $T^*(M \times \mathbb{R})$ agrees with the Floer homology of Γ_0 and Γ_1 in T^*L. Hence, by Corollary 2.7,

$$HF_*(L_0, L_1; \mathbb{Z}_2) \approx H_*(L; \mathbb{Z}_2) \neq 0. \tag{2.2}$$

Let x be a coordinate in the $\mathbb{R}$-direction of $M \times \mathbb{R}$ and y be the conjugate coordinate in $\mathbb{R}^2$ in the decomposition $T^*(M \times \mathbb{R}) = T^*M \times \mathbb{R}^2$. The Hamiltonian flow Φ_t of the function $H\colon T^*(M \times \mathbb{R}) \to \mathbb{R}$, $H = h \circ \pi$, $h(m, x) = x$ is simply translation in the y-direction. In particular for T large enough, by compactness of L, $\Phi_T(L_1)$ is disjoint from L. Thus the Floer complex of L and $\Phi_T(L_1)$ has no generators and hence

$$HF_*(L_0, L_1; \mathbb{Z}_2) = 0. \tag{2.3}$$

Equations (2.2) and (2.3) contradict the fact that Floer homology is invariant under Hamiltonian deformations. We conclude that $f\colon L \to T^*(M \times \mathbb{R})$ could not have been embedded. $\square$

3. Contact homology, its linearization, and a double point estimate

When the homology of the manifold L in the proof of Theorem 2.8 above is large it seems that the method of proof could give a stronger result than merely one double point. As we shall see below it is possible to get more information out of the argument, provided the lift of the Lagrangian satisfies some extra conditions, by using linearized contact homology.

One may view Legendrian contact homology as the counterpart of Floer homology for projections of exact Lagrangian manifolds. The main difference between the Legendrian case and the case of embedded Lagrangian submanifolds is the appearance of one punctured holomorphic disks with boundary on the immersed exact Lagrangian. In particular these disks appear in limits of 1-parameter families and therefore to define some kind of homology theory one must include disks with an arbitrary number of punctures. However, sometimes many of the extra disks can be disregarded and still a reasonable Floer homology theory can be defined. We will exploit this fact below.

3.1. Legendrian contact homology. We associate a differential graded algebra (DGA) to a Legendrian submanifold of a 1-jet space. This algebra is invariant up to stable tame isomorphism under Legendrian isotopies of the submanifold. In particular, the homology of the algebra is invariant and provides a Legendrian isotopy invariant.

Let M be a smooth n-manifold and let $L \subset J^1(M)$ be a Legendrian submanifold which is generic with respect to the *Lagrangian projection* $\Pi\colon J^1(M) \to T^*M$ in the sense that the only self intersections of $\Pi(L)$ are transverse double points. Note that there is a $1-1$ correspondence between double points of $\Pi(L)$ and segments in the $\mathbb{R}$-direction of $J^1(M) = T^*M \times \mathbb{R}$ which begin and end on L. Such segments are called *Reeb chords* since the vector field $\frac{\partial}{\partial z}$ is the *Reeb vector field* of the contact form $\alpha = dz - \theta_M$. We use this notion to conform with [**8**] and [**9**].

The contact homology algebra $\mathcal{A}(L)$ of L is the free unital algebra over $\mathbb{Z}_2$ generated by the set $\mathcal{C} = \{c_1, \dots, c_m\}$ of L. Thus elements of $\mathcal{A}(L)$ are polynomials in the c_j and the order of the factors of a monomial is important since multiplication is generally not commutative e.g. $c_i c_j \neq c_j c_i$ if $i \neq j$.

The differential of $\mathcal{A}(L)$ is defined by counting certain pseudo holomorphic disks in T^*M with boundary on $\Pi(L)$. We next define these objects. Let D_{m+1} be the unit disk in the complex plane with $m+1$ punctures $p_0, \dots, p_m$ on the boundary. Note that if the puncture p_0 is distinguished then the orientation of ∂D induces an ordering of the punctures $p_1, \dots, p_m$ (which we assume agrees with the order indicated in our notation). Note also that we can distinguish the two sheets of $\Pi(L)$ which intersect one of its double points by looking at their z-coordinates. We say that the sheet with larger z-coordinate is the upper sheet and other one the lower.

Definition 3.1. A J-holomorphic disk with boundary on L, positive puncture at the Reeb chord a and negative punctures at the Reeb chords $b_1, \dots, b_k$ is a map $u\colon D_{k+1} \to T^*M$ such that

- u is J-holomorphic, $du + J \circ du \circ i = 0$,
- $u(\partial D_{k+1}) \subset \Pi(L)$ and $u|\partial D_{k+1}$ has a continuous lift to $L \subset J^1(M)$.
- $\lim_{z\to p_0} u(z) = a$, the part of the boundary near p_0 oriented toward p_0 maps to the lower sheet of L at a, and the part oriented away from p_0 maps to the the upper sheet.
- $\lim_{z\to p_j} = b_j$, the part of the boundary near p_j oriented toward p_j maps to the upper sheet of $\Pi(L)$ at b_j, and the part oriented away from p_j maps to the lower sheet.

Note that by Stokes' theorem the area of a J-holomorphic map with boundary on L and positive puncture a and negative punctures $b_1, \dots, b_k$ is

$$\mathrm{Area}(u) = \int_{D_m} u^*\omega = \int_{\partial D_m} u^* dz = \delta z(a) - \sum \delta z(b_j),$$

where $\delta z(c)$ is the length of the Reeb chord c.

We define $\mathcal{M}(a; b_1, \dots, b_k)$ to be the moduli space of J-holomorphic maps with boundary on L. The following theorem is proved in [**8**].

LEMMA 3.2. *For generic J, $\mathcal{M}(a; b_1, \dots, b_k)$ is a finite collection of finite dimensional manifolds with natural compactifications. In particular, the sub-collection of 0-dimensional manifolds is a finite set of points.*

DEFINITION 3.3. The differential $\partial\colon \mathcal{A}(L) \to \mathcal{A}(L)$ of the contact homology algebra is linear over $\mathbb{Z}_2$, satisfies the Leibniz rule

$$\partial(\alpha\beta) = \partial(\alpha)\beta + \alpha(\partial\beta),$$

where α and β are monomials in the generators, and for generators a it is defined as

$$\partial(a) = \sum_{\dim(\mathcal{M}(a;b_1,\dots,b_k))+0} |\mathcal{M}(a; b_1, \dots, b_k)| b_1 \dots b_k,$$

where $|A|$ denotes the mod 2 number of elements of the finite set $|A|$.

The following theorem is proved in [**8**], [**9**] and the proof is similar to the corresponding proofs in Floer homology.

THEOREM 3.4. *With the notation above:*

(1) *The map ∂ is a well defined differential (i.e., $\partial^2 = 0$).*
(2) *The stable tame isomorphism class of $(\mathcal{A}, \partial)$ is an invariant of L.*
(3) *The homology of $(\mathcal{A}, \partial)$ is an invariant of L.*

3.2. Linearized contact homology. Let $L \subset J^1(M)$ be a Legendrian submanifold and let $\mathcal{A}(L)$ be its contact homology algebra with differential ∂. Write $\mathcal{A}(L) = \bigoplus_{j \geq 0} \mathcal{A}_j(L)$, where $\mathcal{A}_j(L)$ denotes the set of all homogeneous polynomials of degree j. Let $\pi_j\colon \mathcal{A}(L) \to \mathcal{A}_j(L)$ be the corresponding projections. Building of ideas of Chekanov [**3**], we say that $\mathcal{A}(L)$ is augmented if the differential of no generator contains a constant. In other words if

$$\partial\left(\bigoplus_{j>0} \mathcal{A}_j(L)\right) \subset \bigoplus_{j>0} \mathcal{A}_j(L).$$

An *augmentation* of an algebra is a map $\epsilon : \mathcal{A}(L) \to \mathbb{Z}_2$ such that $\epsilon(1) = 1$ and $\epsilon \circ \partial = 0$. Given an augmentation ϵ the graded algebra tame isomorphism $\phi_\epsilon(a) = a + \epsilon(a)$ conjugates $(\mathcal{A}(L), \partial)$ to an augmented algebra $(\mathcal{A}(L), \partial^\epsilon)$. A DGA is called *good* if it admits an augmentation, and is hence tame isomorphic to an augmented DGA. If $\mathcal{A}(L)$ is good then we define the *linearized contact homology* of L as the set of vector spaces over $\mathbb{Z}_2$ which arises as

$$\mathrm{Ker}(\partial_1^\epsilon)/\,\mathrm{Im}(\partial_1^\epsilon),$$

where

$$\partial_1^\epsilon\colon \bigoplus_{j>0} \mathcal{A}_j(L) \Big/ \bigoplus_{j>1} \mathcal{A}_j(L) \quad \to \quad \bigoplus_{j>0} \mathcal{A}_j(L) \Big/ \bigoplus_{j>1} \mathcal{A}_j(L)$$

is the map induced by ∂^ϵ, and where ϵ ranges over the finite set of augmentations.

3.3. Double point estimates of exact Lagrangian immersions. Let M be a smooth manifold and let $f\colon L \to T^*(M \times \mathbb{R})$ be an exact Lagrangian immersion. Then after small perturbation we may assume that the Legendrian lift $\tilde{f}$ of L is an embedding which is chord generic.

THEOREM 3.5. *Let $f\colon L \to T^*(M\times\mathbb{R})$ be an exact Lagrangian immersion and let $(\mathcal{A},\partial)$ be the DGA associated to an embedded chord generic Legendrian lift $\tilde f$ of f. If $(\mathcal{A},\partial)$ is good then f has at least*

$$\frac{1}{2}\dim(H_*(L;\mathbb{Z}_2))$$

double points.

PROOF. To simplify notation we identify L with its image under $\tilde f$ and write $L\subset J^1(M\times\mathbb{R}) = T^*M\times\mathbb{R}$. Let L' be a copy of L shifted a large distance in the z-direction, where as usual z is a coordinate in the $\mathbb{R}$-factor. Then $L\cup L'$ is a Legendrian link. Moreover, assuming that the shifting distance in the z-direction is sufficiently large, shifting L' s units in the x-direction, where x is a coordinate in the $\mathbb{R}$-factor of $M\times\mathbb{R}$, gives a Legendrian isotopy of $L\cup L'_s$. After a large such shift $L\cup L'_s$ projects to two distant copies of $\Pi(L)$ and it is evident that an augmentation for L gives an augmentation for $L\cup L'_s$. Moreover, the linearized contact homology of $L\cup L'$ equals the set of sums of two vector spaces from the linearized contact homology of L.

We will next compute the linearized contact homology of $L\cup L'$ in a different manner. Let $g\colon L\to\mathbb{R}$ be a Morse function on L and use g to perturb L in $U\subset J^1(L)$, where U is a small neighborhood of the 0-section. After identification of U with a neighborhood of L' in $J^1(M)$ (which exists by a theorem of Weinstein [**16**]) we use this isotopy to move L' to L''. The projection of L'' into $T^*(M\times\mathbb{R})$ then agrees (locally) with an exact deformation of L in its cotangent bundle and there is a symplectic map from the cotangent bundle T^*L to a neighborhood of $\Pi(L)$ in $T^*(M\times\mathbb{R})$. Pulling back the complex structure from $T^*(M\times\mathbb{R})$ we get an almost complex structure on T^*L. The intersection points of L and L'' are of three types.

(1) Critical points of g.
(2) Pairs of intersection points between L and L'' near the self-intersections of L.
(3) Self intersection points of L and of L'' near self intersections of L.

Fix augmentations of $\mathcal{A}(L)$ and of $\mathcal{A}(L'')$. If ∂ is the differential of $\mathcal{A}(L\cup L'')$ it is easy to see that any monomial in ∂c, where c is a Reeb chord of type (1) or (2), must contain an odd number of Reeb chords of type (1) and (2). Therefore the augmentations of $\mathcal{A}(L)$ and $\mathcal{A}(L'')$ give an augmentation for $\mathcal{A}(L\cup L'')$ that is trivial on double points of type (1) and (2). Denote by d the linearized differential induced by the augmentations chosen and by E_i the span of the double points of type (i), $i=1,2,3$. Suppose a is a type (3) double point, then ∂a has no constant part and its linear part has no double points of type (1) or (2), since each holomorphic disk with a positive puncture at a must have an even number of negative punctures of type (1) or (2). Thus $d(E_3)\subset E_3$. If b is of type (1) or (2) then the linear part of ∂b involves only double points of type (1) and (2). Denote by π_i the projection onto E_i, $i=1,2$, and $d_i=\pi_i\circ d$. Then $d=d_1+d_2$ on $E_1\oplus E_2$. Consider $d_1\colon E_1\to E_1$. We claim that for a sufficiently small perturbation g, $d_1\circ d_1|E_1=0$. To show this we consider gluing of two (two-punctured) disks contributing to d_1. This gives a 1-parameter family of two-punctured disks. Now, for a sufficiently small perturbation, no Reeb chord of type (2) has length lying between the lengths of two Reeb chords of type (1). Moreover, every Reeb chord of type (3) has length bigger than the difference of the lengths of two Reeb chords of type (1). This shows that the 1-parameter family must end at another pair of broken disks with corners of type (1). It follows that $d_1^2 = 0$. It follows that $d_1|E_1$ agrees with the Floer differential of $\hat{L} \cup \hat{L}_g$, where $\hat{L} \subset T^*L$ is the 0-section and where $\hat{L}_g \subset T^*L$ is the graph of dg. Hence,

$$\operatorname{Ker}(d_1|E_1)/\operatorname{Im}(d_1|E_1) \approx H_*(L;\mathbb{Z}_2).$$

Write $E_1 = W \oplus V$, where $W = \operatorname{Ker} d_1|E_1$ and let W' be a direct complement of $d_1(V) \subset W$. Then $\dim W' = \dim H_*(L;\mathbb{Z}_2)$. Fix the augmentations for L and L' which gives the element of the linearized contact homology of L which has the largest dimension. By the above discussion we find that $\operatorname{Ker}(d_3)/\operatorname{Im}(d_3)$ equals a direct sum of two copies of this maximal dimension vector space. It follows that the contribution to the linearized contact homology involving double points between L'' and L must vanish. We check how double points of type (2) kill off the double points of type (1) that exist in the homology of (E_1, d_1). We compute

$$\begin{aligned} 0 &= d(d(W')) = \pi_1(d(d(W'))) \\ &= \pi_1(d(d_2(W'))) = d_1(d_2(W')), \end{aligned}$$

where the third equality is due to the fact that $W' \subset E_1$ is in $\operatorname{Ker}(d_1|E_1)$. It follows that $\operatorname{Im}(d_2|W') \subset \operatorname{Ker}(d_1|E_2)$. Moreover, notice that an element e in W' is a nonzero element in the linearized contact homology if and only if $d_2 e = 0$ and $e \notin \operatorname{Im}(d_1|E_2)$. Thus if $d_2 e = 0$ then e is in $\operatorname{Im}(d_1|E_2)$, showing that $\operatorname{Ker} d_2|W' \subset \operatorname{Im} d_1|E_2$. We find

$$\begin{aligned} \dim(E_2) = \dim(\operatorname{Ker} d_1|E_2) + \dim(\operatorname{Im} d_1|E_2) \geq \\ \dim(\operatorname{Im} d_2|W') + \dim(\operatorname{Ker} d_2|W') = \dim(W'), \end{aligned}$$

and conclude that

$$2 \cdot \sharp\{\text{double points}\} = \dim(E_2) \geq \dim(W') = \dim(H_*(L;\mathbb{Z}_2)).$$

□

3.4. Improving double point estimates. In this section we show that to prove Conjecture 1.2 it is sufficient to prove a seemingly weaker estimate using a certain stabilization procedure which we discuss first.

Lemma 3.6. *Let $f\colon L \to \mathbb{C}^n \times \mathbb{R}$ be a chord generic Legendrian embedding with $R(f)$ Reeb chords. Then, for any $k \geq 1$ there exists a Legendrian embedding $F_k\colon L \times S^k \to \mathbb{C}^{n+k} \times \mathbb{R}$ with $2R(f)$ Reeb chords.*

Proof. For $q \in L$, let $f(q) = (x(q), y(q), z(q))$. Note that translations in the x_j-direction, $j = 1, \dots, n$ and that the scalings $x \mapsto kx$, $z \mapsto kz$, $k \geq 0$ are Legendrian isotopies which preserve the number of Reeb chords. We may thus assume that $f(L)$ is contained in $\{(x,y,z)\colon |x| \leq \epsilon\}$, where ϵ is very small. For convenience, we write $\mathbb{R}^n = \mathbb{R} \times \mathbb{R}^{n-1}$ with coordinates $x = (x_0, x_1)$ and corresponding coordinates (x_0, y_0, x_1, y_1) in $T^*\mathbb{R}^n = \mathbb{C}^n$. Consider the embedding $S^k \subset \mathbb{R}^{k+1} \subset \mathbb{R}^{k+n}$, where S^k is the unit sphere in $\mathbb{R}^{k+1}$. Let $(\sigma, x_0, x_1) \in S^k \times \mathbb{R}_+ \times \mathbb{R}^{n-1}$

$$(\sigma, x_0, x_1) \mapsto x_0 \cdot \sigma + x_1,$$

be polar coordinates on $\mathbb{R}^{k+n}$. Fix a Morse function ϕ, with one maximum and one minimum on S^k which is an approximation of the constant function with value 1.

Define $F\colon S^k \times L \to \mathbb{C}^{n+k} \times \mathbb{R}$, $F = (F_x, F_y, F_z)$ as follows:

$$\begin{aligned}
F_x(\sigma, q) &= (1 + x_0(q)) \cdot \sigma + x_1(q), \\
F_y(\sigma, q) &= (1 + x_0(q))^{-1} \nabla_{S^k} \phi(\sigma) + \phi(\sigma)(y_0(q) \cdot \sigma + y_1(q)), \\
F_z(\sigma, q) &= \phi(\sigma) z(q),
\end{aligned}$$

where we think of the gradient $\nabla_{S^k}\phi(\sigma)$ as a vector in $\mathbb{R}^{k+1}$ tangent to S^k at σ. It is then easily verified that F is a Legendrian embedding. Moreover, the Reeb chords of F occur between points (q, σ) and (q', σ') such that $\sigma = \sigma'$, $x_j(q) = x_j(q')$, $y_j(q) = y_j(q')$, $j = 0, 1$, and either $z(q) = z(q')$ or $\nabla_{S^k}\phi(\sigma) = 0$. However, these conditions are incompatible with f being an embedding unless $\nabla_{S^k}\phi(\sigma) = 0$ and we conclude that the number of double points of F are as claimed. □

THEOREM 3.7. *If there exists a constant $K > 0$ such that any exact Lagrangian immersion $f\colon L \to \mathbb{C}^n$, has at least*

$$\frac{1}{2}\dim(H_*(L; \mathbb{Z}_2)) - K \tag{3.1}$$

double points, then (3.1) *holds also with $K = 0$. In other words, the weaker estimate* (3.1) *for all exact Lagrangian immersions implies that, in fact, any exact Lagrangian immersion has at least*

$$\frac{1}{2}\dim(H_*(L; \mathbb{Z}_2))$$

double points.

PROOF. Given K in the statement of the theorem, choose l so that $2^l > K$. For any immersed exact Lagrangian $f\colon L \to \mathbb{C}^n$ lift f to an embedded Legendrian in $\mathbb{C}^n \times \mathbb{R}$ and apply the construction in Lemma 3.6 l times. The Lagrangian projection of the resulting Legendrian gives a new exact Lagrangian immersion $F_k : L \times S^k \times \ldots \times S^k \to \mathbb{C}^{n+lk}$. Since Reeb chords correspond to double points F_k has 2^l time as many double points as f. We have

$$\begin{aligned}
2^l R(f) = R(F_k) &\geq \frac{1}{2}\dim(H_*(L \times S^k \times \ldots \times S^k; \mathbb{Z}_2)) - K \\
&= \frac{1}{2}(2^l \dim(H_*(L; \mathbb{Z}_2))) - K.
\end{aligned}$$

Thus

$$R(f) \geq \frac{1}{2}\dim(H_*(L; \mathbb{Z}_2)).$$

□

References

[1] M. Chaperon and E. Zehnder,*Quelques résultats globaux en géométrie symplectique*, South Rhone seminar on geometry, III (Lyon, 1983), 51–121, Travaux en Cours, Hermann, Paris, (1984).

[2] Y. Chekanov, *Critical points of quasifunctions, and generating families of Legendrian manifolds*, (Russian) Funktsional. Anal. i Prilozhen. 30 (1996), no. 2, 56–69, 96; translation in Funct. Anal. Appl. 30 (1996), no. 2, 118–128.

[3] Y. Chekanov, *Differential algebras of Legendrian links*, Invent. Math. **150** (2002), no. 3, 441–483.

[4] Y. Eliashberg, *Invariants in contact topology*, Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998). Doc. Math. 1998, Extra Vol. II, 327–338.

[5] Y. Eliashberg, A. Givental, and H. Hofer, *Introduction to symplectic field theory*, GAFA 2000 (Tel Aviv, 1999). Geom. Funct. Anal. 2000, Special Volume, Part II, 560–673.

[6] Y. Eliashberg and M. Gromov, *Lagrangian intersection theory: finite-dimensional approach*, Geometry of differential equations, 27–118, Amer. Math. Soc. Transl. Ser. 2, 186, Amer. Math. Soc., Providence, RI, 1998.

[7] T. Ekholm, J. Etnyre and M. Sullivan *Non-isotopic Legendrian Submanifolds in* $\mathbb{R}^{2n+1}$, math.SG/0210124, to appear in J. Differential Geom.

[8] T. Ekholm, J. Etnyre and M. Sullivan *The Contact Homology of Legendrian Submanifolds in* $\mathbb{R}^{2n+1}$,math.SG/0210124, to appear in J. Differential Geom.

[9] T. Ekholm, J. Etnyre and M. Sullivan *Orientations in Legendrian Contact Homology and Exact Lagrangian Immersions*, Internat. J. Math. **16** (2005), no. 5, 453–532

[10] T. Ekholm, J. Etnyre and M. Sullivan *Legendrian contact homology in* $P \times \mathbb{R}$, math.SG/0505451.

[11] A. Floer, *The unregularized gradient flow of the symplectic action*, Comm. Pure Appl. Math. **41** (1988), no. 6, 775–813.

[12] A. Floer, *Morse theory for Lagrangian intersections*, J. Differential Geom. **28** (1988) 513-547.

[13] A. Floer, *Wittens complex and infinite dimansional Morse theory*, J. Differential Geom. **30** (1989) 207-221.

[14] M. Gromov, *Partial differential relations*, Springer-Verlag (1986).

[15] M. Gromov, *Pseudo-holomorphic curves in symplectic manifolds*, Invent. Math. **82** (1985) 307-347.

[16] A. Weinstein, *Contact surgery and symplectic handlebodies*, Hokkiado Math. Journal **20** (1991), 241–251.

Department of Mathematics, University of Southern California, Los Angeles, CA 90089-1113

E-mail address: `tekholm@usc.edu`

Symplectic 4–manifolds and Seiberg–Witten Invariants

Clay Mathematics Proceedings
Volume 5, 2006

Knot Surgery Revisited

Ronald Fintushel

ABSTRACT. We give an introduction to the topology of smooth 4-manifolds by studying three different proofs of the "knot surgery theorem".

Introduction

This survey is comprised of lectures given at the 2004 Clay Mathematics Institute Summer School in Budapest. My task was to give a general introduction to 4-manifolds in five lectures. (A paraphrasing of this might have been a more clever title for this article.) Since the stated goal seemed to me to be impossible, I instead tried to concentrate on one theorem — relating the Seiberg-Witten invariant of the result of knot surgery to the Alexander polynomial. This theorem has had several proofs from different points of view, and I thought that talking about them would give a nice overview of some of the techniques used in 4-manifold theory.

This article begins with a section which gives a 'user's guide' to Seiberg-Witten theory, concentrating on gluing theorems. Section 2 describes knot surgery and some simple applications. It then outlines the proof due to Ron Stern and myself of the knot surgery theorem: that knot surgery with a knot K has the effect of multiplying the Seiberg-Witten invariant by the Alexander polynomial of K. This proof is based on the relationship of the 'macareña' technique for calculating the Alexander polynomial with surgery formulas for the Seiberg-Witten invariant.

The knot surgery theorem is closely related to the Meng-Taubes Theorem, which relates the Seiberg-Witten invariant of a 3-manifold to its Milnor torsion. This theorem and its relationship to knot surgery is discussed in Section 3, where we give an introduction to the beautiful paper [**D**] of Simon Donaldson. Donaldson's proof relates the Seiberg-Witten invariant of a 3-manifold Y which has the homology of $S^2 \times S^1$ to the abelian vortex equations on a Riemann surface using ideas from topological quantum field theory. Our notes cover the case where Y is fibered over the circle. (There is also a nice exposition of this in unpublished notes of Ivan Smith.)

In Section 4, we have given a short introduction to the Taubes-Gromov theory approach to calculating Seiberg-Witten invariants of symplectic 4-manifolds. After some general comments concerning the definition of Gromov invariants and Taubes'

2000 *Mathematics Subject Classification.* 14J26, 53D05, 57R55, 57R57.
The author was partially supported by NSF Grant DMS0305818.

theorem on the their equivalence to Seiberg-Witten invariants, we discuss a proof of the Meng-Taubes formula from this point of view following one given by Taubes in [**T3**].

In the final section we discuss joint work with Ron Stern which applies knot surgery to the problem of constructing exotic embedded surfaces in 4-manifolds. The two techniques which are covered are 'rim surgery' which allows the exotic reimbedding of smooth surfaces in a fixed topological type (X, Σ), and braiding, which allows the construction of exotic symplectic tori in a fixed homology class.

I hope that no one will misconstrue this survey as being definitive in any sense. One can always learn more by going back to the papers that I have cited. If these notes or my lectures have convinced anyone to do that, they will have more than served their purpose.

These notes were helped by conversations with many people. I would like to thank Tom Mark, Doug Park, Slaven Jabuka, Elly Ionel, Tom Parker, Michael Usher, Olga Buse, and Ron Stern. Thanks also go to Jongil Park for his sterling work on notes for my lectures, and to the participants of the Summer School for their interest (and patience). Finally, I'd like express my deep gratitude to Peter Ozsváth, András Stipsicz, and Zoltan Szabó for making the Clay Institute Summer School so much fun.

1. Seiberg-Witten Invariants and Gluing

The main tool used to understand smooth structures on 4-manifolds is the Seiberg-Witten invariant. The goal of this lecture is to provide a 'user's guide' to these invariants. For more detailed explanations one should see [**W, KM, M, N**].

Consider a smooth compact oriented 4-manifold X with tangent bundle TX. The choice of a Riemannian metric on X reduces the structure group of TX to SO(4), which may be equivalently taken as the structure group of PX, the bundle of tangent frames of X. The double covering group of SO(4) is $\mathrm{Spin}(4) \cong \mathrm{SU}(2) \times \mathrm{SU}(2)$, and a spin-structure on X is a lift of PX to a principal Spin(4)-bundle $\tilde{P}X$ over X such that in the diagram

$$\begin{tikzcd} \tilde{P}X \arrow[r] \arrow[dr] & PX \arrow[d] \\ & X \end{tikzcd}$$

the horizontal map is a double cover on each fiber of PX.

A spin structure gives rise to spinor bundles $S^{\pm} = \tilde{P}X \times_{\mathrm{SU}(2)} \mathbf{C}^2$, where the action of SU(2) on $\tilde{P}X$ arises from one of the two factors of $\mathrm{Spin}(4) \cong \mathrm{SU}(2)\times\mathrm{SU}(2)$. From the point of view of algebraic topology, one can think of a spin structure on X as a lift

$$\begin{tikzcd} & B\mathrm{Spin}(4) \arrow[d] \\ X \arrow[r] \arrow[ur, dotted] & B\mathrm{SO}(4) \end{tikzcd}$$

The obstruction to finding such a lift is the second Stiefel-Whitney class $w_2(X) \in H_2(X; \mathbf{Z}_2)$. One may alternatively think in terms of the transition functions

$$\{\varphi_{i,j} : U_i \cap U_j \to \mathrm{SO}(4)\}$$

of PX. A spin structure on X consists of lifts $\tilde{\varphi}_{i,j} : U_i \cap U_j \to \mathrm{Spin}(4)$. In order to give a bundle $\tilde{P}X$, these lifts must satisfy the cocycle condition $\tilde{\varphi}_{i,j} \circ \tilde{\varphi}_{j,k} = \tilde{\varphi}_{i,k}$. From this point of view, $\tilde{P}X$ corresponds to an element $\tilde{\xi}$ of the Čech cohomology group $H^1(X;\mathrm{Spin}(4))$ such that in the sequence

$$\cdots \to H^1(X;\mathbf{Z}_2) \xrightarrow{i_*} H^1(X;\mathrm{Spin}(4)) \xrightarrow{p_*} H^1(X;\mathrm{SO}(4)) \xrightarrow{\delta} H^2(X;\mathbf{Z}_2) \to \dots$$

$p_*\tilde{\xi} = \xi$, the element which corresponds to PX. Note that $\delta\xi = w_2(X)$, affirming our comment above. Also note that if X admits a spin structure (i.e. a lift of ξ), such lifts are in 1-1 correspondence with $H^1(X;\mathbf{Z}_2)$. To each spin structure there is associated a Dirac operator $D : \Gamma(S^+) \to \Gamma(S^-)$, an elliptic operator which plays an important role in topology and geometry.

In case $w_2(X) \neq 0$, X admits no spin structure, but it can still admit a spin^c -structure. A spin^c structure is given by a pair of rank 2 complex vector bundles $W^\pm$ over X with isomorphisms $\det(W^+) = \det(W^-) = L$, some complex line bundle over X, so that locally $W^\pm = S^\pm \otimes L^{\frac{1}{2}}$. To make sense of this, consider the transition maps $\{\varphi_{i,j} : U_i \cap U_J \to \mathrm{SO}(4)\}$ for PX. We can assume that our charts have overlaps $U_i \cap U_J$ which are contractible, so that we can always get lifts $\tilde{\varphi}_{i,j} : U_i \cap U_J \to \mathrm{Spin}(4)$. However, if $w_2(X) \neq 0$, we can never find lifts satisfying the cocycle condition.

Similarly, suppose that we are given a complex line bundle L with transition functions $\{g_{i,j} : U_i \cap U_j \to \mathrm{U}(1)\}$. Locally these functions have square roots $(g_{i,j})^{\frac{1}{2}}$. The obstruction to finding a system of square roots which satisfy the cocycle condition, i.e. to finding a global bundle $L^{\frac{1}{2}}$ over X such that $L^{\frac{1}{2}} \otimes L^{\frac{1}{2}} \cong L$ is $c_1(L)$ (mod 2) in $H^2(X;\mathbf{Z}_2)$. Now suppose that L is characteristic, i.e. that $w_2(X) = c_1(L)$ (mod 2). The statement that $W^\pm$ should locally be $S^\pm \otimes L^{\frac{1}{2}}$ means that the tensor products $\tilde{\varphi}_{i,j} \otimes (g_{i,j})^{\frac{1}{2}}$ should satisfy the cocycle condition. This function has values in $(\mathrm{U}(1)\times\mathrm{SU}(2)\times\mathrm{SU}(2))/\{\pm 1\} = \mathrm{Spin}^c(4)$, and the corresponding obstruction is $2w_2(X) = 0$; so spin^c structures exist provided we can find characteristic line bundles L over X. A theorem of Hirzebruch and Hopf states that these exist on any oriented 4-manifold [**HH**]. Spin^c structures on X are classified by lifts of $w_2(X)$ to $H^2(X;\mathbf{Z})$ up to the action of $H^1(X;\mathbf{Z}_2)$. (Spin structures correspond to $0 \in H^2(X,\mathbf{Z})$ up to this action.)

The group $\mathrm{Spin}^c(4) \cong (\mathrm{U}(1) \times \mathrm{SU}(2) \times \mathrm{SU}(2))/\{\pm 1\}$ fibers over $SO(4) \cong (\mathrm{SU}(2) \times \mathrm{SU}(2))/\{\pm 1\}$ with fiber $S^1 \cong \mathrm{U}(1)$. A spin^c structure s on X is a lift of PX to a principal $\mathrm{Spin}^c(4)$ bundle $\hat{P}_X$ over X. Since $\mathrm{U}(2) \cong (\mathrm{U}(1)\times\mathrm{SU}(2))/\{\pm 1\}$, we get representations $s^\pm : \mathrm{Spin}^c(4) \to \mathrm{U}(2)$, and associated rank 2 complex vector bundles

$$W^\pm = \hat{P}_X \times_{s^\pm} \mathbf{C}^2$$

called spinor bundles, and referred to above, and $L = \det(W^\pm)$. We sometimes write $c_1(s)$ for $c_1(L)$.

As for ordinary spin structures, one has Clifford multiplication

$$c : T^*X \otimes W^\pm \to W^\mp$$

written $c(v,w) = v.w$ and satisfying $v.(v.w) = -|v|^2 w$. Thus c induces a map

$$c : T^*X \to \mathrm{Hom}(W^+, W^-)$$

A connection A on L together with the Levi-Civita connection on the tangent bundle of X forms a connection $\nabla_A : \Gamma(W^+) \to \Gamma(T^*X \otimes W^+)$ on W^+. This connection, followed by Clifford multiplication, induces the Dirac operator

$$D_A : \Gamma(W^+) \to \Gamma(W^-)$$

Thus D_A depends both on the connection A and the Riemannian metric on X. The case where $L = \det(W^+)$ is trivial corresponds to a usual spin structure on X, and in this case we may choose A to be the trivial connection and then $D_A = D : \Gamma(S^+) \to \Gamma(S^-)$, the usual Dirac operator.

Fix a spin^c structure s on X with determinant line bundle L, and let $\mathcal{A}_L$ denote the affine space of connections on the line bundle L. Let $F_A \in \Omega^2(X)$ denote the curvature of a connection A on L. The Hodge star operator acts as an involution on $\Omega^2(X)$. Its ± 1 eigenspaces are $\Omega^2_\pm(X)$, the spaces of self-dual and anti-self-dual 2-forms. We have $F_A = F_A^+ + F_A^-$. The bundle of self-dual 2-forms $\Omega^2_+(X)$ is also associated to $\hat{P}X$ by $\Omega^2_+(X) \cong \hat{P}X \times_{\mathrm{SU}(2)} \mathfrak{su}(2)$ where SU(2) acts on its Lie algebra $\mathfrak{su}(2) \cong \mathbf{C} \oplus \mathbf{R}$ via the adjoint action. The map

$$\mathbf{C} \oplus \mathbf{C} \to \mathbf{C} \oplus \mathbf{R} \qquad (z, w) \to (z\bar{w}, |z|^2 - |w|^2)$$

is SU(2)-equivariant, and so it induces a map

$$q : \Gamma(W^+) \to \Omega^2_+(X)$$

Given a pair $(A, \psi) \in \mathcal{A}_X(L) \times \Gamma(W^+)$, i.e. A a connection in $L = \det(W^\pm)$ and ψ a section of W^+, the Seiberg-Witten equations [**W**] are:

$$D_A\psi = 0$$
$$F_A^+ = iq(\psi)$$

The gauge group $\mathrm{Aut}(L) = \mathrm{Map}(X, S^1)$ acts on the space of solutions to these equations via

$$g \cdot (A, \psi) = (A - g^{-1}dg, g\psi)$$

and its orbit space is the *Seiberg-Witten moduli space* $M_X(s)$.

Some important features of the Seiberg-Witten equations are

(1) If (A, ψ) is a solution to the Seiberg-Witten equations with $\psi \neq 0$ then its stabilizer in $\mathrm{Aut}(L)$ is trivial. Such solutions are called *irreducible*. The stabilizer of a *reducible* solution $(A, 0)$ consists of the constant maps in $\mathrm{Map}(X, S^1)$. This is a copy of S^1.

(2) $(A, 0)$ is a reducible solution if and only if A is an anti-self-dual connection on the complex line bundle L (i.e. if its curvature $F_A = F_A^-$, is anti-self-dual). If $b_X^+ > 0$ and $c_1(L)$ is nontorsion, a generic metric on X admits no such connections.

(3) The formal dimension of the Seiberg-Witten moduli space is calculated by the Atiyah-Singer theorem to be

$$\dim M_X(s) = \frac{1}{4}(c_1(L)^2 - (3\,\mathrm{sign}(X) + 2\,\mathrm{e}(X))$$

where $\mathrm{e}(X)$ is the Euler number of X and $\mathrm{sign}(X)$ is its signature. Especially interesting is the case where $\dim M_X(s) = 0$, since this is precisely the condition for X to admit an almost-complex structure with first Chern class equal to $c_1(L)$.

(4) An anti-self-dual 2-form η on X gives us a perturbation of the Seiberg-Witten equations:

$$D_A\psi = 0$$
$$F_A^+ = iq(\psi) + i\eta,$$

and for a generic perturbation η, the corresponding moduli space of solutions $M_X(s,\eta)$ is an orientable manifold whose dimension is $\dim M_X(s)$, provided $M_X(s,\eta)$ contains at least one irreducible solution. (As in (2), if $b^+(X) > 0$ and $c_1(L) \neq 0$, all solutions will be irreducible for a generic choice of metric or perturbation η.) For simplicity we let the notation ignore this perturbation and write $M_X(s)$ for $M_X(s,\eta)$. An orientation is given to $M_X(s)$ by fixing a 'homology orientation' for X, that is, an orientation of $H^1(X) \oplus H^2_+(X)$.

(5) There is a Lichnerowicz-type theorem, proved, as usual, with an application of the Weitzenböck formula [**W, KM**]: If X carries a metric of positive scalar curvature, then the only solutions of the Seiberg-Witten equations are reducible (of the form $(A,0)$). Hence, if $b_X^+ > 0$, for a generic metric of positive scalar curvature, $M_X(s) = \emptyset$.

(6) For each s, the Seiberg-Witten moduli space $M_X(s)$ is compact.

(7) There are only finitely many characteristic line bundles L on X for which both $M_X(s) \neq \emptyset$ and $\dim M_X(s) \geq 0$.

Items (6) and (7) are also proved by using the Weitzenböck formula [**W, KM**].

In case $\dim M_X(s) = 0$, items (4) and (6) imply that that generically, $M_X(s)$ is a finite set of signed points (once a homology orientation has been chosen). In this case one defines the *Seiberg-Witten invariant* $\mathrm{SW}_X(s)$ to be the signed count of these points. Generally, $(\mathcal{A}_L \times \Gamma(W^+))/\mathrm{Aut}(L)$ is homotopy equivalent to $\mathbf{CP}^\infty \times \mathbf{T}^{b_1(X)}$, and its homology can be utilized to define $\mathrm{SW}_X(s)$. The Seiberg-Witten invariant is a diffeomorphism invariant provided $b^+ > 1$.

In case $b_X^+ = 1$, one still gets invariants, but there are some complications. For simplicity, we consider the case where X is simply connected. So suppose that X is a simply connected oriented 4-manifold with $b_X^+ = 1$ with a given orientation of $H^2_+(X;\mathbf{R})$. The Seiberg-Witten invariant depends on the metric g and a self-dual 2-form η as follows. There is a unique g-self-dual harmonic 2-form $\omega_g \in H^2_+(X;\mathbf{R})$ with $\omega_g^2 = 1$ and corresponding to the positive orientation. Fix a spin^c structure s on X with determinant line bundle L. Given a pair (A,ψ), where A is a connection in L and ψ a section of the bundle W^+, we have the perturbed Seiberg-Witten equations as given above for each perturbation (self-dual) 2-form η. Write $\mathrm{SW}_{X,g,\eta}(s)$ for the count of solutions with signs. As the pair (g,η) varies, $\mathrm{SW}_{X,g,\eta}(s)$ can change only at those pairs (g,η) for which there are solutions with $\psi = 0$. These solutions occur for pairs (g,η) satisfying $(2\pi c_1(L) + \eta) \cdot \omega_g = 0$. This last equation defines a wall in $H^2(X;\mathbf{R})$.

The point ω_g determines a component of the double cone consisting of elements of $H^2(X;\mathbf{R})$ which have positive square. If we have $(2\pi c_1(L) + \eta) \cdot \omega_g \neq 0$ for a generic η, then $\mathrm{SW}_{X,g,\eta}(s)$ is well-defined, and its value depends only on the sign of $(2\pi c_1(L) + \eta) \cdot \omega_g$. This means that given a simply connected oriented 4-manifold X with $b^+ = 1$ and with a given orientation of $H^2_+(X;\mathbf{R})$, there are two well-defined Seiberg-Witten invariants $\mathrm{SW}_X^\pm$ defined by: $\mathrm{SW}_X^+(s) = \mathrm{SW}_{X,g,\eta}(s)$ for any (g,η)

such that $(2\pi c_1(L)+\eta)\cdot\omega_g > 0$, and $\mathrm{SW}^-_X(s) = \mathrm{SW}_{X,g,\eta}(s)$ for any (g,η) such that $(2\pi c_1(L)+\eta)\cdot\omega_g < 0$.

One of the most important consequences of the Seiberg-Witten equations is the Adjunction Inequality.

THE ADJUNCTION INEQUALITY. [**KM**] *Suppose $b^+(X) > 1$ and $SW_X(s) \neq 0$. Let β be the Poincaré dual of $c_1(s)$. If Σ is an embedded closed surface in X with self-intersection ≥ 0 and genus $g \geq 1$ then $2g-2 \geq \Sigma\cdot\Sigma + |\beta\cdot\Sigma|$.*

A Kähler surface is a complex surface with a metric g such that $g(Jx,y) = \omega(x,y)$ is a symplectic form. Each simply connected complex surface admits a Kähler structure. A Kähler surface has a distinguished spin^c structure s_K with $c_1(s_K) = K_X$, the canonical class of X. ($K_X = -c_1(TX)$.)

THEOREM 1. [**W**] *If X is a minimal Kähler surface with $b^+(X) > 1$ then for its canonical class $|SW_X(\pm s_K)| = 1$. Furthermore, if $c_1^2(X) > 0$ then $SW_X(s) = 0$ for all other spin^c structures.*

('Minimal' means that X contains no embedded holomorphic 2-spheres with self-intersection equal to -1.)

Another important basic fact is that $\mathrm{SW}_X(-s) = (-1)^{(e+\mathrm{sign})/4}\,\mathrm{SW}_X(s)$. It is convenient to view the Seiberg-Witten invariant as an element of the integral group ring $\mathbf{Z}H_2(X;\mathbf{Z})$, where for each $\alpha \in H_2(X;\mathbf{Z})$ we let t_α denote the corresponding element in $\mathbf{Z}H_2(X;\mathbf{Z})$. (Note that $t_\alpha^{-1} = t_{-\alpha}$ and $t_0 = 1$.) We view the Seiberg-Witten invariant of X as the Laurent polynomial

$$\mathcal{SW}_X = \mathrm{SW}_X(0) + \sum \mathrm{SW}_X(\beta)\, t_\beta$$

where the sum is taken over all characteristic elements β of $H_2(X;\mathbf{Z})$ and where $\mathrm{SW}_X(\beta) = \sum\limits_{c_1(s)=\mathrm{PD}(\beta)} \mathrm{SW}_X(s)$. For example, for a minimal Kähler surface with $b^+ > 1$ and $c_1^2 > 0$ we have $\mathcal{SW}_X = t_K \pm t_K^{-1}$. The $K3$-surface is a Kähler surface with $b^+ = 3$ and $c_1 = 0$. Hence $\mathrm{SW}_{K3}(0) = 1$. Adjunction inequality arguments can be used to show that there are no other nontrivial Seiberg-Witten invariants of the $K3$-surface; so $\mathcal{SW}_{K3} = 1$.

Our goal will be to use Seiberg-Witten invariants to study constructions of 4-manifolds. The techniques will involve cutting and pasting along 3-tori. Seiberg-Witten invariants can be defined for 4-manifolds whose boundary is a disjoint union of 3-tori. In this case the invariant is an element of $\mathbf{Z}[[H_2(X;\mathbf{Z})]]$, the ring of formal power series. For example, $\mathcal{SW}_{T^2\times D^2} = (t_T - t_T^{-1})^{-1}$; There is an important gluing theorem due to Morgan, Mrowka, and Szabó [**MMS**], B.D. Park [**P**], and in its most general form to Taubes [**T3**]:

THEOREM 2 (Taubes). *Suppose that $\partial X_1 = \partial X_2$, and that $X = X_1 \cup_{T^3} X_2$ has $b^+ \geq 1$. Also suppose that there is a class $\varpi \in H^2(X;\mathbf{R})$ restricting nontrivially to $H^2(T^3;\mathbf{R})$. Let $j_i : X_i \to X$ be the inclusions. Then*

$$\mathcal{SW}_X = (j_1)_*\mathcal{SW}_{X_1}\cdot (j_2)_*\mathcal{SW}_{X_2}$$

When $b^+_X = 1$, one gets an orientation of $H^2_+(X;\mathbf{R})$ from ϖ: Since the restriction $i^*(\varpi) \in H^2(T^3;\mathbf{R})$ is nonzero, there is a nonzero class $v \in H_2(T^3;\mathbf{R})$ such that $\langle i^*(\varpi), v\rangle > 0$. Then the condition $\langle \alpha, i_*(v)\rangle > 0$ orients $H^2_+(X;\mathbf{R})$. Now it makes sense to speak of $\mathcal{SW}^\pm_X$, and in Taubes' theorem, one takes $\mathcal{SW}^-_X$.

As an example of the use of Taubes' theorem, let T be a homologically nontrivial torus of self-intersection 0 in X with a tubular neighborhood $N_T = T \times D^2$; then

$$\mathcal{SW}_X = \mathcal{SW}_{X \setminus N_T} \cdot \frac{1}{t_T - t_T^{-1}}$$

Furthermore, B.D. Park has proved that if $H^1(X \setminus N_T) \to H^1(\partial N_T)$ has a cokernel which has no torsion, then $\mathcal{SW}_{X \setminus N_T} = \mathcal{SW}_X \cdot (t_T - t_T^{-1})$.

We apply this gluing theorem to calculate the Seiberg-Witten invariants of the elliptic surfaces $E(n)$. These manifolds can be defined inductively as follows. $E(1) = \mathbf{CP}^2 \# 9\overline{\mathbf{CP}}^2$. It admits a holomorphic map to S^2 whose generic fiber is a self-intersection 0 torus, F. Then $E(n)$ is the fiber sum $E(n-1)\#_F E(1)$. This means that $E(n) = (E(n-1) \setminus N_F) \cup_{T^3} (E(1) \setminus N_F)$. In this case, each inclusion j_i is the identity. $E(2)$ is the $K3$-surface; so $\mathcal{SW}_{E(2)} = 1$. Hence

$$1 = \mathcal{SW}_{E(2)} = (\mathcal{SW}_{E(1) \setminus N_F})^2$$

so $\mathcal{SW}_{E(1) \setminus N_F} = 1$ (up to sign). Also $E(2) = (E(2) \setminus N_F) \cup N_F$. This means $1 = \mathcal{SW}_{E(2) \setminus N_F} \cdot \mathcal{SW}_{N_F} = \mathcal{SW}_{E(2) \setminus N_F} \cdot (t_F - t_F^{-1})^{-1}$; so $\mathcal{SW}_{E(2) \setminus N_F} = t_F - t_F^{-1}$.

We then get $\mathcal{SW}_{E(3)} = \mathcal{SW}_{E(2) \setminus N_F} \cdot \mathcal{SW}_{E(1) \setminus N_F} = (t_F - t_F^{-1})$. Inductively, we see that $\mathcal{SW}_{E(n)} = (t_F - t_F^{-1})^{n-2}$, provided $n > 1$. Since $b^+(E(1)) = 1$, one must be more careful in this case. See §3 for a related discussion.

Internal fiber sum follows the same ideas. If T_1 and T_2 are self-intersection 0 tori embedded in X and if the conditions of Taubes and Park are satisfied for $X \setminus (N_{T_1} \cup N_{T_2})$ then let X_{T_1,T_2} be the result of removing the interiors of N_{T_1} and N_{T_2} and gluing up the boundaries so that the boundaries of the normal disks of T_1 and T_2 are matched. Then

$$\mathcal{SW}_{X_{T_1,T_2}} = \mathcal{SW}_X \cdot (t - t^{-1})^2, \qquad t = t_{T_i}$$

Finally, we need a formula for the effect of surgery on the Seiberg-Witten invariants. Let T be a self-intersection 0 torus embedded in X with tubular neighborhood $N_T = T \times D^2$. Given a diffeomorphism $\varphi : \partial(T \times D^2) \to \partial(X \setminus N_T)$ form $X_\varphi = (X \setminus N_T) \cup_\varphi (T \times D^2)$. The manifold X_φ is determined by the homology class $\varphi_*[\partial D^2] \in H_1(\partial(X \setminus N_T); \mathbf{Z})$. Fix a basis $\{\alpha, \beta, [\partial D^2]\}$ for $H_1(\partial(X \setminus N_T); \mathbf{Z})$, then there are integers p, q, r, such that $\varphi_*[\partial D^2] = p\alpha + q\beta + r[\partial D^2]$. We write $X_\varphi = X_T(p,q,r)$. (With this notation, note that $X_T(0,0,1) = X$.) We have the following important formula of Morgan, Mrowka, and Szabó:

Theorem 3. [**MMS**] *Given a class $k \in H_2(X)$:*

$$\sum_i SW_{X_T(p,q,r)}(k_{(p,q,r)} + i[T]) = p \sum_i SW_{X_T(1,0,0)}(k_{(1,0,0)} + i[T]) + \\ + q \sum_i SW_{X_T(0,1,0)}(k_{(0,1,0)} + i[T]) + r \sum_i SW_{X_T(0,0,1)}(k_{(0,0,1)} + i[T])$$

and there are no other nontrivial Seiberg-Witten invariants of $X_T(p,q,r)$.

In this formula, T denotes the torus $T_{(a,b,c)}$ which is the core $T^2 \times 0 \subset T^2 \times D^2$ in each specific manifold $X_T(a,b,c)$ in the formula, and $k_{(a,b,c)} \in H_2(X_T(a,b,c))$ is any class which agrees with the restriction of k in $H_2(X \setminus T \times D^2, \partial)$ in the diagram:

$$
\begin{tikzcd}
H_2(X_T(a,b,c)) \arrow[r] & H_2(X_T(a,b,c), T\times D^2) \arrow[d, "\cong"] \\
 & H_2(X\setminus T\times D^2, \partial) \\
H_2(X) \arrow[r] & H_2(X, T\times D^2) \arrow[u, "\cong"']
\end{tikzcd}
$$

Furthermore, unless the homology class $[T]$ is 2-divisible, in each term, each i must be even since the classes $k_{(a,b,c)} + i[T]$ must be characteristic in $H_2(X_T(a,b,c))$.

Often this formula simplifies. For example, suppose that $\gamma = \varphi_*[\partial D^2]$ is indivisible in $H_1(X \setminus N_T)$. Then there is a dual class $A \in H_3(X \setminus N_T, \partial)$ such that $A \cdot \gamma = 1$. This means that ∂A generates $H_2(N_{T(p,q,r)})$; so

$$H_3(X_T(p,q,r), N_{T(p,q,r)}) \xrightarrow{\text{onto}} H_2(N_{T(p,q,r)}) \xrightarrow{0} H_2(X_T(p,q,r))$$

So $T(p,q,r)$ is nullhomologous in $X_T(p,q,r)$. Hence in the Morgan, Mrowka, Szabó formula, the left hand side has just one term.

A second condition which simplifies the formula uses the adjunction inequality. Suppose that there is an embedded torus Σ of self-intersection 0, such that $\Sigma \cdot T(1,q,r) = 1$ and $\Sigma \cdot k_{(1,q,r)} = 0$. Then the adjunction inequality implies that at most one of the classes $k_{(1,q,r)} + iT_{(1,q,r)}$ can have SW nonzero.

These gluing formulas can often be used to quickly calculate invariants. For example, a log transform of order r on an elliptic surface is a surgery of type $(1,q,r)$ where the basis above is chosen so that α and β project trivially under the map to S^2. (In the typical situation, for example $E(n)$, $n \geq 1$, where there is a fibration with a cusp fiber, the surgery is independent of the choice of q.) Suppose we perform such a log transform of order r on $E(n)$; then the result is $E(n;r) = (E(n) \setminus N_F) \cup_j (T^2 \times D^2)$. Let t be the class in the integral group ring corresponding to $T_r = T_{(1,q,r)} \in H_2(E(n;r))$. T_r is a multiple torus in the sense that in $H_2(E(n;r))$, the generic fiber is $F = r\,T_r$; so $j_*(t_F) = t^r$. Hence

$$\mathcal{SW}_{E(n;r)} = j_*(\mathcal{SW}_{E(n)\setminus N_F}) \cdot \frac{1}{t - t^{-1}} = \frac{(t^r - t^{-r})^{n-1}}{t - t^{-1}}$$

Each simply connected elliptic surface is the result of 0, 1, or 2 log transforms on $E(n)$, $n \geq 1$, of relatively prime orders. Thus we complete the calculation of Seiberg-Witten invariants of simply connected elliptic surfaces with $b^+ > 1$ by noting that a similar argument shows (for $n > 1$)

$$\mathcal{SW}_{E(n;r,s)} = \frac{(t^{rs} - t^{-rs})^n}{(t^r - t^{-r})(t^s - t^{-s})}$$

2. Knot Surgery

Let X be a 4-manifold containing an embedded torus T of self-intersection 0, and let K be a knot in S^3. *Knot surgery* on T is the result of replacing a tubular neighborhood $T \times D^2$ of T with S^1 times the exterior $S^3 \setminus N_K$ of the knot [**FS1**]:

$$X_K = \big(X \setminus (T \times D^2)\big) \cup \big(S^1 \times (S^3 \setminus N_K)\big)$$

where ∂D^2 is identified with a longitude of K. This description doesn't necessarily determine X_K up to diffeomorphism; however, under reasonable hypotheses, all manifolds obtained from the same (X, T) and $K \subset S^3$ will have the same Seiberg-Witten invariant. Knot surgery is a homological variant of surgery in the sense that surgery is the process that removes a $T^2 \times D^2$ and reglues it, whereas knot surgery removes a $T^2 \times D^2$ and replaces it with a homology $T^2 \times D^2$.

Here is an alternative description of knot surgery: Consider a knot K in S^3, and let m denote a meridional circle to K. Let M_K be the 3-manifold obtained by performing 0-framed surgery on K. The effect of such a surgery is to span a longitude of K with a disk. The meridian m can also be viewed as a circle in M_K. In $S^1 \times M_K$ we have the smooth torus $T_m = S^1 \times m$ of self-intersection 0. Let X_K denote the fiber sum

$$X_K = X \#_{T=T_m} (S^1 \times M_K) = \big(X \setminus (T \times D^2)\big) \cup \big((S^1 \times M_K) \setminus (T_m \times D^2)\big)$$

As above, the two pieces are glued together so as to preserve the homology class $[\mathrm{pt} \times \partial D^2]$. Because M_K has the homology of $S^2 \times S^1$ with the class of m generating H_1, the complement $(S^1 \times M_K) \setminus (T \times D^2)$ has the homology of $T^2 \times D^2$. Thus X_K has the same homology (and intersection pairing) as X.

Let us make the additional assumption that $\pi_1(X) = 1 = \pi_1(X \setminus T)$. Then, since the class of m normally generates $\pi_1(M_K)$; the fundamental group of $M_K \times S^1$ is normally generated by the image of $\pi_1(T)$, and it follows from Van Kampen's Theorem that X_K is simply connected. Thus X_K is homotopy equivalent to X. Also, in order to define Seiberg-Witten invariants, the oriented 4-manifold X must also be equipped with an orientation of $H^2_+(X; \mathbf{R})$. The manifold X_K inherits an orientation as well as an orientation of $H^2_+(X_K; \mathbf{R})$ from X.

For example, consider knot surgery on a fiber F of the elliptic surface $E(2)$ (the $K3$-surface). Recall that $\mathcal{SW}_{E(2)} = 1$. The elliptic fibration $E(2) \to S^2$ has a section S which is a sphere of square -2. The homology class $S + F$ is represented by a torus of square 0 which intersects a generic fiber once. Apply the adjunction inequality to this class to see that mF cannot have a nonzero Seiberg-Witten invariant unless $m = 0$. (Of course we already knew this, but the point is that it is the apparatus of the adjunction inequality that is forcing $\mathcal{SW}_{E(2)} = 1$.) Now do knot surgery on F with a knot K of genus g. In $E(2)_K$ we no longer have the section S; the normal disk D to F has been removed from S. In its place there is a surface S' of genus g formed from $S \setminus D$ together with a Seifert surface of the knot K. The class S' still has self-intersection -2, and $S' + F$ is a class represented by a genus $g+1$ surface of self-intersection 0. The fiber F still intersects $S' + F$ once. Apply the adjunction inequality to this class to test whether we can now have $\mathrm{SW}_{E(2)_K}(mF) \neq 0$:

$$2(g+1) - 2 \geq (S' + F) \cdot (S' + F) + |mF \cdot (S' + F)| = |m|$$

Thus mF has the possibility of having a nonzero Seiberg-Witten invariant if $|m| \leq 2g$ (and is even since $E(2)_K$ is spin). Thus performing knot surgery gives us the possibility of constructing 4-manifolds with interesting Seiberg-Witten invariants. In fact:

THEOREM 4. [**FS1**] *Suppose that $b^+(X) > 1$ and $\pi_1(X) = 1 = \pi_1(X \setminus T)$ and that T is a homologically essential torus of self-intersection 0. Then X_K is homeomorphic to X and*

$$\mathcal{SW}_{X_K} = \mathcal{SW}_X \cdot \Delta_K(t^2)$$

where $t = t_T$ and Δ_K is the symmetrized Alexander polynomial of K.

In particular, $\mathcal{SW}_{E(2)_K} = \Delta_K(t_F^2)$. It was shown by Seifert that any symmetric Laurent polynomial $p(t) = a_0 + \sum_{i=1}^{n} a_i(t^i + t^{-i})$ whose coefficient sum $p(1) = \pm 1$ is the Alexander polynomial of some knot in S^3. It follows that the family of smooth 4-manifolds homeomorphic to the $K3$-surface is at least as rich as this family of Alexander polynomials. Also note that since $\mathcal{SW}_{E(2)_K}(1) = \pm 1$ and $\mathcal{SW}_{E(2;r)}(1) = r$, these manifolds are not diffeomorphic to a log transform, or any number of log transforms, of $K3$.

Note that if $\bar{K}$ is the mirror image knot to K in S^3 then $S^1 \times (S^3 \setminus N_K) \cong S^1 \times (S^3 \setminus N_{\bar{K}})$ since we may view this construction as revolving the knot exterior about an axis. At 180° in $S^1 \times (S^3 \setminus N_K)$ we see $S^3 \setminus N_{\bar{K}}$. Thus $X_{\bar{K}} \cong X_K$. There are currently no other known examples of inequivalent knots which give diffeomorphic manifolds via knot surgery.

The rest of this section will be devoted to a presentation of the proof of the knot surgery theorem as given in [**FS1**]. This proof depends on the description of the Alexander polynomial of a knot in terms of the 'knot theory macareña':

$$\Delta_{K_+}(t) = \Delta_{K_-}(t) + (t^{1/2} - t^{-1/2}) \cdot \Delta_{K_0}(t)$$

where K_+ is an oriented knot or link, K_- is the result of changing a single oriented positive (right-handed) crossing in K_+ to a negative (left-handed) crossing, and K_0 is the result of resolving the crossing as shown in Figure 1.

Note that if K_+ is a knot, then so is K_-, and K_0 is a 2-component link. If K_+ is a 2-component link, then so is K_-, and K_0 is a knot.

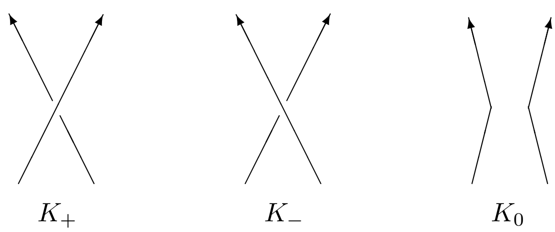


Figure 1

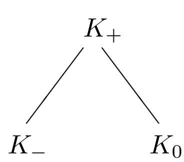
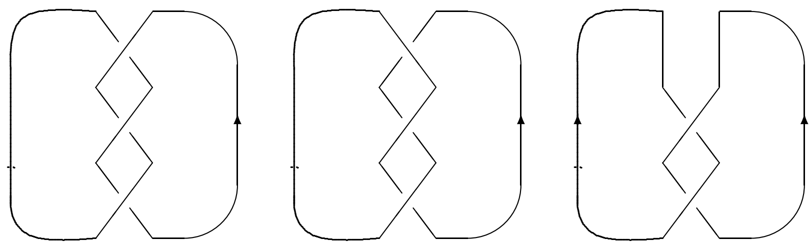


Figure 2

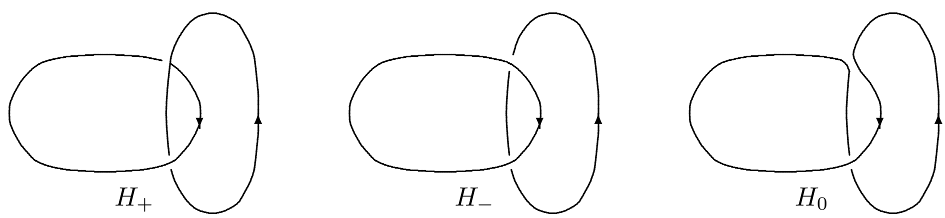
It is proved in [**FS1**] that one can start with a knot K and perform macareña moves so as to build a tree starting from K and at each stage adding the bifurcation of Figure 2, where each K_+, K_-, K_0 is a knot or 2-component link, and so that at the bottom of the tree we obtain only unknots and split links. Then, because for an unknot U we have $\Delta_U(t) = 1$, and for a split link S (of more than one component) we have $\Delta_S(t) = 0$, we can work backwards using the macareña relation to calculate $\Delta_K(t)$.

For example, we compute the Alexander polynomial of the trefoil knot:



K_+ K_- K_0

In the figure above, $K_+ = K$ is the trefoil knot, K_- is the unknot, and $K_0 = H$ is the Hopf link. Thus we have $\Delta_K = 1 + (t^{1/2} - t^{-1/2}) \cdot \Delta_H$. We see from the figure below that H_- is the unlink and H_0 is the unknot; hence $\Delta_H = 0 + (t^{1/2} - t^{-1/2}) \cdot 1$, and $\Delta_K(t) = 1 + (t^{1/2} - t^{-1/2})^2 = t - 1 + t^{-1}$.



H_+ H_- H_0

We next need to describe a method for constructing 3-manifolds which was first studied by W. Brakes [**Br**] and extended by J. Hoste [**Ho**]. Let L be a link in S^3 with two *oriented* components C_1 and C_2. Fix tubular neighborhoods $N_i \cong S^1 \times D^2$ of C_i with $S^1 \times$ (pt on ∂D^2) a longitude of C_i, i.e. nullhomologous in $S^3 \setminus C_i$. For $n \in \mathbf{Z}$, let $A_n = \left(\begin{smallmatrix} -1 & 0 \\ n & 1 \end{smallmatrix}\right)$. Note that A_n takes a meridian to a meridian. We get a 3-manifold

$$s(L;n) = (S^3 \setminus (N_1 \cup N_2))/A_n$$

called a 'sewn-up link exterior' by identifying ∂N_1 with ∂N_2 via a diffeomorphism inducing A_n in homology. A simple calculation shows that $H_1(s(L;n);\mathbf{Z}) = \mathbf{Z} \oplus \mathbf{Z}_{2\ell - n}$ where ℓ is the linking number in S^3 of the two components C_1, C_2, of L. (See [**Br**].) The second summand is generated by the meridian to either component.

J. Hoste [**Ho**, p.357] has given a recipe for producing Kirby calculus diagrams for $s(L;n)$. Consider a portion of L consisting of a pair of strands, oriented in opposite directions, and separated by a band B as in Figure 3.

Proposition 5. [**Ho**] *Let $L = C_1 \cup C_2$ be an oriented link in S^3. Consider a portion of L consisting of a pair of strands as in Figure 3. The band sum of C_1 and C_2 is a knot K, and the sewn-up link exterior $s(L;n)$ is obtained from the framed surgery on the the 2-component link on the right hand side of Figure 3.*

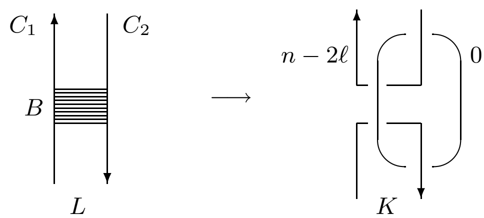


Figure 3

Now we can outline a proof of the knot surgery theorem. Begin with the resolution tree for a given oriented knot K. Each vertex of the tree corresponds to an oriented knot or oriented 2-component link. Replace each knot K' in the tree with the 4-manifold $X_{K'}$, and replace each 2-component link L with the fiber sum

$$X_L = X\#_{T=S^1\times m}(S^1 \times s(L;2\ell))$$

where m is a meridian to either component.

Suppose first that K_- is a knot (and therefore so is K_+). We see in Figure 4 that K_+ is the result of $+1$ surgery on the circle C. The circle C is nullhomologous; it bounds a punctured torus. In Figure 4 there is an obvious disk which is punctured twice by K_-. The punctured torus bounded by C consists of this disk, punctured at the points of intersection with K_- together with an annulus running 'halfway around K_-'.

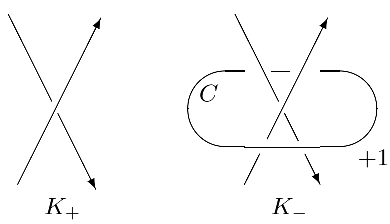


Figure 4

Take the product of this with S^1, and glue into X along $S^1\times m$ to obtain, on the one hand, X_{K_+}, and on the other, the result of $(+1)$-surgery on the nullhomologous torus $T_C = S^1 \times C$.

$$X_{K_+} = (X_{K_-})_{T_C}(0,1,1) \equiv X_{K_-}(1)$$

where the basis for $H_1(\partial(T_C\times D^2)$ consists of $S^1\times\mathrm{pt}$, a pushoff of C in the punctured torus, and $\partial D^2 = m_C$. The Morgan-Mrowka-Szabó formula implies that

$$\mathrm{SW}_{X_{K_+}}(\alpha) = \mathrm{SW}_{X_{K_-}}(\alpha) + \sum_i \mathrm{SW}_{X_{K_-}}(0)(\alpha + 2i[T_0])$$

where $X_{K_-}(0) = (X_{K_-})_{T_C}(0,1,0)$ is the result of 0-surgery on T_C. (Note that T_C is also nullhomologous in X_{K_+}.) As in the concluding comments of §1, only one of the terms $\mathrm{SW}_{X_{K_-}}(0)(\alpha + 2i[T_0])$ in the sum can be nonzero. To see this, we show that there is a torus Λ of self-intersection 0 in $X_{K_-}(0)$ such that $\Lambda\cdot T_0 = 1$ (note $T_0 = S^1 \times m_C$), and such that $\Lambda\cdot\alpha = 0$ for all $\alpha \in H_2(X \setminus T)$. In fact, Λ is formed from the union of the punctured torus bounded by C and the core disk of the 0-framed surgery giving $X_{K_-}(0)$. We thus have

$$\mathcal{SW}_{X_{K_+}} = \mathcal{SW}_{X_{K_-}} + \mathcal{SW}_{X_{K_-}(0)}$$

The manifold $X_{K_-}(0) = X\#_{T=S^1\times m}(S^1 \times Y)$ where Y is the 3-manifold obtained from 0-framed surgery on both components of the link $K_- \cup C$ in S^3 as in Figure 5.

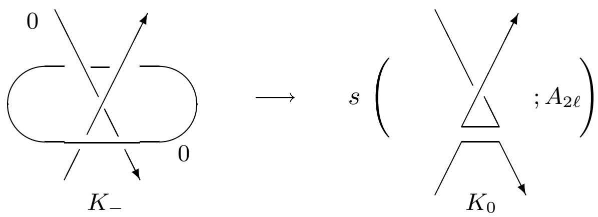


Figure 5

Hoste's recipe tells us that Y is the sewn-up manifold $s(K_0; 2\ell)$. Hence, by definition, $X_{K_-}(0) = X_{K_0}$. We thus get

$$\mathcal{SW}_{X_{K_+}} = \mathcal{SW}_{X_{K_-}} + \mathcal{SW}_{X_{K_0}}$$

The other case to consider is where L_- is an oriented 2-component link (so also L_+ is a 2-component link, and L_0 is a knot.). We get L_+ from L_- by a single surgery on a loop U as in Figure 6.

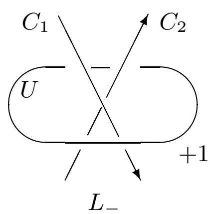


Figure 6

Let ℓ_- denote the linking number of the two components C_1 and C_2 of L_-. In the sewn-up manifold $s(L_-; 2\ell_-)$, the loop U becomes nullhomologous, because according to Hoste's recipe $s(L_-; 2\ell_-)$ is:

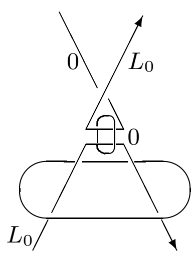


Figure 7

Thus

$$X_{L_+} = X\#_{T=S^1\times m}(S^1 \times s(L_+; 2\ell_+)) = X_{L_-}(1)$$

where $X_{L_-}(1)$ is shorthand for the surgery manifold $(X_{L_-})_{S^1\times U}(0,1,1)$. The first equality is the definition of X_{L_+} and the last equality is an exercise in Kirby calculus. Similarly we let $X_{L_-}(0)$ denote $(X_{L_-})_{S^1\times U}(0,1,0) = X\#_{T=S^1\times m}(S^1\times s(L_-;2\ell_-)_0)$ where $s(L_-;2\ell_-)_0$ stands for 0 (rather than +1) surgery on U in Figure 6.

Proposition 6. $\mathcal{SW}_{X_{L_-}(0)} = \mathcal{SW}_{X_{L_0}} \cdot (t - t^{-1})^2$

Proof. Cut open $s(L_-;2\ell_-)$ to get $S^3 \setminus N_{L_-}$ (where N_{L_-} denotes a tubular neighborhood of L_-). Similarly we can cut open $s(L_-;2\ell_-)_0$ to get a link exterior in $S^1\times S^2$. Take a product with S^1 and glue into X giving $X\#_{T=S^1\times m}(S^1\times(S^1\times S^2))$ with a pair of tori of self-intersection 0 removed. If we sew up the boundary of this manifold using the map $(1)\oplus A_{2\ell_-}$, we re-obtain $X_{L_-}(0)$.

Instead, fill in the boundary components with copies of $T^2 \times D^2$ to get a new manifold, Z. We wish to do this in such a way that when we remove a neighborhood of the new link $T^2 \times \{0\} \cup T^2 \times \{0\} \subset Z$ and sew up the boundaries using $(1)\oplus A_0$, we get $X_{L_-}(0)$. (We want to be able to sew up with this particular matrix because A_0 identifies $S^1\times C_1$ with $S^1\times C_2$.) We can accomplish this by gluing each $T^2\times D^2$ to a boundary component using the matrix $(1)\oplus\left(\begin{smallmatrix}0 & 1\\ 1 & -l_-\end{smallmatrix}\right)$. This matrix corresponds to $S^1\times$ $((-\ell)$-framed surgery). Then, using the internal fiber sum formula of §1,

$$\mathcal{SW}_{X\#_{T=S^1\times m}S^1\times s(L_-;2\ell_-)_0} = \mathcal{SW}_Z \cdot (t - t^{-1})^2$$

Now $Z = X\#_{T=S^1\times m}(S^1 \times Y)$ where Y is the 3-manifold of Figure 8

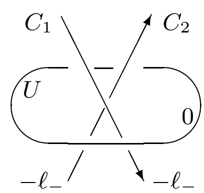


Figure 8

Now slide C_1 over C_2. We get 0-surgery on L_0 together with a cancelling pair of handles. Thus $Y = M_{L_0}$, and the proposition is proved. □

It follows from the proposition and the Morgan-Mrowka-Szabó theorem that

$$\mathcal{SW}_{X_{L_+}} = \mathcal{SW}_{X_{L_-}} + \mathcal{SW}_{X_{L_0}} \cdot (t - t^{-1})^2$$

We are now able to finish the proof of the knot surgery theorem. For a knot K, or an oriented 2-component link L and fixed X, we define a formal Laurent series Θ. For a knot K, define Θ_K to be the quotient, $\Theta_K = \mathcal{SW}_{X_K}/\mathcal{SW}_X$, and for a 2-component link define $\Theta_L = (t^{1/2} - t^{-1/2})^{-1}\cdot\mathcal{SW}_{X_L}/\mathcal{SW}_X$. It follows from from our calculations that in either case Θ satisfies the relation

$$\Theta_{K_+} = \Theta_{K_-} + (t - t^{-1}) \cdot \Theta_{K_0}.$$

Furthermore, for the unknot U, the manifold X_U is just $X\#_T(S^2 \times T^2) = X$, and so $\Theta_U = 1$. If L is a 2 component oriented split link, construct from L the knots K_+ and K_- as shown in Figure 9. Note that in this situation, $K_+ = K_-$ and $K_0 = L$. It follows from $\Theta_{K_+} = \Theta_{K_-} + (t - t^{-1}) \cdot \Theta_L$ that $\Theta_L = 0$.

Subject to these initial values, the resolution tree and the macareña relation determine $\Delta_K(t)$ for any knot K. It follows that Θ_K is a Laurent polynomial in a single variable t, and $\Theta_K(t) = \Delta_K(t)$, completing the proof of the knot surgery theorem.

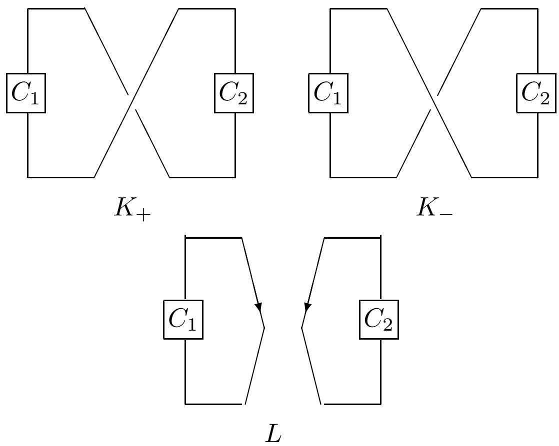


Figure 9

3. The Meng-Taubes Formula

Let Y be an oriented Riemannian 3-manifold. A spin^c structure on Y consists of a rank 2 complex vector bundle S over Y with a hermitian metric together with a bundle map $T^*X \to End(S)$. Up to 2-torsion, spin^c structures on Y correspond to elements of $H^2(Y;\mathbf{Z})$. If s is a spin^c structure on Y, then write $c_1(s)$ for $c_1(S) = c_1(\det S)$. We have $c_1(s+e) = c_1(s) + 2e$. Once again, for each connection A on $\det S$, there is a Dirac operator $D_A : \Gamma(S) \to \Gamma(S)$.

Proposition 7. *Let Y be an oriented Riemannian 3-manifold, and consider a spin^c structure on $S^1 \times Y$ whose determinant line bundle is pulled back from Y. Then any solution of the Seiberg-Witten equations for this spin^c structure is S^1-invariant.*

Proof. Suppose that we are given a spin^c structure on $S^1 \times Y$ which is pulled back from a spin^c structure s on Y. This means that the bundles $W^{\pm}$ corresponding to the spin^c structure over $S^1 \times Y$ are given by pulling back the spinor bundle S on Y. In particular, the first Chern class of the pulled back structure is the pullback of $c_1(L)$ where $L = \det S$.

For $(A, \psi) \in \mathcal{A}_L \times \Gamma(S)$ define the *Chern-Simons-Dirac functional*

$$\Phi(A,\psi) = \frac{1}{2}\int_Y (A - A_0) \wedge (F_A + F_{A_0}) - \frac{1}{2}\int_Y \langle \psi, D_A\psi\rangle d\mathrm{vol} + \int_Y (A - A_0) \wedge i\mu$$

(The extra fixed connection A_0 is necessary to define Φ, but changing it will only change Φ by a constant.) One can check that the gradient flow equations of Φ are precisely the Seiberg-Witten equations for $\mathbf{R} \times Y$.

The gauge group $\mathrm{Map}(Y, S^1)$ acts on $\mathcal{A}_L \times \Gamma(S)$ via $u(A, \psi) = (A - u^{-1}du, u \cdot \psi)$, and it is an exercise that

$$\Phi(u(A, \psi)) = \Phi(A, \psi) - 4\pi^2([u] \cup c_1(L))[Y]$$

where $[u] \in H^1(Y; \mathbf{Z})$ corresponds to the homotopy class of u in $[Y, S^1]$. This means that Φ is not well-defined as a real-valued map on $\mathcal{B}_Y = (\mathcal{A}_L \times \Gamma(S))/\mathrm{Map}(Y, S^1)$; however, it does give a well-defined map $\mathcal{B}_Y \to S^1 = \mathbf{R}/4\pi^2\mathbf{Z}$.

View $S^1 \times Y$ as $([0,1] \times Y)/\sim$. A solution of the Seiberg-Witten equations on $S^1 \times Y$ which is not constant is a (downward) gradient flow line of Φ. In particular, its values at 0 and at 1 must be different. The difference between $(A(0), \psi(0))$ and $(A(1), \psi(1))$ must be such that the path closes up to form a loop in $\mathcal{B}_Y$. Thus $(A(0), \psi(0)) = u(A(1), \psi(1))$ for some nontrivial $u \in \mathrm{Map}(Y, S^1)$. A line bundle over $S^1 \times Y$ corresponding to such a loop has first Chern class equal to $c_1 = c_1(L) + [S^1] \otimes [u] \in H^2(S^1 \times Y) = H^2(Y) \oplus H^1(S^1) \otimes H^1(Y)$. Thus if $\det W^+$ is pulled back from Y then we must have $[u] = 0$.

This means that for our gradient flow line $(A(t), \psi(t))$ we have $\Phi(A(1), \psi(1)) = \Phi(A(0), \psi(0))$, and so by our comments above, the gradient flow 'line' is constant; i.e. the solution $(A(t), \psi(t))$ is constant in t. $\square$

We see that the solutions of the Seiberg-Witten equations for spin^c structures on $S^1 \times Y$ pulled back from Y correspond to stationary solutions, i.e. to critical points, of Φ. The variational equations of Φ are

$$D_A\psi = 0$$
$$F_A = iq(\psi) - i\mu$$

(plus a finite energy condition in case Y is noncompact). These are the Seiberg-Witten equations for 3-manifolds. We get a Seiberg-Witten invariant by counting the critical points with signs. The signs are determined by arbitrarily fixing one critical point and assigning to any critical point of Φ $(-1)^{SF}$ where SF is the spectral flow of the Hessian of Φ from the fixed critical point to the one whose sign is being determined. ('Spectral flow' is the signed number of eigenvalues which go from negative to positive along a path of operators.) The Seiberg-Witten invariant $\mathrm{SW}_Y(s)$ is a diffeomorphism invariant if $b_1(Y) > 1$ or if Y is noncompact with ends of the form $T^2 \times [0, \infty)$.

In case Y is compact and $b_1(Y) = 1$, one gets invariants $\mathrm{SW}_Y^{\pm}(s)$ depending on whether $\pm(2\pi c_1(s) - \mu) \cdot \lambda > 0$ where we fix an orientation of $H_1(Y; \mathbf{R})$ and λ is the dual generator of $H^1(Y; \mathbf{R})$. Li and Liu have calculated the difference between the two Seiberg-Witten invariants:

The Wall-Crossing Formula. [**LL**] *For a spin^c structure s on a 3-manifold Y with $b_1(Y) = 1$, one has* $\mathrm{SW}_Y^-(s) = \mathrm{SW}_Y^+(s) + \frac{1}{2}c_1(s) \cdot \lambda$.

We may now state the theorem of Meng and Taubes. We restrict to the case of a 3-manifold obtained from 0-surgery on a knot. The actual Meng-Taubes theorem gives the Seiberg-Witten invariant of any 3-manifold with $b_1 > 0$.

Theorem 8 (Meng - Taubes [**MT**]). *Let M_K be the homology $S^2 \times S^1$ obtained from 0-framed surgery on a knot $K \subset S^3$. Then*

$$\mathcal{SW}_{M_K}^- \cdot (t - t^{-1})^2 = \Delta_K(t^2)$$

where $t = t_T$ for the generator T of $H^2(M_K; \mathbf{Z}) = \mathbf{Z}$ satisfying $T \cdot \lambda = 1$.

Note that since $H_*(M_K;\mathbf{Z})$ contains no 2-torsion, spin^c structures on M_K correspond to elements of $H^2(M_K;\mathbf{Z}) = \mathbf{Z}$. For each $k \in \mathbf{Z}$ there is a unique spin^c structure s_k on M_K such that $c_1(s_k) = 2kT$. Also, the meridian m to K is the generator of $H_1(M_K)$ (which gives the orientation corresponding to λ).

Before proceeding with the proof of the theorem, it will be useful to compare $\mathrm{SW}^{\pm}_{M_K}$ with the Seiberg-Witten invariants of $S^1 \times M_K$. As for M_K, there are Seiberg-Witten invariants $\mathrm{SW}^{\pm}_{S^1\times M_K}$ determined by perturbation terms η such that $(2\pi c_1(s)+\eta)^+\cdot\Lambda$ is >0 or <0, where Λ is Poincaré dual to $[S^1]\times\lambda$. If the spin^c structure is pulled back from M_K, and the perturbation term is pulled back as well, $\eta = -\mu$, then $(2\pi c_1(s)+\eta)^+\cdot\Lambda = (2\pi c_1(s)-\mu)\cdot\lambda$. Thus we have

$$\mathrm{SW}^{\pm}_{S^1\times M_K}(2kT) = \mathrm{SW}^{\pm}_{M_K}(2kT)$$

(For a more detailed proof see [**OT**].)

Now consider the situation of the knot surgery theorem. Suppose that $b^+(X) > 1$ and $\pi_1(X) = 1 = \pi_1(X \setminus T_0)$ where T_0 is a homologically essential torus of self-intersection 0. We form $X_K = X\#_{T_0=T_m}(S^1\times M_K)$. Recall that $T_m = S^1\times m$. Therefore T_m represents the Poincaré dual of T, and $H_2(S^1\times M_K)$ is generated by the classes of T_m and Σ, a capped off Seifert surface for K. If $\alpha\in H^2(S^1\times M_K)$ is not a class which is pulled back from $H^2(M_K;\mathbf{Z})$, then the Poincaré dual β of α is represented by $r\Sigma + sT_m$, for $r\neq 0$. Since $m\cdot\Sigma = 1$ in M_K, $T_m\cdot\Sigma = 1$ in $S^1\times M_K$. If $\zeta\in H_2(X;\mathbf{Z})$ is any class such that $\zeta+\beta$ is Poincaré dual in X_K to a class with a nontrivial Seiberg-Witten invariant, then the adjunction inequality applied to T_m gives $0\geq T_m\cdot T_m + |T_m\cdot(\zeta+\beta)| = 0+|r|$. Thus $r=0$, so we see that any class in $H^2(X_K;\mathbf{Z})$ with a nonzero Seiberg-Witten invariant must be Poincaré dual to a class of the form $\zeta + 2kT_m$.

This means that it suffices for the purpose of calculating $\mathcal{SW}_{X_K}$ to consider only the restricted Seiberg-Witten invariants

$$\mathcal{SW}^{\pm}_{S^1\times M_K,T} = \sum \mathrm{SW}^{\pm}_{S^1\times M_K}(2kT)t^{2k}$$

We have seen that $\mathcal{SW}^{\pm}_{S^1\times M_K,T} = \mathcal{SW}^{\pm}_{M_K}$. The coefficient of t^{2k} in the product $(t-t^{-1})^2\cdot\mathcal{SW}^{-}_{S^1\times M_K,T}$ is

$$\mathrm{SW}^{-}_{S^1\times M_K}((2k+2)T) - 2\,\mathrm{SW}^{-}_{S^1\times M_K}(2kT) + \mathrm{SW}^{-}_{S^1\times M_K}((2k-2)T)$$

The wall-crossing formula for $S^1\times M_K$ is the same as that for M_K, and it implies that this coefficient is the same as the coefficient of t^{2k} in $(t-t^{-1})^2\cdot\mathcal{SW}^{+}_{S^1\times M_K,T}$. Thus

$$(t-t^{-1})^2\cdot\mathcal{SW}^{+}_{S^1\times M_K,T} = (t-t^{-1})^2\cdot\mathcal{SW}^{-}_{S^1\times M_K,T}$$

The gluing theorems of §1 can thus be seen to apply to the calculation of $\mathcal{SW}_{X_K}$:

$$\begin{aligned}\mathcal{SW}_{X_K} &= \mathcal{SW}_{X\setminus N_{T_0}}\cdot\mathcal{SW}^{\pm}_{S^1\times(M_K\setminus N_m),T}\\ &= \mathcal{SW}_X\cdot(t-t^{-1})\cdot\mathcal{SW}^{\pm}_{S^1\times M_K,T}\cdot(t-t^{-1})\\ &= \mathcal{SW}_X\cdot(t-t^{-1})^2\cdot\mathcal{SW}^{\pm}_{M_K} = \mathcal{SW}_X\cdot\Delta_K(t^2)\end{aligned}$$

by the Meng-Taubes theorem. So we see that the Meng-Taubes theorem gives another proof of the knot surgery theorem.

There is yet another Seiberg-Witten invariant for compact 3-manifolds with $b_1 = 1$ which satisfies the usual properties of the Seiberg-Witten invariant of 3-manifolds with $b_1 > 1$. This is the 'small perturbation invariant' $\mathrm{SW}^0_{M_K}$ which is defined by using an exact perturbation term μ. Its values are:

$$\mathrm{SW}^0_{M_K}(2dT) = \begin{cases} \mathrm{SW}^+_{M_K}(2dT), & d > 0 \\ \mathrm{SW}^-_{M_K}(2dT), & d < 0 \end{cases}$$

The wall-crossing formula implies that if $c_1(s) = 0$, then $\mathrm{SW}^+_{M_K}(s) = \mathrm{SW}^-_{M_K}(s) = \mathrm{SW}^0_{M_K}(s)$. Among the usual properties that the small-perturbation Seiberg-Witten invariant satisfies is that $\mathrm{SW}^0_{M_K}(-2dT) = \mathrm{SW}^0_{M_K}(2dT)$.

Write the Alexander polynomial of K as $\Delta_K(t) = \sum\limits_{i=-n}^{n} a_i t^i$ where $a_{-i} = a_i$ and $\sum a_i = 1$. Applying the wall-crossing formula, we see that the Meng-Taubes formula is equivalent to

$$\mathrm{SW}^0_{M_K}(2dT) = a_{1+|d|} + 2a_{2+|d|} + 3a_{3+|d|} + \dots$$

and because of the symmetry of SW^0, we may restrict to $d \geq 0$. We will give the proof only in the case where K is a fibered knot, i.e. where the exterior $S^3 \setminus N_K$ is a fiber bundle over the circle with fiber a surface Σ_0 with a single boundary component.

The proof that we present is based on a paper of Simon Donaldson [**D**]. Donaldson's proof is both more general and more beautiful than the proof which we will outline, and the reader is strongly recommended to read his paper.

Let W be a 3-manifold with cylindrical ends $\Sigma \times [0, \infty)$ where Σ is an oriented surface of genus g. Fix a spin structure and a metric on Σ. (A spin structure on a surface is a square root of the canonical line bundle $K_\Sigma = K$ where $c_1(K) = 2g - 2$.) The spinor bundle S of a spin^c structure on $\Sigma \times \mathbf{R}$ restricts over Σ as $(K^{\frac{1}{2}} - K^{-\frac{1}{2}}) \otimes E$ and the restriction of its determinant line bundle L over Σ is $L|_\Sigma = E^2$. (So E is a square root of $L|_\Sigma$.) The Dirac operator on Σ can be identified with $\bar\partial : \Gamma(K^{\frac{1}{2}}) \to \Gamma(K^{-\frac{1}{2}})$ and then on $\Sigma \times \mathbf{R}$

$$D_{\Sigma\times\mathbf{R}} = \begin{pmatrix} i\frac{\partial}{\partial t} & \bar\partial^* \\ \bar\partial & -i\frac{\partial}{\partial t} \end{pmatrix}$$

For $A \in \Gamma(L)$ (t-dependent) and $\psi = (\alpha, \beta) \in \Gamma(S) = \Gamma((K^{\frac{1}{2}} \otimes E) \oplus (K^{-\frac{1}{2}} \otimes E))$ the Seiberg-Witten equations on $\Sigma \times \mathbf{R}$ become

$$\begin{aligned} iF_A &= \tfrac{1}{2}(|\beta|^2 - |\alpha|^2) \cdot vol_\Sigma \\ -2i\,\bar\partial_A \alpha &= \dot\beta \\ 2i\,\bar\partial^* \beta &= \dot\alpha \\ \dot A &= \alpha\bar\beta \end{aligned}$$

Using these equations, one may prove the following lemma.

Lemma 9. *Any finite energy solution of the Seiberg-Witten equations on $\Sigma \times \mathbf{R}$ is stationary (constant in t).*

Suppose that $c_1(L) = 2dT$, $d > 0$. In de Rham cohomology, the class $c_1(L)$ is represented by the differential form $\frac{1}{2\pi i} F_A$. So $\alpha = 0$ and the equations simplify to

$$iF_A = \tfrac{1}{2}|\beta|^2 \cdot vol_\Sigma$$
$$\bar{\partial}^* \beta = 0$$

These equations are called the *vortex equations* on Σ for $K^{\frac{1}{2}} \otimes E^*$. The solutions are holomorphic sections of holomorphic bundles over Σ, therefore, to obtain sections we need $0 \leq \deg(K^{\frac{1}{2}} \otimes E^*) = g - 1 - d$. (Recall that E is a square root of K, with degree, $c_1(K) \cdot [\Sigma] = 2g - 2$.) Thus d must satisfy $0 \leq d \leq g - 1$.

The moduli space of solutions of the vortex equations on $K^{\frac{1}{2}} \otimes E^*$ mod equivalence can be identified with positive divisors, i.e. with $\mathrm{Sym}^k(\Sigma)$, $k = g - 1 - d$. Since we are assuming that K is a fibered knot, M_K is a fiber bundle over the circle with fiber a closed surface Σ of genus g, i.e. $M_K = \Sigma \times_f S^1 = M_K \times I/\sim$ where $(x, 1) \sim (f(x), 0)$. Any solution of the Seiberg-Witten equations on $\Sigma \times I$ is stationary; so it corresponds to a solution of the vortex equations on Σ. Such a solution is constant in t; but it must close up to a solution of the Seiberg-Witten equations on M_K; so it must be a fixed point of f acting on the moduli space $\mathrm{Sym}^k(\Sigma)$. (Salamon points out in [**S**] that the induced map $f^{(k)}$ on $\mathrm{Sym}^k(\Sigma)$ is in general not smooth, but only Lipschitz continuous.) Counting solutions (and believing that the signs counted by the Lefschetz number $L(f^{(k)})$ agree with those in the definition of SW) we get $\mathcal{SW}^0_{M_K}(2dT) = L(f^{(k)})$ (where $k = g - 1 - d$). The parenthetical worry can be taken care of rigorously [**S**].

The proof of the Meng-Taubes theorem in the case of a fibered knot K will be completed by calculating $L(f^{(k)})$; so we must understand $H_*(\mathrm{Sym}^k(\Sigma); \mathbf{Q})$. Set $\Lambda^j = \Lambda^j(H_1(\Sigma; \mathbf{Q}))$ and $\Lambda(i) = \Lambda^{g-i} = \Lambda^{g+i} = \Lambda_{(-i)}$. (The second of these equalities follows because the intersection form of Σ raised to the gth power is a volume form on $H_1(\Sigma; \mathbf{Q})$; so the equality is given by the corresponding $*$-operator.) Then it is known that:

$$H_k(\mathrm{Sym}^k(\Sigma); \mathbf{Q}) = \Lambda_{(d+1)} + \Lambda_{(d+3)} + \Lambda_{(d+5)} + \cdots$$
$$H_{k\pm1}(\mathrm{Sym}^k(\Sigma); \mathbf{Q}) = \Lambda_{(d+2)} + \Lambda_{(d+4)} + \Lambda_{(d+6)} + \cdots$$
$$H_{k\pm2}(\mathrm{Sym}^k(\Sigma); \mathbf{Q}) = \Lambda_{(d+3)} + \Lambda_{(d+45)} + \Lambda_{(d+7)} + \cdots$$

etc. [**MD**].

This means that, up to sign,

$$L(f^{(k)}) = \mathrm{Tr}(f^{(k)}|_{\Lambda(d+1)}) - 2\,\mathrm{Tr}(f^{(k)}|_{\Lambda(d+2)}) + 3\,\mathrm{Tr}(f^{(k)}|_{\Lambda(d+3)}) - + \cdots$$

On the other hand, the Alexander polynomial for fibered knots is

$$\Delta_K = t^{-g} \det(f_* - t\,\mathrm{Id})$$

where K has genus g and f_* denotes the action of f on $H_1(\Sigma)$. The coefficient of t^m in this polynomial is $a_m = (-1)^m \mathrm{Tr}(f^{(k)}|_{\Lambda(m)})$. Thus

$$\mathcal{SW}^0_{M_K}(2dT) = L(f^{(k)}) = a_{d+1} + 2a_{d+2} + 3a_{d+3} + \cdots$$

(up to an overall sign, which works out correctly), and this is the Meng-Taubes theorem.

4. Gromov Invariants

The purpose of this lecture is to introduce the work of Cliff Taubes on symplectic 4-manifolds. For more comprehensive surveys covering this material, see [**T3, McD**]. Let (X, ω) be a symplectic 4-manifold. Recall that this means that ω is a closed 2-form on X which is nondegenerate: $\omega \wedge \omega$ is nowhere 0. Every symplectic manifold admits a compatible almost-complex structure J. 'Compatible' means that $\omega(Jv_1, Jv_2) = \omega(v_1, v_2)$ and that $\omega(v, Jv) > 0$ for all v. (This means that $\omega(-, J-)$ is a J-invariant Riemannian metric on X.)

A 2-dimensional submanifold $\Sigma \subset X$ (not necessarily connected) is called J-*holomorphic* if J preserves $T\Sigma$. In this case J restricts to Σ to make it a complex curve. Furthermore, $\omega|_\Sigma$ is a volume form on Σ, and so it induces an orientation on the surface. Σ is called a J-holomorphic curve. Since J preserves the tangent bundle $T\Sigma$ of a J-holomorphic curve, it also preserves the normal bundle N_Σ with respect to the metric $\omega(-, J-)$. Then Σ satisfies the adjunction formula:

$$2g - 2 = \Sigma \cdot \Sigma + K \cdot \Sigma$$

where g is the genus of Σ and $K = K_\omega$ is the canonical class $-PD(c_1(TX))$ with respect to a compatible almost-complex structure. ('PD' denotes 'Poincaré dual'.) Notice that, in particular, this means that the genus of a connected J-holomorphic curve is determined by its homology class.

Fix $A \in H_2(X; \mathbf{Z})$. The moduli space $\mathcal{M}(A, J)$ of J-holomorphic curves has formal dimension $2d(A) = A \cdot A - K \cdot A$, as calculated by the Index Theorem. For a generic compatible J, the moduli space is a compact manifold, except at multiply covered tori. From this information we can extract the Gromov invariant as follows: For simplicity we assume that X is minimal, i.e. it contains no symplectic sphere of self-intersection -1.

If $d(A) = 0$ then $Gr_X(A)$ is a signed count of elements of $\mathcal{M}(A, J)$ provided that the genus

$$g(A) = \frac{1}{2}(A \cdot A + K \cdot A) + 1$$

is ≥ 2 or $g(A) = 1$ and A is a primitive (indivisible) class. Signs are obtained from the spectral flow of an operator D related to $\bar{\partial}$. (Flow to a complex operator.)

If $d(A) > 0$, fix $d(A)$ generic points in X and count only J-holomorphic curves which contain all these points. The result will be a finite set of curves which $Gr_X(A)$ counts with signs.

For $A = 0$, $d(A) = 0$, and by convention, $Gr_X(0) = 1$.

Finally, if $d(A) = 0$, $g(A) = 1$, and $A = mB$, if A is represented by a J-holomorphic curve, this curve might be a multiple torus mT where T represents B and $T \cdot T = 0$. Also note that the adjunction equality then implies that $K \cdot T = 0$. In this case, it is not enough to count embedded J-holomorphic curves in the class $A = mB$, because the count for B might contribute to the count for A via m-fold covers.

This situation has been completely analyzed by Taubes. Also the paper [**IP1**] of Ionel and Parker contains a very clear description of how one deals with multiple tori. Each J-holomorphic torus of square 0 contributes to its own Gromov invariant and to those of its multiples according to one of eight simple functions (often called

'Taubes counting functions'), $f_{(\pm,i)}$ with $i = 0, 1, 2, 3$ and with $f_{(-,i)} = 1/f_{(+,i)}$ and

$$f_{(+,0)} = \frac{1}{1-t}, \quad f_{(+,1)} = 1+t, \quad f_{(+,2)} = \frac{1+t}{1-t}, \quad f_{(+,3)} = \frac{(1+t)(1-t^2)}{1+t^2}$$

In this case, the operator D used to calculate the above spectral flow can be tensored with a real line bundle over T. Such bundles are in $1-1$ correspondence with $H^1(T;\mathbf{Z}_2) \cong \mathbf{Z}_2 \oplus \mathbf{Z}_2$. Thus we get the trivial real line bundle and three nontrivial bundles and correspondingly, D as well as three new operators D_j. In the symbol $(\pm, i)$, the integer i denotes the number of D_j, $j = 1, 2, 3$ with nontrivial mod 2 spectral flow, and the '$\pm$' is the mod 2 spectral flow (i.e. the sign) of the original operator D.

Each such torus T lives in a compact 0-dimensional moduli space, and the contribution of T to $Gr_X(m[T])$ is the coefficient of t^m in the expansion of the appropriate $f_{(\pm,i)}$. A torus of square 0 which has a neighborhood in which J is integrable has type $(+, 0)$. Since

$$f_{(+,0)} = \frac{1}{1-t} = \sum_{m=0}^{\infty} t^m$$

it contributes '1' to the count for $m[T]$ for each $m \geq 1$. If X contains a symplectic torus T of self-intersection 0 and a tubular neighborhood N_T, it is possible to perturb the almost-complex structure to a generic J so that N_T contains a (cancelling) pair of J-holomorphic tori of type $(\pm, 0)$.

As an example, we calculate the Gromov invariant of $E(n)$. By the above comment, we may assume that there is a J-holomorphic torus T of type $(+, 0)$ representing the fiber class of the elliptic fibration $E(n) \to S^2$. Let N_T be a neighborhood of T which is a symplectic product $T \times D^2$. Then a gluing theorem (see [**IP2**]) implies

$$Gr_{E(n)} = Gr_{E(n)\setminus N_T} \cdot Gr_{N_T} = Gr_{E(n)\setminus N_T} \cdot \frac{1}{1-t}$$

(The notation here is similar to that used for the Seiberg-Witten invariant, $Gr_X = \sum Gr_X(A)\, t_A$.) We shall see later that the only elements $A \in H_2(E(n))$ with $Gr_X(A) \neq 0$ are multiples of the fiber class.

It is known in the theory of complex surfaces that a generic complex structure on $E(2) = K3$ admits no holomorphic curves. This fact, along with a limiting argument, implies that $Gr_{E(2)} = 1$ i.e. $Gr_{E(2)}(A) = 0$ for $A \neq 0$ and $Gr_{E(2)}(0) = 1$. Hence

$$Gr_{E(1)\setminus N_T} \cdot Gr_{E(1)\setminus N_T} = Gr_{E(2)} = 1$$

so $Gr_{E(1)\setminus N_T} = 1$. The calculation becomes completely analogous to that of $\mathcal{SW}_{E(n)}$ and we obtain

$$Gr_{E(n)} = (1-t)^{n-2}$$

Next we discuss Taubes' work on symplectic 4-manifolds. Taubes [**T1, T2**] proved the following important basic facts concerning the Gromov invariant:

THEOREM 10 (Taubes). *Let (X, ω) be a symplectic 4-manifold with $b^+ > 1$. Then*
(a) $Gr_X(0) = 1$ *and* $Gr_X(K) = \pm 1$ *where K is the canonical class of (X, ω).*
(b) *If $Gr_X(A) \neq 0$ then $d(A) = 0$.*

(c) $Gr_X(A) = \pm Gr_X(K - A)$.
(d) *If* $Gr_X(A) \neq 0$ *then* $0 \leq A \cdot [\omega] \leq K \cdot [\omega]$ *with equality only if* $A = 0$ *or* K.

In the $b^+ = 1$ case, it still follows that $Gr_X(0) = 1$ and that $Gr_X(A) \neq 0$ implies that $0 \leq A \cdot [\omega]$ with equality only for $A = 0$.

The key fact underpinning this theorem is one of the great achievements of 4-manifold theory:

Theorem 11 (Taubes). *After choosing appropriate orientations, for* $b_X^+ > 1$ *and for all* $A \in H_2(X;\mathbf{Z})$, $Gr_X(A) = \mathrm{SW}_X(2A - K)$. *For* $b_X^+ = 1$, $Gr_X(A) = \mathrm{SW}_X^-(2A - K)$ *for all* A *such that* $A \cdot E \geq -1$ *for each* E *represented by an embedded symplectic* 2*-sphere of square* -1.

Taubes carried out the proof of his theorem in a series of papers that are cited in [**T3**]. This theorem shows that the calculation of $Gr_{E(n)}$ is complete, since the adjunction equality for Seiberg-Witten invariants rules out other possible classes.

From (b) in Taubes' theorem, if $Gr_X(A) \neq 0$ then $d(A) = 0$, or equivalently $A^2 = K \cdot A$. The adjunction formula implies that $2g(A) - 2 = A^2 + K \cdot A = 2A^2$, which means that $A^2 = g(A) - 1$. We also know that $3\,\mathrm{sign}(X) + 2\,\mathrm{e}(X) = c_1^2(X) = K^2$. Thus the dimension of the Seiberg-Witten moduli space for $2A - K$ is

$$\frac{1}{4}\left((2A-K)^2 - (3\,\mathrm{sign}(X) + 2\,\mathrm{e}(X))\right) = \frac{1}{4}(4A^2 - 4A \cdot K + K^2 - K^2) = 0$$

It follows that any generic Seiberg-Witten moduli space for a β with $\mathrm{SW}_X(\beta) \neq 0$ has dimension 0.

Now let's use this point of view to consider the Meng-Taubes Theorem as in [**T3**]. Let Y be a 3-manifold which fibers over the circle, $p : Y \to S^1$, with fiber the surface Σ and with monodromy f. Let λ be a volume form on S^1. It is possible to choose a metric on Y so that p is a harmonic map. This means that $p^*(\lambda)$ is a harmonic 1-form. Let ϑ be the pullback to $S^1 \times Y$ of a volume form on the left-hand S^1. Define a symplectic form on $S^1 \times Y$ by $\omega = \vartheta \wedge p^*(\lambda) + *_3(p^*(\lambda))$.

We restrict to the case where $Y = M_K$ and K is a fibered knot with monodromy f. Then $H_2(S^1 \times M_K;\mathbf{Z}) = \mathbf{Z} \oplus \mathbf{Z}$ is generated by $T = S^1 \times m$ and Σ, a fiber of $M_K \to S^1$ with genus g. In the last section we saw that to understand $\mathcal{SW}^-_{S^1 \times M_K}$, we must calculate $\mathrm{SW}^-_{S^1 \times M_K}(2rT)$, $r \in \mathbf{Z}$. Taubes' theorem relating Gromov to Seiberg-Witten invariants gives the correspondence:

$$\begin{array}{rcl} Gr & \longleftrightarrow & \mathrm{SW} \\ A & \longleftrightarrow & 2A - K \\ nT & \longleftrightarrow & (2n - (2g-2))T \end{array}$$

Theorem 12 (Taubes). *The Seiberg-Witten invariant of* $S^1 \times M_K$ *is*

$$\mathcal{SW}^-_{S^1 \times M_K} = \sum_n \mathcal{SW}_{S^1 \times M_K}(2nT)\, t^n = \frac{\det(t^{-1} - tf_*)}{(t^{-1} - t)^2}$$

where f_* *is the action of the monodromy on* $H_1(\Sigma)$.

Note that

$$\frac{\det(t^{-1} - tf_*)}{(t^{-1} - t)^2} = t^{2-2g}\frac{\det(\mathrm{Id} - t^2 f_*)}{(1 - t^2)^2}$$

whose numerator is $\Delta_K(t^2)$, the symmetrized Alexander polynomial of K evaluated on t^2. So this theorem implies the Meng-Taubes formula.

We now outline Taubes' argument for this theorem. (See also [**IP2, S**].) Choose a metric on $S^1 \times M_K$ which is a product in the ϑ and λ directions. Metrically, $M_K = (\Sigma \times [0,1])/f$. For $\alpha \in T^*(S^1 \times M_K)$, $J\alpha = -*(\omega \wedge \alpha)$ for the ω defined above. It is an exercise to see that a J-holomorphic representative of $[nT]$ must have the form $S^1 \times \gamma$ where $\gamma \subset M_K$ is a (perhaps disconnected) loop formed from n lines $\{x_i\} \times [0,1] \subset \Sigma \times [0,1]$ which get closed up to γ by the monodromy relation $(x,1) = (f(x),0)$. Thus the collection of points $\{x_1, \dots, x_n\}$ is a union of finite orbits of f acting on Σ. Furthermore, each corresponding torus (or union of tori) in $S^1 \times M_K$ has Taubes type $(\pm, 0)$.

Let $S(n)$ consist of all points of Σ which are fixed by f^n but by no f^m, $m < n$. Each corresponding torus has a Taubes function $f_{(\pm,0)}$, so it counts as $(1-t^n)^{\mp 1}$. We wish to count the contribution to the Gromov invariant of $S^1 \times M_K$ corresponding to a fixed orbit of f in $S(n)$. Each of the n points in this orbit gives rise to the same J-holomorphic torus (representing nT) in $S^1 \times M_K$, so the contribution of each point is $(1-t)^{-\varepsilon/n}$, where ε is the sign in $(\pm, 0)$. If we count over all $q \in S(n)$, we get $\prod_{q \in S(n)} (1-t^n)^{-\varepsilon(q)/n}$.

Lemma 13. *$\varepsilon(q) = d(\mathrm{Id} - f^n)_q$, the differential at q.*

To see this, let's consider the case where $n = 1$. Suppose that $x \in \Sigma$ is a fixed point of f. We need to calculate the spectral flow of the linearization of the $\bar\partial$ operator on $S^1 \times [0,1] \times \{x\}$. A normal perturbation is given by

$$S^1 \times [0,1] \to S^1 \times [0,1] \times T_x\Sigma, \quad (s,u) \to (s,u,\eta(s,u))$$

where $\eta : S^1 \times [0,1] \to T_x\Sigma = \mathbf{C}$ satisfies $\eta(s,1) = df(\eta(s,0))$. The linearization is $L(s,u)(\eta) = (\partial\eta/\partial s, (\mathrm{Id} - df)\eta)$. We need to calculate the mod 2 spectral flow of the path $\Lambda_t(s,u) = (\partial\eta/\partial s, (\mathrm{Id} - t\,df)\eta)$, $0 \le t \le 1$, from $\Lambda_1 = L$ to the complex operator Λ_0. Neither Λ_1 nor Λ_0 has a nontrivial kernel, and the mod 2 spectral flow is the count of the number of $0 < t < 1$ whose corresponding Λ_t has a kernel. This will occur when there is an η satisfying $\partial\eta/\partial s = 0$ and $(\mathrm{Id} - t\,df)\eta = 0$. Since η is holomorphic, the first equality implies that η is constant, say $\eta = z \in \mathbf{C}$. So the spectral flow counts the number of t in $0 \le t \le 1$ for which $\mathrm{Id} - t\,df$ has a nontrivial kernel.

Both eigenvalues of Λ_0 are positive. Thus the only way for the spectral flow to be odd is for $L = \Lambda_1$ to have one positive and one negative eigenvalue. This is detected precisely by the sign of $\det(\mathrm{Id} - df)$. The argument in the case where $n > 1$ is similar.

Since the translation from Gr to $\mathcal{SW}$ given by Taubes' theorem is given by $nT \longrightarrow (2n - 2g + 2)T$, we get

$$\begin{aligned}
\mathcal{SW}^-_{S^1\times M_K}(t) &= \sum_k \mathrm{SW}^-_{S^1\times M_K}(2kT)\, t^{2k} \\
&= \sum_k Gr_{S^1\times M_K}((2k+2g-2)T)\, t^{2k} \\
&= t^{2-2g}\sum_k Gr_{S^1\times M_K}((2k+2g-2)T)\, t^{2k+2g-2} \\
&= t^{2-2g}\sum_n Gr_{S^1\times M_K}(2n\,T)\, t^{2n} \\
&= t^{2-2g}\cdot Gr_{S^1\times M_K}(t^2) \\
&= t^{2-2g}\cdot \prod_{n\geq 0}\prod_{q\in S(n)} (1-t^{2n})^{-\varepsilon(q)/n}
\end{aligned}$$

Taubes attributes the rest of this argument to Bott:

$$\begin{aligned}
\ln(\mathcal{SW}^-_{S^1\times M_K}) + (2g-2)\ln t &= \sum_{n\geq 0}\sum_{q\in S(n)} -\frac{\varepsilon(q)}{n}\cdot \ln(1-t^{2n}) \\
&= -\sum_{n\geq 0}\sum_{q\in S(n)}\sum_{k\geq 0}\varepsilon(q)\frac{t^{2nk}}{nk}
\end{aligned}$$

Reordering the summation and setting $m = nk$ we get:

$$\ln(\mathcal{SW}^-_{S^1\times M_K}) + (2g-2)\ln t = -\sum_{m\geq 0} a_m \frac{t^{2m}}{m}$$

where a_m is the sum of all fixed points of f^m weighted by the sign of the determinant of the differential of $(\mathrm{Id} - f^m)$.

The Lefschetz Fixed Point Theorem implies that a_m is the Lefschetz number of f^m, hence $a_m = 2 - \mathrm{Trace}(f^m_*)$. Thus

$$\begin{aligned}
\ln(\mathcal{SW}^-_{S^1\times M_K}) + (2g-2)\ln t &= -2\sum_{m\geq 0}\frac{t^{2m}}{m} + \sum_{m\geq 0}\mathrm{Trace}(f^m_*)\frac{t^{2m}}{m} \\
&= -\ln((1-t^2)^2) + \ln(\det(\mathrm{Id} - t^2 f_*))
\end{aligned}$$

which is exactly the formula we are looking for.

5. Knot Surgery and Embedded Surfaces

This final section is devoted to the discussion of some applications of knot surgery to study the diversity of embedded surfaces in a smooth 4-manifold. Of course, by summing with surfaces in a 4-ball one can change the isotopy or diffeomorphism type of an embedded surface without changing its homology class. This technique includes the construction which connect sums a surface with a 2-knot, a knotted S^2 in S^4. We are not interested in such cheap tricks since our main interest is the relationship of the categories TOP and DIFF. So we fix the homeomorphism type of a pair (X, Σ) and ask if we can find other surfaces $\Sigma' \subset X$ such that the pair (X, Σ') is homeomorphic to (X, Σ) but not diffeomorphic to it.

We restrict our discussion to the case where X is simply connected. Say that the surface $\Sigma \subset X$ is *primitively embedded* if $\pi_1(X\setminus\Sigma) = 1$. Of course this condition implies that the homology class $[\Sigma]$ is primitive (indivisible) in $H_2(X; \mathbf{Z})$. In general, any smoothly embedded (connected) surface S in a simply connected smooth 4-manifold X with $[S] \neq 0$ has the property that the surface Σ which represents the homology class $[S]-[E]$ in $X\#\overline{\mathbf{CP}}^2$ and which is obtained by tubing together the surface S with the exceptional sphere E of $\overline{\mathbf{CP}}^2$ is primitively embedded (since the surface Σ transversally intersects the sphere E in one point).

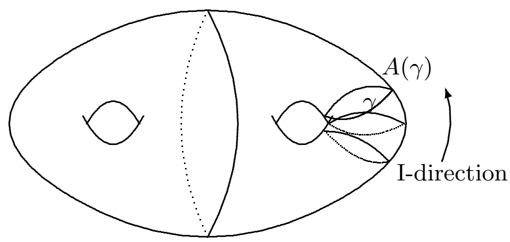


Figure 10

Assume that $\Sigma \cdot \Sigma = 0$. We leave to the reader to see that by blowing up our argument will take care of the case where $\Sigma \cdot \Sigma > 0$. Since $\Sigma^2 = 0$, we may identify a tubular neighborhood N_Σ with $\Sigma \times D^2$. Let γ be any loop on Σ and let T_γ be the torus which is the total space of the normal circle bundle $\gamma \times \partial D^2$ over γ. The torus T_γ is an example of a 'rim torus'. The homology classes of such tori generate the subgroup $\ker[H_2(X \setminus N_\Sigma; \mathbf{Z}) \to H_2(X; \mathbf{Z})]$.

Let $A(\gamma)$ be an annular neighborhood of γ in Σ as shown in Figure 10. Then $A(\gamma) \times D^2 = (S^1 \times I) \times D^2 = S^1 \times (I \times D^2)$ as in Figure 11. The operation we wish to consider replaces this with Figure 12, where the 'K' denotes that the knot K has been tied into the vertical central arc.

This operation is called 'rim surgery'. It was first described and studied in joint work with Ron Stern [**FS2**]. It replaces each arc in the I-direction in $A(\gamma)$ in the original surface Σ with a knotted arc, where the knotting occurs as in Figure 12.

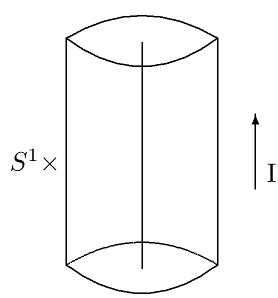


Figure 11

Note that the knotting is parametrized by $S^1 = \gamma$. Call the resultant manifold Σ_K. (To be precise, the notation should also include 'γ'.) Think of Σ_K as a reimbedding of Σ in X.

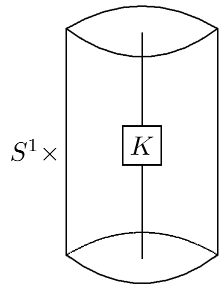


Figure 12

One can tie a knot in an arc as follows: Consider Figure 13. The solid torus V in this figure can be viewed as the complement of a neighborhood of the unknot in $S^3 = (D^2 \times S^1) \cup V$. Replacing V with $S^3 \setminus N(K)$, the exterior of a knot K, the result is again S^3, viewed as $(D^2 \times S^1) \cup (S^3 \setminus N(K))$. That is to say, $D^2 \times S^1$ is now the neighborhood of the knot K.

So let us review rim surgery. From X we remove

$$S^1 \times V = S^1 \times ((\text{meridian to } \Sigma) \times D^2) = T_\gamma \times D^2$$

and replace it with $S^1 \times (S^3 \setminus N(K))$. This is precisely knot surgery on the rim torus T_γ. Of course, T_γ is nullhomologous in X — otherwise the knot surgery formula would imply that the operation would change $\mathcal{SW}_X$, but X is not changed. It is clear that Σ and Σ_K represent the same homology class in X, and topological surgery [**F, Bo**] can be applied to prove:

PROPOSITION 14 ([**FS2**]). *Let X be a simply connected smooth 4-manifold containing a primitively embedded surface Σ. Then for each knot K in S^3, the above construction produces a surface Σ_K satisfying:*

(a) *Σ_K is homologous to Σ*
(b) *There is a homeomorphism of pairs $(X, \Sigma_K) \to (X, \Sigma)$.*

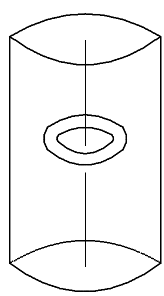


Figure 13

In order to utilize Seiberg-Witten invariants to distinguish the Σ_K we need to use an auxiliary construction which makes the rim tori essential. This can be accomplished by taking fiber sums with a 'standard pair' (Y_g, S_g) where Y_g is a Kähler surface and S_g is an embedded complex curve of self-intersection 0 such that for a normal disk D^2, ∂D^2 is nullhomologous in $H_1(Y_g \setminus N_{S_g}; \mathbf{Z})$. (This last condition implies primitivity.) Temporarily assume that we have such manifolds for all g.

THEOREM 15 ([**FS2**]). *Let X be a simply connected smooth 4-manifold containing a primitively embedded surface Σ. Suppose further that $\mathcal{SW}_{X\#_{\Sigma=S_g}Y_g} \neq 0$. Then one can find an infinite family of knots $\{K_i\}$ such that for any $i \neq j$ there is no diffeomorphism of pairs $(X, \Sigma_{K_1}) \to (X, \Sigma_{K_2})$.*

PROOF. For example, choose the sequence of knots K_i so that $\Delta_{K_i}(t)$ has $2i+1$ nontrivial terms. In $X\#_{\Sigma=S_g}Y_g$, the rim torus T_γ becomes nontrivial in homology. Since T_γ is disjoint from the surface Σ, first performing rim surgery and then taking the fiber sum gives the same result as first taking the fiber sum and then doing knot surgery: $X\#_{\Sigma_K=S_g}Y_g = (X\#_{\Sigma=S_g}Y_g)_K$ so

$$\mathcal{SW}_{X\#_{\Sigma_K=S_g}Y_g} = \mathcal{SW}_{(X\#_{\Sigma=S_g}Y_g)_K} = \mathcal{SW}_{(X\#_{\Sigma=S_g}Y_g)} \cdot \Delta_K(t^2) \quad (\text{where } t = t_{T_\gamma})$$

and our assumption is that $\mathcal{SW}_{X\#_{\Sigma=S_g}Y_g} \neq 0$; so the theorem follows, by counting the number of nonzero terms in $\mathcal{SW}_{X\#_{\Sigma_{K_i}=S_g}Y_g}$ □

One might ask what pairs (X, Σ) satisfy the hypothesis of this theorem. Because the manifolds Y_g are symplectic and the surfaces S_g are symplectic submanifolds, if (X, Σ) is also a symplectic pair, then Gompf's theorem [**G**] implies that the fiber sum $X\#_{\Sigma=S_g}Y_g$ is a symplectic manifold, and Taubes' theorem [**T1**] quoted in §4 implies that $\mathcal{SW}_{X\#_{\Sigma=S_g}Y_g} \neq 0$.

We next need to construct the standard pairs (Y_g, S_g). Start by building a generalized cusp neighborhood $C(g)$. Let $T(2, 2g+1)$ denote the corresponding torus knot. This is a fibered knot with genus g and monodromy φ_g of order $4g+2$. Attach a 2-handle to B^4 along $T(2, 2g+1)$ with framing 0 to get $C(g)$. Because we have used 0-framing, the fibration of the complement of the torus knot extends over $C(g)$. We get a map $C(g) \to D^2$ which is a fiber bundle projection over $D^2 \setminus \{0\}$ with fiber a closed surface S_g of genus g. Over $0 \in D^2$ there is a singular fiber which is a topological 2-sphere, embedded as a core disk of the 2-handle together with the cone in B^4 of the torus knot $T(2, 2g+1)$.

We next need to glue together several copies of $C(g)$ via 'boundary fiber sum'. This means that we take two copies $p_i : C(g)_i \to D_i^2$ and identify an interval I_1 in ∂D_1^2 with another interval I_2 in ∂D_2^2, and then form

$$C(g) \natural C(g) = C(g)_1 \cup p_1^{-1}(I_1) = p_2^{-1}(I_2) C(g)_2$$

where the gluing preserves fibers. There is an induced fiber bundle over the new boundary $\partial D_1^2 \# \partial D_2^2 \cong S^1$. This fiber bundle extends to a fibration over the disk $D_1^2 \natural D_2^2$ with two singular fibers.

The monodromy of the bundle over the boundary is the composition φ_g^2 of φ_g with itself. The original bundle over the boundary circle had monodromy of order $4g+2$. This means that if we take the boundary fiber sum $C(g) \natural C(g) \natural \cdots \natural C(g)$ of $4g+2$ copies of $C(g)$ over the boundary circle of the base disk we will have a fiber bundle with monodromy $\varphi_g^{4g+2} =$ id. This means that $\partial(C(g) \natural C(g) \natural \cdots \natural C(g)) \cong \Sigma_g \times S^1$. Take the union with the trivial fibration $\Sigma_g \times D^2 \to D^2$ to get $Y_g \to S^2$, an S_g-fibration with $4g+2$ singular fibers (all 2-spheres as above). The primitivity condition is satisfied because the fibration admits a section. For example, $g = 1$ gives the elliptic (torus) fibration $Y_1 = E(1) \to S^2$. In fact, this construction always gives rise to a holomorphic fibration, and Y_g is a complex surface.

REMARK 16. *Suppose that (X,Σ) is a symplectic pair where Σ is a primitively embedded surface of square 0 with positive genus. If $\Delta_K(t)\neq 1$, then Σ_K is not smoothly isotopic to a symplectic submanifold of X.*

PROOF. Suppose that Σ_K is isotopic to a symplectic submanifold Σ' of X. This isotopy carries the rim torus T to a rim torus T' of Σ'. We have

$$\mathcal{SW}_{X\#_{\Sigma'=S_g}Y_g}=\mathcal{SW}_{X\#_{\Sigma_K=S_g}Y_g}=\mathcal{SW}_{X\#_{\Sigma=S_g}Y_g}\cdot\Delta_K(t^2)$$

Symplectic forms ω_X on X_n (with respect to which Σ_n' is symplectic) and ω_Y on Y_g induce a symplectic form ω on the symplectic fiber sum $X\#_{\Sigma'=S_g}Y_g$ which agrees with ω_X and ω_Y away from the region where the manifolds are glued together. In particular, since T' is nullhomologous in X, we have $\langle\omega,T'\rangle=\langle\omega_X,T'\rangle=0$. Our equation above implies that the homology classes of $X\#_{\Sigma'=S_g}Y_g$ with nonzero Seiberg-Witten invariants are exactly the classes $b+2mT'$ where b is a class of $X\#_{\Sigma=S_g}Y_g$ with a nonzero Seiberg-Witten invariant and t^m has a nonzero coefficient in $\Delta_K(t)$. Thus these classes for $X\#_{\Sigma'=S_g}Y_g$ can be grouped into collections $\mathcal{C}_b=\{b+2mT'\}$, and if $\Delta_K(t)\neq 1$ then each $\mathcal{C}_b$ contains more than one such class. Note, however, that $\langle\omega,b+2mT'\rangle=\langle\omega,b\rangle$. Now Taubes has shown [**T2**] that the canonical class κ of a symplectic manifold with $b^+>1$ is the class with nontrivial Seiberg-Witten invariant which is characterized by the condition $\langle\omega,\kappa\rangle>\langle\omega,b'\rangle$ for any other such class b'. But this is impossible for $X_n\#_{\Sigma_n=S_g}Y_g$ since each $\mathcal{C}_b$ contains more than one class. $\square$

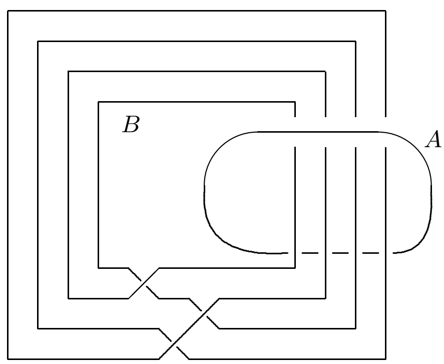


Figure 14

How does one go about finding inequivalent symplectic submanifolds in the same homology class? The most useful technique involves braiding [**FS3**]. (See also papers of Etgü and Park and of Vidussi such as [**EP, V**] for further applications, and see [**FS4**] for other approaches to this problem.)

Here is one example of an application of this technique. Let T be a symplectic torus of self-intersection 0 embedded in a (say) simply connected symplectic 4-manifold X, and consider its tubular neighborhood $N_T=S^1\times(S^1\times D^2)$. Let B be a $2m$-strand braid in $S^1\times D^2$ parallel to $S^1\times\{0\}$ and which becomes an unknot in S^3 when we view $S^1\times D^2$ as an unknotted solid torus in S^3. (For example, see Figure 14. In this figure, A is the axis of the braid, i.e. the core circle of the solid torus in S^3 which is complementary to the $S^1\times D^2$ containing B.) We obtain another self-intersection 0 torus

$$T_B = S^1 \times B \subset S^1 \times (S^1 \times D^2) = N_T$$

The torus T_B represents $2m[T]$ in $H_2(X; \mathbf{Z})$, and it is a symplectic submanifold of X. (The symplectic tubular neighborhood theorem implies that the symplectic structure on N_T is equivalent to $dx \wedge dy + r\,dr \wedge d\theta$, and it follows that T_B is symplectic.) By choosing different such braids, we get many examples. For each fixed m, there is an infinite family of braids $\{B_{m,k}\}$ such that there is no isotopy of X which takes $T_{B_{m,k}}$ to $T_{B_{m,k'}}$ for $k \neq k'$.

Each braid B that we use is unknotted in S^3; so the double cover of S^3 branched over B is again S^3. The axis A links B $2m$ times (see Figure 14), and this means that A lifts to a 2-component link L_B in the double cover S^3. Now A is an unknot and therefore is fibered (with genus 0 fiber). Each disk fiber meets the branch set B in $2m$ points. The double cover of a disk branched over $2m$ points is a twice-punctured genus $(m-1)$-surface. hence, in the double cover, L_B is a fibered 2-component link with a genus $(m-1)$-fiber.

To distinguish different tori T_B, consider the double branched cover $\tilde{X}$ of X branched over T_B. We have

$$\tilde{X} = (X \setminus N_T) \cup \tilde{N}_T \cup (X \setminus N_T)$$

since any loop not in N_T links T_B an even number of times. Now

$$N_T = S^1 \times (S^1 \times D^2) = S^1 \times (S^3 \setminus \mathrm{nbd}(A))$$

so $\tilde{N}_T = S^1 \times (S^3 \setminus \mathrm{nbd}(L_B))$. We thus have

$$\tilde{X} = (X \setminus N_T) \cup (S^1 \times (S^3 \setminus \mathrm{nbd}(L_B))) \cup (X \setminus N_T)$$

This is similar to knot surgery, except with a link, L_B, rather than a knot K. Just as for knot surgery, there is a formula for link surgery developed in [**FS1**]:

$$\mathcal{SW}_{\tilde{X}} = \Delta_{L_B}(t_1^2, t_2^2) \cdot \mathcal{SW}_{X_1} \cdot (t_1 - t_1^{-1}) \cdot \mathcal{SW}_{X_2} \cdot (t_2 - t_2^{-1})$$

where $\mathcal{SW}_{X_i} = (j_i)_* \mathcal{SW}_X$, j_i being the two inclusions of $X \setminus N_T$ in $\tilde{X}$. If B_1, B_2 are braids whose corresponding links L_{B_i} have different reductions $\Delta_{L_{B_1}}(t,t) \neq \Delta_{L_{B_2}}(t,t)$ to 1-variable polynomials, then one can use the above formula to show that there can be no isotopy of T_{B_1} to T_{B_2} in X. (The idea is that setting $t_1 = t_2$ corresponds to passing to the base of the double cover, i.e., back to X. See [**FS3**] for a detailed argument.)

The precise collection of families of braids is presented in [**FS3**], where it is checked that for any fixed m, $\Delta_{L_{B_{m,i}}}(t,t) \neq \Delta_{L_{B_{m,j}}}(t,t)$, and this gives the promised families.

References

[Bo] S. Boyer, *Simply-connected 4-manifolds with a given boundary*, Trans. Amer. Math. Soc., **298**, (1986), 331–357.

[Br] W. Brakes, *Sewing-up link exteriors*, in 'Low-Dimensional Topology:1', L.M.S Lecture Note Series no.48 (1982), pp. 27–37.

[D] S. Donaldson, *Topological field theories and formulae of Casson and Meng-Taubes*, Geom. Topol. Monogr. **2** (1999), 87–102.

[EP] T. Etgü and B. D. Park, *Non-isotopic symplectic tori in the same homology class*, Trans. Amer. Math. Soc. **356** (2004), 3739–3750.

[FS1] R. Fintushel and R. Stern, *Knots, links, and 4-manifolds*, Invent. Math. **134** (1998), 363–400.

[FS2] R. Fintushel and R. Stern, *Surfaces in* 4*-manifolds*, Math. Res. Letters **4** (1997), 907–914.

[FS3] R. Fintushel and R. Stern, *Symplectic surfaces in a fixed homology class*, J. Differential Geom. **52** (1999), no. 2, 203–222.

[FS4] R. Fintushel and R. Stern, *Tori in symplectic* 4*-manifolds*, Geometry and Topology Monographs, Vol. 7 (2004), Proceedings of the Casson Fest, 311–333.

[F] M. Freedman, *The topology of four-dimensional manifolds*, J. Diff. Geo. **17** (1982), 357–432.

[G] R. Gompf, *A new construction of symplectic manifolds*, Ann. Math. **142** (1995), 527–595.

[HH] F. Hirzebruch and H. Hopf, *Felder von Flächenelementen in 4-dimensionalen Mannigfaltigkeiten*, Math. Ann. **136** (1958), 156–172.

[Ho] J. Hoste, *Sewn-up r-link exteriors*, Pacific J. Math. **112** (1984), 347–382.

[IP1] E. Ionel and T. Parker, *The Gromov invariants of Ruan-Tian and Taubes*, Math. Res. Lett. **4** (1997), 521–532.

[IP2] E. Ionel and T. Parker, *Gromov invariants and symplectic maps* Math. Ann. **314** (1999), 127–158.

[KM] P. Kronheimer and T. Mrowka, *The genus of embedded surfaces in the projective plane*, Math. Research Letters **1** (1994), 797–808.

[LL] T. J. Li and A. Liu, *General wall-crossing formula*, Math. Res. Lett. **2** (1995), 797–810.

[MD] I. Macdonald, *Symmetric products of an algebraic curve*, Topology **1** (1962), 319–343.

[McD] D. McDuff, *Lectures on Gromov invariants for symplectic* 4*-manifolds*, in 'Gauge Theory and Symplectic Geometry' (Montreal, 1995), 175–210.

[MT] G. Meng and C. Taubes, $\underline{\mathcal{SW}}$ = *Milnor Torsion*, Math. Research Letters **3** (1996), 661–674.

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

[MMS] J. Morgan, T. Mrowka, Z. Szabó, *Product formulas along T^3 for Seiberg-Witten invariants*, Math. Res. Lett. **4** (1997), 915–929.

[N] L. Nicolaescu, *Notes on Seiberg-Witten theory*, Graduate Studies in Mathematics, 28, American Mathematical Society, Providence, RI, 2000.

[OT] C. Okonek and A. Teleman, 3*-dimensional Seiberg-Witten invariants and non-Kählerian geometry*, Math. Ann. **312** (1998), 261–288.

[P] B. D. Park, *A gluing formula for the Seiberg-Witten invariant along T^3*, Michigan Math. J. **50** (2002), no. 3, 593–611.

[S] D. Salamon, *Seiberg-Witten invariants of mapping tori, symplectic fixed points, and Lefschetz numbers*, Turkish J. Math. **23** (1999), no. 1, 117–143.

[T1] C. Taubes, *The Seiberg-Witten invariants and symplectic forms*, Math. Res. Letters **1** (1994), 809–822.

[T2] C. Taubes, *More constraints on symplectic manifolds from Seiberg-Witten invariants*, Math. Res. Letters **2** (1995), 9–14.

[T3] C. Taubes, *The geometry of the Seiberg-Witten invariants*, in *Surveys in differential geometry, Vol. III (Cambridge, MA, 1996)*, 299–339, Int. Press, Boston, MA.

[T3] C. Taubes, *The Seiberg-Witten invariants and 4-manifolds with essential tori*, Geom. Topol. **5** (2001), 441–519 (electronic).

[V] S. Vidussi, *Symplectic tori in homotopy E(1)'s*, Proc. Amer. Math. Soc. (to appear).

[W] E. Witten, *Monopoles and four-manifolds*, Math. Research Letters **1** (1994), 769–796.

Department of Mathematics, Michigan State University East Lansing, Michigan 48824

E-mail address: `ronfint@math.msu.edu`

Clay Mathematics Proceedings
Volume 5, 2006

Will We Ever Classify Simply-Connected Smooth 4-manifolds?

Ronald J. Stern

Abstract. These notes are adapted from two talks given at the 2004 Clay Institute Summer School on *Floer homology, gauge theory, and low dimensional topology* at the Alfred Rényi Institute. We will quickly review what we do and do not know about the existence and uniqueness of smooth and symplectic structures on closed, simply-connected 4-manifolds. We will then list the techniques used to date and capture the key features common to all these techniques. We finish with some approachable questions that further explore the relationship between these techniques and whose answers may assist in future advances towards a classification scheme.

1. Introduction

Despite spectacular advances in defining invariants for simply-connected smooth and symplectic 4-dimensional manifolds and the discovery of important qualitative features about these manifolds, we seem to be retreating from any hope to classify simply-connected smooth or symplectic 4-dimensional manifolds. The subject is rich in examples that demonstrate a wide variety of disparate phenomena. Yet it is precisely this richness which, at the time of these lectures, gives us little hope to even conjecture a classification scheme. In these notes, adapted from two talks given at the 2004 Clay Institute Summer School on *Floer homology, gauge theory, and low dimensional topology* at the Alfred Rényi Institute, we will quickly review what we do and do not know about the existence and uniqueness of smooth and symplectic structures on closed, simply-connected 4-manifolds. We will then list the techniques used to date and capture the key features common to all these techniques. We finish with some approachable questions that further explore the relationship between these techniques and whose answers may assist in future advances towards a classification scheme.

Algebraic Topology. The critical algebraic topological information for a closed, simply-connected, smooth 4-manifold X is encoded in its Euler characteristic $e(X)$, its signature $\sigma(X)$, and its type $t(X)$ (either 0 if the intersection form of X is even and 1 if it is odd). These invariants completely classify the homeomorphism

1991 *Mathematics Subject Classification.* Primary 57R55, 57R57, 14J26; Secondary 53D05.
Key words and phrases. 4-manifold, Seiberg-Witten invariant, symplectic, Lagrangian.
The author was partially supported by NSF Grant DMS0204041.

type of X ([3, 12]). We recast these algebraic topological invariants by defining $\chi_h(X) = (e(X) + \sigma(X))/4$, which is the holomorphic Euler characteristic in the case that X is a complex surface, and $c(X) = 3\sigma(X) + 2e(X)$, which is the self-intersection of the first Chern class of X in the case that X is complex.

Analysis. To date, the critical analytical information for a smooth, closed, simply-connected 4-manifold X is encoded in its Seiberg-Witten invariants [30]. When $\chi_h(X) > 1$ this integer-valued function SW_X is defined on the set of *spin*^{c} structures over X. Corresponding to each *spin*^{c} structure $\mathfrak{s}$ over X is the bundle of positive spinors $W_{\mathfrak{s}}^+$ over X. Set $c(\mathfrak{s}) \in H_2(X)$ to be the Poincaré dual of $c_1(W_{\mathfrak{s}}^+)$. Each $c(\mathfrak{s})$ is a characteristic element of $H_2(X;\mathbf{Z})$ (i.e. its Poincaré dual $\hat{c}(\mathfrak{s}) = c_1(W_{\mathfrak{s}}^+)$ reduces mod 2 to $w_2(X)$). The sign of SW_X depends on a homology orientation of X, that is, an orientation of $H^0(X;\mathbf{R}) \otimes \det H_+^2(X;\mathbf{R}) \otimes \det H^1(X;\mathbf{R})$. If $\mathrm{SW}_X(\beta) \neq 0$, then β is called a *basic class* of X. It is a fundamental fact that the set of basic classes is finite. Furthermore, if β is a basic class, then so is $-\beta$ with $\mathrm{SW}_X(-\beta) = (-1)^{\chi_h(X)}\,\mathrm{SW}_X(\beta)$. The Seiberg-Witten invariant is an orientation-preserving diffeomorphism invariant of X (together with the choice of a homology orientation). We recast the Seiberg-Witten invariant as an element of the integral group ring $\mathbf{Z}H_2(X)$, where for each $\alpha \in H_2(X)$ we let t_α denote the corresponding element in $\mathbf{Z}H_2(X)$. Suppose that $\{\pm\beta_1, \dots, \pm\beta_n\}$ is the set of nonzero basic classes for X. Then the Seiberg-Witten invariant of X is the Laurent polynomial

$$\mathcal{SW}_X = \mathrm{SW}_X(0) + \sum_{j=1}^{n} \mathrm{SW}_X(\beta_j)\cdot(t_{\beta_j} + (-1)^{\chi_h(X)}\, t_{\beta_j}^{-1}) \in \mathbf{Z}H_2(X).$$

When $\chi_h = 1$ the Seiberg-Witten invariant depends on a given orientation of $H_+^2(X;\mathbf{R})$, a given metric g, and a self-dual 2-form as follows. There is a unique g-self-dual harmonic 2-form $\omega_g \in H_+^2(X;\mathbf{R})$ with $\omega_g^2 = 1$ and corresponding to the positive orientation. Fix a characteristic homology class $k \in H_2(X;\mathbf{Z})$. Given a pair (A, ψ), where A is a connection in the complex line bundle whose first Chern class is the Poincaré dual $\widehat{k} = \frac{i}{2\pi}[F_A]$ of k and ψ a section of the bundle W^+ of self-dual spinors for the associated *spin*^{c} structure, the perturbed Seiberg-Witten equations are:

$$D_A\psi = 0$$
$$F_A^+ = q(\psi) + i\eta$$

where F_A^+ is the self-dual part of the curvature F_A, D_A is the twisted Dirac operator, η is a self-dual 2-form on X, and q is a quadratic function. Write $\mathrm{SW}_{X,g,\eta}(k)$ for the corresponding invariant. As the pair (g, η) varies, $\mathrm{SW}_{X,g,\eta}(k)$ can change only at those pairs (g, η) for which there are solutions with $\psi = 0$. These solutions occur for pairs (g, η) satisfying $(2\pi\widehat{k} + \eta)\cdot\omega_g = 0$. This last equation defines a wall in $H^2(X;\mathbf{R})$.

The point ω_g determines a component of the double cone consisting of elements of $H^2(X;\mathbf{R})$ of positive square. We prefer to work with $H_2(X;\mathbf{R})$. The dual component is determined by the Poincaré dual H of ω_g. (An element $H' \in H_2(X;\mathbf{R})$ of positive square lies in the same component as H if $H'\cdot H > 0$.) If $(2\pi\widehat{k} + \eta)\cdot\omega_g \neq 0$ for a generic η, $\mathrm{SW}_{X,g,\eta}(k)$ is well-defined, and its value depends only on the sign of $(2\pi\widehat{k} + \eta)\cdot\omega_g$. Write $\mathrm{SW}_{X,H}^+(k)$ for $\mathrm{SW}_{X,g,\eta}(k)$ if $(2\pi\widehat{k} + \eta)\cdot\omega_g > 0$ and $\mathrm{SW}_{X,H}^-(k)$ in the other case.

The invariant $\mathrm{SW}_{X,H}(k)$ is defined by $\mathrm{SW}_{X,H}(k) = \mathrm{SW}^+_{X,H}(k)$ if $(2\pi\widehat{k})\cdot\omega_g > 0$, or dually, if $k \cdot H > 0$, and $\mathrm{SW}_{X,H}(k) = \mathrm{SW}^-_{X,H}(k)$ if $H \cdot k < 0$. The wall-crossing formula [15, 16] states that if H', H'' are elements of positive square in $H_2(X;\mathbf{R})$ with $H' \cdot H > 0$ and $H'' \cdot H > 0$, then if $k \cdot H' < 0$ and $k \cdot H'' > 0$,

$$\mathrm{SW}_{X,H''}(k) - \mathrm{SW}_{X,H'}(k) = (-1)^{1+\frac{1}{2}d(k)}$$

where $d(k) = \frac{1}{4}(k^2 - (3\,\mathrm{sign} + 2\,e)(X))$ is the formal dimension of the Seiberg-Witten moduli spaces.

Furthermore, in case $b^- \leq 9$, the wall-crossing formula, together with the fact that $\mathrm{SW}_{X,H}(k) = 0$ if $d(k) < 0$, implies that $\mathrm{SW}_{X,H}(k) = \mathrm{SW}_{X,H'}(k)$ for any H' of positive square in $H_2(X;\mathbf{R})$ with $H \cdot H' > 0$. So in case $b^+_X = 1$ and $b^-_X \leq 9$, there is a well-defined Seiberg-Witten invariant, $\mathrm{SW}_X(k)$.

Possible Classification Schemes. From this point forward and unless otherwise stated all manifolds will be closed and simply-connected. In order to avoid trivial constructions we consider *irreducible* manifolds, i.e. those that cannot be represented as the connected sum of two manifolds except if one factor is a homotopy 4-sphere. (We still do not know if there exist smooth homotopy 4-spheres not diffeomorphic to the standard 4-sphere S^4).

So the existence part of a classification scheme for irreducible smooth (symplectic) 4-manifolds could take the form of determining which $(\chi_h, c, t) \in \mathbf{Z} \times \mathbf{Z} \times \mathbf{Z}_2$ can occur as $(\chi_h(X), c(X), t(X))$ for some smooth (symplectic) 4-manifold X. This is referred to as *the geography problem.* The game plan would be to create techniques to realize all possible lattice points. The uniqueness part of the classification scheme would then be to determine all smooth (symplectic) 4-manifolds with a fixed $(\chi_h(X), c(X), t(X))$ and determine invariants that would distinguish them. Again, the game plan would be to create techniques that preserve the homeomorphism type yet change these invariants.

In the next two sections we will outline what is and is not known about the existence (geography) and uniqueness problems without detailing the techniques. Then we will list the techniques used, determine their interplay, and explore questions that may yield new insight. A companion approach, which we will also discuss towards the end of these lectures, is to start with a particular well-understood class of 4-manifolds and determine how all other smooth (symplectic) 4-manifolds can be constructed from these.

2. Existence

Our current understanding of the geography problem is given by Figure 1 where all known simply-connected smooth irreducible 4-manifolds are plotted as lattice points in the (χ_h, c)-plane. In particular, all known simply-connected irreducible smooth or symplectic 4-manifolds have $0 \leq c < 9\chi_h$ and every lattice point in that region can be realized by a symplectic (hence smooth) 4-manifold.

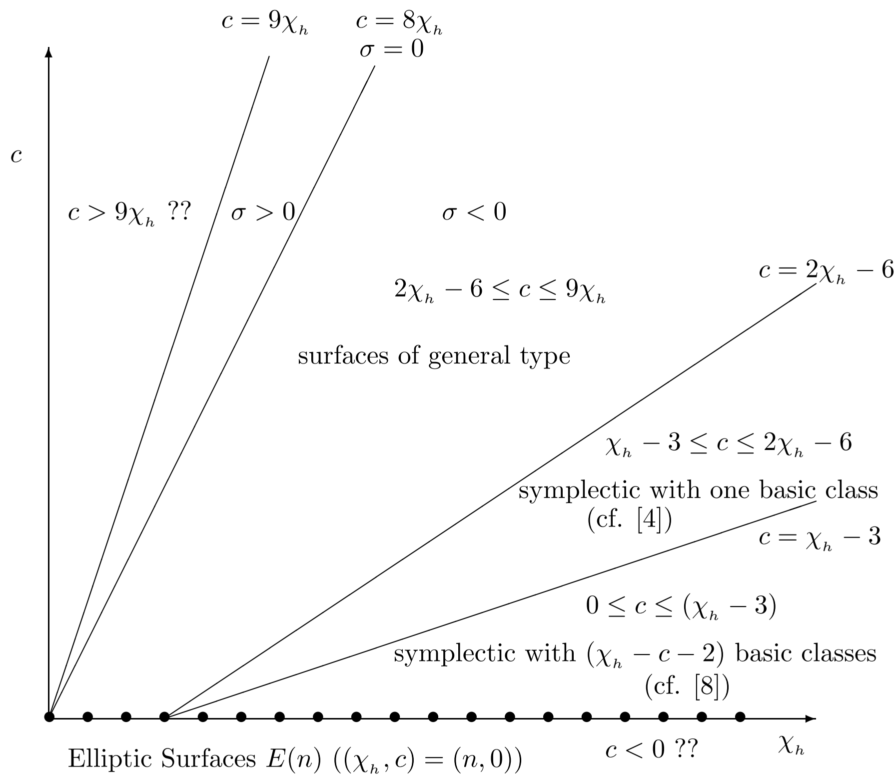


Figure 1

An irreducible 4-manifold need not lie on a lattice point. The issue here is whether $\chi_h \in \mathbf{Z}$ or $\chi_h \in \mathbf{Z}[\frac{1}{2}]$. Note that $\chi_h(X) \in \mathbf{Z}$ iff X has an almost-complex structure. In addition, the Seiberg-Witten invariants are only defined for manifolds with $\chi_h \in \mathbf{Z}$. Since our only technique to determine if a 4-manifold is irreducible is to use the fact that the Seiberg-Witten invariants of a reducible 4-manifold vanish, all known irreducible 4-manifolds have $\chi_h \in \mathbf{Z}$.

Problem 1. *Do there exist irreducible smooth 4-manifolds with $\chi_h \notin \mathbf{Z}$?*

Here the work of Bauer and Furuta [2] on stable homotopy invariants derived from the Seiberg-Witten equations may be useful. To expose our ignorance, consider two copies of the elliptic surface $E(2)$. Remove the neighborhood of a sphere with self-intersection -2 from each and glue together the resulting manifolds along their boundary $\mathbf{RP}^3$ using the orientation reversing diffeomorphism of $\mathbf{RP}^3$. The result has $\chi_h \notin \mathbf{Z}$ and it is unknown if it is irreducible.

All complex manifolds with $c = 9\chi_h > 9$ are non-simply-connected, in particular they are ball quotients. Thus obvious problems are:

Problem 2. *Do there exist irreducible simply connected smooth or symplectic manifolds with $c = 9\chi_h > 9$?*

Problem 3. *Does there exist an irreducible non-complex smooth or symplectic manifold X with $\chi_h > 1$, $c = 9\chi_h$ (with any fundamental group), $\mathcal{SW}_X \neq 0$, and which is not a ball-quotient?*

Problem 4. *Do there exist irreducible smooth or symplectic manifolds with $c > 9\chi_h$?*

The work of Taubes [28] on the relationship between Seiberg-Witten and Gromov-Witten invariants shows that $c \geq 0$ for an irreducible symplectic 4-manifold.

Problem 5. *Do there exist simply connected irreducible smooth manifolds with $c < 0$?*

There appears to be an interesting relationship between the number of Seiberg-Witten basic classes and the pair (χ_h, c). In particular, all known smooth 4-manifolds with $0 \leq c \leq \chi_h - 3$ have at least $\chi_h - c - 2$ Seiberg-Witten basic classes [4]. So

Problem 6. *Does there exist an irreducible smooth manifold X with $0 \leq c(X) \leq \chi_h(X) - 3$ and with fewer than $\chi_h(X) - c(X) - 2$ Seiberg-Witten basic classes? (There is a physics proof that there are no such examples [17].)*

Figure 1 contains no information about the geography of spin 4-manifolds, i.e. manifolds with $t = 0$. For a spin 4-manifold there is the relation $c = 8\chi_h \mod 16$. Almost every lattice point with $c = 8\chi_h \mod 16$ and $0 \leq c < 9\chi_h$ can be be realized by an irreducible spin 4-manifold [21]. Surprisingly not all of the lattice points with $2\chi_h \leq 3(\chi_h - 5)$ can be realized by complex manifolds with $t = 0$ [24], so spin manifolds with $2\chi_h \leq 3(\chi_h - 5)$ provide several examples of smooth irreducible 4-manifolds with $2\chi_h - 6 \leq c < 9\chi_h$ that support no complex structure (cf. [9]). There remains a better understanding of manifolds close to the $c = 9\chi_h$ line, in particular those with $9\chi_h > c \geq 8.76\chi_h$ and not on the lines $c = 9\chi_h - k$ with $k \leq 121$ (cf. [24]).

The techniques used in all these constructions are an artful application of the generalized fiber sum construction (cf. [13]) and the rational blowdown construction [6], which we will discuss later in this lecture.

3. Uniqueness

Here is where we begin to lose control of the classification of smooth 4-manifolds. If a topological 4-manifold admits an irreducible smooth (symplectic) structure that has a smoothly (symplectically) embedded torus with self-intersection zero and with simply-connected complement, then it also admits infinitely many distinct smooth (symplectic) structures and also admits infinitely many distinct smooth structures with no compatible symplectic structure. The basic technique here is the knot-surgery construction of Fintushel-Stern [7], i.e. remove a neighborhood $T^2 \times D^2 = S^1 \times S^1 \times D^2$ of this torus and replace it with $S^1 \times S^3 \setminus K$ where K is a knot in S^3. As we will point out later, the resulting smooth structures are distinguished by the Alexander polynomial of the knot K. There are no known examples of (simply-connected) smooth or symplectic 4-manifolds with $\chi_h > 1$ that do not admit such tori. Hence, there are no known smooth or symplectic 4-manifolds with $\chi_h > 1$ that admit finitely many smooth or symplectic structures. Thus, Problem 7. *Do there exist irreducible smooth (symplectic) 4-manifolds with $\chi_h > 1$ that do not admit a smoothly (symplectically) embedded torus with self-intersection 0 and simply-connected complement?*

All of the constructions used for the geography problem with $\chi_h > 1$ naturally contain such tori, so the only hope is to find manifolds where these constructions have yet to work, i.e. those with $8.76 < c \leq 9\chi_h$, that do not contain such tori.

Problem 8. *Do manifolds with $c = 9\chi_h$ admit exotic smooth structures?*

The situation for $\chi_h = 1$ is potentially more interesting and may yield phenomena not shared by manifolds with $\chi_h > 1$. For example, the complex projective plane $\mathbf{CP}^2$ has $c = 9\chi_h = 9$ and is simply-connected. It is also known that $\mathbf{CP}^2$ as a smooth manifold has a unique symplectic structure [27,28]. Thus, a fundamental question that still remains is

Problem 9. *Does the complex projective plane $\mathbf{CP}^2$ admit exotic smooth structures?*

Problem 10. *What is the smallest m for which $\mathbf{CP}^2 \# m\, \overline{\mathbf{CP}}^2$ admits an exotic smooth structure?*

The primary reason for our ignorance here is that for $c > 1$ (i.e. $m < 9$), these manifolds do not contain homologically essential tori with zero self-intersction. Since the rational elliptic surface $E(1) \cong \mathbf{CP}^2 \# 9\, \overline{\mathbf{CP}}^2$ admits tori with self-intersection zero, it has infinitely many distinct smooth structures. In the late 1980's Dieter Kotschick [14] proved that the Barlow surface, which was known to be homeomorphic to $\mathbf{CP}^2 \# 8\overline{\mathbf{CP}}^2$, is not diffeomorphic to it. In the following years the subject of simply connected smooth 4-manifolds with $m < 8$ languished because of a lack of suitable examples. However, largely due to a beautiful paper of Jongil Park [22], who found the first examples of exotic simply connected 4-manifolds with $m = 7$, interest was revived. Shortly after this conference ended, Peter Ozsváth and Zoltán Szabó proved that Park's manifold is minimal [20] by computing its Seiberg-Witten invariants. Then András Stipsicz and Zoltán Szabó used a technique similar to Park's to construct an exotic manifold with $m = 6$ [25]. The underlying technique in these constructions is an artful use of the rational blowdown construction.

Since $\mathbf{CP}^2 \# m\, \overline{\mathbf{CP}}^2$ for $m < 9$ does not contain smoothly embedded tori with self-intersection zero, it has not been known whether it can have an infinite family of smooth structures. Most recently, Fintushel and Stern [11], introduced a new technique which was used to show that for $6 \leq m \leq 8$, $\mathbf{CP}^2 \# m\, \overline{\mathbf{CP}}^2$ does indeed have an infinite family of smooth structures, and, in addition, none of these smooth structures support a compatible symplectic structure. These are the first examples of manifolds that do not contain homologically essential tori, yet have infinitely many distinct smooth structures. Park, Stipsicz, and Szabó [23], and independently Fintushel and Stern [11] used this construction to show that $m = 5$ also has an infinite family of smooth structures none of which support a compatible symplectic structure (cf. [11]). The basic technique in these constructions is a prudent blend of the knot surgery and rational blowdown constructions.

As is pointed out in [25], the Seiberg-Witten invariants will never distinguish more than two distinct irreducible symplectic structures on $\mathbf{CP}^2 \# m\, \overline{\mathbf{CP}}^2$ for $m <$

9. Basically, this is due to the fact that if there is more than one pair of basic classes for a $\chi_h = 1$ manifold, then it is not minimal. So herein lies one of our best hopes for finiteness in dimension 4.

PROBLEM 11. *Does* $\mathbf{CP}^2 \# m\, \overline{\mathbf{CP}}^2$ *for* $m < 9$ *support more than two irreducible symplectic structures that are not deformation equivalent?*

4. The techniques used for the construction of all known simply-connected smooth and symplectic 4-manifolds

The construction of simply-connected smooth or symplectic 4-manifolds sometimes takes the form of art rather than science. This is exposed by the lack of success in proving structural theorems or uncovering any finite phenomena in dimension 4. In this lecture we will list all the constructions used in building the 4-manifolds necessary for the results of the first two sections and try to bring all the unusual phenomena in dimension 4 into a framework that will allow us to at least understand those surgical operations that one performs to go from one smooth structure on a given simply-connected 4-manifold to any other smooth structure. This will take the form of understanding a variety of cobordisms between 4-manifolds.

Here is the list of constructions used in the first two sections.

generalized fiber sum: Assume two 4-manifolds X_1 and X_2 each contain an embedded genus g surface $F_j \subset X_j$ with self-intersection 0. Identify tubular neighborhoods νF_j of F_j with $F_j \times D^2$ and fix a diffeomorphism $f : F_1 \to F_2$. Then the fiber sum $X = X_1 \#_f X_2$ of (X_1, F_1) and (X_2, F_2) is defined as $X_1 \setminus \nu F_1 \cup_\phi X_2 \setminus \nu F_2$, where ϕ is $f\times$ (complex conjugation) on the boundary $\partial(X_j \setminus \nu F_j) = F_j \times S^1$.

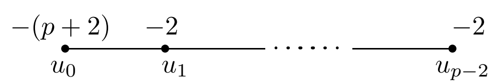
generalized logarithmic transform: Assume that X contains a homologically essential torus T with self-intersection zero. Let νT denote a tubular neighborhood of T. Deleting the interior of νT and regluing $T^2 \times D^2$ via a diffeomorphism $\phi : T^2 \times D^2 \to \partial(X - \text{int } \nu T) = \partial \nu T$ we obtain a new manifold X_ϕ, the generalized logarithmic transform of X along T.

If p denotes the absolute value of the degree of the map $\pi \circ \phi : \{pt\} \times S^1 \to \pi(\partial \nu T) = S^1$, then X_ϕ is called a generalized logarithmic transformation of multiplicity p.

If the complement of T is simply-connected and $t(X) = 1$, then X_ϕ is homeomorphic to X. If the complement of T is simply-connected and $t(X) = 0$, then X_ϕ is homeomorphic to X if p is odd, otherwise X_ϕ has the same c and χ_h but with $t(X_\phi) = 1$.

blowup: Form $X \# \overline{\mathbf{CP}}^2$.

rational blowdown : Let C_p be the smooth 4-manifold obtained by plumbing $(p-1)$ disk bundles over the 2-sphere according to the diagram



Then the classes of the 0-sections have self-intersections $u_0^2 = -(p+2)$ and $u_i^2 = -2$, $i = 1, \ldots, p-2$. The boundary of C_p is the lens space $L(p^2, 1 -$

p) which bounds a rational ball B_p with $\pi_1(B_p) = \mathbf{Z}_p$ and $\pi_1(\partial B_p) \to \pi_1(B_p)$ surjective. If C_p is embedded in a 4-manifold X then the rational blowdown manifold $X_{(p)}$ is obtained by replacing C_p with B_p, i.e., $X_{(p)} = (X \setminus C_p) \cup B_p$ (cf. [6]). If $X \setminus C_p$ is simply connected, then so is $X_{(p)}$

knot surgery: Let X be a 4-manifold which contains a homologically essential torus T of self-intersection 0, and let K be a knot in S^3. Let $N(K)$ be a tubular neighborhood of K in S^3, and let $T \times D^2$ be a tubular neighborhood of T in X. Then the knot surgery manifold X_K is defined by

$$X_K = (X \setminus (T \times D^2)) \cup (S^1 \times (S^3 \setminus N(K))$$

The two pieces are glued together in such a way that the homology class $[\mathrm{pt} \times \partial D^2]$ is identified with $[\mathrm{pt} \times \lambda]$ where λ is the class of a longitude of K. If the complement of T in X is simply-connected, then X_K is homeomorphic to X.

The Seiberg-Witten invariants are sensitive to all of these operations.

generalized logarithmic transform: If T is contained in a node neighborhood, then

$$\mathcal{SW}_{X_\phi} = \mathcal{SW}_X \cdot (s^{-(p-1)} + s^{-(p-3)} + \cdots + s^{(p-1)})$$

where $s = \exp(T/p)$, p the order of the generalized logarithmic transform (cf. [6]).

blowup: The relationship between the Seiberg-Witten invariants of X and its blowup $X \# \overline{\mathbf{CP}}^2$ is referred to as the blowup formula and was given in Witten's original article [30] (cf. [5]). In particular, if e is the homology class of the exceptional curve and $\{B_1, \ldots, B_n\}$ are the basic classes of X, then the basic classes of $X \# \overline{\mathbf{CP}}^2$ are $\{B_1 \pm E, \ldots, B_n \pm E\}$ and $\mathrm{SW}_{X \# \overline{\mathbf{CP}}^2}(B_j \pm E) = \mathrm{SW}_X(B_j)$.

rational blowdown: The Seiberg-Witten invariants of X and $X_{(p)}$ can be compared as follows. The homology of $X_{(p)}$ can be identified with the orthogonal complement of the classes u_i, $i = 0, \ldots, p-2$ in $H_2(X; \mathbf{Z})$, and then each characteristic element $k \in H_2(X_{(p)}; \mathbf{Z})$ has a lift $\widetilde{k} \in H_2(X; \mathbf{Z})$ which is characteristic and for which the dimensions of moduli spaces agree, $d_{X_{(p)}}(k) = d_X(\widetilde{k})$. It is proved in [6] that if $b_X^+ > 1$ then $\mathrm{SW}_{X_{(p)}}(k) = \mathrm{SW}_X(\widetilde{k})$. In case $b_X^+ = 1$, if $H \in H_2^+(X; \mathbf{R})$ is orthogonal to all the u_i then it also can be viewed as an element of $H_2^+(X_{(p)}; \mathbf{R})$, and $\mathrm{SW}_{X_{(p)},H}(k) = \mathrm{SW}_{X,H}(\widetilde{k})$.

knot surgery: If, for example, T is contained in a node neighborhood and $\chi_h(X) > 1$ then the Seiberg-Witten invariant of the knot surgery manifold X_K is given by

$$\mathcal{SW}_{X_K} = \mathcal{SW}_X \cdot \Delta_K(t)$$

where $\Delta_K(t)$ is the symmetrized Alexander polynomial of K and $t = \exp(2[T])$. When $\chi_h = 1$, the Seiberg-Witten invariants of X_K are still completely determined by those of X and the Alexander polynomial $\Delta_K(t)$ [7].

Here T contained in a node neighborhood means that an essential loop on $\partial\nu T$ bounds a disk in the complement with relative self-intersection -1. We sometimes refer to this disk as a vanishing cycle.

In many circumstances, there are formulas for determining the Seiberg-Witten invariants of a fiber sum in terms of the Seiberg-Witten invariants of X_1 and X_2 and how the basic classes intersect the surfaces F_1 and F_2.

Interaction of the operations. While knot surgery appears to be a new operation, the constructions in [7] point out that the knot surgery construction is actually a series of ± 1 generalized logarithmic transformations on *null-homologous tori*. To see this, note that any knot can be unknotted via a sequence of crossing changes, which in turn can be realized as a sequence of ± 1 surgeries on unknotted curves $\{c_1, \dots, c_n\}$ that link the knot algebraically zero times and geometrically twice. When crossed with S^1 this translates to the fact that X can be obtained from X_K via a sequence of ± 1 generalized logarithmic transformations on the *null-homologous tori* $\{S^1 \times c_1, \dots, S^1 \times c_n\}$ in X_K. So the hidden mechanism behind the knot surgery construction is generalized logarithmic transformations on null-homologous tori. The calculation of the Seiberg-Witten invariants is then reduced to understanding how the Seiberg-Witten invariants change under a generalized logarithmic transformation on a null-homologous torus. This important formula is due to Morgan, Mrowka, and Szabó [18] (see also [29]). For this formula fix simple loops α, β, δ on $\partial N(T)$ whose homology classes generate $H_1(\partial N(T))$. If $\omega = p\alpha + q\beta + r\delta$ write $X_T(p,q,r)$ instead of $X_T(\omega)$. Given a class $k \in H_2(X)$:

$$\begin{aligned}(1)\quad \sum_i \mathrm{SW}_{X_T(p,q,r)}(k_{(p,q,r)} + 2i[T]) &= p\sum_i \mathrm{SW}_{X_T(1,0,0)}(k_{(1,0,0)} + 2i[T]) + \\ &+ q\sum_i \mathrm{SW}_{X_T(0,1,0)}(k_{(0,1,0)} + 2i[T]) + r\sum_i \mathrm{SW}_{X_T(0,0,1)}(k_{(0,0,1)} + 2i[T])\end{aligned}$$

In this formula, T denotes the torus which is the core $T^2 \times 0 \subset T^2 \times D^2$ in each specific manifold $X(a,b,c)$ in the formula, and $k_{(a,b,c)} \in H_2(X_T(a,b,c))$ is any class which agrees with the restriction of k in $H_2(X \setminus T \times D^2, \partial)$ in the diagram:

$$\begin{array}{ccc} H_2(X_T(a,b,c)) & \longrightarrow & H_2(X_T(a,b,c), T \times D^2) \\ & & \downarrow \cong \\ & & H_2(X \setminus T \times D^2, \partial) \\ & & \uparrow \cong \\ H_2(X) & \longrightarrow & H_2(X, T \times D^2) \end{array}$$

Let $\pi(a,b,c) : H_2(X_T(a,b,c)) \to H_2(X \setminus T \times D^2, \partial)$ be the composition of maps in the above diagram, and $\pi(a,b,c)_*$ the induced map of integral group rings. Since we are often interested in invariants of the pair (X,T), it is sometimes useful to work with

$$\overline{\mathcal{SW}}_{(X_T(a,b,c),T)} = \pi(a,b,c)_*(\mathcal{SW}_{X_T(a,b,c)}) \in \mathbf{Z}H_2(X \setminus T \times D^2, \partial).$$

The indeterminacy due to the sum in (1) is caused by multiples of $[T]$; so passing to $\overline{\mathcal{SW}}$ removes this indeterminacy, and the Morgan-Mrowka-Szabó formula becomes

$$(2)\quad \overline{\mathcal{SW}}_{(X_T(p,q,r),T)} = p\overline{\mathcal{SW}}_{(X_T(1,0,0),T)} + q\overline{\mathcal{SW}}_{(X_T(0,1,0),T)} + r\overline{\mathcal{SW}}_{(X_T(0,0,1),T)}.$$

So if we expand the notion of generalized logarithmic transformation to include both homologically essential and null-homologous tori, then we can eliminate the knot surgery construction from our list of essential surgery operations. Thus our list is of essential operations is reduced to

- generalized fiber sum
- generalized logarithmic transformations on a torus with trivial normal bundle
- blowup
- rational blowdown

There are further relationships between these operations. In [6] it is shown that if T is contained in a node neighborhood, then a generalized logarithmic transformation can be obtained via a sequence of blowups and rational blowdowns. (This together with work of Margaret Symington [26] shows that logarithmic transformations ($p \neq 0$) on a symplectic torus results in a symplectic manifold. We do not know of any other proof that a generalized logarithmic transformations on a symplectic torus in a node neighborhood results in a symplectic manifold.) However, it is not clear that a rational blowdown is always the result of blowups and logarithmic transforms.

Rational blowdown changes the topology of the manifold X; while χ_h remains the same, c is decreased by $p-3$. So, an obvious problem would be

Problem 12. *Are any two homeomorphic simply-connected smooth 4-manifolds related via a sequence of generalized logarithmic transforms on tori?*

As already pointed out, there are two cases.

(1) T is essential in homology
(2) T is null-homologous

This leads to:

Problem 13. *Can a generalized logarithmic transform on a homologicaly essential torus be obtained via a sequence of generalized logarithmic transforms on null-homologous tori?*

For the rest of the lecture we will discuss these last two problems.

5. Cobordisms between 4-manifolds

.

Let X_1 and X_2 be two homeomorphic simply-connected smooth 4-manifolds. Early results of C.T.C. Wall show that there is an h–cobordism W^5 between X_1 and X_2 obtained from $X_1 \times I$ by attaching n 2−handles and n 3−handles. A long standing problem that still remains open is:

Problem 14. *Can W^5 can be chosen so that $n = 1$.*

Let's explore the consequences if we can assume $n = 1$. We can then describe the h−cobordism W^5 as follows. First, let W_1 be the cobordism from X_1 to $X_1 \# S^2 \times S^2$ given by attaching the 2−handle to X_1. To complete W^5 we then would add the 3−handle. Dually, this is equivalent to attaching a 2−handle to X_2. So let W_2 be the cobordism from X_2 to $X_2 \# S^2 \times S^2$ given by attaching this 2−handle to X_2. Then $W^5 = W_1 \cup_f (-W_2)$ for a suitable diffeomorphism $f\colon X_1 \# S^2 \times S^2 \to X_2 \# S^2 \times S^2$. Let A be any of the standard spheres in $S^2 \times S^2$.

Then the complexity of the h-cobordism can be measured by the type k, which is half the minimum of the number of intersection points between A and $f(A)$ (as $A \cdot f(A) = 0$ there are k positive intersection points and k negative intersection points). This complexity has been studied in [19]. A key observation is that if $k = 1$, then a neighborhood of $A \cup f(A)$ is diffeomorphic to an embedding of twin spheres in S^4 and that its boundary is the three-torus T^3. A further observation is that X_2 is then obtained from X_1 by removing a neighborhood of a null-homologous torus T embedded in X_1 (with trivial normal bundle) and sewing it back in differently. Thus when $k = 1$, X_2 is obtained from X_1 by a generalized generalized logarithmic transform on a *null-homologous* torus.

This points out that the answers to Problems 12 and 13 are related to the complexity k of h-cobordisms. We expect that the answer to Problem 13 is NO and that ordinary generalized logarithmic transforms on homologically essential tori will provide examples of homeomorphic X_1 and X_2 that require h–cobordisms with arbitrarily large complexity.

Independent of this, an important next step is to study complexity $k > 1$ h-cobordisms. Here, new surgical techniques are suggested. In particular, the neighborhood of $A \cup f(A)$ above is diffeomorphic to the neighborhood N' of two $2-$spheres embedded in S^4 with $2k$ points of intersection. Let N be obtained from N' with one of the $2-$spheres surgered out. Then it can be shown that X_1 is obtained from X_2 by removing an embedding of N' and regluing along a diffeomorphism of its boundary. This could lead to a useful generalization of logarithmic transforms along null-homologous tori. It would then be important to compute its effect on the Seiberg-Witten invariants, and reinterpret generalized logarithmic transforms from this point of view.

Round handlebody cobordisms. Suppose that X_1 and X_2 are two manifolds with the same c and χ_h. It follows from early work of Asimov [1] that there is a round handlebody cobordism W between X_1 and X_2. Thus X_1 can be obtained from X_2 by attaching a sequence of round $1-$handles and round $2-$handles. A round handle is just S^1 times a handle in one lower dimension. So for us a round $r-$handle is a copy of $S^1 \times (D^r \times D^{4-r})$ attached along $S^1 \times (S^{r-1} \times D^{4-r})$ (see [1] for definitions).

Problem 15. *Can W be chosen so that there are no round $1-$handles?*

For a moment, suppose that the answer to Problem 15 is Yes. Then W would consist of only round $2-$handles. It then follows that X_2 would be obtained from X_1 via a sequence of generalized logarithmic transforms on tori. Thus the answer to Problem 15 is tightly related to Problem 12.

Note that if X_1 and X_2 are round handlebody cobordant, then the only invariant preventing them from being homeomorphic is whether $t(X_1) = t(X_2)$. So suppose $t(X_1) = 0$ and $t(X_2) = 1$. If the answer to Problem 15 were yes, then one could change the second Stiefel-Whitney class via a sequence of generalized logarithmic transforms on tori. By necessity these tori cannot be null-homologous. So understanding new surgical operations that will change t without changing c, χ_h, and preserving the Seiberg-Witten invariants should provide new insights.

Problem 16. *Suppose two simply-connected smooth 4-manifolds have the same c, χ_h, number of Seiberg-Witten basic classes, and different t. Determine surgical operations that will transform one to the other.*

There are explicit examples of this phenomena amongst complex surfaces, e.g. two Horikawa surfaces with the same c and χ_h, but different t.

6. Modifying symplectic 4-manifolds

To finish up this lecture, we point out that all known constructions of (simply-connected) non-symplectic 4-manifolds can be obtained from symplectic 4-manifolds by performing logarithmic transforms on null-homologous Lagrangian tori with non-vanishing framing defect (cf. [10]). Let's look at a specific example of this phenomena. In particular, let's consider $E(n)_K$.

The elliptic surface $E(n)$ is the double branched cover of $S^2 \times S^2$ with branch set equal to four disjoint copies of $S^2 \times \{\mathrm{pt}\}$ together with $2n$ disjoint copies of $\{\mathrm{pt}\} \times S^2$. The resultant branched cover has $8n$ singular points (corresponding to the double points in the branch set), whose neighborhoods are cones on $\mathbf{RP}^3$. These are desingularized in the usual way, replacing their neighborhoods with cotangent bundles of S^2. The result is $E(n)$. The horizontal and vertical fibrations of $S^2 \times S^2$ pull back to give fibrations of $E(n)$ over $\mathbf{CP}^1$. A generic fiber of the vertical fibration is the double cover of S^2, branched over 4 points — a torus. This describes an elliptic fibration of $E(n)$. The generic fiber of the horizontal fibration is the double cover of S^2, branched over $2n$ points, and this gives a genus $n-1$ fibration on $E(n)$. This genus $n-1$ fibration has four singular fibers which are the preimages of the four $S^2 \times \{\mathrm{pt}\}$'s in the branch set together with the spheres of self-intersection -2 arising from desingularization. The generic fiber T of the elliptic fibration meets a generic fiber Σ_{n-1} of the horizontal fibration in two points, $\Sigma_{n-1} \cdot T = 2$.

Now let K be a fibered knot of genus g, and fix a generic elliptic fiber T_0 of $E(n)$. Then in the knot surgery manifold

$$E(n)_K = (E(n) \setminus (T_0 \times D^2)) \cup (S^1 \times (S^3 \setminus N(K)),$$

each normal 2-disk to T_0 is replaced by a fiber of the fibration of $S^3 \setminus N(K)$ over S^1. Since T_0 intersects each generic horizontal fiber twice, we obtain a 'horizontal' fibration

$$h : E(n)_K \to \mathbf{CP}^1$$

of genus $2g + n - 1$.

This fibration also has four singular fibers arising from the four copies of $S^2 \times \{\mathrm{pt}\}$ in the branch set of the double cover of $S^2 \times S^2$. Each of these gets blown up at $2n$ points in $E(n)$, and the singular fibers each consist of a genus g surface Σ_g of self-intersection $-n$ and multiplicity 2 with $2n$ disjoint 2-spheres of self-intersection -2, each meeting Σ_g transversely in one point. The monodromy around each singular fiber is (conjugate to) the diffeomorphism of Σ_{2g+n-1} which is the deck transformation η of the double cover of Σ_g, branched over $2n$ points. Another way to describe η is to take the hyperelliptic involution ω of Σ_{n-1} and to connect sum copies of Σ_g at the two points of a nontrivial orbit of ω. Then ω extends to the involution η of Σ_{2g+n-1}.

The fibration which we have described is not Lefschetz since the singularities are not simple nodes. However, it can be perturbed locally to be Lefschetz.

So in summary, if K is a fibered knot whose fiber has genus g, then $E(n)_K$ admits a locally holomorphic fibration (over $\mathbf{CP}^1$) of genus $2g + n - 1$ which has exactly four singular fibers. Furthermore, this fibration can be deformed locally to be Lefschetz.

There is another way to view these constructions. Consider the branched double cover of $\Sigma_g \times S^2$ whose branch set consists of two disjoint copies of $\Sigma_g \times \{\mathrm{pt}\}$ and $2n$ disjoint copies of $\{\mathrm{pt}\} \times S^2$. After desingularizing as above, one obtains a complex surface denoted $M(n,g)$. Once again, this manifold carries a pair of fibrations. There is a genus $2g+n-1$ fibration over S^2 and an S^2 fibration over Σ_g.

Consider first the S^2 fibration. This has $2n$ singular fibers, each of which consists of a smooth 2-sphere E_i, $i = 1, \dots, 2n$, of self-intersection -1 and multiplicity 2, together with a pair of disjoint spheres of self-intersection -2, each intersecting E_i once transversely. If we blow down E_i we obtain again an S^2 fibration over Σ_g, but the ith singular fiber now consists of a pair of 2-spheres of self-intersection -1 meeting once, transversely. Blowing down one of these gives another S^2 fibration over Σ_g, with one less singular fiber. Thus blowing down $M(n,g)$ $4n$ times results in a manifold which is an S^2 bundle over Σ_g. This shows that (if $n > 0$) $M(n,g)$ is diffeomorphic to $(S^2 \times \Sigma_g)\#4n\,\overline{\mathbf{CP}}^2$.

The genus $2g+n-1$ fibration on $M(n,g)$ has 2 singular fibers. As above, these fibers consist of a genus g surface Σ_g of self-intersection $-n$ and multiplicity 2 with $2n$ disjoint 2-spheres of self-intersection -2, each meeting Σ_g transversely in one point. The monodromy of the fibration around each of these fibers is the deck transformation of the double branched cover of Σ_g. This is just the map η described above.

Let φ be a diffeomorphism of $\Sigma_g \setminus D^2$ which is the identity on the boundary. For instance, φ could be the monodromy of a fibered knot of genus g. There is an induced diffeomorphism Φ of $\Sigma_{2g+n-1} = \Sigma_g\#\Sigma_{n-1}\#\Sigma_g$ which is given by φ on the first Σ_g summand and by the identity on the other summands. Consider the twisted fiber sum

$$M(n,g)\#_\Phi M(n,g) = \{M(n,g)\setminus(D^2\times\Sigma_{2g+n-1})\}\cup_{\mathrm{id}\times\Phi}\{M(n,g)\setminus(D^2\times\Sigma_{2g+n-1})\}$$

where fibered neighborhoods of generic fibers Σ_{2g+n-1} have been removed from the two copies of $M(n,g)$, and they have been glued by the diffeomorphism $\mathrm{id}\times\Phi$ of $S^1 \times \Sigma_{2g+n-1}$.

In the case that φ is the monodromy of a fibered knot K, it can be shown that $M(n,g)\#_\Phi M(n,g)$ is the manifold $E(n)_K$ with the genus $2g+n-1$ fibration described above. To see this, we view S^2 as the base of the horizontal fibration. Then it suffices to check that the total monodromy map $\pi_1(S^2\setminus 4\,\text{points}) \to \mathrm{Diff}(\Sigma_{2g+n-1})$ is the same for each. It is not difficult to see that if we write the generators of $\pi_1(S^2 \setminus 4\,\text{points})$ as α, β, γ with α and β representing loops around the singular points of, say, the image of the first copy of $M(n,g)$ and basepoint in this image, and γ a loop around a singular point in the image of the second $M(n,g)$ then the monodromy map μ satisfies $\mu(\alpha) = \eta$, $\mu(\beta) = \eta$ and $\mu(\gamma)$ is $\varphi \oplus \omega \oplus \varphi^{-1}$, expressed as a diffeomorphism of $\Sigma_g\#\Sigma_{n-1}\#\Sigma_g$. That this is also the monodromy of $E(n)_K$ follows directly from its construction.

Now let $E(n)_g$ denote $E(n)$ fiber summed with $T^2 \times \Sigma_g$ along an elliptic fiber. The penultimate observation is that $E(n)_K$, viewed as $M(n,g)\#_\Phi M(n,g)$, is then the result of a sequence of generalized logarithmic transforms on null-homologous Lagrangian tori in $E(n)_g$. The effect of these surgeries is to change the monodromy of the genus $n+2g-1$ Lefschetz fibration (over $\mathbf{CP}^1$) on $E(n)_g$. This is accomplished by doing a $1/n$, with respect to the natural Lagrangian framing, generalized logarithmic transform on these Lagrangian tori (cf. [9, 10]). The final observation is that if the Lagrangian framing of these tori differs from the null-homologous framing (cf. [10]), then a $1/n$ log transformations on T with respect to the null-homologous framing can be shown, by computing Seiberg-Witten invariants, to result in non-symplectic 4-manifolds. Careful choices of these tori and framings will result in manifolds homotopy equivalent to $M(n,g)\#_{\Phi}M(n,g)$ (cf. [9]).

References

[1] Daniel Asimov, *Round handles and homotopy of nonsingular vector fields*, Bull. Amer. Math. Soc. **81** (1975), 417–419. MR0370658 (51 #6884)

[2] Stefan Bauer and Mikio Furuta, *A stable cohomotopy refinement of Seiberg-Witten invariants. I*, Invent. Math. **155** (2004), 1–19. MR2025298

[3] S. K. Donaldson, *An application of gauge theory to four-dimensional topology*, J. Differential Geom. **18** (1983), 279–315. MR710056 (85c:57015)

[4] Ronald J. Fintushel, Jongil Park, and Ronald J. Stern, *Rational surfaces and symplectic 4-manifolds with one basic class*, Algebr. Geom. Topol. **2** (2002), 391–402 (electronic). MR1917059 (2003f:57063)

[5] Ronald Fintushel and Ronald J. Stern, *Immersed spheres in* 4*-manifolds and the immersed Thom conjecture*, Turkish J. Math. **19** (1995), 145–157. MR1349567 (96j:57036)

[6] ———, *Rational blowdowns of smooth* 4*-manifolds*, J. Differential Geom. **46** (1997), 181–235. MR1484044 (98j:57047)

[7] ———, *Knots, links, and* 4*-manifolds*, Invent. Math. **134** (1998), 363–400. MR1650308 (99j:57033)

[8] ———, *The canonical class of a symplectic 4-manifold*, Turkish J. Math. **25** (2001), 137–145. MR1829084 (2002g:57053)

[9] ———, *Families of simply connected 4-manifolds with the same Seiberg-Witten invariants*, Topology **43** (2004), 1449–1467. MR2081432

[10] ———, *Invariants for Lagrangian Tori*, Geometry and Topology **8** (2004), 949–968.

[11] ———, *Double node neighborhoods and families of simply connected 4-manifolds with* $b^+=1$ (2004), http://front.math.ucdavis.edu/math.GT/0412126 (preprint).

[12] Michael Hartley Freedman, *The topology of four-dimensional manifolds*, J. Differential Geom. **17** (1982), 357–453. MR679066 (84b:57006)

[13] Robert E. Gompf and András I. Stipsicz, 4*-manifolds and Kirby calculus*, Graduate Studies in Mathematics, vol. 20, American Mathematical Society, Providence, RI, 1999, ISBN 0-8218-0994-6. MR1707327 (2000h:57038)

[14] Dieter Kotschick, *On manifolds homeomorphic to* $\mathbf{C}\mathrm{P}^2\#8\overline{\mathrm{CP}}^2$, Invent. Math. **95** (1989), 591–600. MR979367 (90a:57047)

[15] P. B. Kronheimer and T. S. Mrowka, *The genus of embedded surfaces in the projective plane*, Math. Res. Lett. **1** (1994), 797–808. MR1306022 (96a:57073)

[16] T. J. Li and A. Liu, *Symplectic structure on ruled surfaces and a generalized adjunction formula*, Math. Res. Lett. **2** (1995), 453–471. MR1355707 (96m:57052)

[17] Marcos Mariño, Gregory Moore, and Grigor Peradze, *Superconformal invariance and the geography of four-manifolds*, Comm. Math. Phys. **205** (1999), 691–735. MR1711332 (2001h:57037)

[18] John W. Morgan, Tomasz S. Mrowka, and Zoltán Szabó, *Product formulas along* T^3 *for Seiberg-Witten invariants*, Math. Res. Lett. **4** (1997), 915–929. MR1492130 (99f:57039)

[19] John W. Morgan and Zoltán Szabó, *On h-cobordisms and Seiberg-Witten invariants*, Topics in Symplectic 4-Manifolds (Irvine, CA, 1996), First Int. Press Lect. Ser., I, Internat. Press, Cambridge, MA, 1998, pp. 117–124. MR1635699 (99h:57061)

[20] Peter Ozsváth and Zoltán Szabó, *On Park's exotic smooth four-manifolds* (2004), http://front.math.ucdavis.edu/math.GT/0411218 (preprint).

[21] Jongil Park, *The geography of Spin symplectic 4-manifolds*, Math. Z. **240** (2002), 405–421. MR1900318 (2003c:57030)

[22] ———, *Simply connected symplectic* 4*-manifolds with* $b_2^+ = 1$ *and* $c_1^2 = 2$ (2004), http://front.math.ucdavis.edu/math.GT/0311395 (to appear in Invent. Math.)

[23] Jongil Park, Andras Stipsicz, and Zoltán Szabó, *Exotic smooth structures on* $\mathbf{CP}^2\#5\,\overline{\mathbf{CP}}^2$ (2004), http://front.math.ucdavis.edu/math.GT/0412216 (preprint).

[24] Ulf Persson, Chris Peters, and Gang Xiao, *Geography of spin surfaces*, Topology **35** (1996), 845–862. MR1404912 (98h:14046)

[25] Andras Stipsicz and Zoltán Szabó, *An exotic smooth structure on* $\mathbf{CP}^2\#6\overline{\mathbf{CP}}^2$ (2004), http://front.math.ucdavis.edu/math.GT/0411258 (preprint).

[26] Margaret Symington, *Symplectic rational blowdowns*, J. Differential Geom. **50** (1998), 505–518. MR1690738 (2000e:57043)

[27] Clifford Henry Taubes, *The Seiberg-Witten and Gromov invariants*, Math. Res. Lett. **2** (1995), 221–238. MR1324704 (96a:57076)

[28] ———, *Seiberg-Witten and Gromov invariants for symplectic* 4*-manifolds*, First International Press Lecture Series, vol. 2, International Press, Somerville, MA, 2000, ISBN 1-57146-061-6, Edited by Richard Wentworth. MR1798809 (2002j:53115)

[29] ———, *The Seiberg-Witten invariants and 4-manifolds with essential tori*, Geom. Topol. **5** (2001), 441–519 (electronic). MR1833751 (2002d:57025)

[30] Edward Witten, *Monopoles and four-manifolds*, Math. Res. Lett. **1** (1994), 769–796. MR1306021 (96d:57035)

Department of Mathematics, University of California, Irvine, CA 92697
E-mail address: rstern@uci.edu

Clay Mathematics Proceedings
Volume 5, 2006

A Note on Symplectic 4-manifolds with $b_2^+ = 1$ and $K^2 \geq 0$

Jongil Park

ABSTRACT. In this article we survey recent results on the existence problem of simply connected symplectic 4-manifolds with $b_2^+ = 1$ and $K^2 \geq 0$. We also investigate exotic smooth structures on rational surfaces $\mathbf{C}P^2 \sharp n \overline{\mathbf{C}P}^2$.

1. Introduction

Since S. Donaldson introduced gauge theory in the study of smooth 4-manifolds ([**DK**]), various techniques have been developed to produce new families of symplectic 4-manifolds which were not known before. For example, R. Gompf constructed many symplectic 4-manifolds using fiber-sum surgery ([**G**]), and R. Fintushel and R. Stern also constructed a family of symplectic 4-manifolds using rational blow-down surgery and 0-framed surgery ([**FS1**], [**FS2**]). Recently, using these techniques, the author constructed new simply connected symplectic 4-manifolds with $b_2^+ = 1$ and $K^2 \geq 0$ ([**P1**], [**P2**]).

The aim of this article is to survey these constructions which appeared in [**P1**] and [**P2**]. Let us start with classifying symplectic 4-manifolds with $b_2^+ = 1$. It is the usual convention that the set of symplectic 4-manifolds with $b_2^+ = 1$ is classified by the sign of the square K^2 of the canonical class K associated to a compatible almost complex structure on a given symplectic 4-manifold. In contrast to the fact that every minimal symplectic 4-manifold with $b_2^+ > 1$ satisfies $K^2 \geq 0$ ([**T**]), there are many symplectic 4-manifolds with $b_2^+ = 1$ satisfying $K^2 < 0$, $K^2 = 0$ and $K^2 > 0$ respectively. It was known that only irrational ruled surfaces are minimal symplectic 4-manifolds with $K^2 < 0$ ([**MS**]). Next, in the category of $K^2 = 0$, most known symplectic 4-manifolds are complex surfaces such as rational or ruled surfaces, Dolgachev surfaces and Enriques surface. Even though there are some non-simply connected and non-complex symplectic 4-manifolds such as some torus bundles over the torus and $S^1 \times M$ with a fibered 3-manifold $M(\neq S^1 \times S^2)$, little has been known about simply connected minimal symplectic 4-manifolds which do not admit a complex structure. In Section 3 we confirm that most homotopy elliptic surfaces $\{E(1)_K \,|\, K \text{ is a fibered knot in } S^3\}$ constructed by R. Fintushel and R. Stern in [**FS2**] are simply connected minimal symplectic 4-manifolds which cannot admit a complex structure. The main technique involved

2000 *Mathematics Subject Classification.* Primary 57R17, 57R57; Secondary 14J26.

The author was supported by grant No. R14-2002-007-01002-0 from KOSEF.

in the proof is a computation of the Seiberg-Witten invariant obtained by a small generic perturbation of the Seiberg-Witten equations. Finally, in the case when $K^2 > 0$, until now the only known simply connected symplectic 4-manifolds with $b_2^+ = 1$ were rational surfaces such as $\mathbf{CP}^2, S^2 \times S^2$ and $\mathbf{C}P^2 \sharp n \overline{\mathbf{C}P}^2 (n \leq 8)$ and the Barlow surface. In Section 4 we present new simply connected symplectic 4-manifolds with $b_2^+ = 1$ and $1 \leq K^2 \leq 2$. The main technique involved in the construction is a rational blow-down surgery.

2. Preliminaries

In this section we briefly review Seiberg-Witten theory for smooth 4-manifolds. In particular, we focus on the Seiberg-Witten invariants of 4-manifolds with $b_2^+ = 1$ (see [**M**] for details).

Let X be a closed, oriented smooth 4-manifold with $b_2^+ > 0$ and a fixed metric g, and let L be a characteristic line bundle on X, i.e. $c_1(L)$ is an integral lift of $w_2(X)$ (We assume that $H_1(X; \mathbf{Z})$ has no 2-torsion.) Then L determines a $Spin^c$-structure on X which induces a complex spinor bundle $W \cong W^+ \oplus W^-$, where $W^{\pm}$ are the associated $U(2)$-bundles on X such that $\det(W^{\pm}) \cong L$. Note that the Levi-Civita connection on TX together with a unitary connection A on L induces a connection $\nabla_A : \Gamma(W^+) \to \Gamma(T^*X \otimes W^+)$. This connection, followed by Clifford multiplication, induces a $Spin^c$-Dirac operator $D_A : \Gamma(W^+) \to \Gamma(W^-)$. Then, for each self-dual 2-form $h \in \Omega^2_{+_g}(X : \mathbf{R})$ the following pair of equations for a unitary connection A on L and a section Ψ of $\Gamma(W^+)$ are called the *perturbed Seiberg-Witten equations*:

$$(2.1) \qquad (SW_{g,h}) \begin{cases} D_A \Psi &= 0 \\ F_A^{+_g} &= i(\Psi \otimes \Psi^*)_0 + ih \, . \end{cases}$$

Here $F_A^{+_g}$ is the self-dual part of the curvature of A with respect to a metric g on X and $(\Psi \otimes \Psi^*)_0$ is the trace-free part of $(\Psi \otimes \Psi^*)$. The gauge group $\mathcal{G} := Aut(L) \cong Map(X, S^1)$ acts on the space $\mathcal{A}_X(L) \times \Gamma(W^+)$ by

$$g \cdot (A, \Psi) = (g \circ A \circ g^{-1}, g \cdot \Psi).$$

Since the set of solutions is invariant under the action, it determines an orbit space, called the *Seiberg-Witten moduli space*, denoted by $M_{X,g,h}(L)$, whose formal dimension is

$$\dim M_{X,g,h}(L) = \frac{1}{4}(c_1(L)^2 - 3\sigma(X) - 2e(X))$$

where $\sigma(X)$ is the signature of X and $e(X)$ is the Euler characteristic of X. Note that if $b_2^+(X) > 0$ and $M_{X,g,h}(L)$ is not empty then for a generic self-dual 2-form h on X the moduli space $M_{X,g,h}(L)$ contains no reducible solutions, hence it is a compact smooth manifold of the given dimension.

Definition 2.1. The *Seiberg-Witten invariant* (*in brief, SW-invariant*) of a smooth 4-manifold X with $b_2^+ > 0$ is a function $SW_X : Spin^c(X) \to \mathbf{Z}$ defined by

$$(2.2) \qquad SW_X(L) := \begin{cases} <\beta^{d_L}, [M_{X,g,h}]> & \text{if } \dim M_{X,g,h}(L) := 2d_L \geq 0 \\ & \text{is nonnegative and even} \\ 0 & \text{otherwise}\, . \end{cases}$$

Here β is a generator of $H^2(\mathcal{B}_X^*(L); \mathbf{Z})$ which is the first Chern class of the S^1-bundle

$$\widetilde{\mathcal{B}}_X^*(L) = \mathcal{A}_X(L) \times (\Gamma(W^+) - \{0\})/Aut^0(L) \longrightarrow \mathcal{B}_X^*(L)$$

where $Aut^0(L)$ consists of gauge transformations which are the identity on the fiber of L over a fixed base point in X.

If $b_2^+(X) > 1$, the SW-invariant, denoted by $SW_X = \sum SW_X(L) \cdot e^{c_1(L)}$, is a diffeomorphism invariant, i.e. SW_X does not depend on the choice of a metric on X or a generic perturbation. Furthermore, only finitely many $Spin^c$-structures on X have non-zero Seiberg-Witten invariant. We say that the characteristic line bundle L or equivalently, its Chern class $c_1(L) \in H^2(X; \mathbf{Z})$, is a *SW-basic class* of X if $SW_X(L) \neq 0$.

When $b_2^+(X) = 1$, the SW-invariant $SW_X(L)$ defined in (2.2) above depends not only on a metric g but also on a self-dual 2-form h. Because of this fact, there are several types of Seiberg-Witten invariants for a smooth 4-manifold with $b_2^+ = 1$ depending on how the Seiberg-Witten equations are perturbed. We introduce three types of SW-invariants and investigate how they are related. In (2.1) we first allow all metrics g and self-dual 2-forms h. Then the SW-invariant $SW_X(L)$ defined in (2.2) above has generically two values which are determined by the sign of $(2\pi c_1(L) + [h]) \cdot [\omega_g]$, where ω_g is the unique g-self-dual harmonic 2-form of norm one lying in the (preassigned) positive component of $H^2_{+_g}(X; \mathbf{R})$. We denote the SW-invariant for the metric g and generic self-dual 2-form h satisfying $(2\pi c_1(L) + [h]) \cdot [\omega_g] > 0$ by $SW_X^+(L)$ and denote the other one by $SW_X^-(L)$. Secondly one may perturb the Seiberg-Witten equations by adding only a small generic self-dual 2-form $h \in \Omega^2_{+_g}(X; \mathbf{R})$, so that one can define the SW-invariants as in (2.2) above. In this case we denote the SW-invariant for a metric g satisfying $(2\pi c_1(L)) \cdot [\omega_g] > 0$ by $SW_X^{\circ,+}(L)$ and we denote the other one by $SW_X^{\circ,-}(L)$. Note that, if it exists, $SW_X^{\circ,\pm}(L) = SW_X^{\pm}(L)$. But it sometimes happens that the sign of $(2\pi c_1(L)) \cdot [\omega_g]$ is the same for all metrics, so that there exists only one SW-invariant obtained by a small generic perturbation of the Seiberg-Witten equations. In such a case we define the SW-invariant of L on X by

$$SW_X^\circ(L) := \begin{cases} SW_X^{\circ,+}(L) & \text{if } 2\pi c_1(L) \cdot [\omega_g] > 0 \\ SW_X^{\circ,-}(L) & \text{if } 2\pi c_1(L) \cdot [\omega_g] < 0\,. \end{cases}$$

If $SW_X^\circ(L) \neq 0$, we call the corresponding $c_1(L)$ (or L) a *SW-basic class* of X. Then the Seiberg-Witten invariant of X, denoted by $SW_X^\circ = \sum SW_X^\circ(L) \cdot e^{c_1(L)}$, will also be a diffeomorphism invariant. Furthermore we can extend many results obtained for smooth 4-manifolds with $b_2^+ > 1$ to this case. For example, if X is a simply connected closed smooth 4-manifold with $b_2^+ = 1$ and $b_2^- \leq 9$, then there are only finitely many characteristic line bundles L on X such that $SW_X^\circ(L) \neq 0$. Finally we introduce one more type of Seiberg-Witten invariants for $b_2^+ = 1$. Given a fixed cohomology class $[x] \in H^2(X; \mathbf{Z})$ with $[x] \cdot [x] \geq 0$, one may divide the set of metrics and self-dual 2-forms into two classes according to the sign of $proj_{+_g}(2\pi c_1(L) + [h]) \cdot [x]$, where $proj_{+_g}$ is the projection of $\Omega^2(X; \mathbf{R})$ onto the space $H^2_{+_g}(X; \mathbf{R})$ of g-self-dual harmonic 2-forms. In this case we denote the SW-invariant for a metric g and a generic self-dual 2-form h satisfying $proj_{+_g}(2\pi c_1(L) + [h]) \cdot [x] > 0$ by $SW_X^{[x],+}(L)$ and we denote the other one by $SW_X^{[x],-}(L)$. R. Fintushel and R. Stern used this type of SW-invariants with $[x] = [T]$ for $b_2^+ = 1$ in [**FS2**]. Note that $SW_X^{[T],\pm}(L) = SW_X^{\pm}(L)$.

3. Non-complex symplectic 4-manifolds with $b_2^+ = 1$ and $K^2 = 0$

As mentioned in the Introduction, most known simply connected minimal symplectic 4-manifolds with $b_2^+ = 1$ are complex surfaces such as rational or ruled surfaces. Note that there are several ways to characterize rational or ruled surfaces. One way to characterize them is to compute Seiberg-Witten invariants obtained by adding a small generic perturbation to the Seiberg-Witten equations. Explicitly, using the fact that they admit a metric of positive scalar curvature, one can prove the following

Proposition 3.1. *Suppose X is a minimal symplectic 4-manifold with $b_2^+ = 1$ such that its canonical class K is a torsion-free class of non-negative square. Then X is rational or ruled if and only if its Seiberg-Witten invariant SW_X° vanishes.*

Hence, in order to find a minimal symplectic 4-manifold with $b_2^+ = 1$ which is neither rational nor ruled, we first need to find symplectic 4-manifolds whose Seiberg-Witten invariants are non-zero. For such candidates, we choose a family of homotopy elliptic surfaces $\{E(1)_K \,|\, K$ is a fibered knot in $S^3\}$ constructed by Fintushel and Stern in [**FS2**] and we compute their Seiberg-Witten invariants $SW^\circ_{E(1)_K}$. First we briefly review their constructions. Suppose that K is a fibered knot in S^3 with a punctured surface Σ_g° of genus g as fiber. Let M_K be a 3-manifold obtained by performing 0-framed surgery on K, and let m be a meridional circle to K. Then the 3-manifold M_K can be considered as a fiber bundle over the circle with a closed Riemann surface Σ_g as fiber, and there is a smoothly embedded torus $T_m := m \times S^1$ of square 0 in $M_K \times S^1$. Thus $M_K \times S^1$ fibers over $S^1 \times S^1$ with Σ_g as fiber and with $T_m = m \times S^1$ as section. It is a theorem of Thurston that such a 4-manifold $M_K \times S^1$ has a symplectic structure with symplectic section T_m. Thus, if X is a symplectic 4-manifold with a symplectically embedded torus T of square 0, then the fiber sum $X_K := X \sharp_{T=T_m} (M_K \times S^1)$, obtained by taking a fiber sum along $T = T_m$, is symplectic. R. Fintushel and R. Stern proved that X_K is homotopy equivalent to X under a mild condition on X and computed the SW-invariant of X_K. (In the case when $b_2^+ = 1$, they computed the relative SW-invariant of X_K). For example, applying the construction above on an elliptic surface $E(1)$, they get a family of homotopy elliptic surfaces $\{E(1)_K \,|\, K$ is a fibered knot in $S^3\}$ and computed the *relative SW-invariant* $SW^{\pm}_{E(1)_K, T} := \sum_{L\cdot[T]=0} SW^{[T],\pm}_{E(1)_K}(L) \cdot e^L$ of $E(1)_K$:

Theorem 3.2 ([**FS2**]). *For each fibered knot K in S^3, a homotopy elliptic surface $E(1)_K$ is a simply connected symplectic 4-manifold whose $[T]$-relative SW-invariants are*

$$\begin{aligned}\sum_{L\cdot[T]=0} SW^{[T],\pm}_{E(1)_K}(L)\cdot e^L &= \sum_{L\cdot[T]=0} SW^{[T],\pm}_{E(1)}(L)\cdot e^L \cdot \Delta_K(e^{2[T]}) \\ &= \sum_{n=0}^{\infty} (\mp 1)\cdot e^{\mp(2n+1)[T]}\cdot \Delta_K(e^{2[T]})\end{aligned}$$

where Δ_K is the Alexander polynomial of K and T is a symplectically embedded torus induced from a standard torus fiber lying in $E(1)$.

Then, using a relation between various types of Seiberg-Witten invariants for smooth 4-manifolds with $b_2^+ = 1$ and using Theorem 3.2 above, we are able to compute the SW-invariant $SW^\circ_{E(1)_K}$ of $E(1)_K$:

THEOREM 3.3 ([**P1**]). *For each fibered knot K in S^3, $E(1)_K$ has a SW-invariant denoted by $SW^{\circ}_{E(1)_K} = P^{+}_{E(1)_K} + P^{-}_{E(1)_K}$ such that $P^{\pm}_{E(1)_K}$ is the partial sum consisting of only positive (negative) multiples of $[T]$ in the exponent of the series $\mp \sum_{n=0}^{\infty} e^{\mp(2n+1)[T]} \cdot \Delta_K(e^{2[T]})$, where Δ_K is the Alexander polynomial of K and T is an embedded torus induced from a standard torus fiber lying in $E(1)$.*

Note that Proposition 3.1 and Theorem 3.3 above imply that, if K is a non-trivial fibered knot, then $E(1)_K$ is neither rational nor ruled. Furthermore, since the set of SW-basic classes of $E(1)_K$ is of the form $\{\lambda_i[T] \mid \lambda_1, \ldots, \lambda_n \text{ are some integers}\}$, $E(1)_K$ should be minimal.

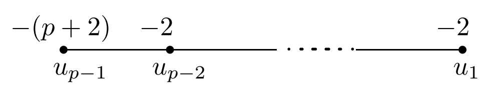
COROLLARY 3.4. *For each fibered knot K with a non-trivial Alexander polynomial in S^3, a homotopy elliptic surface $E(1)_K$ is a minimal symplectic 4-manifold which is neither rational nor ruled.*

Finally, since the set of all simply connected complex surfaces satisfying $b_2^+ = 1$ and $K^2 = 0$ is classified as a blowing up of other complex surfaces and Dolgachev surfaces, we can conclude that most homotopy rational surfaces $E(1)_K$ are not diffeomorphic to any of such complex surfaces by comparing their Seiberg-Witten invariants. Explicitly, we have

THEOREM 3.5. *If K is a fibered knot in S^3 whose Alexander polynomial is non-trivial and is different from that of any (p, q)-torus knot, then $E(1)_K$ is a simply connected minimal symplectic 4-manifold which cannot admit a complex structure.*

4. Symplectic 4-manifolds with $b_2^+ = 1$ and $K^2 > 0$

In this section we present new symplectic 4-manifolds with $b_2^+ = 1$ and $K^2 > 0$ which are not diffeomorphic to rational surfaces. These manifolds are constructed by applying the rational blow-down surgery to rational surfaces. We first review the *rational blow-down surgery* introduced by R. Fintushel and R. Stern (see [**FS1**] for details). Let C_p be the smooth 4-manifold obtained by plumbing $(p-1)$ disk bundles over the 2-sphere according to the following diagram



where each vertex u_i represents a disk bundle over the 2-sphere with Euler class labeled above and an interval between vertices indicates plumbing the disk bundles corresponding to the vertices. Label the homology classes represented by the 2-spheres in C_p by $u_1, \ldots, u_{p-1}$ so that the self-intersections are $u_{p-1}^2 = -(p+2)$ and $u_i^2 = -2$ for $1 \leq i \leq p-2$. Then the configuration C_p is a negative definite simply connected smooth 4-manifold whose boundary is the lens space $L(p^2, 1-p)$ which bounds a rational ball B_p with $\pi_1(B_p) \cong \mathbf{Z}_p$. Furthermore, the intersection form on $H^2(C_p; \mathbf{Q})$ with respect to the dual basis $\{\gamma_i : 1 \leq i \leq p-1\}$ (i.e. $< \gamma_i\ , u_j >= \delta_{ij}$) is given by

$$Q_p := (\gamma_i \cdot \gamma_j) = P_p^{-1},$$

where P_p is the plumbing matrix for C_p with respect to the basis $\{u_i : 1 \leq i \leq p-1\}$.

DEFINITION 4.1. Suppose X is a smooth 4-manifold which contains the configuration C_p. Then we may construct a new smooth 4-manifold X_p, called the *rational blow-down* of X, by replacing C_p with the rational ball B_p. Note that this process is well-defined, that is, a new smooth 4-manifold X_p is uniquely constructed (up to diffeomorphism) from X because each diffeomorphism of $\partial B_p = L(p^2, 1-p)$ extends over the rational ball B_p. Furthermore, M. Symington has proved that a rational blow-down manifold X_p admits a symplectic structure in some cases.

THEOREM 4.2 ([**Sy**]). *Suppose (X,ω) is a symplectic 4-manifold containing the configuration C_p. If all 2-spheres u_i in C_p are symplectically embedded and intersect positively, then the rational blow-down manifold $X_p = X_0 \cup_{L(p^2,1-p)} B_p$ admits a symplectic 2-form ω_p such that $(X_0, \omega_p|_{X_0})$ is symplectomorphic to $(X_0, \omega|_{X_0})$.*

Next, we review the rational surface $E(1) = \mathbf{C}P^2 \sharp 9\overline{\mathbf{C}P}^2$. One way to describe $E(1)$ is to view it as a Lefschetz fibration over $\mathbf{C}P^1$ whose generic fiber is an elliptic curve, say f, and which has one $\widetilde{E_6}$-singular fiber and four fishtail singular fibers. Note that a neighborhood of the $\widetilde{E_6}$-fiber in $E(1)$ is a smooth 4-manifold obtained by plumbing disk bundles over the holomorphically embedded 2-spheres $S_i (1 \le i \le 7)$ of square -2 according to the Dynkin diagram of $\widetilde{E_6}$ (see [**HKK**] for details).

PROPOSITION 4.3. *For an integer k with $2 \le k \le 4$, there exists a configuration C_{2k-1} in a rational surface $E(1)\sharp k\overline{\mathbf{C}P}^2$ such that all 2-spheres u_i lying in C_{2k-1} are symplectically embedded.*

PROOF. Since the homology class $[f]$ of the elliptic fiber f in $E(1)$ can be represented by an immersed 2-sphere with one positive double point (equivalently, a fishtail fiber) and since $E(1)$ contains at least 4 such immersed 2-spheres, for an integer k with $2 \le k \le 4$ we can blow up $E(1)$ k times at these double points so that there exist embedded 2-spheres, $f - 2E_{10}, \dots, f - 2E_{9+k}$, in $E(1)\sharp k\overline{\mathbf{CP}}^2$ which intersect a section E_9 of $E(1)$ positively at one point each. And then, resolving symplectically all the intersection points between $f - 2E_{10}, \dots, f - 2E_{9+k}$ and E_9, we have a symplectically embedded 2-sphere, denoted by S, in $E(1)\sharp k\overline{\mathbf{CP}}^2$ which represents the homology class $kf + E_9 - 2(E_{10} + \cdots + E_{9+k})$ of square $-(1+2k)$. Now, using a plumbing manifold consisting of $(2k-3)$ disk bundles $\{S_1, S_2, \dots, S_{2k-3}\}$ lying in a neighborhood of the $\widetilde{E_6}$-singular fiber, we obtain a configuration $C_{2k-1} \subset E(1)\sharp k\overline{\mathbf{CP}}^2$ so that $u_{2k-2} = S = kf + E_9 - 2(E_{10} + \cdots + E_{9+k})$ and $u_i = S_i$ for $1 \le i \le 2k-3$. $\square$

Now, by rationally blowing down along a configuration C_{2k-1} lying in a rational surface $E(1)\sharp k\overline{\mathbf{C}P}^2$ (see Proposition 4.3 above), we get a new smooth 4-manifold which in fact admits a symplectic structure due to Theorem 4.2. Furthermore it is easily proved that the manifold obtained by a rational blow-down surgery is simply connected and it has $K^2 = k - 2$. Hence it is homeomorphic to a rational surface $\mathbf{C}P^2\sharp(11-k)\overline{\mathbf{C}P}^2$ due to M. Freedman's classification theorem. But a computation shows that the induced canonical class K and a compatible symplectic 2-form ω of the manifold satisfies $K \cdot [\omega] > 0$ (refer to Theorem 3 in [**P2**] for details). This means that it is not diffeomorphic to the rational surface $\mathbf{C}P^2\sharp(11-k)\overline{\mathbf{C}P}^2$. For example, we get the simply connected symplectic 4-manifolds P and Q by choosing $k = 4$ and $k = 3$, respectively.

THEOREM 4.4. *There exists a simply connected symplectic 4-manifold P with $b_2^+ = 1$ and $K^2 = 2$ which is homeomorphic, but not diffeomorphic, to $\mathbf{C}P^2\sharp 7\overline{\mathbf{C}P}^2$.*

COROLLARY 4.5. *There exists a simply connected symplectic 4-manifold Q with $b_2^+ = 1$ and $K^2 = 1$ which is homeomorphic, but not diffeomorphic, to $\mathbf{C}P^2\sharp 8\overline{\mathbf{C}P}^2$.*

REMARKS

1. Recently, P. Ozsváth and Z. Szabó confirmed that the symplectic 4-manifolds P and Q constructed above are minimal ([**OS**]).
2. At the time of writing this article, A. Stipsicz and Z. Szabó constructed a simply connected symplectic 4-manifold with $b_2^+ = 1$ and $K^2 = 3$ which is not diffeomorphic to $\mathbf{C}P^2\sharp 6\overline{\mathbf{C}P}^2$ by using a similar configuration ([**SS**]).

Finally, we close this paper by mentioning exotic smooth structures on rational surfaces $\mathbf{C}P^2\sharp n\overline{\mathbf{C}P}^2$. We say that a smooth 4-manifold admits an *exotic smooth structure* if it has more than one distinct smooth structure. It has long been a very intriguing question to find the smallest positive integer n such that a rational surface $\mathbf{C}P^2\sharp n\overline{\mathbf{C}P}^2$ admits an exotic smooth structure. By Theorem 4.4 above, we can at least conclude that

COROLLARY 4.6. *The rational surface $\mathbf{C}P^2\sharp 7\overline{\mathbf{C}P}^2$ admits an exotic smooth structure.*

References

[DK] S. Donaldson and P. Kronheimer, *The geometry of 4-manifolds*, Oxford University Press, 1990

[FS1] R. Fintushel and R. Stern, *Rational blowdowns of smooth 4-manifolds*, Jour. Diff. Geom. **46** (1997), 181–235.

[FS2] R. Fintushel and R. Stern, *Knots, links and 4-manifolds*, Invent. Math. **134** (1998), 363–400.

[G] R. Gompf, *A new construction of symplectic manifolds*, Annals of Math. **142** no **3** (1995), 527–595.

[HKK] J. Harer, A. Kas and R. Kirby, *Handlebody decompositions of complex surfaces*, Memoirs of the AMS **62** (1986), AMS

[M] J. Morgan, *The Seiberg-Witten equations and applications to the topology of smooth four-manifolds*, Mathematical Notes **44**, Princeton University Press, 1996

[MS] D. McDuff and D. Salamon, *A survey of symplectic 4-manifolds with $b_2^+ = 1$*, Turkish Jour. Math. **20** (1996), 47–61

[OS] P. Ozsváth and Z. Szabó, *On Park's exotic smooth four-manifolds*, math.GT/0411218

[P1] J. Park, *Non-complex symplectic 4-manifolds with $b_2^+ = 1$*, Bull. London Math. Soc. **36** (2004), 231–240

[P2] J. Park, *Simply connected symplectic 4-manifolds with $b_2^+ = 1$ and $c_1^2 = 2$*, Invent. Math. **159** (2005), 657–667.

[SS] A. Stipsicz and Z. Szabó, *An exotic smooth structure on $\mathbf{C}P^2\sharp 6\overline{\mathbf{C}P}^2$*, math.GT/0411258

[Sy] M. Symington, *Symplectic rational blowdowns*, Jour. Diff. Geom. **50** (1998), 505–518

[T] C. Taubes, *The Seiberg-Witten invariants and symplectic forms*, Math. Res. Letters **1** (1994), 809–822.

DEPARTMENT OF MATHEMATICAL SCIENCES, SEOUL NATIONAL UNIVERSITY, SAN 56-1, SILLIM-DONG, GWANAK-GU, SEOUL 151-747, KOREA

E-mail address: `jipark@math.snu.ac.kr`

Clay Mathematics Proceedings
Volume 5, 2006

The Kodaira Dimension of Symplectic 4-manifolds

Tian-Jun Li

Abstract. This survey is concerned with the classification of symplectic 4–manifolds. The Kodaira dimension κ divides the symplectic 4–manifolds into 4 classes, each with distinct features. We give an overview of what is known for each class of manifolds.

1. Introduction

Ever since Thurston [**63**] discovered that any T^2-bundle over T^2 with $b_1 = 3$ admits symplectic structures but no Kähler structures, many constructions of closed non-Kähler symplectic $4-$manifolds have appeared. For instance, Gompf [**19**] used the fiber–sum construction to build, for any finitely presented group G, a closed non-Kähler symplectic $4-$manifold M_G with $\pi_1(M_G) = G$ (see also [**3**] for a systematic approach to comparing symplectic 4-manifolds with Kähler surfaces). As a result, it is impossible to classify all symplectic $4-$manifolds. Nevertheless one could attempt to devise a coarse classification scheme. In this regard, the notion of the Kodaira dimension is a perfect place to start.

The Kodaira dimension κ for a Kähler surface is a measure of how positive the canonical bundle is in terms of the growth of plurigenera. The extension of this notion to closed symplectic $4-$manifolds (M, ω) measures the positivity of the *symplectic* canonical class K_ω, and, as for the case of Kähler surfaces, it also takes four values: $-\infty, 0, 1$ and 2. More specifically, for a *minimal* symplectic 4–manifold (M, ω), its Kodaira dimension is defined in terms of the positivity of $K_\omega \cdot [\omega]$ and $K_\omega \cdot K_\omega$. To extend it to general symplectic 4-manifolds, one needs to use results on existence (and uniqueness) of minimal models.

When examined under the lens of the Kodaira dimension, except for T^2-bundles over T^2, all the known non-Kähler symplectic $4-$manifolds have positive values. And the bigger κ is the less we know about the manifolds in that class. The 4–manifolds of $\kappa = -\infty$ have been classified up to symplectomorphisms. There is a conjectured classification for those of $\kappa = 0$. Progress has been made towards bounding the Betti numbers and there is hope of determining their homology types. It is impossible to classify manifolds of positive κ. Instead there are various geography problems and the focus has been on the simply connected ones. We think

1991 *Mathematics Subject Classification.* Primary 57R57.

The author is supported in part by NSF and the McKnight fellowship.

it is also interesting to consider general 4–manifolds of this kind, in particular taking into account the degeneracy and the nullity in the geography problems. The structure of the Gromov-Taubes invariants is also an important problem here.

We would like to thank the referee, A. Greespoon and A. Stipsicz for their careful readings of this paper.

2. Definition and basic properties

Let (M, ω) be a closed symplectic 4–manifold. Associated with it is the contractible space of $\omega-$compatible almost complex structures. Thus we can define the symplectic Chern classes $c_i(M, \omega) = c_i(M, J)$ where J is any $\omega-$compatible almost complex structure. In particular, $-c_1(M, \omega) \in H^2(M; \mathbf{Z})$ is called the symplectic canonical class, and is denoted by K_ω.

As mentioned in the introduction, we will first define the Kodaira dimension of (M, ω) when it is minimal, so we need to recall the notion of minimality. Let $\mathcal{E}_M$ be the set of homology classes which have square -1 and are represented by smoothly embedded spheres. M is said to be smoothly minimal if $\mathcal{E}_M$ is empty. Let $\mathcal{E}_{M,\omega}$ be the subset of $\mathcal{E}_M$ which are represented by embedded $\omega-$symplectic spheres. (M, ω) is said to be symplectically minimal if $\mathcal{E}_{M,\omega}$ is empty. When (M, ω) is non-minimal, one can blow down some of the symplectic -1 spheres to obtain a minimal symplectic $4-$manifold (N, μ), which is called a symplectic minimal model of (M, ω) ([**45**]). Now we summarize the basic facts about the minimal models.

Proposition 2.1. *([**28**], [**38**], [**45**], [**60**]) Let M be a closed oriented smooth $4-$manifold and ω a symplectic form on M compatible with the orientation of M.*

1. *M is smoothly minimal if and only if (M, ω) is symplectically minimal. In particular, the underlying smooth manifold of the symplectic minimal model of (M, ω) is smoothly minimal.*
2. *If (M, ω) is not rational nor ruled, then it has a unique symplectic minimal model. Furthermore, for any other symplectic form ω' on M compatible with the orientation of M, the symplectic minimal models of (M, ω) and (M, ω') are diffeomorphic as oriented manifolds.*
3. *If (M, ω) is rational or ruled, then its symplectic minimal models are diffeomorphic to $\mathbf{CP}^2$ or an S^2-bundle over a Riemann surface.*

Here a rational symplectic $4-$manifold is a symplectic $4-$manifold whose underlying smooth manifold is $S^2 \times S^2$ or $\mathbf{CP}^2 \# k\overline{\mathbf{CP}^2}$ for some non-negative integer k. A ruled symplectic $4-$manifold is a symplectic $4-$manifold whose underlying smooth manifold is the connected sum of a number of (possibly zero) $\overline{\mathbf{CP}^2}$ with an S^2-bundle over a Riemann surface.

Now we are ready to define the symplectic Kodaira dimension.

Definition 2.2. ([**47**], [**30**]) For a minimal symplectic $4-$manifold (M, ω) with symplectic canonical class K_ω, the Kodaira dimension of (M, ω) is defined in the following way:

$$\kappa(M, \omega) = \begin{cases} -\infty & \text{if } K_\omega \cdot [\omega] < 0 \text{ or } K_\omega \cdot K_\omega < 0, \\ 0 & \text{if } K_\omega \cdot [\omega] = 0 \text{ and } K_\omega \cdot K_\omega = 0, \\ 1 & \text{if } K_\omega \cdot [\omega] > 0 \text{ and } K_\omega \cdot K_\omega = 0, \\ 2 & \text{if } K_\omega \cdot [\omega] > 0 \text{ and } K_\omega \cdot K_\omega > 0. \end{cases}$$

The Kodaira dimension of a non-minimal manifold is defined to be that of any of its symplectic minimal models.

REMARK 2.3. In [**47**] the Kodaira dimension of a minimal symplectic $4-$manifold (M,ω) is defined to be $-\infty$ if $K_\omega\cdot[\omega]<0$, and zero if $K_\omega\cdot[\omega]=0$. Our modification in [**30**] is to take into account the sign of $K_\omega\cdot K_\omega$ as well in these two cases. Since, for any minimal ruled surface with negative $K_\omega\cdot K_\omega$, there are symplectic forms ω with $K_\omega\cdot[\omega]$ non-negative, this slight modification is necessary for the Kodaira dimension to be well-defined for all symplectic $4-$manifolds.

For a minimal symplectic $4-$manifold, its Kodaira dimension has the following properties.

THEOREM 2.4. *Let M be a closed oriented smooth $4-$manifold and ω a symplectic form on M compatible with the orientation of M. If (M,ω) is symplectically minimal, then*

(1) *The Kodaira dimension of (M,ω) is well-defined.*
(2) *(M,ω) has Kodaira dimension $-\infty$ if and only if it is rational or ruled.*
(3) *(M,ω) has Kodaira dimension 0 if and only if K_ω is a torsion class.*

Furthermore $\kappa(M,\omega)$ is well-defined for any symplectic 4-manifold (M,ω).

All the properties are based on the Taubes-Seiberg-Witten theory (cf. [**59**], [**60**] and [**38**]). To show (1) amounts to showing that any minimal symplectic 4-manifold must satisfy one and only one of the four conditions above, i.e. there is no minimal manifold (M,ω) with $K_\omega\cdot K_\omega>0$ and $K_\omega\cdot[\omega]=0$. This is an immediate consequence of the following fact proved in [**30**]: If (M,ω) is minimal with $K_\omega\cdot[\omega]=0$ and $K_\omega\cdot K_\omega\geq 0$, then K_ω is a torsion class and hence $K_\omega\cdot K_\omega=0$. Notice that property (3) also follows from this fact. Property (2) follows from [**41**]. The last property is now a consequence of Proposition 2.1. If (M,ω) is not rational or ruled, it has a unique symplectic minimal model by Proposition 2.1 (2), so $\kappa(M,\omega)$ is well defined by property (1). If (M,ω) is rational or ruled, it has non-diffeomorphic symplectic minimal models. However the different minimal models are still rational or ruled by Proposition 2.1 (3), so all have Kodaira dimension $-\infty$ by property (2).

There are two additional properties for $\kappa(M,\omega)$. It is not hard to verify that the holomorphic Kodaira dimension of a Kähler surface coincides with the Kodaira dimension of the underlying symplectic 4-manifold. Furthermore, the Kodaira dimension of (M,ω) only depends on the oriented diffeomorphism type of M, i.e. if ω' is another symplectic form on M compatible with the orientation of M, then $\kappa(M,\omega)=\kappa(M,\omega')$.

REMARK 2.5. We would like to see whether it is possible to define $\kappa(M,\omega)$ for higher dimensional symplectic manifolds. Again we would first define it for 'minimal' manifolds of dimension $2n$ as follows: $\kappa(M,\omega)$ is defined to be $-\infty$ if $K^i_\omega\cdot[\omega]^{n-i}$ is negative for some i; and $\kappa(M,\omega)=i$ if $K^j_\omega\cdot[\omega]^{n-j}=0$ for any $j\geq i+1$ and $K^j_\omega\cdot[\omega]^{n-j}>0$ for any $j<i+1$. To show it is well-defined we need to exclude other possibilities of the $n+1$ numbers $\{K^i_\omega[\omega]^{n-i}\}^n_{i=0}$. Then we would extend it to general manifolds by requiring 'birational' invariance. Of course this is just a speculation since there are many issues to be settled here, one of which is that different minimal models of a manifold should have the same Kodaira dimension.

3. Kodaira dimension $-\infty$

As already mentioned, (M, ω) has Kodaira dimension $-\infty$ if and only if M is rational or ruled. Notice that rational or ruled manifolds all admit Kähler structures. There are other beautiful characterizations, one of which is the existence of metrics of positive scalar curvature ([**41**], [**51**]). Embedded symplectic spheres can be used to characterize such manifolds. It is shown in [**45**] that a symplectic 4–manifold is rational or ruled if M has an embedded symplectic sphere with non-negative square (or even a smoothly embedded essential sphere with non-negative square, cf. [**28**]), and the converse is shown in [**59**] and [**36**]. There is also a symplectic Castlenuovo criterion of rationality in [**38**]: If (M, ω) is minimal with $b_1 = 0$, and $2K_\omega$ is a non-torsion class not represented by a symplectic surface, then M is rational.

Let $\Omega(M)$ denote the space of orientation$-$compatible symplectic forms on M. $\Omega(M)$ is an infinite dimensional manifold modeled on the space of closed $2-$forms. The image of $\Omega(M)$ in $H^2(M;\mathbf{R})$ under the map of taking the cohomology class, denoted by $\mathcal{C}(M)$, is called the symplectic cone of M. The quotient $\mathcal{M}(M) = \Omega(M)/\mathrm{Diff}^+(M)$ is called the moduli space of symplectic structures on M, where $\mathrm{Diff}^+(M)$ is the group of orientation$-$preserving diffeomorphisms of M. Symplectic structures on manifolds with $\kappa = -\infty$ are unique in the sense that the moduli space $\mathcal{M}(M)$ is connected and diffeomorphic to $\mathcal{C}(M)/D(M)$ ([**26**], [**38**]). Here $D(M)$ is the group of automorphisms of the cohomology lattice induced by orientation-preserving diffeomorphisms. Moreover, both $\mathcal{C}(M)$ and $D(M)$ can be explicitly determined in terms of the set $\mathcal{E}_M$ (for the minimal case see [**26**], and for the general case see [**14**], [**38**], [**33**] and work in progress [**34**]). The topology of the symplectomorphism group is also rather well understood, at least in the minimal case (see [**21**], [**1**], [**2**]).

It is interesting to study the uniqueness of symplectic structures on rational and ruled symplectic orbifolds. This might be useful for the study of birational geometry in dimension 6. A good example to start with is the nodal quadric surface. In this case its blowup is $\mathbf{CP}^1 \times \mathbf{CP}^1$ with a symplectic -2 sphere; thus the connectedness of the space of symplectic -2 spheres shown in [**1**] can be used here. For the general case, one should analyze the moduli spaces of embedded orbifold rational curves and the associated evaluation maps. To show the existence of such curves one would need the orbifold version of Taubes-Seiberg-Witten theory, especially the SW$\Rightarrow$Gr part, which has been developed recently in [**7**].

To end this section we mention the question of Yau: If (M, ω) is a symplectic 4–manifold with M homotopic to $\mathbf{CP}^2$, then is (M, ω) symplectomorphic to $(\mathbf{CP}^2, \lambda\omega_{FS})$? Here ω_{FS} is the standard Fubini-Study form and λ is a positive scalar. The corresponding question in the complex world was affimatively answered by Yau as a consequence of the solution to the Calabi conjecture. Notice that, since M is assumed to be smooth, by Freedman's celebrated classification of simply connected topological 4–manifolds, M is homeomorphic to $\mathbf{CP}^2$. And since the symplectic structure on $\mathbf{CP}^2$ is unique, this question is equivalent to

QUESTION 3.1. *Let (M, ω) be a symplectic 4–manifold homeomorphic to $\mathbf{CP}^2$, is M diffeomorphic to $\mathbf{CP}^2$?*

Notice that if there is a counterexample it must have Kodaira dimension 2. Notice also that, for any $l \geq 6$, there are symplectic 4–manifolds homeomorphic to $\mathbf{CP}^2 \# l\overline{\mathbf{CP}^2}$ but not diffeomorphic to it (see [**53**], [**58**]).

4. Kodaira dimension 0

This is the case where a classification is still feasible. As already seen in §2, the minimal ones are those with torsion canonical classes. We first collect some general properties of such symplectic 4−manifolds.

PROPOSITION 4.1. *Let (M, ω) be a symplectic 4−manifold with torsion canonical class K_ω. Then*

(1) $2\chi(M) + 3\sigma(M) = 0$, *where χ and σ are the Euler number and the signature respectively.*
(2) *M has even intersection form.*
(3) *K_ω is either trivial, or of order two, which only occurs when M is an integral homology Enriques surface.*
(4) *M is spin except when M is an integral homology Enriques surface.*

The first statement follows from $K_\omega \cdot K_\omega = 2\chi + 3\sigma$. The second statement follows from another property of K_ω: for any class $e \in H^2(M; \mathbf{Z})$, $e \cdot e = e \cdot K_\omega$ modulo 2. Since M is spin if and only if the second Stiefel-Whitney class $w_2(M)$ is trivial and $w_2(M)$ is the mod 2 reduction of K_ω, the last statement follows from the third statement. Finally the third statement follows from [**25**], [**38**] and [**60**], as observed in [**47**].

Any Kähler surface with (holomorphic) Kodaria dimension 0 also has $\kappa = 0$. Such Kähler surfaces have been classified: the K3 surface, the Enriques surface and the hyperelliptic surfaces. It is not hard to find non-Kähler ones with $\kappa = 0$. In fact the first example of a non-Kähler symplectic manifold, the Kodaira-Thurston manifold, has $\kappa = 0$. The Kodaira-Thurston manifold is an example of a T^2−bundle over T^2. In fact we have

LEMMA 4.2. *Let M be an oriented T^2−bundle over T^2. Then M is minimal and admits symplectic structures. Moreover, there exists a symplectic form ω on M with K_ω a torsion class.*

M is minimal since it is a $K(\pi, 1)$ manifold. The fact that all T^2−bundles over T^2 admit symplectic structures is observed in [**18**] and we briefly sketch the argument here. On the one hand, the Thurston construction gives rise to symplectic forms on any surface bundle over a surface as long as the fibers are homology essential. On the other hand, any M with homology inessential fibers is shown in [**18**] to be a principal S^1−bundle over a 3–manifold which itself is a principal S^1−bundle over T^2, and thus admits symplectic structures by the construction in [**11**]. To show that a T^2−bundle over T^2 has torsion canonical class, we use the explicit representation of M as a geometric manifold $\Gamma\backslash\mathbf{R}^4$ in [**64**] to find a basis of $H_2(M; \mathbf{R})$ generated by symplectic tori of square zero and then apply the adjunction formula.

To our knowledge, no potentially new minimal symplectic 4−manifolds with $\kappa = 0$ have been constructed so far. For instance, Fintushel and Stern's knot surgery ([**16**]) on a fibered knot is a powerful technique to produce infinitely many families of homeomorphic but non-diffeomorphic symplectic 4−manifolds. In order to get one with torsion canonical class, one however has to start with such a manifold which is known, e.g. the K3 surface, and apply this surgery to a fibred knot with trivial Alexander polynomial. Though there are many knots with trivial Alexander polynomial, the only fibered one is the trivial knot. Therefore the knot surgery produces nothing new in this context. In fact if (M, ω) has torsion canonical class and admits a genus g Lefschetz fibration structure, then the adjunction formula applied to the fiber class leads to the conclusion that $g = 1$. Now it follows from the classification of genus one Lefschetz fibrations of Moishezon and Matsumoto (cf. [**44**]) that M is either K3 or a torus bundle over a torus. There is additional evidence that no new symplectic manifolds with $\kappa = 0$ exist, e.g. when M is simply connected, it follows from [**50**] that M is homeomorphic to the $K3$ surface. Using results in [**22**] and [**54**], it is also shown in [**30**] that if $\pi_1(M)$ is a non-trivial finite group then $\pi_1(M) = \mathbf{Z}_2$ and M is homeomorphic to the Enriques surface, and if $b_1 = 4$ then $H^*(M; \mathbf{R})$ is generated by $H^1(M; \mathbf{R})$ and hence isomorphic to $H^*(T^4; \mathbf{R})$ as a ring.

Here is the table of Kähler surfaces with $\kappa = 0$ and T^2-bundles over T^2 according to the homology type.

TABLE 1

class	b^+	b_1	χ	σ	b^-	known as
a)	3	0	24	-16	19	$K3$
b)	3	4	0	0	3	4-torus
c)	2	3	0	0	2	primary Kodaira surface
d)	1	0	12	-8	9	Enriques surface
e)	1	2	0	0	1	hyperelliptic surface if complex

Notice that the values of their Euler numbers are either 0, 12 or 24, and they satisfy the Betti number bounds:

$$(*) \qquad b_1 \leq 4, \quad b^+ \leq 3, \quad b^- \leq 19.$$

Thus it is natural to conjecture in [**30**] that a minimal symplectic 4–manifold with $\kappa = 0$ satisfies $(*)$. We are able to obtain in [**30**] the bounds on b^+ and b^- if we assume the *desired* bound of b_1.

THEOREM 4.3. *Let M be a closed minimal symplectic 4–manifold with $\kappa = 0$ and $b_1 \leq 4$. Then $b^+ \leq 3$, $b^- \leq 19$, and the Euler number of M is* 0, 12 *or* 24.

In fact we can show that the manifolds in our situation share more topological properties, like the real homology group and the intersection form, with the manifolds in Table 1.

The proof is divided into two cases: case 1. $b^+ \leq b_1+1$; case 2. $b^+ \geq b_1+2$. The first case is the easier one. The relation on Betti numbers $2\chi+3\sigma = 0$, together with Rokhlin's congruence on signature coming from the fact that M is spin if $b^+ > 1$, readily gives the theorem in this case. The approach to the second case is to show that, on a closed smooth oriented 4−manifold with $2\chi + 3\sigma = 0$, $b_1 \leq 4$ and $b^+ > \max\{3, b_1+1\}$, the mod 2 Seiberg-Witten invariant of any reducible Spin^c structure vanishes. Here a Spin^c structure is called reducible if it admits a reduction to a spin structure. Such a vanishing result was first proved in [**50**] with the assumption that $b_1 = 0$ and $b^+ > 3$. To obtain our vanishing result we use the refinement of the Seiberg-Witten invariants in [**6**] and [**12**], which is a Pin(2)−equivariant stable homotopy class between sphere bundles over the torus T^{b_1}. Under the assumption $b^+ \geq b_1 + 2$ one can pass from the stable homotopy Seiberg-Witten invariants to the ordinary Seiberg-Witten invariants. Under the same assumption we mimic the construction in [**13**] to construct the *unoriented bordism* Seiberg-Witten invariants from the stable homotopy Seiberg-Witten invariants. In the case of interest to us, they coincide with the ordinary Seiberg-Witten invariants modulo 2, while it can be shown that they are trivial using techniques in [**13**]. On the other hand, the fundamental result in [**59**] implies that, on a closed symplectic 4−manifold with trivial canonical class, a certain canonical reducible Spin^c structure has Seiberg-Witten invariant one. The theorem in this case then follows by comparing the vanishing result with Taubes' non-vanishing result.

Furthermore we are now close to proving the following statement: If M is a minimal symplectic 4–manifold with $\kappa = 0$ and $\chi(M) > 0$, then $\chi(M) = 12$ or 24, and M is either a homology Enriques surface or a rational homology K3 respectively. A consequence is that the symplectic Noether inequality $(**)$ of §6 holds for symplectic 4–manifolds with $\kappa = 0$. As in [**30**] the stable homotopy SW invariants in [**6**] and [**12**] play a crucial role here.

To identify the diffeomorphism types, we believe the parametrized SW theory in [**37**] should play a pivotal role, at least when M has a winding family. A winding family of symplectic forms is an S^{b^+-1} family of symplectic forms which represents the generator of $\pi_{b^+-1}(\mathcal{P})$, where $\mathcal{P}$ is the cone of classes with positive square. Every known manifold of $\kappa = 0$ carries such a family. We speculate that this is always the case. For the manifolds in Table 1, another important feature is that they are all 'fibered' by tori. Those in classes b), c) and e) are T^2-bundles over T^2, and those in classes a) and d) fiber over S^2. An ambitious goal here is to see whether it is possible to use the parametrized theory to construct an elliptic fibration assuming the existence of a winding family. If the answer is yes then such a manifold is either Kähler or a T^2-bundle as remarked before. Assuming M has a winding family, the following result can be proved using the wall crossing formula and SW⇒GT in the family setting: if $b^+ > 1$ and $\chi(M) > 0$, then $H_1(M; \mathbf{Z})$ is torsion-free. It then follows that if M has $b^+ > 1, b_1 = 0$ and a winding family, then M is an *integral* homology K3.

For a minimal manifold M with $\kappa = 0$, there is a unique symplectic canonical class, and according to [**38**], for all the manifolds listed in Table 1, the cone of classes of symplectic forms is the positive cone $\mathcal{P}$. We speculate that there is a unique deformation class of symplectic forms and that the moduli space of symplectic structures is similarly given by $\mathcal{P}/D(M)$ as in the case of $\kappa = \infty$.

5. Kodaira dimension 1

For symplectic 4–manifolds with $\kappa = 1$ we cannot expect to have a classification: it is shown in [**19**] that any finitely presented group is the fundamental group of a symplectic 4–manifold with $\kappa = 1$ (same for $\kappa = 2$). Instead one is interested in the geography problem of minimal manifolds subject to the Noether condition and the conjecture of Gompf. The Noether condition, due to the existence of an almost complex structure, says that $2\chi + 3\sigma \equiv \sigma \pmod 8$, or equivalently, $b^+ - b_1$ is odd. The conjecture of Gompf says that the Euler number of a symplectic 4–manifold with $\kappa \geq 0$ is non-negative. Notice that

$$2\chi + 3\sigma = 4 - 4b_1 + 5b^+ - b^- = 4 + 4b^+ - 4b_1 + \sigma = 0.$$

Therefore b_1 is at least $2+\frac{\sigma}{4}$, as b^+ has to be positive for a symplectic 4–manifold.

If we are only interested in the simply connected situation this has a simple positive answer. In this case we just need to find simply connected symplectic 4–manifolds with $\kappa = 1$ and $\sigma = 8k$ with $k \leq 0$. The Dolgachev surfaces and elliptic surfaces $E(k)$, $k \geq 2$ are such manifolds with signature -8 and $-8k$, $k \geq 2$ respectively.

Taking into account the first Betti number b_1, a pair (a, b) of integers corresponding to (σ, b_1) is called admissible if $a = 8k$ for some non-positive integer k and $b \geq \max\{0, 2+a/4\}$. Again it is not hard to show that, for any admissible pair (a, b), there exists a minimal symplectic 4$-$manifold (M, ω) with $\kappa = 1$ and with $(\sigma(M), b_1(M)) = (a, b)$. When $a = 0$ we have that b is at least equal to 2. In fact we can simply use $S^1 \times N$, where N is a fibered 3$-$manifold with $b_1(N) = b-1$ and with fibers of genus at least 2. When $a = -8k$ with $k > 0$, b is allowed to be any non-negative integer. For the case $b = 0$, we can take a simply connected manifold as above. For the case $b \geq 1$, we combine the previous two cases by fiber–summing a simply connected elliptic fibration with $\sigma = a$ and $S^1 \times N$ with $b_1(N) = b + 1$ along a fiber and a torus which is the product of S^1 and a section. The resulting manifolds are minimal, as it is proved in [**40**] that fiber–sums of minimal manifolds are minimal.

Symplectic 4–manifolds (M, ω) are said to be of *Lefschetz type* if $[\omega] \in H^2(M; \mathbf{R})$ satisfies the conclusion of the Hard Lefschetz Theorem, namely, that the linear map $\cup[\omega] : H^1(M; \mathbf{R}) \longrightarrow H^3(M; \mathbf{R})$ is an isomorphism. For a general symplectic 4–manifold (M, ω), the rank of the kernel of this map is called the degeneracy and denoted by $d(M, \omega)$. In [**5**] we include the degeneracy as an extra parameter in the geography problem. A triple $(a, b, c) \in \mathbf{Z}^3$ corresponding to (σ, b_1, d) is called *Lefschetz admissible* if $a = 8k$ where k is a non-positive number, $0 \leq c \leq b$, $b - c$ even, and $b \geq \max\{0, 2 + a/4\}$. The Gompf conjecture says that for every minimal symplectic 4–manifold (M, ω) with $\kappa = 1$, the triple $(\sigma(M), b_1(M), d(M, \omega))$ is admissible. The converse is shown to be true as well in [**5**].

THEOREM 5.1. *For any Lefschetz admissible triple* (a, b, c) *there exists a minimal symplectic 4–manifold* (M, ω) *with* $\kappa = 1$ *and* $(a, b, c) = (\sigma(M), b_1(M), d(M, \omega))$.

Our construction centers around the notion of bundle manifolds, which are certain S^1-bundles over a base which is itself a surface bundle over S^1. They are uniquely specified by four integers: the genus g of the surface bundle, two weights d and k describing the holonomy of the surface bundle in terms of the Dehn twists, where $0 \leq d \leq k \leq g$, and a number $e = 0, 1, 2$ based upon the Euler class of the S^1-bundle. Such a manifold admits S^1-invariant symplectic structures if and only if the Euler class of the S^1-bundle pairs trivially with the fibers ([**11**]). Using [**4**] we are able to derive the cohomology ring explicitly and in particular calculate the degeneracy.

There still remains the nullity question. For any $\alpha \in H^1(M; \mathbf{R})$, and $i = 1, 2$, consider the map $i_\alpha = \cup\alpha : H^i(M; \mathbf{R}) \longrightarrow H^{i+1}(M; \mathbf{R})$. The dimension of the linear space $\{\alpha | i_\alpha = 0\}$, denoted by $n_i(M)$, is called the $i-$nullity of M. By Poincaré Duality, $n_i(M) = n_2(M)$. Thus we can speak simply of nullity $n(M)$. It also follows from Poincaré Duality that the nullity is a lower bound for the degeneracy, i.e. $d(M, \omega) \geq n(M)$. Any triple (a, b, c) of integers is called null admissible if $a = 8k$ where k is a non-positive number, $0 \leq c \leq b, c \neq b-1$, and $b \geq$max$\{0, 2 + a/4\}$. Here a, b and c correspond to the signature, the 1st Betti number and the nullity respectively. Observe that it is required that $c \neq b-1$, for if the nullity of a manifold M was $b_1(M)-1$, there would be an element in $H^1(M;\mathbf{R})$ whose cup product square would not be zero. It is found in [**5**] that many null admissible triples can be realized via the construction in [**5**]. But many triples including $(0,3,1)$ remain uncharted.

Finally let us comment on the conjecture of Gompf. Any counterexample would have nonzero b_1. On the other hand it is easy to see that a Lefschetz fibration over a positive genus surface, or more generally, a local holomorphic fibration over a positive genus surface (see [**20**]), either has Kodaira dimension $-\infty$ or has non-negative Euler number. Thus it is interesting to analyze when a manifold with $\kappa \geq 0$ has such a structure. Donaldson's construction of Lefschetz pencils [**9**] might be useful here. However there is an obstruction coming from the cohomology ring structure, and hence this investigation should restrict to manifolds of Lefschetz type.

Another approach is to see whether a symplectic 4–manifold with $\kappa = 0$ always arises as a toridal fiber–sum of some building blocks known to have non-negative Euler number. (Here we call a fiber–sum along square zero tori a toridal fiber–sum.) If this is the case, we would immediately have a positive answer to the conjecture of Gompf, as the Euler number is additive under the toridal fiber–sum. Indeed, the manifolds M_G mentioned in the introduction are toridal fiber–sums of $T^2 \times \Sigma_k$ and (a number of) $E(1)$. Similarly, the manifolds in [**5**] are toridal fiber–sums of bundle manifolds and some manifolds with $b^+ = 1$ including $E(1)$ and the Dolgachev surfaces. Thus it seems that the building blocks might be fiber bundles and 4–manifolds with $b^+ = 1$. Notice that the Euler number of a symplectic 4–manifold with $\kappa = 0$ and $b^+ = 1$ is always non-negative as it satisfies $b_1 \leq 2$ by [**41**], while the Euler number of a T^2-bundle or a bundle manifold is zero. A nice feature of a toridal fiber–sum is that the torus T_i which is summed along in each summand M_i reappears as a square zero torus T in the new manifold. Thus the natural problem is, given a symplectic 4–manifold M with $\kappa = 0$, $b^+ \geq 2$ and which is not a fiber bundle, whether we are able to locate a torus $T \subset M$ along which M is decomposed. Observe that such a torus T represents a toridal GT basic class due to [**23**] and [**39**] (in fact this observation can be proved directly). Thus an interesting and important question is whether the reverse is also true.

6. Kodaira dimension 2

In this case one of the main questions is whether the manifolds satisfy the Bogomolov-Miyaoka-Yau inequality $\chi \geq 3\sigma$, or equivalently, $K_\omega^2 \leq 3\chi$ (for minimal symplectic 4–manifolds with $\kappa = 0$ or 1, since $K_\omega^2 = 0$, it is equivalent to the conjecture of Gompf). And the standard geography problem is to realize all pairs $(\chi(M), \sigma(M))$ of a simply connected minimal symplectic 4–manifold M subject to the Noether condition and the (conjectured) Bogomolov-Miyaoka-Yau inequality $\chi \geq 3\sigma$ (see e.g. [**19**], [**48**], [**52**], [**57**]).

The structure of the GT basic classes of manifolds with $\kappa = 2$ and $b^+ \geq 2$ is also an interesting problem. Recall that the GT invariant of a class e in $H^2(M;\mathbf{Z})$ is a Gromov type invariant defined by Taubes (cf. [**61**]) counting embedded (but not necessarily connected) symplectic surfaces representing the Poincaré dual to e, and e is called a GT basic class if its GT invariant is nonzero. In [**17**], given any configuration of surfaces of genus at least 2, Fintushel and Stern construct a symplectic

4–manifold with $\kappa = 2$ whose canonical class is represented by such a configuration of embedded symplectic surfaces, each representing an indecomposatble GT basic class. On the other hand, the question whether there is a unique way to decompose K_ω into the sum of indecomposable GT basic classes is raised in [**46**]. Here is the notion of a GT decomposition (which is slightly finer than the one in [**46**]).

Definition 6.1. A (fine) GT decomposition of a nonzero class e is an unordered set of *pairwise orthogonal* nonzero GT basic classes $\{A_1, ..., A_m\}$ such that $e = A_1 + \cdots + A_m$. m is called the length of the decomposition. The GT length $l(e)$ of the class e is the maximal length among all such decompositions, and it is defined to be zero if the class is not a GT basic class. A GT basic class e is said to be indecomposable if the only GT decomposition is $\{e\}$.

Some initial progress has been made in [**65**] and [**66**] by identifying the GT invariants with the Donaldson–Smith standard surface count in [**10**] and [**56**], and applying the family enumeration techniques in [**42**]. A consequence of this unique decomposition would be an upper bound of the number of symplectic canonical classes for a minimal manifold in terms of b^+.

Lemma 6.2. *Let M be a closed oriented smooth minimal 4–manifold with $b^+ \geq 2$. Suppose there is a symplectic form ω on M such that K_ω has a unique GT decomposition into indecomposable classes. Then there are at most 2^{b^+-1} symplectic canonical classes up to sign.*

When there is no toridal GT class, we have $l(K_\omega) \leq b^+$. By the uniqueness of decomposition there are at most b^+ indecomposable GT basic classes and hence there are at most 2^{b^+} SW basic classes. In the general case, consider the rays generated by indecomposable GT classes. The non-maximal GT points on each ray cannot correspond to or lead to another symplectic canonical class by [**62**]. Thus we arrive at the same conclusion, as there are at most b^+ rays by the uniqueness of decomposition. In fact even when M is not minimal, we can make the same conclusion as long as we take into account the action of $D(M)$. Notice that there are examples with more than one symplectic canonical class up to sign, the first appearing in [**49**] (see also [**27**], [**55**], [**67**]). Notice also that in the case $b^+ = 1$, it is shown in [**38**] that there is a unique symplectic canonical class up to sign. Interestingly the bound 1 in this case fits with the general bound 2^{b^+-1}.

The GT length $l(K_\omega)$ should also be related to the conjectured symplectic Noether inequality of Fintushel and Stern (a weaker version also appears in [**43**]): A symplectic 4–manifold (M, ω) with $\kappa \geq 2$ satisfies

$$(**) \qquad K_\omega^2 \geq \frac{\chi(M) + \sigma(M)}{4} - (2 + c(K_\omega)),$$

where $c(K_\omega)$ is the maximal number of connected components of an embedded symplectic representative of K_ω. The case $c(K_\omega) = 1$ is studied in [**15**]. We believe it also applies to 4–manifolds with $\kappa = 0$ or $\kappa = 1$. For example, for $E(n)$, $(**)$ is actually an equality. Observe that the χ and σ terms can be replaced by b^+ and b_1, as $[\chi(M) + \sigma(M)]/4 = [b^+(M) - b_1(M) + 1]/2$. We speculate further that the b_1 term can be dropped and the last term $c(K_\omega)$ can be replaced by the GT length $l(K_\omega)$, namely the *stronger* inequality

$$(***) \qquad K_\omega^2 \geq \frac{b^+(M) - 3}{2} - l(K_\omega)$$

should hold for any symplectic 4–manifold (M,ω) with $\kappa \geq 0$. Of course $l(K_\omega) \leq c(K_\omega)$, though it may well be in fact that $l(K_\omega)$ and $c(K_\omega)$ are the same. Notice also that in the case $\kappa = 0$, the conjectured bound $b^+ \leq 3$ is the same as $(***)$ as $K_\omega^2 = c(K_\omega)$ in this case.

Finally let us mention that in the presence of a Lefschetz fibration structure there are bounds for both K_ω^2 and $l(K_\omega)$. For surface bundles with positive genus fiber and positive genus base, it is shown in [**24**] that $K_\omega^2 \geq \chi/2$. This is generalized to the bound $K_\omega^2 \geq 2(g-1)(h-1)$ in [**29**] for any relatively minimal genus g Lefschetz fibration over a genus h surface. Regarding $l(K_\omega)$, there is a sharp bound $l(K_\omega) \leq g-1$ for a minimal genus g Lefschetz fibration over S^2 in [**32**] (this was known to Fintushel as well).

References

[1] M. Abreu, *Topology of symplectomorphism groups of $S^2 \times S^2$*, Invent. Math. 131 (1998), no. 1, 1–23.

[2] M. Abreu, D. McDuff, *Topology of symplectomorphism groups of rational ruled surfaces*, J. Amer. Math. Soc. 13 (2000), no. 4, 971–1009.

[3] D. Auroux, *Some open questions about symplectic 4-manifolds, singular plane curves, and braid group factorizations*, preprint math.GT/0410119.

[4] S. Baldridge, *New symplectic 4–manifolds with $b_+ = 1$*, preprint.

[5] S. Baldridge, T. J. Li, *Geography of symplectic 4-manifolds with Kodaira dimension one*, to appear in Algebr. Geom. Topol..

[6] S. Bauer, M. Furuta, *A stable cohomotopy refinement of Seiberg-Witten invariants: I*, Invent. Math. 155 (2004), No.1, 1-19.

[7] W. Chen, *Pseudoholomorphic curves in four-orbifolds and some applications*, preprint math.SG/0410608.

[8] S. Donaldson, *Symplectic submanifolds and almost-complex geometry*, J. Diff. Geom. 44 (1996), 666-705.

[9] S. Donaldson, *Lefschetz fibrations in symplectic geometry*, Doc. Math. J. DMV., ICM II (1998), 309-314.

[10] S. Donaldson, I. Smith, *Lefschetz pencils and the canonical class for symplectic four-manifolds*, Topology 42 (2003), 743-785.

[11] M. Fernández, A. Gray, J. Morgan, *Compact symplectic manifolds with free circle actions, and Massey products*, Michigan Math. J. 38 (1991), 271–283.

[12] M. Furuta, *Stable homotopy version of Seiberg-Witten invariant*, preprint.

[13] M. Furuta, M. Furuta, Y. Kametani, H. Matsue, N. Minami, *Stable-homotopy Seiberg-Witten invariants and Pin bordisms*, preprint.

[14] R. Friedman, J. Morgan, *On the diffeomorphism types of certain algebraic surfaces I*, J. Differential Geom., 27 (1988).297-369.

[15] R. Fintushel, J. Park, R. Stern, *Rational surfaces and symplectic 4-manifolds with one basic class*, Algebr. Geom. Topol. 2 (2002), 391–402.

[16] R. Fintushel, R. Stern, *Knots, links, and 4-manifolds*, Inv. Math. 134 (1998), 363-400.

[17] R. Fintushel, R. Stern, *The canonical class of a symplectic 4-manifold*, Turkish J. Math. 25 (2001), 137-145.

[18] H. Geiges, *Symplectic structures on T^2-bundles over T^2*, Duke Math. J. 67 (1992), 539–555.

[19] R. Gompf, *A new construction of symplectic manifolds*, Ann. of Math. (2) 142 (1995), 527–595.

[20] R. Gompf, *Locally holomorphic maps yield symplectic structures*, preprint.

[21] M. Gromov, *Pseudo holomorphic curves in symplectic manifolds*, Inv. Math., 82 (1985), 307–347.

[22] I. Hambleton, M. Kreck, *Cancellation, elliptic surfaces and the topology of certain four-manifolds*, J. Reine Angew. Math. 444 (1993), 79–100.

[23] E-N. Ionel and T. Parker, *Gromov invariants of symplectic sums*, Math. Res. Lett. 5 (1998), no. 5, 563–576.

[24] D. Kotschick, *Signatures, monopoles and mapping class groups*, Math. Res. Lett. 5 (1998), no. 1-2, 227–234.
[25] P. Kronheimer and T. Mrowka, *The genus of embedded surfaces in the projective plane*, Math. Res. Letters 1 (1994) 797-808.
[26] F. Lalonde, D. McDuff, *J-curves and the classification of rational and ruled symplectic* 4-*manifolds*, Contact and symplectic geometry Publ. Newton Inst. 8, (Cambridge, 1994), 3–42, Cambridge Univ. Press, Cambridge, 1996
[27] C. LeBrun, *Diffeomorphisms, symplectic forms and Kodaira fibrations*, Geom. Topol. 4 (2000), 451–456.
[28] T. J. Li, *Smoothly embedded spheres in symplectic four manifolds*, Proc. AMS. Vol. 127 (1999) no.2 609-613.
[29] T. J. Li, *Symplectic Parshin-Arakelov inequality*, Internat. Math. Res. Notices, No. 18 (2000) 941-954.
[30] T. J. Li, *Symplectic 4-manifolds with Kodaira dimension zero*, submitted.
[31] T. J. Li, *Existence of symplectic surfaces*, to appear in the Fields-McMaster Proc. of a conference on Geometry and Topology of manifolds.
[32] T. J. Li, *Fibersums of Lefschetz fibrations*, preprint.
[33] B. H. Li, T. J. Li, *Symplectic genus, minimal genus and diffeomorphisms*, Asian J. Math. 26 (2002), 123-144.
[34] B. H. Li, T. J. Li, *On the diffeomorphism groups of rational and ruled 4-manifolds*, in preparation.
[35] T. J. Li and A. Liu, *General wall crossing formula*, Math. Res. Letters 2 (1995), 797-810.
[36] T. J. Li and A. Liu, *Symplectic structures on ruled surfaces and a generalized adjunction inequality*, Math. Res. Letters 2 (1995), 453-471.
[37] T. J. Li and A. Liu, *Family Seiberg-Witten invariants and wall crossing formulas*, Comm. in Analysis and Geometry Vol. 9 No. 4 (2001) 777-823.
[38] T. J. Li, A. Liu, *Uniqueness of symplectic canonical class, surface cone and symplectic cone of* $4-$*manifolds with* $b^+=1$, J. Differential Geom. Vol. 58 No. 2 (2001) 331-370.
[39] A-M. Li, Y. Ruan, *Symplectic surgery and Gromov-Witten invariants of Calabi-Yau 3-folds I*, Invent. Math. 145 (2001), no. 1, 151–218.
[40] T. J. Li, A. Stipsicz, *Minimality of certain normal connected sums*, Turkish J. Math. 26 (2002) 75-81.
[41] A. Liu, *Some new applications of the general wall crossing formula*, Math. Res. Letters 3 (1996), 569-585.
[42] A. Liu, *Family blowup formula, admissible graphs and the enumeration of singular curves (I)*, J. Diff. Geom. 56 (2001), 381-579.
[43] M. Marino, G. Moore, G. Peradze, *Four-manifold geography and superconformal symmetry*, Math. Res. Lett. 6 (1999), 429–437.
[44] Y. Matsumoto, *On 4-manifolds fibered by tori*, Proc. Japan Acad. 58 (1982), 298-301.
[45] D. McDuff, *The structure of rational and ruled symplectic* $4-$*manifolds*, J. AMS.v.1. no.3. (1990), 679-710.
[46] D. McDuff, *Lectures on Gromov invariants for symplectic 4-manifolds*, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 488, Gauge theory and symplectic geometry (Montreal, PQ, 1995) 175-210, Kluwer Acad. Publ., Dordrecht, 1997. the Proc. NATO Summer School, Montreal.
[47] D. McDuff and D. Salamon, *A survey of symplectic* $4-$*manifolds with* $b^+=1$, Proc. Gökova Conference, 1995, 47-60.
[48] J. McCarthy, J. Wolfson, *Symplectic normal connected sum*, Topology 33 (1994), 729-764.
[49] C. McMullen, C. Taubes, $4-$*manifolds with inequivalent symplectic forms and* $3-$*manifolds with inequivalent fibrations*, Math. Res. Lett. 6 (1999), no. 5-6, 681–696.
[50] J. Morgan, Z. Szabó, *Homotopy K3 surfaces and Mod 2 Seiberg-Witten invariants*, Math. Res. Lett. 4 (1997), 17-21.
[51] H. Ohta and K. Ono, *Notes on symplectic* 4-*manifolds with* $b_2^+=1$. II, Internat. J. Math. 7 (1996), no. 6, 755–770.
[52] J. Park, *The geography of Spin symplectic 4-manifolds*, Math. Z. 240 (2002), no. 2, 405–421.
[53] J. Park, *Simply connected symplectic 4-manifolds with* $b_2^+=1$ *and* $c_1^2=2$, to appear in Inv. Math..
[54] D. Ruberman, S. Strle, *Mod 2 Seiberg-Witten invariants of homology tori*, Math. Res. Lett. 7 (2000), 789-799.

[55] I. Smith, *On moduli spaces of symplectic forms*, Math. Res. Lett. 7 (2000), no. 5-6, 779–788.
[56] I. Smith, *Serre-Taubes Duality for pseudoholomorphic curves*, Topology 42 (2003), 931-979.
[57] A. Stipsicz, *Simply connected symplectic* 4*-manifolds with positive signature*, Proceedings of 6th Gökova Geometry-Topology Conference, Turkish J. Math. **23** (1999), no. 1, 145–150.
[58] A. Stipsicz, Z. Szabo, *An exotic smooth structure on* $\mathbf{CP}^2\#6(-\mathbf{CP}^2)$, preprint math.GT/0411258.
[59] C. H. Taubes, The Seiberg-Witten invariants and symplectic forms, Math. Res. Letters 1 (1994), 809-822.
[60] C. H. Taubes, *SW⇒Gr: From the Seiberg-Witten equations to pseudo-holomorphic curves*, Jour. Amer. Math. Soc. 9(1996), 845-918.
[61] C. H. Taubes, *Counting pseudo-holomorphic submanifolds in dimension four*, J. Diff. Geom. 44 (1996), no. 4, 818-893.
[62] C. H. Taubes, *More constraints on symplectic forms from Seiberg-Witten invariants*, Math. Res. Lett. 2 (1995), no. 1, 9–13.
[63] W. Thurston, *Some simple examples of symplectic manifolds*, Proc. Amer. Math. Soc. 55 (1976), 467-468.
[64] M. Ue, *Geometric 4-manifolds in the sense of Thurston and Seifert 4-manifolds I*, J. Math. Soc. Japan 42 (1990), 511-540.
[65] M. Usher, *The Gromov invariants and the Donaldson-Smith standard surface count*, Geom. Topol. 8 (2004), 565-610.
[66] M. Usher, *Standard surfaces and nodal curves in symplectic 4-manifolds*, preprint math.SG/0407494.
[67] S. Vidussi, *Homotopy* $K3$*'s with several symplectic structures*, Geom. Topol. 5 (2001), 267–285

School of Mathematics, University of Minnesota, Minneapolis, MN 55455
E-mail address: `tjli@math.umn.edu`

Clay Mathematics Proceedings
Volume 5, 2006

Symplectic 4-manifolds, Singular Plane Curves, and Isotopy Problems

Denis Auroux

Abstract. We give an overview of various recent results concerning the topology of symplectic 4-manifolds and singular plane curves, using branched covers and isotopy problems as a unifying theme. While this paper does not contain any new results, we hope that it can serve as an introduction to the subject, and will stimulate interest in some of the open questions mentioned in the final section.

1. Introduction

An important problem in 4-manifold topology is to understand which manifolds carry symplectic structures (i.e., closed non-degenerate 2-forms), and to develop invariants that can distinguish symplectic manifolds. Additionally, one would like to understand to what extent the category of symplectic manifolds is richer than that of Kähler (or complex projective) manifolds. Similar questions may be asked about singular curves inside, e.g., the complex projective plane. The two types of questions are related to each other via symplectic branched covers.

A branched cover of a symplectic 4-manifold with a (possibly singular) symplectic branch curve carries a natural symplectic structure. Conversely, using approximately holomorphic techniques it can be shown that every compact symplectic 4-manifold is a branched cover of the complex projective plane, with a branch curve presenting nodes (of both orientations) and complex cusps as its only singularities (cf. §3). The topology of the 4-manifold and that of the branch curve are closely related to each other; for example, using braid monodromy techniques to study the branch curve, one can reduce the classification of symplectic 4-manifolds to a (hard) question about factorizations in the braid group (cf. §4). Conversely, in some examples the topology of the branch curve complement (in particular its fundamental group) admits a simple description in terms of the total space of the covering (cf. §5).

In the language of branch curves, the failure of most symplectic manifolds to admit integrable complex structures translates into the failure of most symplectic branch curves to be isotopic to complex curves. While the symplectic isotopy

2000 *Mathematics Subject Classification.* 57R17 (53D35).

Partially supported by NSF grant DMS-0244844.

problem has a negative answer for plane curves with cusp and node singularities, it is interesting to investigate this failure more precisely. Various partial results have been obtained recently about situations where isotopy holds (for smooth curves; for curves of low degree), and about isotopy up to stabilization or regular homotopy (cf. §6). On the other hand, many known examples of non-isotopic curves can be understood in terms of twisting along Lagrangian annuli (or equivalently, Luttinger surgery of the branched covers), leading to some intriguing open questions about the topology of symplectic 4-manifolds versus that of Kähler surfaces.

2. Background

In this section we review various classical facts about symplectic manifolds; the reader unfamiliar with the subject is referred to the book [**19**] for a systematic treatment of the material.

Recall that a *symplectic form* on a smooth manifold is a 2-form ω such that $d\omega = 0$ and $\omega \wedge \cdots \wedge \omega$ is a volume form. The prototype of a symplectic form is the 2-form $\omega_0 = \sum dx_i \wedge dy_i$ on $\mathbb{R}^{2n}$. In fact, one of the most classical results in symplectic topology, Darboux's theorem, asserts that every symplectic manifold is locally symplectomorphic to $(\mathbb{R}^{2n}, \omega_0)$: hence, unlike Riemannian metrics, symplectic structures have no local invariants.

Since we are interested primarily in compact examples, let us mention compact oriented surfaces (taking ω to be an arbitrary area form), and the complex projective space $\mathbb{CP}^n$ (equipped with the Fubini-Study Kähler form). More generally, since any submanifold to which ω restricts non-degenerately inherits a symplectic structure, all complex projective manifolds are symplectic. However, the symplectic category is strictly larger than the complex projective category, as first evidenced by Thurston in 1976 [**36**]. In 1994 Gompf obtained the following spectacular result using the *symplectic sum* construction [**14**]:

Theorem 1 (Gompf). *Given any finitely presented group G, there exists a compact symplectic 4-manifold (X, ω) such that $\pi_1(X) \simeq G$.*

Hence, a general symplectic manifold cannot be expected to carry a complex structure; however, we can equip it with a compatible *almost-complex* structure, i.e. there exists $J \in \mathrm{End}(TX)$ such that $J^2 = -\mathrm{Id}$ and $g(\cdot,\cdot) := \omega(\cdot, J\cdot)$ is a Riemannian metric. Hence, at any given point $x \in X$ the tangent space (T_xX, ω, J) can be identified with $(\mathbb{C}^n, \omega_0, i)$, but there is no control over the manner in which J varies from one point to another (J is not *integrable*). In particular, the $\bar\partial$ operator associated to J does not satisfy $\bar\partial^2 = 0$, and hence there are no local holomorphic coordinates.

An important problem in 4-manifold topology is to understand the hierarchy formed by the three main classes of compact oriented 4-manifolds: (1) complex projective, (2) symplectic, and (3) smooth. Each class is a proper subset of the next one, and many obstructions and examples are known, but we are still very far from understanding what exactly causes a smooth 4-manifold to admit a symplectic structure, or a symplectic 4-manifold to admit an integrable complex structure.

One of the main motivations to study symplectic 4-manifolds is that they retain some (but not all) features of complex projective manifolds: for example the structure of their Seiberg-Witten invariants, which in both cases are non-zero and count certain embedded curves [**31, 32**]. At the same time, every compact oriented smooth 4-manifold with $b_2^+ \geq 1$ admits a "near-symplectic" structure, i.e. a closed 2-form which vanishes along a union of circles and is symplectic over the complement of its zero set [**13, 16**]; and it appears that some structural properties of symplectic manifolds carry over to the world of smooth 4-manifolds (see e.g. [**33, 5**]).

Many new developments have contributed to improve our understanding of symplectic 4-manifolds over the past ten years (while results are much scarcer in higher dimensions). Perhaps the most important source of new results has been the study of pseudo-holomorphic curves in their various incarnations: Gromov-Witten invariants, Floer homology, ...(for an overview of the subject see [**20**]). At the same time, gauge theory (mostly Seiberg-Witten theory, but also more recently Ozsváth-Szabó theory) has made it possible to identify various *obstructions* to the existence of symplectic structures in dimension 4 (cf. e.g. [**31, 32**]). On the other hand, various new constructions, such as link surgery [**11**], symplectic sum [**14**], and symplectic rational blowdown [**30**] have made it possible to exhibit interesting families of non-Kähler symplectic 4-manifolds. In a slightly different direction, approximately holomorphic geometry (first introduced by Donaldson in [**9**]) has made it possible to obtain various structure results, showing that symplectic 4-manifolds can be realized as symplectic Lefschetz pencils [**10**] or as branched covers of $\mathbb{CP}^2$ [**2**]. In the rest of this paper we will focus on this latter approach, and discuss the topology of *symplectic branched covers* in dimension 4.

3. Symplectic branched covers

Let X and Y be compact oriented 4-manifolds, and assume that Y carries a symplectic form ω_Y.

DEFINITION 2. *A smooth map $f : X \to Y$ is a* symplectic branched covering *if given any point $p \in X$ there exist neighborhoods $U \ni p$, $V \ni f(p)$, and local coordinate charts $\phi : U \to \mathbb{C}^2$ (orientation-preserving) and $\psi : V \to \mathbb{C}^2$ (adapted to ω_Y, i.e. such that ω_Y restricts positively to any complex line in $\mathbb{C}^2$), in which f is given by one of:*

(i) $(x, y) \mapsto (x, y)$ *(local diffeomorphism),*
(ii) $(x, y) \mapsto (x^2, y)$ *(simple branching),*
(iii) $(x, y) \mapsto (x^3 - xy, y)$ *(ordinary cusp).*

These local models are the same as for the singularities of a generic holomorphic map from $\mathbb{C}^2$ to itself, except that the requirements on the local coordinate charts have been substantially weakened. The *ramification curve* $R = \{p \in X,\ \det(df) = 0\}$ is a smooth submanifold of X, and its image $D = f(R)$ is the *branch curve*, described in the local models by the equations $z_1 = 0$ for $(x, y) \mapsto (x^2, y)$ and $27z_1^2 = 4z_2^3$ for $(x, y) \mapsto (x^3 - xy, y)$. The conditions imposed on the local coordinate charts imply that D is a *symplectic curve* in Y (i.e., $\omega_{Y|TD} > 0$ at every point of D). Moreover the restriction of f to R is an immersion everywhere except at the cusps. Hence, besides the ordinary complex cusps imposed by the local model, the only generic singularities of D are transverse double points ("nodes"), which may occur with either the complex orientation or the anti-complex orientation.

We have the following result [**2**]:

PROPOSITION 3. *Given a symplectic branched covering $f : X \to Y$, the manifold X inherits a natural symplectic structure ω_X, canonical up to isotopy, in the cohomology class $[\omega_X] = f^*[\omega_Y]$.*

The symplectic form ω_X is constructed by adding to $f^*\omega_Y$ a small multiple of an exact form α with the property that, at every point of R, the restriction of α to $\mathrm{Ker}(df)$ is positive. Uniqueness up to isotopy follows from the convexity of the space of such exact 2-forms and Moser's theorem.

Conversely, we can realize every compact symplectic 4-manifold as a symplectic branched cover of $\mathbb{CP}^2$ [**2**], at least if we assume *integrality*, i.e. if we require that $[\omega] \in H^2(X, \mathbb{Z})$, which does not place any additional restrictions on the diffeomorphism type of X:

THEOREM 4. *Given an integral compact symplectic 4-manifold (X^4, ω) and an integer $k \gg 0$, there exists a symplectic branched covering $f_k : X \to \mathbb{CP}^2$, canonical up to isotopy if k is sufficiently large.*

Moreover, the natural symplectic structure induced on X by the Fubini-Study Kähler form and f_k (as given by Proposition 3) agrees with ω up to isotopy and scaling (multiplication by k).

The main tool in the construction of the maps f_k is *approximately holomorphic geometry* [**9, 10, 2**]. Equip X with a compatible almost-complex structure, and consider a complex line bundle $L \to X$ such that $c_1(L) = [\omega]$: then for $k \gg 0$ the line bundle $L^{\otimes k}$ admits many approximately holomorphic sections, i.e. sections such that $\sup|\bar\partial s| \ll \sup|\partial s|$. Generically, a triple of such sections (s_0, s_1, s_2) has no common zeroes, and determines a projective map $f : p \mapsto [s_0(p) : s_1(p) : s_2(p)]$. Theorem 4 is then proved by constructing triples of sections which satisfy suitable transversality estimates, ensuring that the structure of f near its critical locus is the expected one [**2**]. (In the complex case it would be enough to pick three generic holomorphic sections, but in the approximately holomorphic context one needs to work harder and obtain uniform transversality estimates on the derivatives of f.)

Because for large k the maps f_k are canonical up to isotopy through symplectic branched covers, the topology of f_k and of its branch curve D_k can be used to define invariants of the symplectic manifold (X, ω). The only generic singularities of the plane curve D_k are nodes (transverse double points) of either orientation and complex cusps, but in a generic one-parameter family of branched covers pairs of nodes with opposite orientations may be cancelled or created. However, recalling that a node of D_k corresponds to the occurrence of two simple branch points in the same fiber of f_k, the creation of a pair of nodes can only occcur in a manner compatible with the branched covering structure, i.e. involving disjoint sheets of the covering. Hence, for large k the sequence of branch curves D_k is, up to isotopy (equisingular deformation among symplectic curves), cancellations and admissible creations of pairs of nodes, an invariant of (X, ω).

The ramification curve of f_k is just a smooth connected symplectic curve representing the homology class Poincaré dual to $3k[\omega] - c_1(TX)$, but the branch curve D_k becomes more and more complicated as k increases: in terms of the symplectic volume and Chern numbers of X, its degree (or homology class) d_k, genus g_k, and number of cusps κ_k are given by

$$d_k = 3k^2\,[\omega]^2 - k\,c_1 \cdot [\omega], \qquad 2g_k - 2 = 9k^2\,[\omega]^2 - 9k\,c_1 \cdot [\omega] + 2c_1^2,$$
$$\kappa_k = 12k^2\,[\omega]^2 - 9k\,c_1 \cdot [\omega] + 2c_1^2 - c_2.$$

It is also worth mentioning that, to this date, there is no evidence suggesting that negative nodes actually do occur in these high degree branch curves; our inability to rule our their presence might well be a shortcoming of the approximately holomorphic techniques, rather than an intrinsic feature of symplectic 4-manifolds. So in the following sections we will occasionally consider the more conventional problem of understanding isotopy classes of curves presenting only positive nodes and cusps, although most of the discussion applies equally well to curves with negative nodes.

Assuming that the topology of the branch curve is understood (we will discuss how to achieve this in the next section), one still needs to consider the branched covering f itself. The structure of f is determined by its *monodromy morphism* $\theta : \pi_1(\mathbb{CP}^2 - D) \to S_N$, where N is the degree of the covering f. Fixing a base point $p_0 \in \mathbb{CP}^2 - D$, the image by θ of a loop γ in the complement of D is the permutation of the fiber $f^{-1}(p_0)$ induced by the monodromy of f along γ. (Since viewing this permutation as an element of S_N depends on the choice of an identification between $f^{-1}(p_0)$ and $\{1, \dots, N\}$, the morphism θ is only well-defined up to conjugation by an element of S_N.) By Proposition 3, the isotopy class of the branch curve D and the monodromy morphism θ completely determine the symplectic 4-manifold (X, ω) up to symplectomorphism.

Consider a loop γ which bounds a small topological disc intersecting D transversely once: such a loop plays a role similar to the meridian of a knot, and is called a *geometric generator* of $\pi_1(\mathbb{CP}^2 - D)$. Then $\theta(\gamma)$ is a transposition (because of the local model near a simple branch point). Since the image of θ is generated by transpositions and acts transitively on the fiber (assuming X to be connected), θ is a surjective group homomorphism. Moreover, the smoothness of X above the singular points of D imposes certain compatibility conditions on θ. Therefore, not every singular plane curve can be the branch curve of a smooth covering; moreover, the morphism θ, if it exists, is often unique (up to conjugation in S_N). In the case of algebraic curves, this uniqueness property, which holds except for a finite list of well-known counterexamples, is known as Chisini's conjecture, and was essentially proved by Kulikov a few years ago [**18**].

The upshot of the above discussion is that, in order to understand symplectic 4-manifolds, it is in principle enough to understand singular plane curves. Moreover, if the branch curve of a symplectic covering $f : X \to \mathbb{CP}^2$ happens to be a complex curve, then the integrable complex structure of $\mathbb{CP}^2$ can be lifted to an integrable complex structure on X, compatible with the symplectic structure; this implies that X is a complex projective surface. So, considering the branched coverings constructed in Theorem 4, we have:

Corollary 5. *For $k \gg 0$ the branch curve $D_k \subset \mathbb{CP}^2$ is isotopic to a complex curve (up to node cancellations) if and only if X is a complex projective surface.*

This motivates the study of the *symplectic isotopy problem*, which we will discuss in §6. For now we focus on the use of braid monodromy invariants to study the topology of singular plane curves. In the present context, the goal of this approach is to reduce the classification of symplectic 4-manifolds to a purely algebraic problem, in a manner vaguely reminiscent of the role played by Kirby calculus in the classification of smooth 4-manifolds; as we shall see below, representing symplectic

4-manifolds as branched covers of $\mathbb{CP}^2$ naturally leads one to study the calculus of factorizations in braid groups.

4. The topology of singular plane curves

The topology of singular algebraic plane curves has been studied extensively since Zariski. One of the main tools is the notion of *braid monodromy* of a plane curve, which has been used in particular by Moishezon and Teicher in many papers since the early 1980s in order to study branch curves of generic projections of complex projective surfaces (see [**34**] for a detailed overview). Braid monodromy techniques can be applied to the more general case of *Hurwitz curves* in ruled surfaces, i.e. curves which behave in a generic manner with respect to the ruling. In the case of $\mathbb{CP}^2$, we consider the projection $\pi : \mathbb{CP}^2 - \{(0:0:1)\} \to \mathbb{CP}^1$ given by $(x : y : z) \mapsto (x : y)$.

Definition 6. *A curve $D \subset \mathbb{CP}^2$ (not passing through $(0:0:1)$) is a Hurwitz curve (or braided curve) if D is positively transverse to the fibers of π everywhere except at finitely many points where D is smooth and non-degenerately tangent to the fibers.*

The projection π makes D a singular branched cover of $\mathbb{CP}^1$, of degree $d = \deg D = [D] \cdot [\mathbb{CP}^1]$. Each fiber of π is a complex line $\ell \simeq \mathbb{C} \subset \mathbb{CP}^2$, and if ℓ does not pass through any of the singular points of D nor any of its vertical tangencies, then $\ell \cap D$ consists of d distinct points. We can trivialize the fibration π over an affine subset $\mathbb{C} \subset \mathbb{CP}^1$, and define the *braid monodromy morphism*

$$\rho : \pi_1(\mathbb{C} - \mathrm{crit}(\pi_{|D})) \to B_d.$$

Here B_d is the Artin braid group on d strings (the fundamental group of the configuration space $\mathrm{Conf}_d(\mathbb{C})$ of d distinct points in $\mathbb{C}$), and for any loop γ the braid $\rho(\gamma)$ describes the motion of the d points of $\ell \cap D$ inside the fibers of π as one moves along the loop γ.

Equivalently, choosing an ordered system of arcs generating the free group $\pi_1(\mathbb{C} - \mathrm{crit}(\pi_{|D}))$, one can express the braid monodromy of D by a *factorization*

$$\Delta^2 = \prod_i \rho_i$$

of the central element Δ^2 (representing a full rotation by 2π) in B_d, where each factor ρ_i is the monodromy around one of the special points (cusps, nodes, tangencies) of D.

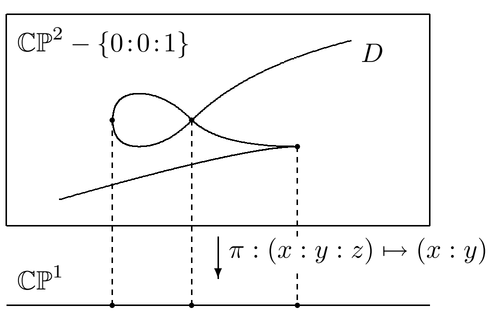


Figure 1. A Hurwitz curve in $\mathbb{CP}^2$

The same Hurwitz curve can be described by different factorizations of Δ^2 in B_d: namely, switching to a different ordered system of generators of $\pi_1(\mathbb{C} - \mathrm{crit}(\pi_{|D}))$ affects the collection of factors $\langle \rho_1, \dots, \rho_r \rangle$ by a sequence of *Hurwitz moves*, i.e. operations of the form

$$\langle \rho_1, \cdots, \rho_i, \rho_{i+1}, \cdots, \rho_r \rangle \longleftrightarrow \langle \rho_1, \cdots, (\rho_i \rho_{i+1} \rho_i^{-1}), \rho_i, \cdots, \rho_r \rangle;$$

and changing the identification between the reference fiber $(\ell, \ell \cap D)$ of π and the base point in $\mathrm{Conf}_d(\mathbb{C})$ affects braid monodromy by a *global conjugation*

$$\langle \rho_1, \cdots, \rho_r \rangle \longleftrightarrow \langle b^{-1} \rho_1 b, \cdots, b^{-1} \rho_r b \rangle.$$

For Hurwitz curves whose only singularities are cusps and nodes (of either orientation), or more generally curves with A_n (and $\overline{A}_n$) singularities, the braid monodromy factorization determines the isotopy type completely (see for example [**17**]). Hence, determining whether two given Hurwitz curves are isotopic among Hurwitz curves is equivalent to determining whether two given factorizations of Δ^2 coincide up to Hurwitz moves and global conjugation.

It is easy to see that any Hurwitz curve in $\mathbb{CP}^2$ can be made symplectic by an isotopy through Hurwitz curves: namely, the image of any Hurwitz curve by the rescaling map $(x : y : z) \mapsto (x : y : \lambda z)$ is a Hurwitz curve, and symplectic for $|\lambda| \ll 1$. On the other hand, a refinement of Theorem 4 makes it possible to assume without loss of generality that the branch curves $D_k \subset \mathbb{CP}^2$ are Hurwitz curves [**7**]. So, from now on we can specifically consider symplectic coverings with Hurwitz branch curves. In this setting, braid monodromy gives a purely combinatorial description of the topology of compact (integral) symplectic 4-manifolds.

The braid monodromy of the branch curves D_k given by Theorem 4 can be computed explicitly for various families of complex projective surfaces (non-Kähler examples are currently beyond reach). In fact, in the complex case the branched coverings f_k are isotopic to generic projections of projective embeddings. Accordingly, most of these computations rely purely on methods from algebraic geometry, using the degeneration techniques extensively developed by Moishezon and Teicher (see [**1, 21, 22, 24, 26, 34, 35**] and references within); but approximately holomorphic methods can be used to simplify the calculations and bring a whole new range of examples within reach [**6**]. This includes some complex surfaces of general type which are mutually homeomorphic and have identical Seiberg-Witten invariants but of which it is unknown whether they are symplectomorphic or even diffeomorphic (the *Horikawa surfaces*).

However, the main obstacle standing in the way of this approach to the topology of symplectic 4-manifolds is the intractability of the so-called "Hurwitz problem" for braid monodromy factorizations: namely, there is no algorithm to decide whether two given braid monodromy factorizations are identical up to Hurwitz moves. Therefore, since we are unable to compare braid monodromy factorizations, we have to extract the information contained in them by indirect means, via the introduction of more manageable (but less powerful) invariants.

5. Fundamental groups of branch curve complements

The idea of studying algebraic plane curves by determining the fundamental groups of their complements is a very classical one, which goes back to Zariski and Van Kampen. More recently, Moishezon and Teicher have shown that fundamental groups of branch curve complements can be used as a major tool to further our understanding of complex projective surfaces (cf. e.g. [**21, 25, 34**]). By analogy with the situation for knots in S^3, one expects the topology of the complement to carry a lot of information about the curve; however in this case the fundamental group does not determine the isotopy type. For an algebraic curve in $\mathbb{CP}^2$, or more generally for a Hurwitz curve, the fundamental group of the complement is determined in an explicit manner by the braid monodromy factorization, via the Zariski-Van Kampen theorem. Hence, calculations of fundamental groups of complements usually rely on braid monodromy techniques.

A close examination of the available data suggests that, contrarily to what has often been claimed, in the specific case of generic projections of complex surfaces projectively embedded by sections of a sufficiently ample linear system (i.e. taking $k \gg 0$ in Theorem 4), the fundamental group of the branch curve complement may be determined in an elementary manner by the topology of the surface (see below).

In the symplectic setting, the fundamental group of the complement of the branch curve D of a covering $f : X \to \mathbb{CP}^2$ is affected by node creation or cancellation operations. Indeed, adding pairs of nodes (in a manner compatible with the monodromy morphism $\theta : \pi_1(\mathbb{CP}^2 - D) \to S_N$) introduces additional commutation relations between geometric generators of the fundamental group. Hence, it is necessary to consider a suitable "symplectic stabilization" of $\pi_1(\mathbb{CP}^2 - D)$ [**6**]:

Definition 7. *Let K be the normal subgroup of $\pi_1(\mathbb{CP}^2 - D)$ generated by the commutators $[\gamma, \gamma']$ for all pairs γ, γ' of geometric generators such that $\theta(\gamma)$ and $\theta(\gamma')$ are disjoint commuting transpositions. Then the symplectic stabilization of $\pi_1(\mathbb{CP}^2 - D)$ is the quotient $\bar{G} = \pi_1(\mathbb{CP}^2 - D)/K$.*

Considering the branch curves D_k of the coverings given by Theorem 4, we have the following result [**6**]:

Theorem 8 (A.-Donaldson-Katzarkov-Yotov). *For $k \gg 0$, the stabilized group $\bar{G}_k(X, \omega) = \pi_1(\mathbb{CP}^2 - D_k)/K_k$ is an invariant of the symplectic manifold (X^4, ω).*

The fundamental group of the complement of a plane branch curve $D \subset \mathbb{CP}^2$ comes naturally equipped with two morphisms: the symmetric group valued monodromy homomorphism θ discussed above, and the abelianization map $\delta : \pi_1(\mathbb{CP}^2 - D) \to H_1(\mathbb{CP}^2 - D, \mathbb{Z})$. Since we only consider irreducible branch curves, we have $H_1(\mathbb{CP}^2 - D, \mathbb{Z}) \simeq \mathbb{Z}_d$, where $d = \deg D$, and δ counts the linking number (mod d) with the curve D. The morphisms θ and δ are surjective, but the image of $(\theta, \delta) : \pi_1(\mathbb{CP}^2 - D) \to S_N \times \mathbb{Z}_d$ is the index 2 subgroup consisting of all pairs (σ, p) such that the permutation σ and the integer p have the same parity (note that d is always even). The subgroup K introduced in Definition 7 lies in the kernel of (θ, δ); therefore, setting $G^0 = \mathrm{Ker}(\theta, \delta)/K$, we have an exact sequence

$$1 \longrightarrow G^0 \longrightarrow \bar{G} \stackrel{(\theta,\delta)}{\longrightarrow} S_N \times \mathbb{Z}_d \longrightarrow \mathbb{Z}_2 \longrightarrow 1.$$

Moreover, assume that the symplectic 4-manifold X is simply connected, and denote by $L = f^*[\mathbb{CP}^1]$ the pullback of the hyperplane class and by $K_X = -c_1(TX)$ the canonical class. Then we have the following result [**6**]:

Theorem 9 (A.-Donaldson-Katzarkov-Yotov). *If $\pi_1(X) = 1$ then there is a natural surjective homomorphism $\phi : \mathrm{Ab}(G^0) \twoheadrightarrow (\mathbb{Z}^2/\Lambda)^{N-1}$, where*

$$\Lambda = \{(L \cdot C, K_X \cdot C),\ C \in H_2(X, \mathbb{Z})\} \subset \mathbb{Z}^2.$$

The fundamental groups of the branch curve complements have been computed for generic polynomial maps to $\mathbb{CP}^2$ on various algebraic surfaces, using braid monodromy techniques (cf. §4) and the Zariski-Van Kampen theorem. Since in the symplectic setting Theorem 4 gives uniqueness up to isotopy only for $k \gg 0$, we restrict ourselves to those examples for which the fundamental groups have been computed for $\mathbb{CP}^2$-valued maps of arbitrarily large degree.

The first such calculations were carried out by Moishezon and Teicher, for $\mathbb{CP}^2$, $\mathbb{CP}^1 \times \mathbb{CP}^1$ [**22**], and Hirzebruch surfaces ([**24**], see also [**6**]); the answer is also known for some specific linear systems on rational surfaces and K3 surfaces realized as complete intersections (by work of Robb [**26**], see also related papers by Teicher et al). Additionally, the symplectic stabilizations of the fundamental groups have been computed for all double covers of $\mathbb{CP}^1 \times \mathbb{CP}^1$ branched along connected smooth algebraic curves [**6**], which includes an infinite family of surfaces of general type.

In all these examples it turns out that, if one considers projections of sufficiently large degree (i.e., assuming $k \geq 3$ for $\mathbb{CP}^2$ and $k \geq 2$ for the other examples), the structure of G^0 is very simple, and obeys the following conjecture:

Conjecture 10. *Assume that X is a simply connected algebraic surface and $k \gg 0$. Then:* (1) *the symplectic stabilization operation is trivial, i.e. $K = \{1\}$ and $\bar{G} = \pi_1(\mathbb{CP}^2 - D)$;* (2) *the homomorphism $\phi : \mathrm{Ab}(G^0) \to (\mathbb{Z}^2/\Lambda)^{N-1}$ is an isomorphism; and* (3) *the commutator subgroup $[G^0, G^0]$ is a quotient of $\mathbb{Z}_2 \times \mathbb{Z}_2$.*

6. The symplectic isotopy problem

The symplectic isotopy problem asks under which conditions (assumptions on degree, genus, types and numbers of singular points) it is true that any symplectic curve in $\mathbb{CP}^2$ (or more generally in a complex surface) is symplectically isotopic to a complex curve (by isotopy we mean a continuous family of symplectic curves with the same singularities).

The first result in this direction is due to Gromov, who proved that every smooth symplectic curve of degree 1 or 2 in $\mathbb{CP}^2$ is isotopic to a complex curve [**15**]. The argument relies on a careful study of the deformation problem for pseudo-holomorphic curves: starting from an almost-complex structure J for which the given curve C is pseudo-holomorphic, and considering a family of almost-complex structures $(J_t)_{t\in[0,1]}$ interpolating between J and the standard complex structure, one can prove the existence of smooth J_t-holomorphic curves C_t realizing an isotopy between C and a complex curve.

The isotopy property is expected to hold for smooth and nodal curves in all degrees, and also for curves with sufficiently few cusps. For smooth curves, successive improvements of Gromov's result have been obtained by Sikorav (for degree 3), Shevchishin (for degree ≤ 6), and more recently Siebert and Tian [**28**]:

Theorem 11 (Siebert-Tian). *Every smooth symplectic curve of degree ≤ 17 in $\mathbb{CP}^2$ is symplectically isotopic to a complex curve.*

Some results have been obtained by Barraud and Shevchishin for nodal curves of low genus. For example, the following result holds [**27**]:

Theorem 12 (Shevchishin). *Every irreducible nodal symplectic curve of genus $g \leq 4$ in $\mathbb{CP}^2$ is symplectically isotopic to a complex curve.*

Moreover, work in progress by S. Francisco is expected to lead to an isotopy result for curves of low degree with node and cusp singularities (subject to specific constraints on the number of cusps).

If one aims to classify symplectic 4-manifolds by enumerating branched covers of $\mathbb{CP}^2$ according to the degree and number of singularities of the branch curve, then the above cases are those for which the classification is the simplest and does not include any non-Kähler examples. On the other hand, Corollary 5 implies that the isotopy property cannot hold for all curves with node and cusp singularities; in fact, explicit counterexamples have been constructed by Moishezon [**23**] (see below).

Even when the isotopy property fails, the classification of singular plane curves becomes much simpler if one considers an equivalence relation weaker than isotopy, such as *regular homotopy*, or *stable isotopy*. Namely, let D_1, D_2 be two Hurwitz curves (see Definition 6) in $\mathbb{CP}^2$ (or more generally in a rational ruled surface), with node and cusp singularities (or more generally singularities of type A_n). Assume that D_1 and D_2 represent the same homology class, and that they have the same numbers of singular points of each type. Then we have the following results [**8, 17**]:

THEOREM 13 (A.-Kulikov-Shevchishin). *Under the above assumptions, D_1 and D_2 are* regular homotopic *among Hurwitz curves, i.e. they are isotopic up to creations and cancellations of pairs of nodes.*

THEOREM 14 (Kharlamov-Kulikov). *Under the above assumptions, let D'_i ($i \in \{1,2\}$) be the curve obtained by adding to D_i a union of n generic lines (or fibers of the ruling) intersecting D_i transversely at smooth points, and smoothing out all the resulting intersections. Then for all large enough values of n the Hurwitz curves D'_1 and D'_2 are isotopic.*

Unfortunately, Theorem 13 does not seem to have any implications for the topology of symplectic 4-manifolds, because the node creation operations appearing in the regular homotopy need not be admissible: even if both D_1 and D_2 are branch curves of symplectic coverings, the homotopy may involve plane curves for which the branched cover is not smooth.

For similar reasons, the direct applicability of Theorem 14 to branch curves is limited to the case of double covers, i.e. symplectic 4-manifolds which admit *hyperelliptic* Lefschetz fibrations. Nonetheless, a similar stable isotopy property also holds for arbitrary Lefschetz fibrations [**3**]:

THEOREM 15. *For every g there exists a genus g Lefschetz fibration f^0_g with the following property. Let $f_1 : M_1 \to S^2$ and $f_2 : M_2 \to S^2$ be two genus g Lefschetz fibrations, such that: (i) M_1 and M_2 have the same Euler characteristic and signature; (ii) f_1 and f_2 admit sections with the same self-intersection; (iii) f_1 and f_2 have the same numbers of reducible singular fibers of each type. Then, for all large enough values of n, the fiber sums $f_1 \# n\, f^0_g$ and $f_2 \# n\, f^0_g$ are isomorphic.*

In this statement, f^0_g can actually be chosen to be any genus g Lefschetz fibration which admits a section and whose monodromy generates the entire mapping class group $\mathrm{Map}_{g,1}$ of a genus g surface with one boundary component. (The argument in [**3**] relies on a specific choice of f^0_g, but the monodromy factorization of that particular f^0_g embeds into that of the fiber sum of sufficiently many copies of any other genus g Lefschetz fibration whose monodromy generates $\mathrm{Map}_{g,1}$).

Corollary 16. *Let X_1 and X_2 be two integral compact symplectic 4-manifolds with the same $(c_1^2,\ c_2,\ c_1{\cdot}[\omega],\ [\omega]^2)$. Then X_1 and X_2 become symplectomorphic after sufficiently many blowups and fiber sums with a same symplectic 4-manifold (the total space of the fibration f_g^0 for a suitable g).*

This result can be thought of as a symplectic analogue of the classical result of Wall which asserts that any two simply connected smooth 4-manifolds with the same intersection form become diffeomorphic after repeatedly performing connected sums with $S^2 \times S^2$ [**37**].

Returning to the symplectic isotopy problem, a closer look at the known examples of non-isotopic singular plane curves suggests that a statement much stronger than those mentioned above might hold.

It was first observed in 1999 by Fintushel and Stern [**12**] that many symplectic 4-manifolds contain infinite families of non-isotopic smooth connected symplectic curves representing the same homology class (see also [**29**]). The simplest examples are obtained by "braiding" parallel copies of the fiber in an elliptic surface, and are distinguished by comparing the Seiberg-Witten invariants of the corresponding double branched covers. Other examples have been constructed by Smith, Etgü and Park, and Vidussi. However, for singular plane curves the first examples were obtained by Moishezon more than ten years ago [**23**]:

Theorem 17 (Moishezon). *For all $p \geq 2$, there exist infinitely many pairwise non-isotopic singular symplectic curves of degree $9p(p-1)$ in $\mathbb{CP}^2$ with $27(p-1)(4p-5)$ cusps and $\frac{27}{2}(p-1)(p-2)(3p^2+3p-8)$ nodes, not isotopic to any complex curve.*

Moishezon's approach is purely algebraic (using braid monodromy factorizations), and very technical; the curves that he constructs are distinguished by the fundamental groups of their complements [**23**]. However, a much simpler geometric description of this construction can be given in terms of braiding operations, which makes it possible to distinguish the curves just by comparing the canonical classes of the associated branched covers [**4**].

Given a symplectic covering $f : X \to Y$ with branch curve D, and given a Lagrangian annulus A with interior in $Y \setminus D$ and boundary contained in D, we can *braid* the curve D along the annulus A by performing the local operation depicted in Figure 2. Namely, we cut out a neighborhood U of A, and glue it back via a non-trivial diffeomorphism which interchanges two of the connected components of $D \cap \partial U$, in such a way that the product of S^1 with the trivial braid is replaced by the product of S^1 with a half-twist (see [**4**] for details).

Braiding the symplectic curve D along the Lagrangian annulus A affects the branched cover X by a *Luttinger surgery* along a smooth embedded Lagrangian

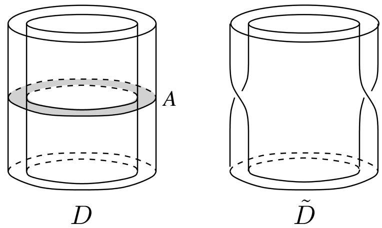


Figure 2. The braiding construction

torus T which is one of the connected components of $f^{-1}(A)$ [**4**]. This operation consists of cutting out from X a tubular neighborhood of T, foliated by parallel Lagrangian tori, and gluing it back via a symplectomorphism wrapping the meridian around the torus (in the direction of the preimage of an arc joining the two boundaries of A), while the longitudes are not affected.

The starting point of Moishezon's construction is the complex curve D_0 obtained by considering $3p(p-1)$ smooth cubics in a pencil, removing balls around the 9 points where these cubics intersect, and inserting into each location the branch curve of a generic degree p polynomial map from $\mathbb{CP}^2$ to itself. By repeatedly braiding D_0 along a well-chosen Lagrangian annulus, one obtains symplectic curves D_j, $j \in \mathbb{Z}$. Moishezon's calculations show that, whereas for the initial curve the fundamental group of the complement $\pi_1(\mathbb{CP}^2 - D_0)$ is infinite, the groups $\pi_1(\mathbb{CP}^2 - D_j)$ are finite for all $j \neq 0$, and of different orders [**23**]. On the other hand, it is fairly easy to check that, as expected from Theorem 9, this change in fundamental groups can be detected by considering the canonical class of the p^2-fold covering X_j of $\mathbb{CP}^2$ branched along D_j. Namely, the canonical class of X_0 is proportional to the cohomology class of the symplectic form induced by the branched covering: $c_1(K_{X_0}) = \lambda[\omega_{X_0}]$, where $\lambda = \frac{6p-9}{p}$. On the other hand, $c_1(K_{X_j}) = \lambda[\omega_{X_j}] + \mu\, j\, [T]^{PD}$, where $\mu = \frac{2p-3}{p} \neq 0$, and the homology class $[T]$ of the Lagrangian torus T is not a torsion element in $H_2(X_j, \mathbb{Z})$ [**4**].

Many constructions of non-Kähler symplectic 4-manifolds can be thought of in terms of twisted fiber sum operations, or Fintushel-Stern surgery along fibered links. However the key component in each of these constructions can be understood as a particular instance of Luttinger surgery; so it makes sense to ask to what extent Luttinger surgery may be responsible for the greater variety of symplectic 4-manifolds compared to complex surfaces. More precisely, we may ask the following questions:

QUESTION 18. *Let D_1, D_2 be two symplectic curves with nodes and cusps in $\mathbb{CP}^2$, of the same degree and with the same numbers of nodes and cusps. Is it always possible to obtain D_2 from D_1 by a sequence of braiding operations along Lagrangian annuli?*

QUESTION 19. *Let X_1, X_2 be two integral compact symplectic 4-manifolds with the same $(c_1^2,\, c_2,\, c_1\cdot[\omega],\, [\omega]^2)$. Is it always possible to obtain X_2 from X_1 by a sequence of Luttinger surgeries?*

This question is the symplectic analogue of a question asked by Ron Stern about smooth 4-manifolds, namely whether any two simply connected smooth 4-manifolds with the same Euler characteristic and signature differ from each other by a sequence of logarithmic transformations. However, here we do not require the manifolds to be simply connected; we do not even require them to have the same fundamental group.

References

[1] M. Amram, D. Goldberg, M. Teicher, U. Vishne, *The fundamental group of the Galois cover of the surface* $\mathbb{CP}^1 \times T$, Alg. Geom. Topol. **2** (2002), 403–432.

[2] D. Auroux, *Symplectic 4-manifolds as branched coverings of* $\mathbb{CP}^2$, Invent. Math. **139** (2000), 551–602.

[3] D. Auroux, *A stable classification of Lefschetz fibrations*, Geom. Topol. **9** (2005), 203–217.

[4] D. Auroux, S. K. Donaldson, L. Katzarkov, *Luttinger surgery along Lagrangian tori and non-isotopy for singular symplectic plane curves*, Math. Ann. **326** (2003), 185–203.

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

[6] D. Auroux, S. K. Donaldson, L. Katzarkov, M. Yotov, *Fundamental groups of complements of plane curves and symplectic invariants*, Topology **43** (2004), 1285–1318.

[7] D. Auroux, L. Katzarkov, *Branched coverings of* $\mathbb{CP}^2$ *and invariants of symplectic 4-manifolds*, Invent. Math. **142** (2000), 631–673.

[8] D. Auroux, V. S. Kulikov, V. V. Shevchishin, *Regular homotopy of Hurwitz curves*, Izv. Math. **68** (2004), 521–542 (`math.SG/0401172`).

[9] S. K. Donaldson, *Symplectic submanifolds and almost-complex geometry*, J. Diff. Geom. **44** (1996), 666–705.

[10] S.K. Donaldson, *Lefschetz pencils on symplectic manifolds*, J. Diff. Geom. **53** (1999), 205–236.

[11] R. Fintushel, R. Stern, *Knots, links, and 4-manifolds*, Invent. Math. **134** (1998), 363–400.

[12] R. Fintushel, R. Stern, *Symplectic surfaces in a fixed homology class*, J. Diff. Geom. **52** (1999), 203–222.

[13] D. T. Gay, R. Kirby, *Constructing symplectic forms on 4-manifolds which vanish on circles*, Geom. Topol. **8** (2004), 743–777.

[14] R. E. Gompf, *A new construction of symplectic manifolds*, Ann. Math. **142** (1995), 527–595.

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

[16] K. Honda, *Transversality theorems for harmonic forms*, Rocky Mountain J. Math. **34** (2004), 629–664.

[17] V. Kharlamov, V. Kulikov, *On braid monodromy factorizations*, Izv. Math. **67** (2003), 79–118.

[18] V. Kulikov, *On a Chisini conjecture*, Izv. Math. **63** (1999), 1139–1170.

[19] D. McDuff, D. Salamon, *Introduction to symplectic topology*, Oxford University Press, Oxford, 1995.

[20] D. McDuff, D. Salamon, *J-holomorphic curves and symplectic topology*, Amer. Math. Soc. Colloquium Publ. **52**, Amer. Math. Soc., Providence, 2004.

[21] B. Moishezon, *Stable branch curves and braid monodromies*, Algebraic Geometry (Chicago, 1980), Lect. Notes in Math. **862**, Springer, Heidelberg, 1981, pp. 107–192.

[22] B. Moishezon, *On cuspidal branch curves*, J. Algebraic Geom. **2** (1993) 309–384.

[23] B. Moishezon, *The arithmetic of braids and a statement of Chisini*, Geometric Topology (Haifa, 1992), Contemp. Math. **164**, Amer. Math. Soc., Providence, 1994, pp. 151–175.

[24] B. Moishezon, A. Robb, M. Teicher, *On Galois covers of Hirzebruch surfaces*, Math. Ann. **305** (1996), 493–539.

[25] B. Moishezon, M. Teicher, *Simply connected algebraic surfaces of positive index*, Invent. Math. **89** (1987), 601–643.

[26] A. Robb, *On branch curves of algebraic surfaces*, Singularities and Complex Geometry (Beijing, 1994), Amer. Math. Soc./Int. Press Stud. Adv. Math. **5**, Amer. Math. Soc., Providence, 1997, pp. 193–221.

[27] V. V. Shevchishin, *On the local Severi problem*, Int. Math. Res. Not. (2004), 211–237 (`math.AG/0207048`).

[28] B. Siebert, G. Tian, *On the holomorphicity of genus two Lefschetz fibrations*, Ann. Math. **161** (2005), 959–1020 (`math.SG/0305343`).

[29] I. Smith, *Symplectic submanifolds from surface fibrations*, Pacific J. Math. **198** (2001), 197–205.

[30] M. Symington, *Symplectic rational blowdowns*, J. Diff. Geom. **50** (1998), 505–518.

[31] C. H. Taubes, *The Seiberg-Witten and the Gromov invariants*, Math. Res. Lett. **2** (1995), 221–238.

[32] C. H. Taubes, *The geometry of the Seiberg-Witten invariants*, Surveys in Differential Geometry, Vol. III (Cambridge, 1996), Int. Press, Boston, 1998, pp. 299–339.

[33] C. H. Taubes, *Seiberg-Witten invariants and pseudo-holomorphic subvarieties for self-dual, harmonic 2-forms*, Geom. Topol. **3** (1999), 167–210.

[34] M. Teicher, *Braid groups, algebraic surfaces and fundamental groups of complements of branch curves*, Algebraic Geometry (Santa Cruz, 1995), Proc. Sympos. Pure Math., **62** (part 1), Amer. Math. Soc., Providence, 1997, pp. 127–150.

[35] M. Teicher, *New invariants for surfaces*, Tel Aviv Topology Conference: Rothenberg Festschrift (1998), Contemp. Math. **231**, Amer. Math. Soc., Providence, 1999, pp. 271–281.

[36] W. Thurston, *Some simple examples of symplectic manifolds*, Proc. Amer. Math. Soc. **55** (1976), 467–468.

[37] C. T. C. Wall, *On simply-connected 4-manifolds*, J. London Math. Soc. **39** (1964), 141–149.

Department of Mathematics, M.I.T., Cambridge MA 02139, USA
E-mail address: auroux@math.mit.edu

Clay Mathematics Proceedings
Volume 5, 2006

Monodromy, Vanishing Cycles, Knots and the Adjoint Quotient

Ivan Smith

Abstract. After reviewing some (mostly standard) material on symplectic fibre bundles, we describe a cohomology theory for oriented links in the three-sphere. This cohomological invariant, introduced in joint work with Paul Seidel, is defined by combining results from Lie algebra theory with Lagrangian Floer cohomology, and conjecturally equals Khovanov cohomology after collapsing the latter's grading.

These notes, are divided into two parts. The first part describes as background some of the geometry of symplectic fibre bundles and their monodromy. The second part applies these general ideas to certain Stein fibre bundles that arise naturally in Lie theory, to construct an invariant of oriented links in the three-sphere (Section 2.8). Despite its very different origins, this invariant is conjecturally equal to the combinatorial homology theory defined by Mikhail Khovanov (Section 2.9). In the hope of emphasising the key ideas, concision has taken preference over precision; there are no proofs, and sharp(er) forms of statements are deferred to the literature.

Much of the first part I learned from, and the second part represents joint work with, Paul Seidel, whose influence and insights generously pervade all that follows.

1. Monodromy, vanishing cycles

Most of the material in this section is well-known; general references are [**16**], [**23**],[**5**].

1.1. Symplectic fibre bundles: We will be concerned with fibrations $p : X \to B$ with symplectic base and fibre, or more precisely where X carries a closed vertically non-degenerate 2-form Ω, for which $d\Omega(u, v, \cdot) = 0$ whenever u, v are vertical tangent vectors. If the fibration is proper, the cohomology class $[\Omega|_{\mathrm{Fibre}}]$ is locally constant, and parallel transport maps are symplectomorphisms. Examples abound:

(1) A surface bundle over any space $\Sigma_g \to X \to B$ with fibre essential in homology can be given this structure; define Ω by picking any 2-form dual to the fibre and whose restriction to each fibre is an area form. The homology constraint is

1991 *Mathematics Subject Classification.* 53D35, 57M25.

Thanks to Aaron Lauda for help with the pictures. This work was partially supported by a grant NUF-NAL/00876/G from the Nuffield Foundation.

automatically satisfied whenever $g \geq 2$ (evaluate the first Chern class of the vertical tangent bundle on a fibre).

(2) Given a holomorphic map $p : X \to B$ defined on a quasiprojective variety and which is smooth over $B^0 \subset B$, the restriction $p^{-1}(B^0) \to B^0$ defines a symplectic fibre bundle, where the 2-form Ω is the restriction of a Kähler form on X. Such examples show the importance of *singular fibres*. Rational maps and linear systems in algebraic geometry provide a plethora of interesting singular fibrations.

(3) Contrastingly, the (singular) fibrations arising from moment maps, cotangent bundle projections and many dynamical systems have Lagrangian fibres, and fall outside the scope of the machinery we'll discuss.

Strictly, it is sensible to make a distinction between Hamiltonian and more general symplectic fibrations; essentially this amounts to the question of whether the vertically non-degenerate 2-form Ω has a closed extension to the total space, as in the cases above. The subtlety will not play any role in what follows, but for discussion see [**16**].

1.2. Parallel transport: A symplectic fibre bundle has a distinguished connexion, where the horizontal subspace at $x \in X$ is the symplectic orthogonal complement to the vertical distribution $\mathrm{Hor}_x = \ker(dp_x)^{\perp_\Omega}$. For Kähler fibrations, since the fibres are complex submanifolds, we can also define this as the orthogonal complement to $\ker(dp_x)$ with respect to the Kähler metric. We should emphasise at once that, in contrast to the Darboux theorem which prevents local curvature-type invariants entering symplectic topology naively, there is no "universal triviality" result for symplectic fibre bundles. The canonical connexion can, and often does, have curvature, and that curvature plays an essential part in the derivations of some of the theorems of the sequel.

Given a path $\gamma : [0,1] \to B$ we can lift the tangent vector $d\gamma/dt$ to a horizontal vector field on $p^{-1}(\gamma)$, and flowing along the integral curves of this vector field defines local *symplectomorphisms* h^γ of the fibres.

(1) If $p : X \to B$ is proper, the horizontal lifts can be globally integrated and we see that p is a fibre bundle with structure group $\mathrm{Symp}(p^{-1}(b))$. Note that, since the connexion isn't flat, the structure group does not in general reduce to the symplectic mapping class group (of components of Symp).

(2) Often there is a group G acting fibrewise and preserving all the structure, in which case parallel transport will be G-equivariant. An example will be given shortly.

1.3. Non-compactness: If the fibres are not compact, the local parallel transport maps may not be globally defined, since the solutions to the differential equations defining the integral curves may not exist for all times. To overcome this, there are several possible strategies. The simplest involves estimating the parallel transport vector fields explicitly (which in turn might rely on choosing the right Ω).

Suppose for instance $p : X \to \mathbb{C}$ where X has a Kähler metric; then for $V \in T_{p(x)}(\mathbb{C})$ the lift $V^{hor} = V.\frac{(\nabla p)_x}{|\nabla p|^2}$. If $p : (\mathbb{C}^n, \omega_{st}) \to \mathbb{C}$ is a homogeneous polynomial, clearly its only critical value is the origin, giving a fibre bundle over $\mathbb{C}^*$. The identity $dp_x(x) = deg(p).p(x)$, together with the previous formula for V^{hor}, shows that the horizontal lift of a tangent vector $V \in T_p(x)(\mathbb{C}) = \mathbb{C}$ has norm $|V^{hor}| \leq \frac{|V|.|x|}{deg(p).|p(x)|}$. On a fixed fibre $p = const$ this grows linearly with $|x|$ and can be globally integrated.

Corollary: For homogeneous polynomials $p : (\mathbb{C}^n, \omega_{st}) \to \mathbb{C}$ parallel transport is globally defined over $\mathbb{C}^*$.

Example: the above applies to the determinant mapping (indeed any single component of the characteristic polynomial or *adjoint quotient* map), $\det : \mathrm{Mat}_n(\mathbb{C}) \to \mathbb{C}$. In this case, parallel transport is invariant under $SU_n \times SU_n$. The monodromy of the associated bundle seems never to have been investigated.

For mappings $p : \mathbb{C}^n \to \mathbb{C}^m$ in which each component is a homogeneous polynomial but the homogeneous degrees differ, the above arguments do not quite apply but another approach can be useful. Since the smooth fibres are Stein manifolds of finite type, we can find a vector field Z which points inwards on all the infinite cones. By flowing with respect to a vector field $V^{hor} - \delta Z$, for large enough δ, and then using the Liouville flows, we can define "rescaled" parallel transport maps $h^{\gamma}_{resc} : p^{-1}(t) \cap B(R) \hookrightarrow p^{-1}(t')$ on arbitrarily large pieces of a fixed fibre $p^{-1}(t)$, which embed such compacta symplectically into another fibre $p^{-1}(t')$. This is not quite the same as saying that the fibres are globally symplectomorphic, but is enough to transport closed Lagrangian submanifolds around (uniquely up to isotopy), and often suffices in applications. For a detailed discussion, see [**29**]. (In fact, if the Stein fibres are finite type and complete one can "uncompress" the flows above to show the fibres really are globally symplectomorphic, cf. [**13**].)

1.4. Vanishing cycles: The local geometry near a singularity (critical fibre of p) shows up in the *monodromy* of the smooth fibre bundle over B^0, i.e. the representation

$$\pi_1(B^0, b) \to \pi_0(\mathrm{Symp}(p^{-1}(b), \Omega)).$$

Consider the ordinary double point (Morse singularity, node...)

$$p : (z_1, \dots, z_n) \mapsto \sum z_i^2.$$

The smooth fibres $p^{-1}(t)$, when equipped with the restriction of the flat Kähler form $(i/2)\sum_j dz_j \wedge d\bar{z}_j$ from $\mathbb{C}^n$, are symplectically isomorphic to $(T^*\mathbb{S}^{n-1}, \omega_{can})$. Indeed, an explicit symplectomorphism can be given in co-ordinates by viewing

$$T^*\mathbb{S}^{n-1} = \{(a, b) \in \mathbb{R}^n \times \mathbb{R}^n \mid |a| = 1, \langle a, b\rangle = 0\}$$

and taking $p^{-1}(1) \ni z \mapsto (\Re(z)/|\Re(z)|, -|\Re(z)|\Im(z))$. There is a distinguished Lagrangian submanifold of the fibre, the zero-section, which can also be defined as the locus of points which flow into the singularity under parallel transport along a radial line in $\mathbb{C}$. Accordingly, this locus – which is $\{(z_1, \dots, z_n) \in \mathbb{R}^n \mid \sum_i |z_i|^2 = 1\}$ in co-ordinates – is also called the *vanishing cycle* of the singularity.

Lemma: The monodromy about a loop encircling $0 \in \mathbb{C}$ is a Dehn twist in the vanishing cycle.

To define the Dehn twist, fix the usual metric on $\mathbb{S}^{n-1} \subset \mathbb{R}^n$, which identifies $T^*\mathbb{S}^{n-1} \cong T\mathbb{S}^{n-1}$. The Dehn twist is the composite of the time π map of the geodesic flow on the unit disc tangent bundle $U(T\mathbb{S}^{n-1})$ with the map induced by the antipodal map; it's antipodal on the zero-section and vanishes on the boundary ∂U. If $n = 2$ this construction is classical, and we get the usual Dehn twist on a curve in an annulus $T^*\mathbb{S}^1$.

1.5. Variant constructions: There are two useful variants of the above model:

(1a) The relative version of the above geometry: If we have $p : X \to \mathbb{C}$ and the fibre X_0 over $0 \in \mathbb{C}$ has smooth singular locus Z^c, with normal data locally holomorphically modelled on the map $\sum_{i=1}^n z_i^2$, we say p has a "fibred A_1-singularity". Then the nearby smooth fibre X_t contains a relative vanishing cycle $\mathbb{S}^{n-1} \to Z \to Z^c$. An open neighbourhood U of $Z \subset X_t$ is of the form $T^*\mathbb{S}^{n-1} \hookrightarrow U \to Z^c$, and the monodromy about 0 is a "fibred Dehn twist" (the above construction in every T^*-fibre).

(1b) Examples: Fix a stable curve over a disc $f : \mathfrak{X} \to \Delta$ with all fibres smooth except for an irreducible curve with a single node over the origin. The relative Picard fibration $\mathrm{Pic}(f) \to \Delta$ has a singular fibre over 0 with a fibred A_1-singularity, and with singular set the Picard of the normalisation $\tilde{C}$ of $C = f^{-1}(0)$. The relative Hilbert scheme $\mathrm{Hilb}^r(f) \to \Delta$ has a fibred A_1-singularity over 0 with singular locus $\mathrm{Hilb}^{r-1}(\tilde{C}) = \mathrm{Sym}^{r-1}(\tilde{C})$. In both these cases $n = 2$. For the relative moduli space of stable rank two bundles with fixed odd determinant, there is again a model for the compactification (symplectically, not yet constructed algebraically) with a fibred A_1-singularity, but this time with $n = 4$ and hence $\mathbb{S}^3 \cong SU(2)$ vanishing cycles [**28**].

(2a) Morsification: If we have a "worse" isolated singular point at the origin of a hypersurface defined e.g. by some polynomial $P(x) = 0$ then often we can perturb to some $P_\varepsilon(x) = 0$ which is an isotopic hypersurface outside a compact neighbourhood of the original singularity but which now only has a collection of finitely many nodes. In particular, the global monodromy of the projection of $P^{-1}(0)$ to the first co-ordinate in $\mathbb{C}$ about a large circle is a *product of Dehn twists in Lagrangian spheres* in the generic fibre.

(2b) Examples: The triple point $x^3 + y^3 + z^3 + t^3 = 0$ can be perturbed to a hypersurface with 16 nodes, or completely smoothed to give a configuration of 16 Lagrangian vanishing cycles, cf. [**34**].

The second result enters into various "surgery theoretic" arguments, along the following lines. Given a symplectic manifold with a tree-like configuration of Lagrangian spheres which matches the configuration of vanishing cycles of the Morsified singularity, we can cut out a (convex) neighbourhood of the tree and replace it with the resolution (full blow-up) of the original singular point, and this is a *symplectic* surgery. Examples are given in [**34**].

1.6. Lefschetz fibrations: A remarkable theorem due to Donaldson asserts that every symplectic manifold admits a Lefschetz pencil. In dimension four, this comprises a map $f : X\backslash\{b_i\} \to \mathbb{S}^2$ submersive away from a finite set $\{p_j\}$, and with f given by z_1/z_2 near b_i and $z_1 z_2$ near p_j. Removing fibres $f^{-1}(p_j)$ gives a symplectic fibre bundle over $\mathbb{S}^2\backslash\{f(p_j)\}$, and the global monodromy is encoded as a word in positive Dehn twists in $\Gamma_g = \pi_0(\mathrm{Symp}(\Sigma_g))$. In general, Donaldson's theory of symplectic linear systems reduces a swathe of symplectic topology to combinatorial group theory, and places issues of monodromy at the centre of the symplectic stage.

Example: The equation

$$\left(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \right)^{6n} = (AB)^{6n} = I$$

encodes the elliptic surface $E(n)$ as a word in $SL(2,\mathbb{Z})$. The fact that all words in matrices conjugate to $A \in SL_2(\mathbb{Z})$ are equivalent by the Hurwitz action of the braid group to a word of this form, for some n, gives an algebraic proof that the $E(n)$ exhaust all elliptic Lefschetz fibrations, hence all such are Kähler. In turn, from this one can deduce that degree 4 symplectic surfaces in $\mathbb{CP}^2$ are isotopic to complex curves.

(1) This kind of algebraic monodromy encoding generalises branched covers of Riemann surfaces and gives (in principle) a classification of integral symplectic 4-manifolds.

(2) The importance of *Lagrangian* intersection theory – i.e. geometric and not algebraic intersections of curves on a 2d surface – already becomes clear.

(3) Donaldson has suggested that the algebraic complexity of Lefschetz fibrations might be successfully married with the algebraic structure of Floer homology [**8**]. Steps in this direction were first taken by Seidel in the remarkable [**25**], see also [**27**],[**4**].

Donaldson's initial ideas have been developed and extended in a host of useful and indicative directions: we mention a few. Lefschetz pencils can be constructed adapted to embedded symplectic submanifolds or Lagrangian submanifolds [**3**] (in the latter case one extends a Morse function on L to a Lefschetz pencil on $X \supset L$); there are higher-dimensional linear systems, leading to iterative algebraic encodings of symplectic manifolds [**1**]; analogues exist in contact topology [**18**] and, most recently, for (non-symplectic) self-dual harmonic 2-forms on four-manifolds [**2**]. In each case, the techniques give an algebraic encoding of some important piece of geometric data.

Challenge: show that symplectomorphism of integral simply-connected symplectic 4-manifolds is (un)decidable.

1.7. Counting sections: A good way to define an invariant for a Lefschetz fibration is to replace the fibres with something more interesting and then count holomorphic sections of the new beast. In other words, one studies the Gromov-Witten invariants for those homology classes which have intersection number 1 with the fibre of the new fibration.

Explicitly, suppose we have a moduli problem on Riemann surfaces in the following sense: $\Sigma \mapsto \mathcal{M}(\Sigma)$ associates to a Riemann surface Σ some projective or quasiprojective moduli space, with a relative version for families of irreducible stable curves. A Lefschetz fibration $f : X \to \mathbb{S}^2$ gives rise to a relative moduli space $F : \mathcal{M}(f) \to \mathbb{S}^2$; our assumptions on the moduli problem should ensure that this is smooth and symplectic, and is either convex at infinity, compact or has a natural compactification. Then we associate $(X, f) \mapsto Gr_A(F)$ where A is the homology class of some fixed section. This follows a philosophy derived from algebraic geometry: holomorphic sections of a family of moduli spaces on the fibres should be "equivalent information" to data about geometric objects on the total space which could, in principle, be defined without recourse to any given fibration structure. Naive as this sounds, the theory is not entirely hopeless in the actual examples.

(1a) For $f : X^4 \to \mathbb{S}^2$, replace X_t by $\mathrm{Sym}^r(X_t)$ and desingularise, forming the relative Hilbert scheme [**9**] to get $F : X_r(f) \to \mathbb{S}^2$. Obviously sections of this new fibration are related to 2-cycles in the original four-manifold. A pretty theorem due to Michael Usher [**36**] makes this intuition concrete and sets the theory in a very satisfactory form: the Gromov-Witten invariants $\mathcal{I}_{X,f}$ counting sections of F, known as the *standard surface count*, are equal to Taubes' $Gr(X)$ [**35**]. In particular, the invariants are independent of f, as algebraic geometers would expect.

(1b) Application [**9**], [**33**]: if $b_+ > 2$ the invariant $\mathcal{I}_{X,f}(\kappa) = \pm 1$, where κ refers to the unique homology class of section for which the cycles defined in X lie in the class Poincaré dual to K_X. This gives a Seiberg-Witten free proof of the fact that, for minimal such manifolds, $c_1^2(X) \geq 0$. The key to the argument is the Abel-Jacobi map $\mathrm{Sym}^r\Sigma \to \mathrm{Pic}^r(\Sigma) \cong \mathbb{T}^{2g}$, which describes $\mathrm{Sym}^r(\Sigma)$ as a family of projective spaces over a torus; for the corresponding fibrations with fibre $\mathbb{P}^n$ or $\mathbb{T}^{2g}$ one can compute moduli spaces of holomorphic sections explicitly, and hence compute Gromov invariants.

(2) One can also count sections of symplectic Lefschetz fibrations over surfaces with boundary, provided suitable Lagrangian boundary conditions are specified. In place of absolute invariants one obtains invariants living in Floer homology groups associated to the boundary, or, formulated differently, morphisms on Floer homology groups. This is reminiscent of the formalism of Topological Quantum Field Theory; such ideas are central to the main theorem of [**23**].

In (1), the fact that the compactifications of the relative moduli spaces exist and are smooth can be understood in terms of the local geometry of fibred A_1-singularities and normal crossings, as in the discussion of Section 1.5, 1.2 above.

1.8. Braid relations: It's harder to get invariants of the total space straight out of the monodromy of a fibre bundle, but it is very natural to study $\pi_0\mathrm{Symp}(Fibre)$ this way. Let $p : \mathbb{C}^{n+1} \to \mathbb{C}$ be given by $\{x^{k+1} + \sum_{j=1}^{n} y_j^2 = \varepsilon\}$. There are $(k+1)$ critical values, and if we fix a path between two of them then we can construct a Lagrangian $\mathbb{S}^{n+1}$ in the total space by "matching" vanishing $\mathbb{S}^n$-cycles associated to two critical points [**27**]. This is just the reverse process of finding a Lefschetz fibration adapted to a given Lagrangian $(n+1)$-sphere, by extending the obvious Morse function from $\mathbb{S}^{n+1}$ to the total space, mentioned above.

Lemma: For two Lagrangian spheres L_1, L_2 meeting transversely in a point, the Dehn twists τ_{L_i} satisfy the braid relation $\tau_{L_1}\tau_{L_2}\tau_{L_1} = \tau_{L_2}\tau_{L_1}\tau_{L_2}$.

The proof of this is by direct computation [**22**]. In the lowest dimensional case $n = 1$ it is completely classical. A disguised version of the same Lemma will underlie central properties of a fibre bundle of importance in our application to knot theory in the second part.

Corollary: If X contains an A_k-chain of Lagrangian spheres, there is a natural homomorphism $Br_k \to \pi_0(\mathrm{Symp}_{ct}(X))$.

These homomorphisms have come to prominence in part because of mirror symmetry, cf. [**31**]. The relevant chains of Lagrangian spheres can be obtained by Morsifying A_{k+1}-singularities. The existence of the homomorphism, of course, gives no information on its non-triviality; we address that next.

1.9. Simultaneous resolution: The map $\mathbb{C}^3 \to \mathbb{C}$ which defines a node $z_1^2 + z_2^2 + z_3^2$ has fibres $T^*\mathbb{S}^2$ and one singular fibre. If we pull back under a double cover $\mathbb{C} \to \mathbb{C}, w \mapsto w^2$ then we get $\sum z_i^2 - w^2 = 0$, i.e. a 3-fold ODP, which has a *small resolution*; replacing the singular point by $\mathbb{C}P^1$ gives a smooth space.

Corollary: the fibre bundle upstairs is *differentiably* trivial, since it completes to a fibre bundle over the disc.

Seidel showed in [**26**] that this is not true symplectically; the Dehn twist in $T^*\mathbb{S}^2$ has infinite order as a symplectomorphism. So the natural map

$$\pi_0(\mathrm{Symp}_{ct}(T^*\mathbb{S}^2)) \to \pi_0(\mathrm{Diff}_{ct}(T^*\mathbb{S}^2))$$

has infinite kernel – the interesting structure is *only visible symplectically.* For a smoothing of the A_{k+1}-singularity above, a similar picture shows that the braid group acts *faithfully* by symplectomorphisms but factors through Sym_k acting by diffeomorphisms (compactly supported in each case). The injectivity is established by delicate Floer homology computations [**13**]. Such a phenomenon is at least possible whenever one considers families with *simultaneous resolutions*; that is, a family $\mathcal{X} \xrightarrow{\phi} B$ for which there is a ramified covering $\tilde{B} \to B$ and a family $\tilde{\mathcal{X}} \xrightarrow{\tilde{\phi}} \tilde{B}$ with a map $\pi : \tilde{\mathcal{X}} \to \mathcal{X}$ and with

$$\pi : \tilde{\mathcal{X}}_t = \tilde{\phi}^{-1}(t) \to \phi^{-1}(\pi(t)) = \mathcal{X}_t$$

a resolution of singularities for every $t \in \tilde{B}$. The small resolution of the 3-fold node will be the first in a sequence of simultaneous resolutions considered in the second section, and in each case the inclusion of Symp_{ct} into Diff_{ct} of the generic fibre will have infinite kernel.

1.10. Long exact sequences: Aside from their role in monodromy, Lagrangian spheres and Dehn twists also give rise to special structures and properties of Floer cohomology. Suppose L_1, L_2 are Lagrangians in X and $L \cong \mathbb{S}^n$ is a Lagrangian sphere. The main theorem of [**23**] is the following:

Theorem: (Seidel) Under suitable technical conditions, there is a long exact triangle of Floer cohomology groups

$$\begin{tikzcd} HF(L_1, L_2) \arrow[r] & HF(L_1, \tau_L(L_2)) \arrow[d] \\ & HF(L_1, L) \otimes HF(L, L_2) \arrow[ul] \end{tikzcd}$$

The technical conditions are in particular valid for exact Lagrangian submanifolds of a Stein manifold of finite type; in this setting there is no bubbling, and the manifold will be convex at infinity which prevents loss of compactness from solutions escaping to infinity. Hence, the Floer homology groups are well-defined; if moreover the Stein manifold has $c_1 = 0$ (for instance is hyperkähler), then the groups in the exact triangle can be naturally $\mathbb{Z}$-graded.

Corollary ([**24**], Theorem 3): For a Lefschetz pencil of K3 surfaces in a Fano 3-fold, the vanishing cycles $\{L_j\}$ "fill" the generic affine fibre: every closed Lagrangian submanifold disjoint from the base locus and with well-defined Floer homology must hit one of the $\{L_j\}$.

Proof: The global monodromy acts as a *shift* on (graded) HF^*, so if K is disjoint from all the spheres then the exact sequence shows $HF(K,K) = HF(K,K)$ $[shifted]$. Iterating, and recalling that $HF(K,K)$ is supported in finitely many degrees, this forces $HF(K,K) = 0$. But this is impossible for any homologically injective Lagrangian submanifold, by general properties of Floer theory, cf. [**10**], which completes the contradiction.

There are simpler proofs that any Lagrangian must intersect one of the vanishing cycles, but this gives a bit more: the vanishing cycles "generate" Donaldson's quantum category of the K3 (the underlying homological category of the Fukaya category). The Corollary above was in part motivated by an older and easier result, specific to the situation for curves in Riemann surfaces, given in [**32**].

In the second part we will focus attention on a Stein manifold Y_m which also contains a distinguished finite collection of Lagrangian submanifolds (cf. Section 2.6 below), which conjecturally generate the quantum category of Y_m in a similar way. However, these arise not as vanishing cycles of a pencil but from the components of a "complex Lagrangian" small resolution, giving another point of contact between the two general themes of the last section.

2. Knots, the adjoint quotient

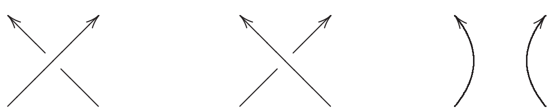
All the material of this section is joint work with Paul Seidel; we were considerably influenced by the ideas of Mikhail Khovanov. References are [**29**],[**30**],[**11**].

2.1. Knot polynomials: The Jones polynomial and Alexander polynomial $V_K(t)$, $\Delta_K(t)$ are powerful knot invariants defined by *skein relations*. They are Laurent polynomials in $t^{\pm 1/2}$ determined by saying $V_K(U) = 1 = \Delta_K(U)$, for U an unknot, and also that

$$t^{-1}V_{L_+} - tV_{L_-} + (t^{-1/2} - t^{1/2})V_{L_0} = 0;$$

$$\Delta_{L_+} - \Delta_{L_-} - (t^{-1/2} - t^{1/2})\Delta_{L_0} = 0.$$

Here the relevant links differ only near a single crossing, where they look as in the picture below:



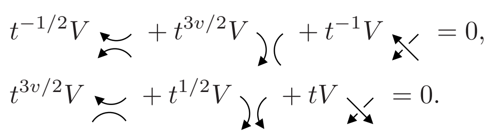
L_+ L_- L_0

The Alexander polynomial is well-understood geometrically, via homology of an infinite cyclic cover $H_1(\widetilde{\mathbb{S}^3\backslash K})$ [**14**]. There is also an interpretation in terms of 3-dimensional Seiberg-Witten invariants, beautifully explained in [**7**]. The Jones polynomial is more mysterious, although it does have certain representation theoretic incarnations in the theory of quantum groups and loop groups. The Jones polynomial solved a host of conjectures immediately after its introduction, one famous one being the following:

Example: (Kauffman) A connected reduced alternating diagram for a knot exhibits the minimal number of crossings of any diagram for the knot. [Reduced: no crossing can be removed by "flipping" half the diagram.]

Before moving on, it will be helpful to rephrase the skein property in the following slightly more involved fashion.

(1)



Here v denotes the signed number of crossings between the arc ending at the top left of the crossing and the other connected components of the diagram. Some of the arcs have no labelled arrow since resolving a crossing in one of the two possible ways involves a non-local change of orientation, but the relations are between the polynomials of *oriented* links. Obviously these two equations together imply the original skein relation. (The Jones polynomial of a *knot* is independent of the choice of orientation, but for links this is no longer true.)

2.2. Khovanov homology: Mikhail Khovanov (circa 1998) "categorified" the Jones polynomial – he defines combinatorially an invariant $K \mapsto Kh^{*,*}(K)$ which is a $\mathbb{Z}\times\mathbb{Z}$-graded abelian group, and such that

(i) $Kh^{0,*}(U_n) = H^*((\mathbb{S}^2)^n)[-n]$, where U_n is an n-component unlink (and the cohomology is concentrated in degrees $(0,*)$);

(ii) Skein-type exact sequence: for oriented links as indicated, there are long exact sequences which play the role of (1) above:

$$\begin{aligned}(2)\qquad &\cdots \longrightarrow \mathrm{Kh}^{i,j}(\quad) \longrightarrow \mathrm{Kh}^{i,j-1}(\quad) \longrightarrow \mathrm{Kh}^{i-v,j-3v-2}(\quad)\\ &\longrightarrow \mathrm{Kh}^{i+1,j}(\quad) \longrightarrow \cdots\end{aligned}$$

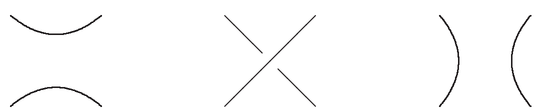
and

$$\begin{aligned}(3)\qquad &\cdots \longrightarrow \mathrm{Kh}^{i,j}(\quad) \longrightarrow \mathrm{Kh}^{i-v+1,j-3v+2}(\quad) \longrightarrow \mathrm{Kh}^{i+1,j+1}(\quad)\\ &\longrightarrow \mathrm{Kh}^{i+1,j}(\quad) \longrightarrow \cdots\end{aligned}$$

(iii) As an easy consequence of (ii), a change of variables recovers Jones:

$$\frac{1}{q+q^{-1}}\sum_{i,j}(-1)^i \mathrm{rk}_{\mathbb{Q}}\mathrm{Kh}^{i,j}(K)q^j \;=\; V_K(t)|_{q=-\sqrt{t}}.$$

Note the exact sequences are not quite skein relations, since they do not involve the crossing change, but rather the two different crossing resolutions (sometimes called the horizontal and vertical resolutions, as in the next picture).



L_{hor} $\qquad$ L_{cross} $\qquad$ L_{vert}

Khovanov homology is known to be a strictly stronger invariant than the Jones polynomial, but its principal interest lies in its extension to a "Topological Quantum Field Theory"; cobordisms of knots and links induce canonical homomorphisms of Khovanov homology. Relying heavily on this structure, at least one beautiful topological application has now emerged:

Example: Rasmussen [**20**] uses $Kh^{*,*}$ to compute the unknotting number, which is also the slice genus, of torus knots, $\mathrm{Unknot}(T_{p,q}) = (p-1)(q-1)/2$.

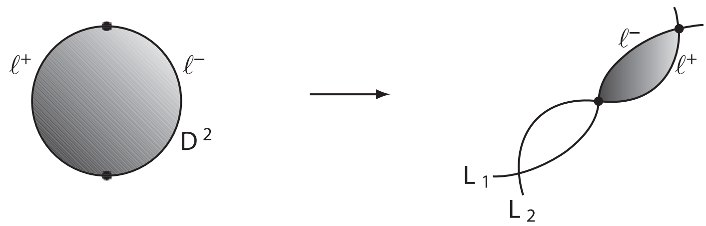
This result, first proved by Kronheimer and Mrowka, was formerly accessible only via adjunction-type formulae in gauge theory (or the rebirth of gauge theory via Ozsváth and Szabó); by all comparisons, Rasmussen's combinatorial proof represents an enormous simplification. One current limitation on Khovanov homology is precisely that its mystery makes it unclear which, comparable or other, problems it could profitably be applied to.

2.3. Invariants of braids: Here is a general way to (try to) define knot invariants using symplectic geometry. We begin with:

(1) a symplectic fibre bundle $\mathcal{Y} \to \mathrm{Conf}_{2n}(\mathbb{C})$ over the configuration space of unordered $2n$-tuples of points in $\mathbb{C}$. Suppose parallel transport is well-defined, or at least its rescaled cousin from Section 1.2.

(2) a distinguished (up to isotopy) Lagrangian submanifold $L \subset \mathcal{Y}_t$ in some distinguished fibre over $t \in \mathrm{Conf}_{2n}(\mathbb{C})$.

Given a braid β on $2n$ strands, i.e. a loop in the base, we can use parallel transport to get Lagrangian submanifolds $L, \beta(L) \subset \mathcal{Y}_t$ and then consider the Lagrangian Floer homology group $\beta \mapsto HF(L, \beta(L))$. This is the homology of a chain complex generated by intersection points, with boundary maps defined by counting pseudoholomorphic discs with boundary on the Lagrangian submanifolds as in the picture below.

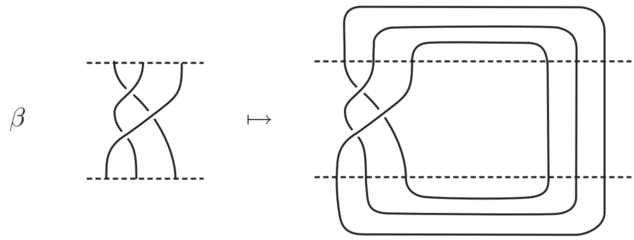


Caution: we're ignoring all technical difficulties. As before, well-definition of Floer homology relies on overcoming compactness problems, but for exact Lagrangians in finite type Stein manifolds this is standard. If the Lagrangians are spin, there are coherent orientations and Floer homology can be defined with $\mathbb{Z}$-coefficients. If the Lagrangians have $b_1 = 0$ (so zero Maslov class) and the ambient space has trivial first Chern class, the Lagrangians can be graded and Floer homology will be $\mathbb{Z}$-graded.

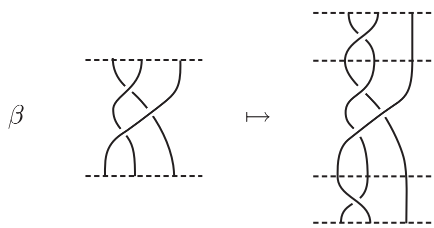
In the discussion so far, we could obtain invariants of braids on any number of strands. The restriction to the even-strand case comes in making the connection to the theory of knots and links, which we do below.

2.4. Markov moves: It is well-known that every oriented link can be obtained as the "closure" of a braid in the fashion given in the following diagram: one goes from $Br_n \ni \beta \mapsto \beta \times \mathrm{id} \in Br_{2n}$ and then caps off top and bottom with a collection of nested horseshoes. Such a representation of oriented links is enormously non-unique, but the equivalence relation on braids that generates this non-uniqueness is well-understood, and generated by the so-called Markov moves. The first is conjugation $\beta \mapsto \sigma\beta\sigma^{-1}$ by any $\sigma \in Br_n$, and the second – which is more interesting, since it changes the number of strands of the braid – involves linking in an additional strand by a single positive or negative half-twist, giving the II^+ and II^- stabilisations. All are pictured below. (To see that the link closure is canonically oriented, put an "upwards" arrow on all the parallel right hand strands.)

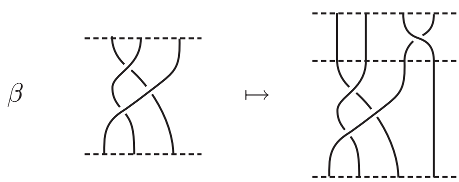
(1) Link closure:



(2) Markov I:



(3) Markov II^+:



It follows that if the association $\beta \mapsto HF(L, (\beta \times \mathrm{id})(L))$ of Section 2.3 is invariant under the Markov moves, then it in fact defines an invariant of oriented links in the three-sphere. We will now turn to a particular case in which precisely this occurs.

2.5. The adjoint quotient: We will get our family of symplectic manifolds from (a cousin of) the characteristic polynomial mapping, also called the adjoint quotient $\chi : \mathfrak{sl}_m \to \mathbb{C}^{m-1}$ which is smooth over $\mathrm{Conf}^0_m(\mathbb{C})$, the space of balanced configurations, i.e. symmetric functions of distinct eigenvalues of trace-free matrices. The following is the content of the Jacobson-Morozov theorem:

Fact: given a nilpotent matrix $N^+ \in \mathfrak{sl}_m(\mathbb{C})$, there is a unique conjugacy class of homomorphisms $\rho : \mathfrak{sl}_2 \to \mathfrak{sl}_m$ such that $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \overset{\rho}{\mapsto} N^+$.

Let N^- be the image of the other standard nilpotent $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ in $\mathfrak{sl}_2$. Then $N^+ + \ker(ad\, N^-)$ is a "transverse slice" to the adjoint action, i.e. it's an affine subspace of $\mathfrak{sl}_m$ which intersects the adjoint orbit of N^+ inside the nilpotent cone only at N^+. (Given the JM theorem, this is an easy fact about $\mathfrak{sl}_2$-representations.)

Lemma: the restriction of χ to such a transverse slice $\mathcal{S}$ is still a fibre bundle over $\mathrm{Conf}^0_m(\mathbb{C})$.

For a suitable Kähler form, the rescaled parallel transport construction for this fibre bundle can be pushed through, and closed Lagrangian submanifolds transported into any desired fibres. This follows the general programme outlined at the end of Section 1.2. Although we will not dwell on the details here, we should say at once that the relevant symplectic forms are *exact*, and are not related to the Kostant-Kirillov forms that also arise when dealing with the symplectic geometry of adjoint orbits.

2.6. Simultaneous resolutions: Grothendieck gave a simultaneous resolution of $\chi|\mathcal{S}$: replace a matrix A by the space of pairs $(A, \mathcal{F})$ where $\mathcal{F}$ is a flag stabilised by A. This orders the eigenvalues, i.e. the resolution involves base-changing by pulling back under the symmetric group. Hence, via Section 1.9, the differentiable monodromy of the fibre bundle $\chi : \mathcal{S} \to \mathbb{C}^{m-1}$ factors through Sym_m; we get a diagram as follows, writing Y_t for a fibre of the slice over some point t:

$$\begin{array}{ccc} Br_m = \pi_1(\mathrm{Conf}^0_m(\mathbb{C})) & \longrightarrow & \pi_0(\mathrm{Symp}(Y_t)) \\ \downarrow & & \downarrow \\ \mathrm{Sym}_m & \longrightarrow & \pi_0(\mathrm{Diff}(Y_t)) \end{array}$$

All representations of symmetric groups (and more generally Weyl groups) arise this way, in what is generally known as the *Springer correspondence*. As in Section 1.9, the symplectic monodromy is far richer (perhaps even faithful?).

Example: $\mathfrak{sl}_2(\mathbb{C})$ and $N^+ = 0$ so $\mathcal{S} = \mathfrak{sl}_2$; then χ is the map $(a, b, c, -a) \mapsto -a^2 - bc$ which after a change of co-ordinates is the usual node, with generic fibre $T^*\mathbb{S}^2$.

Example: $\mathfrak{sl}_{2m}(\mathbb{C})$ and N^+ with two Jordan blocks of equal size. Then the slice $\mathcal{S}_m \cong \mathbb{C}^{4m-1}$ is all matrices of 2×2-blocks with I_2 above the diagonal, any $(A_1, \ldots, A_m)$ in the first column with $tr(A_1) = 0$, and zeroes elsewhere; all $A_i = 0$ gives N^+ back. Explicitly, then, a general member of the slice has the shape

$$A = \begin{pmatrix} A_1 & I & & \\ A_2 & & I & \\ \vdots & & & \\ A_{m-1} & & & I \\ A_m & & \cdots & 0 \end{pmatrix}$$

where the A_k are 2×2 matrices, and with $tr(A_1) = 0$. The characteristic polynomial is $det(\lambda I - A) = det(\lambda^m - A_1\lambda^{m-1} - \cdots - A_m)$. The smooth fibres $Y_{m,t} = \chi^{-1}(t) \cap \mathcal{S}_m$ are smooth complex affine varieties of dimension $2m$.

Caution: for Lie theory purists, this is not in fact a JM slice (there is no suitable N^-), but is orbit-preservingly isomorphic to JM slices, and more technically convenient for our purposes.

Note that the generic fibre of the map (i.e. over a point of configuration space) is unchanged by the simultaneous resolution, so in principle all the topology of these spaces can be understood explicitly in terms of the linear algebra of certain matrices. On the other hand, the resolution of the zero-fibre (the nilpotent cone) is well-known to retract to a compact core which is just the preimage of the matrix N^+ itself; in other words, it's the locus of all flags stabilised by N^+. This core is a union of *complex Lagrangian* submanifolds, described in more detail in [**12**], in particular the number of irreducible components is given by the Catalan number $\frac{1}{m+1}\binom{2m}{m}$.

In nearby smooth fibres, these complex Lagrangian components $L_{\wp}$ give rise to distinguished real Lagrangian submanifolds, and it is plausible to conjecture that this finite set of Lagrangian submanifolds generate Donaldson's quantum category of $Y_{m,t}$ (the underlying homological category of the Fukaya category) in the weak sense that every Floer homologically essential closed Lagrangian submanifold has non-trivial Floer homology with one of the $L_{\wp}$, cf. Section 1(J).

2.7. Inductive geometry: The key construction with this slice is an "inductive scheme", relating the "least singular non-smooth" fibres of $\mathcal{S}_m$ to the smooth fibres of $\mathcal{S}_{m-1}$. Fix $\mu = (\mu_1 = 0, \mu_2 = 0, \mu_3, \ldots, \mu_m)$ a tuple of eigenvalues, the first two of which vanish and with all others being pairwise distinct, and let $\hat{\mu} = (\mu_3, \ldots, \mu_m)$.

Lemma: The fibre of $(\chi|\mathcal{S}_m)^{-1}(\mu)$ has complex codimension 2 smooth singular locus which is canonically isomorphic to $(\chi|\mathcal{S}_{m-1})^{-1}(\hat{\mu})$. Moreover, along the singular locus χ has a fibred A_1-singularity (an open neighbourhood of the singular locus looks like its product with $x^2 + yz = 0$).

Rescaled parallel transport and the vanishing cycle construction give a relative vanishing cycle in smooth fibres of $\chi|\mathcal{S}_m$ which is an $\mathbb{S}^2$-bundle over a fibre of $\chi|\mathcal{S}_{m-1}$. General properties of symplectic parallel transport give that these relative vanishing cycles are not Lagrangian but coisotropic, with the obvious $\mathbb{S}^2$-fibrations being the canonical foliations by isotropic leaves.

The force of the Lemma is that this process can now be *iterated*. Of course, an isotropic fibration restricted to a Lagrangian submanifold gives rise to a Lagrangian submanifold of the total space.

2.8. Symplectic Khovanov homology: Fix a crossingless matching $\wp$ of $2m$ points in the plane; the points specify a fibre Y_m of $\chi|\mathcal{S}_m$. Bringing eigenvalues together in pairs along the paths specified by the matching, and iterating the vanishing cycle construction above, gives a Lagrangian $L_{\wp}$ which is an iterated $\mathbb{S}^2$-bundle inside Y_m. In fact one can show that it is diffeomorphic to $(\mathbb{S}^2)^m$ (hence spin). We care especially about the first case $\wp_+$ below; we remark that the number of crossingless matchings which lie entirely in the upper half-plane, up to isotopy, is given by the Catalan number $\frac{1}{m+1}\binom{2m}{m}$, cf. section 2.6 above.

Crossingless matchings: the nested horseshoe on the left is denoted $\wp_+$.

Given a braid $\beta \in Br_m$ we get a Floer group via thinking of $\beta \times \mathrm{id} \in Br_{2m} = \pi_1(\mathrm{Conf}_{2m}(\mathbb{C}))$ as explained above. The following is the main result of these notes:

THEOREM 1 (Seidel, S.). *The $\mathbb{Z}$-graded Floer cohomology group*

$$Kh^*_{symp}(K_\beta) = HF^{*+m+w}(L_{\wp_+}, (\beta \times id)(L_{\wp_+}))$$

is an oriented link invariant: here m is the number of strands and w the writhe of the braid diagram.

It is important to realise that the loss of information in passing from the bigrading to the single grading is substantial: for instance, Kh_{symp} does not in itself determine the Jones polynomial.

The proof of the Theorem involves verifying invariance of the Floer group under the Markov moves. For the first move, this is relatively straightforward, since – once the machinery of rescaled parallel transport has been carefully set in place – the Lagrangian submanifold $L_{\wp}$ is itself unchanged (up to Hamiltonian isotopy) by effecting a conjugation. For the second Markov move, the proof is more involved since one must compare Floer groups for Lagrangians of different dimension living in different spaces. The key is the fibred A_1-structure along the singular set where two eigenvalues coincide, and a fibred A_2-generalisation to the case where three eigenvalues coincide. Indeed, locally near branches of the discriminant locus of χ where eigenvalues $1, 2$ resp. $2, 3$ coincide, the smooth fibres contain pairs of vanishing cycles which together form an $\mathbb{S}^2 \vee \mathbb{S}^2$-fibration. The fact that the monodromy symplectomorphisms about the two branches of the discriminant satisfy the braid relations can be deduced explicitly from Section 1.8; of course, the fact that we have a fibre bundle over configuration space gives the same result without any appeal to the local structure.

However, a similar local analysis allows one to explicitly identify the Floer complexes for the Lagrangians before and after the Markov II move. (The grading shift in the definition takes care of the difference of the effects of the Markov II^+ and Markov II^- moves on the Maslov class.) The upshot is that very general features of the singularities of the mapping χ encode the local geometric properties which lead to the symplectic Khovanov homology being an invariant.

2.9. Long exact sequences revisited: In a few cases – unlinks, the trefoil – one can compute Kh_{symp} explicitly, and in such cases one finds that the answer agrees with Khovanov's combinatorial theory. Even in these cases, the result is rather surprising, since the methods of computation do not particularly parallel one another. Thus the Main Theorem is complemented by:

Conjecture 2 (Seidel, S.). $Kh^*_{symp} = \oplus_{i-j=*} Kh^{i,j}$.

Main evidence: Kh^*_{symp} should also satisfy the right skein-type exact sequences (in the notation of Section 2.2)

$$Kh^*_{symp}(L_{hor}) \to Kh^*_{symp}(L_{vert}) \to Kh^*_{symp}(L_{cross}).$$

These should come from a version of the LES in Floer theory for a *fibred* Dehn twist, which is just the monodromy of χ corresponding to inserting a single negative crossing. Indeed, one can speculate that appropriate long exact sequences exist for suitable correspondences, as follows.

Suppose in general we are given Lagrangians $L_0, L_1 \subset X$ and $\hat{L}_0, \hat{L}_1 \subset \hat{X}$, and a Lagrangian correspondence $C \subset (X \times \hat{X}, \omega_X \oplus -\omega_{\hat{X}})$ which is an isotropic $\mathbb{S}^a$-fibration over X. Suppose moreover the $\hat{L}_i$ are given by lifting the L_i from X to $\hat{X}$ via the correspondence. One can try to find an exact triangle of the shape

$$HF(C \times C, L_0 \times \Delta \times L_1) \to HF(\hat{L}_0, \hat{L}_1) \to HF(\hat{L}_0, \tau(\hat{L}_1))$$

where τ denotes a fibred Dehn twist along C and the first homology group is taken inside $X \times \hat{X} \times \hat{X} \times X$ with the symplectic form reversed on the second two factors, and with Δ the diagonal. Moreover, if the geometry is sufficiently constrained, one can hope to relate the first group to Floer homology $HF(L_0, L_1)$ taken inside X. Using the relative vanishing cycles inside $\mathcal{S}_{m-1} \times \mathcal{S}_m$ as correspondences, this general picture includes the desired skein-type relation.

If one assumes the existence of the long exact sequence, then following a rather general algebraic strategy one can construct a spectral sequence with $E^2 = Kh^{*,*}$ and converging to $E^\infty = Kh^*_{symp}$ (the model outline is contained in Ozsváth and Szabó's work [**17**], in which they use a similar approach to relate Khovanov homology of a link L with the Heegaard Floer homology of the branched double cover $M(L)$). From this perspective, the above conjecture asserts the vanishing of the higher differentials in this spectral sequence; in the analogous story with Heegaard Floer theory, by contrast, the higher order differentials are often non-zero.

A distinct circle of ideas relating the chain complex underlying symplectic Khovanov homology to the Bigelow-Lawrence homological construction of the Jones polynomial [**6**] has recently been given by Manolescu in [**15**].

2.10. Counting sections revisited: Khovanov's theory is especially interesting since it fits into a TQFT (and we know a lot about knots, but little about surfaces with or without boundary in $\mathbb{R}^4$). A small piece of that is easily visible in Kh^*_{symp}, in the spirit of Section 1.7; (2).

Suppose we have a *symplectic* cobordism (surface in $\mathbb{R}^4$) between two positive braids. By fibring $\mathbb{R}^4 \subset \mathbb{CP}^2$ by $\mathbb{C}$-lines, we get a braid monodromy picture of the surface, which is just a relative version of the Lefschetz fibration story from Section 1.6. Geometrically, the braid monodromy gives an annulus in configuration space whose boundary circles represent the two boundary knots/braids. Now counting holomorphic sections of χ over the annulus, with suitable Lagrangian boundary conditions, gives rise to a morphism on Floer homology groups and hence on symplectic Khovanov homology.

Challenge: detect symplectically knotted surfaces (or families of such with common boundary) this way.

References

[1] Auroux, D. *Symplectic maps to projective spaces and symplectic invariants*, Turkish J. Math. **25**, p.1-42, (2001).

[2] Auroux, D., Donaldson, S. and Katzarkov, L. *Singular Lefschetz pencils*, Preprint, available at math.DG/0410332, (2004).

[3] Auroux, D., Munoz, V. and Presas, F. *Lagrangian submanifolds and Lefschetz pencils*, Preprint, available at math.SG/0407126, (2004).

[4] Auroux, D., Katzarkov, L. and Orlov, D. *Mirror symmetry for weighted projective planes and their non-commutative deformations*, Preprint, available at math.AG/0404281, (2004).

[5] Auroux, D. and Smith, I. *Lefschetz pencils, branched covers and symplectic invariants*, CIME Lectures, available at math.SG/0401021, (2004).

[6] Bigelow, S. *A homological definition of the Jones polynomial*, Geometry & Topology monographs **4**, p.29-41, (2002).

[7] Donaldson, S.K. *Topological field theories and formulae of Casson and Meng-Taubes*, Geom. Topol. Monographs **2**, p.87-102, (1999).

[8] Donaldson, S.K. *Polynomials, vanishing cycles and Floer homology*, In "Mathematics: Frontiers and Perspectives", Amer. Math. Soc., p.55-64, (2000).

[9] Donaldson, S.K. and Smith, I. *Lefschetz pencils and the canonical class for symplectic 4-manifolds*, Topology **42**, p.743-785, (2003).

[10] Fukaya, K. and others, *Lagrangian intersection Floer theory – anomaly and obstruction*, Preprint, (2000).

[11] Khovanov, M. *A categorification of the Jones polynomial*, Duke Math. J. **101**, p.359-426, (2000).

[12] Khovanov, M. *Crossingless matchings and the cohomology of* (n, n) *Springer varieties*, Commun. Contemp. Math. **6**, p.561-577, (2004).

[13] Khovanov, M. and Seidel, P. *Quivers, Floer cohomology and braid group actions*, J. Amer. Math. Soc. **15**, p.203-271, (2002).

[14] Lickorish, R. *An introduction to knot theory*, Springer, (1997).

[15] Manolescu, C. *Nilpotent slices, Hilbert schemes and the Jones polynomial*, Preprint, available at math.SG/0411015, (2004).

[16] McDuff, D. and Salamon, D. *Introduction to symplectic topology*, (2nd ed.), Oxford Univ. Press, (2001).

[17] Ozsváth, P. and Szabó, Z. *On the Heegaard Floer homology of branched double covers*, Preprint, available at math.GT/0309170, (2003).

[18] Presas, F. *Lefschetz type pencils on contact manifolds*, Asian J. Math. **6**, p.277-301, (2002).

[19] Ran, Z. *A note on Hilbert schemes of nodal curves*, Preprint, available at math.AG/0410037, (2004).

[20] Rasmussen, J. *Khovanov homology and the slice genus*, Preprint, available at math.GT/0402131, (2004).

[21] Segal, G. *Definition of conformal field theory*, In "Geometry, topology and physics", (U.Tillmann, ed.), CUP, (2004).

[22] Seidel, P. *Lagrangian 2-spheres can be symplectically knotted*, Jour. Diff. Geom. **52**, p.145-171, (1999).

[23] Seidel, P. *A long exact sequence for symplectic Floer cohomology*, Topology **42**, p.1003-1063, (2003).

[24] Seidel, P. *Fukaya categories and their deformations*, Proceedings of the ICM, Beijing 2002, p.351-360, Higher Ed. Press Beijing, (2002).

[25] Seidel, P. *Vanishing cycles and mutation*, European congress of mathematics, Barcelona 2000, p.65-85, Birkhäuser, (2001).

[26] Seidel, P. *Lectures on four-dimensional Dehn twists*, Preprint, available at math.SG/0309012, (2003).

[27] Seidel, P. *Homological mirror symmetry for the quartic surface*, Preprint, available at math.SG/0310414, (2003).

[28] Seidel, P. *Unpublished notes*, (1997).

[29] Seidel, P. and Smith, I. *A link invariant from the symplectic geometry of nilpotent slices*, Preprint, available at math.SG/0405089, (2004).

[30] Seidel, P. and Smith, I. *In preparation*

[31] Seidel, P. and Thomas, R. *Braid group actions on derived categories of coherent sheaves*, Duke Math. J. **108**, p.37-108, (2001).

[32] Smith, I. *Geometric monodromy and the hyperbolic disc*, Quarterly J. Math. **52**, p.217-228, (2001).

[33] Smith, I. *Serre-Taubes duality for pseudoholomorphic curves*, Topology **42**, p.931-979, (2003).

[34] Smith, I. and Thomas, R. *Symplectic surgeries from singularities*, Turkish J. Math. **27**, p.231-250, (2003).

[35] Taubes, C.H. *Seiberg-Witten and Gromov invariants for symplectic 4-manifolds*, International Press, (2000).

[36] Usher, M. *The Gromov invariant and the Donaldson-Smith standard surface count*, Geometry & Topology **8**, p.565-610, (2004).

Centre for Mathematical Sciences, Cambridge CB3 0WB, U.K.
E-mail address: is200@cam.ac.uk

List of Participants

Akbulut, Selman
Michigan State University
akbulut@math.msu.edu

Akhmedov, Anar
University of California, Irvine
aakhmedo@math.uci.edu

Akhmetiev, Petr
Sojuzpatent
akhmetiev@sojuzpatent.com

Andersson, Ole
Uppsala University, Sweden
ole@ma.uu.se

Arikan, Mehmet
Michigan State University
arikanme@msu.edu

Armstrong, John
Yale University
john.armstrong@math.yale.edu

Auroux, Denis
MIT
auroux@math.mit.edu

Baldwin, John
Columbia University
baldwin@math.columbia.edu

Baricz, Arpad
Babes-Bolyai Egyetem, Romania
baricocsi@kolozsvar.ro

Baykur, Refik
Michigan State University
baykurre@msu.edu

Cha, Jae Choon
Information and Communications
University, Korea
jccha@icu.ac.kr

Chantraine, Baptiste
Universite du Quebec, Montreal
chantrai@math.uqam.ca

Cho, Hyunjoo
Michigan State University
chohyunj@msu.edu

Cotton, Andrew
University of California, Berkeley
acotton@math.berkeley.edu

Crombecque, David
University of Southern California
crombecq@usc.edu

Deruelle, Arnaud
Universite Aix Marseille, France
deruelle@cmi.univ-mrs.fr

Doran, Joan
Yale University
joan.doran@yale.edu

Easson, Vivien
Oxford University, UK
easson@maths.ox.ac.uk

Ekholm, Tobias
University of Southern California
tekholm@usc.edu

Ellwood, David
Clay Mathematics Institute
ellwood@claymath.org

Elson, Ilya
MIT
ielson@math.mit.edu

Etnyre, John
University of Pennsylvania
etnyre@math.upenn.edu

Fawcett, Jennifer Sunny
Cornell University
fawcett@math.cornell.edu

Fintushel, Ronald
Michigan State University
ronfint@math.msu.edu

Gadre, Vaibhav
Brandeis University
vaibhav@brandeis.edu

Ghiggini, Paolo
Universita di Pisa, Italy
ghiggini@mail.dm.unipi.it

Gillam, William
Columbia University
wgillam@math.columbia.edu

Goda, Hiroshi
Tokyo University of Agriculture and Technology, Japan
goda@cc.tuat.ac.jp

Goetz, Benjamin
University of Texas, Austin
benni@math.utexas.edu

Gordon, Cameron
University of Texas, Austin
gordon@math.utexas.edu

Gornik, Bojan
Princeton University
bgornik@math.princeton.edu

Gospodinov, Georgi
Brandeis University
ggospodi@brandeis.edu

Grigsby, Julia
University of California, Berkeley
egrigsby@math.berkeley.edu

Haire, Timoy
University of Cambridge
tjh29@dpmms.cam.ac.uk

Harvey, Shelly
MIT
sharvey@math.mit.edu

Haydys, Andriy
Universitaet Goettingen, Germany
haydys@uni-math.gwdg.de

Hedden, Matthew
Columbia University
meh2003@columbia.edu

Himpel, Benjamin
Indiana University
bhimpel@indiana.edu

Hog-Angeloni, Cynthia
Universitaet Frankfurt, Germany
cyn@math.uni-frankfurt.de

Honda, Ko
University of Southern California
khonda@usc.edu

Jabuka, Stanislav
Columbia University
jabuka@math.columbia.edu

Jiangang, Yao
University of California, Berkeley
jgyao@math.berkeley.edu

Karakurt, Cagri
Middle East Technical University, Turkey
karakurt@metu.edu.tr

Kaufman, Sam
MIT
sam@math.mit.edu

Kerr, Gabriel
University of Chicago
gdkerr@math.uchicago.edu

Kim, Se-Goo
University of California, Santa Barbara
sekim@math.ucsb.edu

Kirby, Rob
University of California, Berkeley
kirby@math.berkeley.edu

Knapp, Adam
Michigan State University
knappa@math.msu.edu

Kricker, Andrew
University of Toronto
akricker@math.toronto.edu

Kutluhan, Cagatay
Middle East Technical University, Turkey
borak@isbank.net.tr

Kwon, Oh Hoon
Michigan State University
kwonoh1@msu.edu

Larchevque, Philippe
Universite Aix Marseille, France
plarchev@cmi.univ-mrs.fr

Lee, Peter
MIT
lee@math.mit.edu

Li, Tian-Jun
University of Minnesota
tjli@math.umn.edu

Lipshitz, Robert
Stanford University
lipshitz@stanford.edu

Lipyanskiy, Maksim
MIT
mlipyan@mit.edu

Lisca, Paolo
Universita di Pisa, Italy
lisca@dm.unipi.it

Lobb, Andrew
Harvard University
lobb@math.harvard.edu

Malmendier, Andreas
MIT
malmendi@mit.edu

Manturov, Vassily
Moscow State University, Russia
vassily@manturov.mccme.ru

Mark, Thomas
Southeastern Louisiana University
tmark@selu.edu

Matsuda, Hiroshi
Hiroshima University, Japan
matsuda@math.sci.hiroshima-u.ac.jp

Melvin, Paul
Bryn Mawr College
pmelvin@brynmawr.edu

Metzler, Wolfgang
Universitaet Frankfurt, Germany
cyn@math.uni-frankfurt.de

Moffatt, Iain
University of Warwick, UK
iainm@maths.warwick.ac.uk

Naot, Gad
University of Toronto
gnaot@math.toronto.edu

Nash, Oliver
Oxford University
nash@maths.ox.ac.uk

Ni, Yi
Princeton University
yni@math.princeton.edu

Niederkrger, Klaus
Universitaet Koeln, Germany
kniederk@math.uni-koeln.de

Niepel, Martin
Princeton University
mniepel@math.princeton.edu

Owens, Brendan
McMaster University
owensb@math.mcmaster.ca

Ozbagci, Burak
Koc University, Turkey
bozbagci@ku.edu.tr

Ozer, Arda Bugra
Michigan State University
ozerarda@msu.edu

Ozsváth, Peter
Columbia University
petero@math.columbia.edu

Ozturk, Ferit
Bogazici University, Turkey
ferit.ozturk@boun.edu.tr

Pamuk, Mehmetcik
McMaster University
pamukm@math.mcmaster.ca

Pamuk, Semra
McMaster University
pamuks@math.mcmaster.ca

Park, Jongil
Michigan State University
jipark@math.msu.edu

Plamenevskaya, Olga
Harvard University
olga@math.harvard.edu

Rajagopalan, Sridhar
Brandeis University
shridhar@brandeis.edu

Rustamov, Raif
Princeton University
rustamov@princeton.edu

Sadykov, Rustam
Universiry of Florida
sadykov@math.ufl.edu

Salepci, Nermin
Middle East Technical University, Turkey
nermins@metu.edu.tr

Schleimer, Saul
University of Illinois, Chicago
saul@math.uic.edu

Schoenenberger, Stephan
University of Pennsylvania
stephans@math.upenn.edu

Shirokova, Nadya
IHES, France
nadya@ihes.fr

Silveira, Candida
SUNY, Stony Brook
candida@math.sunysb.edu

Smith, Ivan
University of Cambridge, UK
is200@cam.ac.uk

Stern, Ron
University of California, Irvine
rstern@uci.edu

Stipsicz, András
Rényi Institute, Budapest
stipsicz@renyi.hu

Strle, Saso
McMaster University
strles@math.mcmaster.ca

Sung, Chanyoung
Korea Institute for Advanced Study
cysung@kias.re.kr

Suvaina, Ioana
SUNY, Stony Brook
sio@math.sunysb.edu

Symington, Margaret
Georgia Institute of Technology
msyming@math.gatech.edu

Szabó, Zoltán
Princeton University
szabo@math.princeton.edu

Tievsky, Aaron
MIT
tievsky@math.mit.edu

Titirez, Adara - Monica
West University of Timisoara, Romania
adaratitirez@yahoo.com

Tomatis, Andrea
Imperial College, UK
andrea.tomatis@imperial.ac.uk

Tripp, James
University of Pennsylvania
jtripp@math.upenn.edu

Tuncer, Nigar
Binghamton University
nigaroz@yahoo.com

Usher, Michael
MIT
usher@math.mit.edu

Van Koert, Otto
universitaet Koeln, Germany
okoert@math.uni-koeln.de

Wagner, Emmanuel
IRMA Strasbourg, France
emmanuel.wagner@ulp.u-strasbG.fr

Wand, Andrew
University of California, Berkeley
wand@math.berkeley.edu

Wang, Jiajun
University of California, Berkeley
wang@math.berkeley.edu

Wehrli, Stephan
University of Basel, Switzerland
wehrli@math-lab.unibas.ch

Welji, Shaffiq
Columbia University
welji@math.columbia.edu

Yamamoto, Minoru
Hokkaido University, Japan
minomoto@math.sci.hokkaido.ac.jp

Yanagisawa, Masuo
University of Chicago
masuoy@math.uchicago.edu

Zehmisch, Kai
Universitaet Leipzig, Germany
zehmisch@math.uni-leipzig.de

Zentner, Raphael
Harvard University
rzentner@math.harvard.edu

Zuddas, Daniele
Scuola Normale Superiore di Pisa, Italy
d.zuddas@sns.it

Mathematical gauge theory studies connections on principal bundles, or, more precisely, the solution spaces of certain partial differential equations for such connections. Historically, these equations have come from mathematical physics, and play an important role in the description of the electro-weak and strong nuclear forces. The use of gauge theory as a tool for studying topological properties of four-manifolds was pioneered by the fundamental work of Simon Donaldson in the early 1980s, and was revolutionized by the introduction of the Seiberg–Witten equations in the mid-1990s. Since the birth of the subject, it has retained its close connection with symplectic topology. The analogy between these two fields of study was further underscored by Andreas Floer's construction of an infinite-dimensional variant of Morse theory that applies in two a priori different contexts: either to define symplectic invariants for pairs of Lagrangian submanifolds of a symplectic manifold, or to define topological invariants for three-manifolds, which fit into a framework for calculating invariants for smooth four-manifolds. "Heegaard Floer homology", the recently-discovered invariant for three- and four-manifolds, comes from an application of Lagrangian Floer homology to spaces associated to Heegaard diagrams. Although this theory is conjecturally isomorphic to Seiberg–Witten theory, it is more topological and combinatorial in flavor and thus easier to work with in certain contexts. The interaction between gauge theory, low-dimensional topology, and symplectic geometry has led to a number of striking new developments in these fields. The aim of this volume is to introduce graduate students and researchers in other fields to some of these exciting developments, with a special emphasis on the very fruitful interplay between disciplines.

This volume is based on lecture courses and advanced seminars given at the 2004 Clay Mathematics Institute Summer School at the Alfréd Rényi Institute of Mathematics in Budapest, Hungary. Several of the authors have added a considerable amount of additional material to that presented at the school, and the resulting volume provides a state-of-the-art introduction to current research, covering material from Heegaard Floer homology, contact geometry, smooth four-manifold topology, and symplectic four-manifolds.

ISBN 0-8218-3845-8

9 780821 838457

CMIP/5

www.ams.org

www.claymath.org