FLOER THEORY AND LOW DIMENSIONAL TOPOLOGY

DUSA MCDUFF

ABSTRACT. The new 3- and 4-manifold invariants recently constructed by Ozsváth and Szabó are based on a Floer theory associated with Heegaard diagrams. The following notes try to give an accessible introduction to their work. In the first part we begin by outlining traditional Morse theory, using the Heegaard diagram of a 3-manifold as an example. We then describe Witten's approach to Morse theory and how this led to Floer theory. Finally, we discuss Lagrangian Floer homology. In the second part, we define the Heegaard Floer complexes, explaining how they arise as a special case of Lagrangian Floer theory. We then briefly describe some applications, in particular the new 4-manifold invariant, which is conjecturally just the Seiberg–Witten invariant.

1. THE FLOER COMPLEX

This section begins by outlining traditional Morse theory, using the Heegaard diagram of a 3-manifold as an example. It then describes Witten's approach to Morse theory, a finite dimensional precursor of Floer theory. Finally, it discusses Lagrangian Floer homology. This is fundamental to Ozsváth and Szabó's work; their Heegaard Floer theory is a special case of this general construction. Readers wanting more detail should consult Ozsváth and Szabó's excellent recent survey article [28]. Since this also contains a comprehensive bibliography, we give rather few references here.

1.1. **Classical Morse theory.** Morse theory attempts to understand the topology of a space X by using the information provided by real valued functions $f : X \to \mathbb{R}$. In the simplest case, X is a smooth m-dimensional manifold, compact and without boundary, and we assume that f is generic and smooth. This means that its critical points p are isolated and there is a local normal form: in suitable local coordinates $x_1, \ldots, x_m$ near the critical point $p = 0$ the function f may be written as

$$f(x) = -x_1^2 - \cdots - x_i^2 + x_{i+1}^2 + \cdots + x_m^2.$$

The number of negative squares occurring here is independent of the choice of local coordinates and is called the **Morse index** $\mathrm{ind}(p)$ of the critical point.

Functions $f : X \to \mathbb{R}$ that satisfy these conditions are called **Morse functions**. One analyses the structure of X by considering the family of sublevel sets

$$X^c := f^{-1}(-\infty, c].$$

These spaces are diffeomorphic as c varies in each interval of regular (i.e. noncritical) values, and their topology changes in a predictable way as c passes a critical level.

One way to prove this is to consider the **negative gradient flow** of f. Choose a generic metric μ on X. Then the gradient vector field ∇f is perpendicular to the level sets $f^{-1}(c)$ at regular values and vanishes only at the critical points. Therefore one can push a regular level $f^{-1}(c)$ down to $f^{-1}(c - \varepsilon)$ by following the flow of ∇f. Moreover one can understand what happens to the sublevel sets as one passes a critical level by looking at the set of downward gradient trajectories emanating from the critical point p. The points on this set of trajectories form the **unstable manifold**

$$W_f^u(p) := \{p\} \cup \big\{u(s) : s \in \mathbb{R}, \; \dot{u}(s) = -\nabla f(u(s)), \lim_{s \to -\infty} u(s) = p\big\}.$$

It is easy to see that $W_f^u(p)$ is diffeomorphic to $\mathbb{R}^d$ where $d = \mathrm{ind}(p)$. Similarly, each critical point has a stable manifold $W_f^s(p)$ consisting of trajectories that converge towards p as $s \to \infty$.
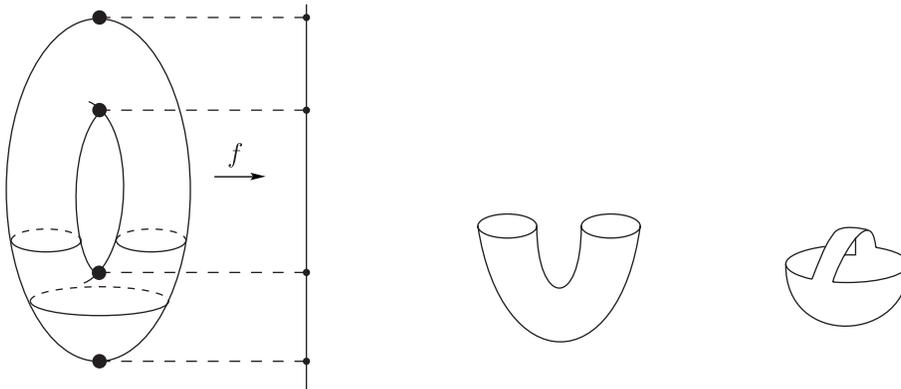

Figure 1. Morse decomposition of the torus.

For example, if c is close to $\min f$ (and we assume that f has a unique minimum), then the sublevel set X^c is diffeomorphic to the closed ball $D^m := \{x \in \mathbb{R}^m : \|x\| \leq 1\}$ of dimension m. When c passes a critical point p of index 1, a one handle (homeomorphic to $[0, 1] \times D^{m-1}$) is added. One should think of this handle as a neighborhood of the unstable manifold $W_f^u(p) \cong \mathbb{R}$. Similarly, when one passes a critical point of index 2, one adds a 2-handle; see Milnor [20]. When $m = 2$, a 2-handle is just a 2-disc, as one can see in the well known decomposition for the 2-torus $\mathbb{T} = S^1 \times S^1$ given by the height function; cf. Fig. 1.

The next example shows how one can use a Morse function to give a special kind of decomposition of an oriented 3-manifold Y that is known as a Heegaard splitting. This description of Y lies at the heart of Ozsváth and Szabó's theory.

Example 1.1. Heegaard diagram of a 3-manifold. Choose the Morse function $f : Y \to \mathbb{R}$ to be *self-indexing*, i.e. so that all the critical points of index i lie on
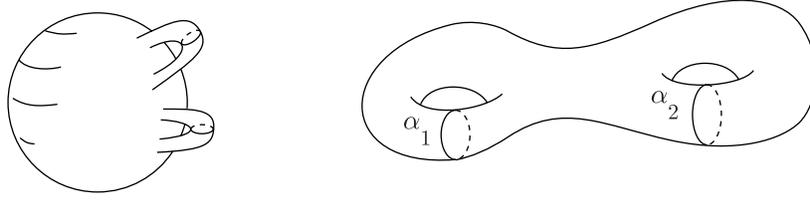
FIGURE 2. Genus two handlebody.

the level $f^{-1}(i)$. Then the cut $f^{-1}(3/2)$ at the half way point is a Riemann surface Σ_g of genus g equal to the number of index 1 critical points of f, and the sublevel set $Y^{3/2}$ is a *handlebody of genus g*, i.e. the union of a 3-ball D^3 with g 1-handles; see Fig. 2. By symmetry, the other half, $f^{-1}([3/2, 3])$, of Y is another handlebody of genus g. Thus Y is built from a single copy of the surface $\Sigma = f^{-1}(3/2)$ by attaching handlebodies $U_\alpha := Y^{3/2}, U_\beta$ to its two sides. When Y is oriented, the surface Σ inherits a natural orientation as the boundary of U_α.

The attaching map of U_α is determined by the loops in Σ that bound discs in U_α: if Σ has genus g, there is an essentially unique collection of g disjoint embedded circles $\alpha_1, \ldots, \alpha_g$ in Σ that bound discs $D_1, \ldots, D_g$ in U_α. These discs are chosen so that when they are cut out the remainder $U_\alpha \smallsetminus \{\alpha_1, \ldots, \alpha_g\}$ of U_α is still connected. Therefore Y can be described by two collections, $\alpha := \{\alpha_1, \ldots, \alpha_g\}$ and $\beta := \{\beta_1, \ldots, \beta_g\}$, of disjoint circles on the Riemann surface Σ_g. See Fig. 3. This description (known as a **Heegaard diagram**) is unique modulo some basic moves.[1] As an example, there is a well known decomposition of the 3-sphere $\{(z_1, z_2) : |z_1|^2 + |z_2|^2 = 1\}$ into two solid tori (handlebodies of genus 1), $U_1 := \{|z_1| \leq |z_2|\}$ and $U_2 := \{|z_1| \geq |z_2|\}$, and the corresponding circles in the 2-torus $\Sigma_1 = \{|z_1| = |z_2|\}$ are

$$\alpha_1 = \left\{ \frac{1}{\sqrt{2}}(e^{i\theta}, 1) : \theta \in [0, 2\pi] \right\}, \qquad \beta_1 = \left\{ (\frac{1}{\sqrt{2}}(1, e^{i\theta}) : \theta \in [0, 2\pi] \right\},$$

with a single intersection point $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. Section 2 in [24] contains a nice description of the properties of Heegaard diagrams.
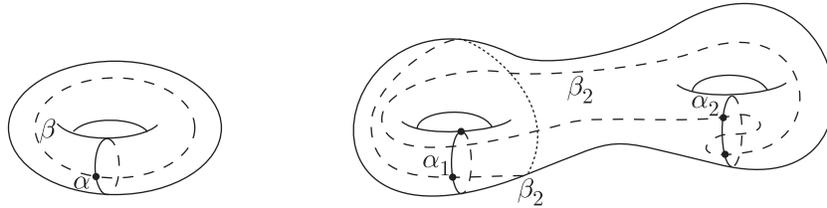
FIGURE 3. Genus 1 and 2 Heegaard diagrams for S^3.

The Oszváth–Szabó invariants capture information about the intersection points $\alpha_j \cap \beta_k$ of these two families. Note that each α_j is the intersection $W^s(p_j) \cap \Sigma$ of the upward gradient trajectories from some index 1 critical point p_j of f with the

---

[1]These are: isotoping the loops in α, β; changing these loops by "handleslides"; and finally stabilizing Σ_g by increasing its genus in a standard way.

level set Σ. Similarly, the β_k are the intersections with Σ of the downward gradient trajectories from the index 2 critical points q_k. Hence, each intersection point $\alpha_j \cap \beta_k$ corresponds to a gradient trajectory from q_k to p_j.

This traditional version of Morse theory is useful in some infinite dimensional cases as well, especially in the study of closed geodesics. Here one looks at the length (or energy) functional $\mathcal{F}$ on the space $\mathcal{X}$ of smooth loops in X. Its critical points are closed geodesics. They may not be isolated, but they have finite index. For further discussion see Bott's wonderful survey article [1].

1.2. **The Morse–Smale–Witten complex.** Much of the geometric information contained in a self-indexing Morse function f can be expressed in terms of the Morse–Smale complex $(C_*(X; f), \partial)$. The k-chains in this complex are finite sums of critical points of index k:

$$(1.1) \qquad C_k(X; f) = \Big\{ \sum_{x \in \mathrm{Crit}_k(f)} a_x \langle x \rangle \, : \, a_x \in \mathbb{Z} \Big\},$$

while the boundary operator is determined by the change in the sublevel sets $f^{-1}(-\infty, c]$ as one passes a critical level. Witten observed in [35] that sublevel sets have little physical meaning. More relevant are the gradient trajectories between critical points, which occur as "tunnelling effects" in which one state (regime at a critical point) affects another. His influential paper pointed out that one could use these trajectories to give the following alternative definition of the boundary operator $\partial : C_k(X; f) \to C_{k-1}(X; f)$ of the Morse–Smale complex:

$$(1.2) \qquad \partial \langle x \rangle = \sum_{y \in \mathrm{Crit}_{k-1}(f)} n(x, y) \, \langle y \rangle,$$

where $n(x, y)$ is the number of gradient trajectories of f from x to y. (Here one either counts mod 2 or counts using appropriate signs that come from suitably defined orientations of the trajectory spaces.) Note that the chain groups depend only on f, but the boundary operator depends on the choice of a generic auxiliary metric μ.

We claim that $C_*(X; f)$ is a **chain complex**, i.e. that $\partial^2 = 0$. To see this, note that

$$\begin{aligned} \partial^2 \langle y \rangle \;\; &= \sum_{y \in \mathrm{Crit}_{k-1}(f)} n(x, y) \, \partial \langle y \rangle \\ &= \sum_{y \in \mathrm{Crit}_{k-1}(f)} \sum_{z \in \mathrm{Crit}_{k-2}(f)} n(x, y) n(y, z) \langle z \rangle. \end{aligned}$$

The coefficient $\sum_y n(x, y) n(y, z)$ of $\langle z \rangle$ in this expression is the number of once-broken gradient trajectories from x to z and vanishes because these occur in cancelling pairs. In fact the space

$$\widehat{\mathcal{M}}(x, z) := \mathcal{M}(x, z)/\mathbb{R} = \big(W_f^u(x) \cap W_f^s(z) \big)/\mathbb{R}$$

of (unparametrized)[2] trajectories from x to z is a union of circles and open intervals whose ends may be identified with the set of once-broken gradient trajectories from x to z. For example, a parametrized trajectory from x to z whose image in $\widehat{\mathcal{M}}(x, z)$

---

[2]The elements $a \in \mathbb{R}$ act on the trajectories $u : \mathbb{R} \to X$ in $\mathcal{M}(x, z)$ by reparametrization: $a * u(s) := u(s + a)$.

is close to a broken trajectory goes fairly quickly from a neighborhood of x to a neighborhood of y but spends a lot of time negotiating the turn near y before moving off and approaching z.

Therefore the homology $H_*(X; f) := \ker \partial / \operatorname{im} \partial$ of this complex is defined. It turns out to be isomorphic to the usual homology $H_*(X)$ of X. In particular, it is independent of the choice of metric μ and function f. The complex as defined in this way is known as the Morse–Smale–Witten complex, or sometimes simply the Morse complex.

Remark 1.2. **Morse–Novikov theory.** There is a variant of this construction whose initial data is a closed 1-form ν on X instead of a Morse function; see Novikov [22]. If ν represents an integral cohomology class, it has the form $\nu = df$ for some circle valued function $f : X \to S^1$, and there is a cover $\mathbb{Z} \to \widetilde{X} \to X$ of X on which f lifts to a real valued function $\widetilde{f}$. Each critical point of f lifts to an infinite number of critical points of $\widetilde{f}$. The Morse–Novikov complex of f is essentially just the Morse complex of $\widetilde{f}$. It supports an action of the group ring $\mathbb{Z}[U, U^{-1}]$ of the group $\{U^n : n \in \mathbb{Z}\}$ of deck transformations of the cover $\widetilde{X} \to X$ and is finitely generated over this ring. One of the Heegaard Floer complexes is precisely of this kind. For another recent application, see Hutchings–Lee [15].

Remark 1.3. **Operations on the Morse complex.** Witten's point of view has proved very fruitful, not only for the applications we discuss later, but also for the understanding of the topology of manifolds and their loop spaces, a topic of central importance in so-called "string topology". Here the aim is to understand various homological operations (e.g. products) at the chain level, and it is very important to have a versatile chain complex to work with. The Morse–Smale–Witten complex fits into such theories very well. For example, given three generic Morse functions f_k, $k = 1, 2, 3$, one can model the homology intersection product $H_i \otimes H_i \to H_{i+j-m}$ on an m-dimensional manifold by defining a chain level homomorphism

$$\phi : C_i(X; f_1) \times C_j(X; f_2) \to C_{i+j-m}(X; f_3).$$

The map ϕ is given by counting Y-shaped graphs from a pair (x_1, x_2) of critical points in $\operatorname{Crit}(f_1) \times \operatorname{Crit}(f_2)$ to a third critical point $x_3 \in \operatorname{Crit}(f_3)$, where the two arms of the Y are gradient trajectories for f_1 and f_2 emanating from x_1, x_2 and the leg is a trajectory for f_3 converging to x_3. Thus

$$\phi(x_1, x_2) = \sum n(x_1, x_2, x_3) \langle x_3 \rangle,$$

where $n(x_1, x_2, x_3)$ is the number of such Y-graphs, counted with signs. If the functions f_k and metric μ are generic, then this number is finite and agrees with the number of triple intersection points of the three cycles $W_{f_1}^s(x_1), W_{f_2}^s(x_2)$ and $W_{f_3}^u(x_3)$, which have dimensions i_1, i_2 and $m - i_3$ respectively, where $i_3 = i_1 + i_2 - m$. In fact, there is a bijection between the set of Y-images and the set of such triple intersection points.

This is just the beginning. The tree graph Y has two inputs at the top and one output at the bottom and so defines a product. Its nonassociativity at the chain level gives rise to a new operation that counts maps of trees in X with three inputs and one output. Continuing this way, one may construct the full Morse–Smale–Witten A^∞-algebra as well as many other homology operations such as the Steenrod squares; see for example Cohen [2].

The fact that the chain complexes of Lagrangian Floer theory support similar maps is an essential ingredient of Ozsváth and Szabó's work.

1.3. **Floer theory.** Inspired partly by Witten's point of view but also by work of Conley [3] and Gromov [14], Floer [9, 10, 11] realised that there are interesting infinite dimensional situations in which a similar approach makes sense. In these cases, the ambient manifold $\mathcal{X}$ is infinite dimensional and the critical points of the function $\mathcal{F} : \mathcal{X} \to \mathbb{R}$ have infinite index and coindex. Therefore one usually cannot get much information from the sublevel sets $\mathcal{F}^{-1}(-\infty, c]$ of $\mathcal{F}$. Also, one may not be able to choose a metric on $\mathcal{X}$ such that the gradient flow of $\mathcal{F}$ is everywhere defined. However, Floer realised that in some important cases one can choose a metric so that the spaces $\mathcal{M}(x, y)$ of gradient trajectories between distinct critical points x, y of $\mathcal{F}$ have properties analogous to those in the finite dimensional case. Hence one can define the **Floer chain complex** using the recipe described in equations (1.1) and (1.2) above.

We now describe the version of Floer theory used by Ozsváth–Szabó. In their situation both the critical points of $\mathcal{F}$ and its gradient flow trajectories have natural geometric interpretations.

Example 1.4. Lagrangian Floer homology. Let M be a $2n$-dimensional manifold with symplectic form ω (i.e. a closed, nondegenerate 2-form) and choose two Lagrangian submanifolds L_0, L_1. These are smooth submanifolds of dimension n on which the symplectic form vanishes identically. (Physicists call them *branes*.) We assume that they intersect transversally and also that their intersection is nonempty, since otherwise the complex we aim to define is trivial.

Denote by $\mathcal{P} := \mathcal{P}(L_0, L_1)$ the space of paths x from L_0 to L_1:

$$x : [0, 1] \to M, \qquad x(0) \in L_0, \ x(1) \in L_1.$$

Pick a base point $x_0 \in L_0 \cap L_1$ considered as a constant path in $\mathcal{P}$ and consider the universal cover $\widetilde{\mathcal{P}}$ based at x_0. Thus elements in $\widetilde{\mathcal{P}}$ are pairs, $(x, [\hat{x}])$, where $[\hat{x}]$ is an equivalence class of maps $\hat{x} : [0, 1] \times [0, 1] \to M$ satisfying the boundary conditions

$$\hat{x}(0, t) = x_0, \quad \hat{x}(s, i) \in L_i, \quad \hat{x}(1, t) = x(t).$$

The function $\mathcal{F}$ is the action functional $\mathcal{A} : \widetilde{\mathcal{P}} \to \mathbb{R}$ given by

$$\mathcal{A}(x, [\hat{x}]) = \int_0^1 \int_0^1 \hat{x}^*(\omega),$$

and its critical points are the lifts to $\widetilde{\mathcal{P}}$ of the points of the intersection $L_0 \cap L_1$. (See Fig. 4. $\mathcal{A}$ does not depend on the homotopy class of the map $\hat{x}$ because ω is closed and vanishes on the L_i.)

Now, let us consider the $\mathcal{A}$-gradient trajectories between the critical points. Since $\mathcal{P}$ is infinite dimensional these depend significantly on the choice of metric. We use a metric on $\widetilde{\mathcal{P}}$ that is pulled back from a metric on $\mathcal{P}$ that, in turn, is determined by a particular kind of Riemannian metric on M, namely a metric g_J given in the form $g_J(v, w) = \omega(v, Jw)$ where $J : TM \to TM$ is an ω-compatible[3] almost complex

---

[3]This means that $\omega(Jv, Jw) = \omega(v, w)$ and $\omega(v, Jv) > 0$ for all nonzero $v, w \in T_pM$. These equations generalise the well known relations between the Kähler metric g_J and Kähler form ω on a Kähler manifold M. The only difference is that the almost complex structure J need not be integrable, i.e. need not come from an underlying complex structure on the manifold M.
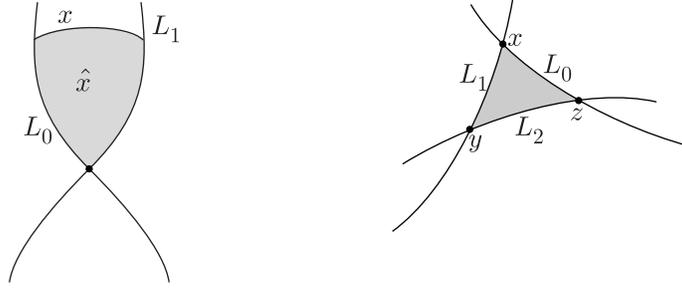
FIGURE 4. The path space $\widetilde{P}$ and a holomorphic triangle.

structure on M. It then turns out[4] that the gradient trajectories $\widetilde{v} : \mathbb{R} \to \widetilde{\mathcal{P}}$ of $\mathcal{A}$ are given by J-holomorphic strips

$$u : \mathbb{R} \times [0,1] \to M, \qquad u(s,t) := \widetilde{v}(s)(t),$$

in M with boundary on L_0 and L_1:

(1.3) $$\partial_s u + J(u)\partial_t u = 0, \quad u(s,0) \in L_0, \quad u(s,1) \in L_1.$$

One cannot always define a Floer complex in this setup because ∂^2 may not always vanish. The basic problem is that it may be impossible to define a good compactification of the 1-dimensional trajectory spaces $\widehat{\mathcal{M}}(x,z)$ simply by adding once-broken trajectories. (There is recent work by Fukaya–Oh–Ohta–Ono that sets up a framework in which to measure the obstructions to the existence of the Floer complex.) However, Ozsváth–Szabó consider a very special case of this construction in which the Lagrangian submanifolds arise from the geometry of the Heegaard diagram. In their case, $\partial^2 = 0$ and so the Floer homology groups $HF_*(L_0, L_1)$ are defined. Moreover they are independent of the choice of almost complex structure J on M and of any perturbations used in their construction.

Just as in the case of the Morse complex where one can define various products on the chain level by counting images of Ys and other trees, one can define topologically interesting chain maps between the complexes $CF_*(L_i, L_j)$ for different Lagrangian pairs by counting holomorphic triangles (think of these as fattened up Ys) or other polygons, with each boundary component mapping to a different Lagrangian submanifold L_i. For short we refer to the collection of such maps as the **naturality properties** of Lagrangian Floer theory. These properties lie at the heart of the proof that the Heegaard Floer groups depend only on the manifold Y rather than on the chosen Heegaard splitting. They can also be used to establish various interesting long exact sequences in the theory. Similar structures appear in

---

[4] The associated L_2-inner product on the tangent bundle of the path space is defined as follows. Given a path $x : [0,1] \to M$, the tangent space $T_x(\mathcal{P})$ consists of all (smooth) sections ξ of the pullback bundle $x^*(TM)$, i.e. $\xi(t) \in T_{x(t)}M$ for all $t \in [0,1]$, and satisfies the boundary conditions $\xi(i) \in T_{x(i)}L_i$ for $i = 0, 1$. Given two such sections ξ, η, we set

$$\langle \xi, \eta \rangle := \int_0^1 g_J(\xi(t), \eta(t)) \, dt.$$

Then the g_J-gradient of $\mathcal{A}$ is the vector field $\nabla \mathcal{A}$ defined by setting the inner product $\langle \nabla \mathcal{A}, \xi \rangle$ equal to $d\mathcal{A}(\xi)$ (the differential $d\mathcal{A}$ evaluated on the tangent vector ξ).

Seidel's work [32] on the **Fukaya category** of a symplectic manifold, the basis of one side of the homological mirror symmetry conjecture.

2. Heegaard Floer theory

In this section we first define the Heegaard Floer complexes. Then we briefly describe some applications.

2.1. **Definition of the invariants.** We saw in Example 1.1 that an oriented 3-manifold Y is completely determined by a triple (Σ, α, β) where Σ is a Riemann surface of genus g and α, β are sets of disjoint embedded circles

$$\alpha = \{\alpha_1, \ldots, \alpha_g\}, \quad \beta = \{\beta_1, \ldots, \beta_g\}.$$

Ozsváth and Szabó's idea is to use this data to construct a symplectic manifold (M, ω) together with a pair of Lagrangian submanifolds $\mathbb{T}_\alpha, \mathbb{T}_\beta$ and then to consider the corresponding Floer complex. This is a very rough version of their idea: in fact the manifold is not quite symplectic, the submanifolds are not quite Lagrangian, and they also put some extra structure on the Floer complex. The most amazing thing about their construction is that it does give interesting 3-manifold invariants.

For simplicity we shall assume throughout the following discussion that Y is a **rational homology sphere**, i.e. that $H_*(Y; \mathbb{Q}) \cong H_*(S^3; \mathbb{Q})$. This means that the abelianization $H_1(Y; \mathbb{Z})$ of the fundamental group $\pi_1(Y)$ is finite, that $H_2(Y; \mathbb{Z}) = 0$, and that Y is orientable. However, the invariants may be defined for all oriented Y.

The manifold M: This is the g-fold symmetric product $M_g := \mathrm{Sym}^g \Sigma_g$ of Σ_g, i.e. the quotient

$$M_g := \prod_g \Sigma / S_g,$$

of the g-fold product $\prod_g \Sigma := \Sigma \times \cdots \times \Sigma$ by the obvious action of the symmetric group S_g on g letters. M_g is smooth: if $\mathbb{C}$ is a local chart in Σ, then the points in $\mathrm{Sym}^g \mathbb{C}$ are unordered sets of g points in $\mathbb{C}$ and hence are the roots of a unique monic polynomial whose coefficients give a local chart on $\mathrm{Sym}^g \mathbb{C}$. However, M_g has no *natural* smooth structure; it inherits a complex structure J_M from the choice of a complex structure j on Σ, but different choices of j give rise to different[5] smooth structures on M_g. Similarly, although (M_g, J_M) is a Kähler manifold and so has symplectic structures, there is no natural choice of symplectic structure on M_g.

The manifold M_g has rather simple homotopy and cohomology. For example, in genus two $M_2 := \mathrm{Sym}^2 \Sigma_2$ is a 1-point blow up of the standard 4-torus $\mathbb{T}^4$; i.e. topologically it is the connected sum of $\mathbb{T}^4$ with a negatively oriented copy of the complex projective plane. In general, $\pi_1(M_g) \cong H_1(M_g; \mathbb{Z})$ is abelian of rank $2g$. In fact the inclusion $\Sigma \times \mathrm{pt} \times \cdots \times \mathrm{pt}$ induces an isomorphism $H_1(\Sigma_g) \cong \pi_1(M_g)$. When $g > 1$ the cohomology ring of M_g has one other generator in $H^2(M_g)$ that is Poincaré dual to the submanifold

$$\{z\} \times \mathrm{Sym}^{g-1}(\Sigma) \subset M_g,$$

---

[5]These smooth structures s_j are diffeomorphic. They are different in the sense that the identity map $(M, s_j) \to (M, s_{j'})$ is not smooth. Readers familiar with complex geometry might note that $\mathrm{Sym}^g \Sigma_g$ is a rather special complex manifold. It is birationally equivalent to the Picard variety $\mathrm{Pic}^g(\Sigma) \cong \mathbb{T}^{2g}$ of Σ: to get a map $\mathrm{Sym}^g \Sigma \to \mathrm{Pic}^g(\Sigma)$ think of the set of g points as a divisor and map it to the point in $\mathrm{Pic}^g(\Sigma)$ given by the corresponding degree g line bundle.

where z is any fixed point in Σ. Further when $g > 2$ $\pi_2(M_g) = \mathbb{Z}$, with generator

$$S^2 \equiv \Sigma/\rho \overset{\iota}{\hookrightarrow} \mathrm{Sym}^2(\Sigma) \to \mathrm{Sym}^g(\Sigma),$$

where we think of the 2-sphere S^2 as the quotient of Σ by a suitable involution ρ (e.g. the hyperelliptic involution) and set

$$\iota(z) := [z, \rho(z)] \in \mathrm{Sym}^2\Sigma, \quad z \in \Sigma.$$

When $g = 2$ the situation is a little more complicated. Nevertheless, in all cases the simple structure of $\pi_2(M_g)$ is one of the reasons why Ozsváth–Szabó's boundary operator ∂ has $\partial^2 = 0$. (Technically, it means that they are working in the **monotone** case.)

The tori $\mathbb{T}_\alpha, \mathbb{T}_\beta$: Because the circles α_i are mutually disjoint, the product

$$\alpha_1 \times \cdots \times \alpha_g \subset \prod_g \Sigma$$

maps bijectively onto a torus $\mathbb{T}_\alpha$ in M_g. This torus is clearly totally real; i.e. its tangent bundle $T\mathbb{T}_\alpha$ intersects $J_M(T\mathbb{T}_\alpha)$ transversally. There is no natural smooth symplectic structure on M_g that makes it Lagrangian, but this does not really matter since its inverse image in the product is Lagrangian for product symplectic forms.

If the α_j and β_k intersect transversally, then the two tori $\mathbb{T}_\alpha, \mathbb{T}_\beta$ also intersect transversally. Each intersection point can be written as

$$\mathbf{x} := (x_1, \ldots, x_g), \qquad x_k \in \alpha_k \cap \beta_{\pi(k)}, \;\; k = 1, \ldots, g, \;\; \pi \in S_g.$$

The trajectory spaces $\mathcal{M}(\mathbf{x}, \mathbf{y})$: Fix a complex structure j on Σ and consider the corresponding complex structure $J = J_M$ on the symmmetric product M_g. Given two intersection points $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ the elements in $\mathcal{M}(\mathbf{x}, \mathbf{y})$ are the J-holomorphic strips $u : \mathbb{R} \times [0, 1] \to M_g$ from $\mathbf{x}$ to $\mathbf{y}$ satisfying the conditions of (1.3). The domain $\mathbb{R} \times [0, 1]$ is conformally equivalent to the closed unit disc $\mathbb{D}$ in $\mathbb{C}$ with the two boundary points $\pm i$ removed. Thus Ozsváth–Szabó think of the strips as continuous maps

$$u : \mathbb{D} \to M_g$$

that are holomorphic in the interior $\mathrm{int}\, \mathbb{D}$ and take the left boundary $\partial\mathbb{D} \cap \{\Re z < 0\}$ to $\mathbb{T}_\alpha$ and the right boundary $\partial\mathbb{D} \cap \{\Re z > 0\}$ to $\mathbb{T}_\beta$. Continuous maps $\phi : \mathbb{D} \to M_g$ that satisfy these boundary conditions but are not necessarily holomorphic are called **Whitney discs** from $\mathbf{x}$ to $\mathbf{y}$.

One can define a complex whose vertices are the intersection points $\mathbb{T}_\alpha \cap \mathbb{T}_\beta$ and whose boundary map is given as in equation (1.2) by counting the number of elements in the 0-dimensional components of $\mathcal{M}(\mathbf{x}, \mathbf{y})/\mathbb{R}$. As we explain in more detail below (cf. equation 2.1)), the maps $u : \mathbb{D} \to M_g$ in this moduli space lift to holomorphic maps $\tilde{u} : F \to \Sigma$, where $F \to \mathbb{D}$ is a suitable g-fold branched cover. Thus the discs u can be understood by looking at their lifts to Σ, a fact which makes it possible to analyse the boundary map from a geometric point of view.

However, this complex contains no interesting information: its homology depends just on $H_*(Y)$. Therefore Ozsváth and Szabó add two pieces of extra structure.

Firstly, they observed that this complex decomposes into a direct sum of subcomplexes that are indexed by the Spin^c-structures[6] $\mathfrak{s}$ on Y. Secondly, they work in a suitable covering $\widetilde{\mathcal{P}}$ of the path space $\mathcal{P}(\mathbb{T}_\alpha, \mathbb{T}_\beta)$ with deck transformation group $\mathbb{Z}$. By taking the action of the generator U of this group into account as in Remark 1.2, they define various different, but related, chain complexes:

$$CF^\infty(Y, \mathfrak{s}), \quad \widehat{CF}(Y, \mathfrak{s}), \quad CF^+(Y, \mathfrak{s}), \quad CF^-(Y, \mathfrak{s}), \quad CF_{\mathrm{red}}(Y, \mathfrak{s}).$$

Whitney discs and Spin^c-structures: Given $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ we denote by $\pi_2(\mathbf{x}, \mathbf{y})$ the set of homotopy classes of Whitney discs from $\mathbf{x}$ to $\mathbf{y}$. Recall from Example 1.1 that each intersection point $\alpha_j \cap \beta_k$ lies on a unique f-gradient trajectory in Y that connects an index 2 critical point q_k to an index 1-critical point p_j. Thus the point $\mathbf{x} \in \mathbb{T}_\alpha$ can be thought of as a g-tuple of such gradient flow lines connecting each p_j to some q_k. The corresponding 1-chain $\gamma_{\mathbf{x}}$ in Y is called a **simultaneous trajectory**.

When $g > 1$ there is a Whitney disc $\phi : \mathbb{D} \to M_g$ from $\mathbf{x}$ to $\mathbf{y}$ only if the 1-cycle $\gamma_{\mathbf{x}} - \gamma_{\mathbf{y}}$ is null homologous. To see this, consider the commutative diagram

(2.1)
$$
\begin{array}{ccc}
F & \xrightarrow{\widetilde{\phi}} & \prod \Sigma \\
\downarrow & & \pi \downarrow \\
\mathbb{D} & \xrightarrow{\phi} & \mathrm{Sym}^g \Sigma,
\end{array}
$$

where $F \to \mathbb{D}$ is a suitable (possibly disconnected) branched g-fold cover (the pullback of π by ϕ.) Denote the component functions of $\widetilde{\phi}$ by $\widetilde{\phi}_\ell : F \to \Sigma$. The inverse images of the points $\pm i \in \partial \mathbb{D}$ divide the boundary of F into arcs that that are taken by the $\widetilde{\phi}_\ell$ alternately into subarcs of the α and β curves joining the intersections in $\mathbf{x}$ to those in $\mathbf{y}$. Each such subarc in an α_j-curve extends to a triangle in Y consisting of f-gradient flow lines in the stable manifold $W^s(p_j)$. Similarly the subarcs in β_k extend to triangles in the unstable manifolds $W^u(q_k)$, and it is not hard to see that the union of these triangles with the surfaces $\widetilde{\phi}_\ell(F)$ form a 2-chain with boundary $\gamma_{\mathbf{x}} - \gamma_{\mathbf{y}}$; see Fig. 5.
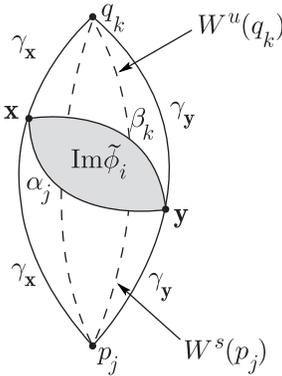

FIGURE 5. Whitney disc gives chain in Y with boundary $\gamma_{\mathbf{x}} - \gamma_{\mathbf{y}}$.

---

[6] This gives a point of contact with the Seiberg–Witten invariants, which depend for their very definition on the choice of a Spin^c-structure.

We say that two intersection points $\mathbf{x}, \mathbf{y}$ are equivalent if $\pi_2(\mathbf{x}, \mathbf{y})$ is nonempty. Using the Mayer–Vietoris sequence for the decomposition $Y = Y_1 \cup Y_2$, one can check that the differences $\gamma_\mathbf{x} - \gamma_\mathbf{y}$ generate $H_1(Y; \mathbb{Z})$. Hence these equivalence classes form an affine space modelled on the finite abelian group $H_1(Y; \mathbb{Z})$. The set of Spin^c structures on Y is also an affine space modelled on $H_1(Y; \mathbb{Z}) \cong H^2(Y; \mathbb{Z})$.

We now explain how the choice of a point $z \in \Sigma$ that does not lie on any α_j or β_k curve determines a natural map

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

such that $s_z(\mathbf{x}) = s_z(\mathbf{y})$ iff $\gamma_\mathbf{x} - \gamma_\mathbf{y} = 0$.

A Spin^c-structure on Y may be thought of as a decomposition of the (trivial) tangent bundle TY into the sum $L \oplus \mathbb{R}$ of a complex line bundle L with a trivial real line bundle,[7] and so corresponds to a nonvanishing vector field ξ on Y (a section of $\mathbb{R}$) that is well defined up to *homology*.[8] Therefore to define $s_z(\mathbf{x})$ we just need to associate a nonvanishing vector field $\sigma_\mathbf{x}$ to $\mathbf{x}$ that is well defined modulo homology. But z lies on a unique f-gradient trajectory γ_z from $\max f$ to $\min f$. This, together with the simultaneous trajectory $\gamma_\mathbf{x}$, pairs up the set of critical points of f. Since each pair has index sum 3, the gradient vector field ∇f of f can be modified near these trajectories to a nonvanishing vector field $\sigma_\mathbf{x}$. Then $\sigma_\mathbf{x} = \nabla f$ outside a union of 3-balls and so is well defined up to homology. We therefore set

$$s_z(\mathbf{x}) = [\sigma_\mathbf{x}] \in \mathrm{Spin}^c(Y).$$

Definition of $CF^\infty(Y, \mathfrak{s})$: Given a Spin^c-structure $\mathfrak{s}$, denote by $\mathcal{S} \subset \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ the corresponding set of intersection points. We define $CF^\infty(Y, \mathfrak{s})$ to be the free abelian group with generators $[\mathbf{x}, i] \in \mathcal{S} \times \mathbb{Z}$ and with relative grading

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

Here ϕ is any Whitney disc from $\mathbf{x}$ to $\mathbf{y}$, $n_z(\phi)$ is its intersection number with the generator $\{z\} \times \mathrm{Sym}^{g-1}(\Sigma)$ of $H_{2n-2}(M_g)$ and $\mu(\phi)$ is its Maslov index, that is, the expected dimension of the set $\mathcal{M}(\mathbf{x}, \mathbf{y}; \phi)$ of all components of the trajectory space $\mathcal{M}(\mathbf{x}, \mathbf{y})$ that contain elements homotopic to ϕ. One can show that the number $\mu(\phi) - 2n_z(\phi)$ is independent of the choice of ϕ. We then define the boundary operator δ^∞ by:

$$\delta^\infty([\mathbf{x}, i]) = \sum_{\mathbf{y} \in \mathcal{S}} \sum_{\phi \in \pi_2(\mathbf{x}, \mathbf{y}) : \mu(\phi) = 1} n(\mathbf{x}, \mathbf{y}; \phi) \, [\mathbf{y}, i - n_z(\phi)],$$

where $n(\mathbf{x}, \mathbf{y}; \phi)$ denotes the (signed) number of elements in

$$\widehat{\mathcal{M}}(\mathbf{x}, \mathbf{y}; \phi) := \mathcal{M}(\mathbf{x}, \mathbf{y}; \phi)/\mathbb{R}.$$

For the reasons outlined in Example 1.4, $(\delta^\infty)^2 = 0$. Hence $CF^\infty(Y, \mathfrak{s})$ is a chain complex.

Definition of $CF^\pm(Y, \mathfrak{s})$ and $\widehat{CF}(Y, \mathfrak{s})$: Since the submanifold $\{z\} \times \mathrm{Sym}^{g-1}(\Sigma)$ is a complex hypersurface, any holomorphic trajectory meets it positively. In other words, $n_z(\phi) \geq 0$ whenever $\mathcal{M}(\mathbf{x}, \mathbf{y}; \phi)$ is nonempty. Therefore the subset $CF^-(Y, \mathfrak{s})$ generated by the elements $[\mathbf{x}, i]$ with $i < 0$ forms a subcomplex of

---

[7]A Spin^c-structure on Y is a lift of the structural group $\mathrm{SO}(3)$ of the tangent bundle TY to the group $\mathrm{Spin}^c(3) := \mathrm{Spin}(3) \times_{\mathbb{Z}/2\mathbb{Z}} S^1 = \mathrm{SU}(2) \times_{\mathbb{Z}/2\mathbb{Z}} S^1$.

[8]Two nonvanishing vector fields are called homologous if one can be homotoped through nonvanishing vector fields to agree with the other except on a finite union of 3-balls.

$CF^\infty(Y, \mathfrak{s})$. We define $CF^+(Y, \mathfrak{s})$ to be the quotient $CF^\infty(Y, \mathfrak{s})/CF^-(Y, \mathfrak{s})$, i.e. the complex generated by $[\mathbf{x}, i]$, $i \geq 0$. All three complexes are $Z[U]$-modules where U acts by

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

reducing grading by 2. Finally we define $\widehat{CF}(Y, \mathfrak{s})$ to be the complex generated by the kernel of the U-action on $CF^+(Y, \mathfrak{s})$. Thus we may think of $\widehat{CF}(Y, \mathfrak{s})$ as generated by the elements $\langle \mathbf{x} \rangle, \mathbf{x} \in \mathcal{S}$, with differential

$$\widehat{\partial} \langle \mathbf{x} \rangle = \sum_{\mathbf{y}} \sum_{\phi \in \pi_2(\mathbf{x}, \mathbf{y}) : \mu(\phi) = 1, n_z(\phi) = 0} n(\mathbf{x}, \mathbf{y}; \phi) \langle \mathbf{y} \rangle,$$

i.e. we count only those trajectories that do not meet $\{z\} \times \mathrm{Sym}^{g-1}(\Sigma)$. The corresponding homology groups are related by exact sequences

$$(2.2) \qquad \begin{array}{ccccccccc} \cdots \longrightarrow & HF^-(Y, \mathfrak{s}) & \xrightarrow{i} & HF^\infty(Y, \mathfrak{s}) & \xrightarrow{\pi} & HF^+(Y, \mathfrak{s}) & \xrightarrow{\delta} & \cdots \\ \cdots \longrightarrow & \widehat{HF}(Y, \mathfrak{s}) & \xrightarrow{j} & HF^+(Y, \mathfrak{s}) & \xrightarrow{U} & HF^+(Y, \mathfrak{s}) & \longrightarrow & \cdots . \end{array}$$

There is yet another interesting group, namely $HF_{\mathrm{red}}(Y, \mathfrak{s})$, the cokernel of the above map π. This vanishes for the 3-sphere and for lens spaces. Later, we will use the fact that there is a pairing $HF^+ \otimes HF^- \to \mathbb{Z}$ that induces a pairing

$$(2.3) \qquad\qquad\qquad \langle \cdot, \cdot \rangle : HF_{\mathrm{red}} \otimes HF_{\mathrm{red}} \to \mathbb{Z}.$$

The following result is proved in [24].

Theorem 2.1. *Each of these relatively $\mathbb{Z}$-graded $\mathbb{Z}[U]$-modules is a topological invariant of the pair $(Y, \mathfrak{s})$.*

The proof that these homology groups are independent of the choice of almost complex structure j on Σ, of isotopy class of the loops α_i, β_j and of basepoint z uses fairly standard arguments from Gromov–Witten–Floer theory. To see that they remain unchanged under handleslides of the curves in α, β, one uses the naturality properties of Lagrangian Floer homology, defining a chain map by counting suitable holomorphic triangles; cf. the discussion following equation (2.4) below. Finally the fact that they are invariant under stabilization of the Heegaard splitting uses a "stretch the neck" argument.

At first glance it is not at all clear why one needs such a variety of homology groups. However, if we ignore the action of U and consider only HF^∞, we get very little information. Thus, for example, it is shown in [24] that when Y is a rational homology 3-sphere

$$HF^\infty(Y, \mathfrak{s}) \cong \mathbb{Z}[U, U^{-1}],$$

for all choices of $\mathfrak{s}$.[9] In fact the different complexes HF are just ways of encoding the subtle information given by the basepoint z, for they may all be defined in terms of the chain complex $CF^-(Y, \mathfrak{s})$ of $\mathbb{Z}[U]$-modules:

- $CF^\infty(Y, \mathfrak{s})$ is the "localization" $CF^-(Y, \mathfrak{s}) \otimes \mathbb{Z}[U, U^{-1}]$,
- $CF^+(Y, \mathfrak{s})$ is the cokernel of the localization map, and
- $\widehat{CF}(Y, \mathfrak{s})$ is the quotient $CF^-(Y, \mathfrak{s})/U \cdot CF^-(Y, \mathfrak{s})$.

---

[9] A similar phenomenon occurs in the Hamiltonian Floer theory of the loop space of a symplectic manifold M. The resulting homology groups $FH_*(M; H, J)$ are always (additively) isomorphic to the homology of M, but one gets interesting information by filtering by the values of the action functional.

The relations between these complexes are reminiscent of those occurring between the different homology groups of a space X with S^1 action. This is no accident: the group $\widehat{HF}$ is intended to resemble the ordinary homology of an S^1-space X with a single fixed point, while HF^+ is analogous to its equivariant homology $H_*^{S^1}(X)$. The second exact sequence in (2.2) then has the form of the Gysin sequence of the fibration $S^1 \to X \to X_{S^1}$, where $X_{S^1} := ES^1 \times_{S^1} X$ is the space whose homology is $H_*^{S^1}(X)$. In the conjectured equivalence between Heegaard Floer homology and the Seiberg–Witten–Floer theory of Y (which is an S^1-equivariant theory), the element U corresponds to the generator of $H_2(BS^1)$, although the underlying geometric reason for this is not yet fully understood; see Lee [17].

Example 2.2. Consider the case $Y = S^3$. We saw in Example 1.1 that this has a Heegaard splitting consisting of a torus $\mathbb{T}^2$, with a single α and a single β curve intersecting once transversally at $\mathbf{x}$. Denote by $\mathfrak{s}$ the unique Spin^c-structure on S^3. Then the complex $CF^-(S^3, \mathfrak{s})$ has generators $[\mathbf{x}, i], i < 0$, and trivial boundary map (this has to vanish since the relative gradings are even). This determines all the other groups; for example, $\widehat{HF}(S^3) \cong \mathbb{Z}$. There are many other Heegaard splittings for S^3. Ozsváth–Szabó give an example in [28, §2.2] of a genus 2 splitting where the differential ∂ *depends on the choice of complex structure j* on Σ_2. This might seem paradoxical. The point is that the differential is given by counting holomorphic discs in $\mathrm{Sym}^g(\Sigma)$, but as in diagram (2.1) these correspond to counting images in Σ of some branched cover F of the disc, and these images can have nontrivial moduli. This shows that Heegaard Floer theory is not entirely combinatorial: the next big advance might be the construction of combinatorial invariants, possibly similar to Khovanov's new knot invariants [16]. For recent work relating Khovanov homology to Lagrangian Floer theory and Heegaard Floer theory, see Dunfield–Gutov–Rasmussen [4], Manolescu [19], and Seidel–Smith [33].

One can make various additional refinements to the theory. For example, when Y is a rational homology sphere it is possible to lift the relative $\mathbb{Z}$-grading to an absolute $\mathbb{Q}$-grading that is respected by the naturality maps we discuss below; see [28, §3.2]. Ozsváth–Szabó also define knot invariants in [26] and use them in [27] to give a new obstruction for a knot to have unknotting number one. Heegaard Floer theory has many other interesting consequences for knot theory; see for example recent papers by Ekterkary [5], Livingston–Naik [18], Plamenevskaya [29], and Rasmussen [30].

2.2. **Properties and Applications of the invariants.** The power of Heegaard Floer theory comes from the fact that it is well adapted to certain natural geometric constructions in 3-manifold theory, such as adding a handle or performing a Dehn surgery on a knot, because these have simple descriptions in terms of Heegaard diagrams. Here is the basic geometric construction.

Suppose we are given three sets α, β, γ of g disjoint curves on the Riemann surface Σ_g that are the attaching circles for the handlebodies $U_\alpha, U_\beta, U_\gamma$. Then there are three associated manifolds

$$Y_{\alpha,\beta} = U_\alpha \cup U_\beta, \quad Y_{\alpha,\gamma} = U_\alpha \cup U_\gamma, \quad Y_{\beta,\gamma} = U_\beta \cup U_\gamma.$$

We now construct a 4-manifold $X = X_{\alpha\beta\gamma}$ with these three manifolds as boundary components. Let Δ be a triangle (or 2-simplex) with vertices $v_\alpha, v_\beta, v_\gamma$ and edges $e_\alpha, e_\beta, e_\gamma$ (where e_α lies opposite v_α), and form X from the four pieces
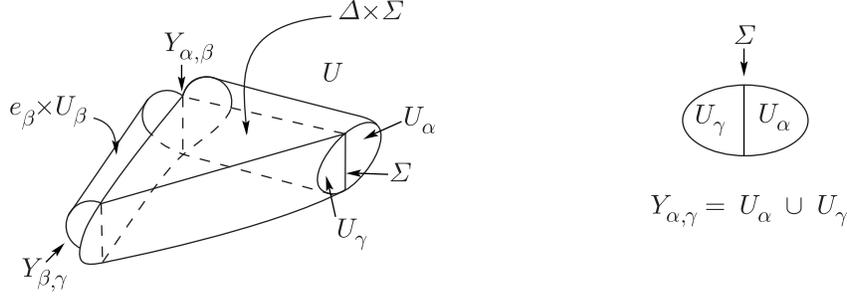
FIGURE 6. The 4-manifold $X_{\alpha\beta\gamma}$ (with Σ represented as an interval).

$$\left(\Delta \times \Sigma\right) \sqcup \left(e_\alpha \times U_\alpha\right) \sqcup \left(e_\beta \times U_\beta\right) \sqcup \left(e_\gamma \times U_\gamma\right)$$

by making the obvious identifications along $\partial\Delta \times \Sigma$ and then smoothing. For example, the part $e_\alpha \times \Sigma$ of $\partial\Delta \times \Sigma$ is identified with $e_\alpha \times \partial U_\alpha$; see Fig. 6. The resulting manifold has three boundary components, one corresponding to each vertex, with $Y_{\alpha,\beta}$ lying over $v_\gamma = e_\alpha \cap e_\beta$ for example. One can orient X so that

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

This elementary cobordism is called a **pair of pants** cobordism. Counting holomorphic triangles in M_g with boundaries on the three tori $\mathbb{T}_\alpha, \mathbb{T}_\beta, \mathbb{T}_\gamma$ gives rise under good circumstances to a map

$$(2.4) \qquad f^\infty : CF^\infty(Y_{\alpha,\beta}, \mathfrak{s}_{\alpha,\beta}) \otimes CF^\infty(Y_{\beta,\gamma}, \mathfrak{s}_{\beta,\gamma}) \to CF^\infty(Y_{\alpha,\gamma}, \mathfrak{s}_{\alpha,\gamma}).$$

(Here the Spin^c structures are assumed to extend to a common Spin^c structure $\mathfrak{s}$ on X.)

There are some interesting special cases of this construction. For example, the proof of handleslide invariance follows by taking γ to be a set of Heegaard circles obtained from α by handleslides. The construction can also be used to obtain a **long exact surgery sequence** which is very useful in analysing the effect of rational Dehn surgeries on Y. Here we shall concentrate on explaining some of the corresponding naturality properties of the theory.

Maps induced by cobordisms: Suppose that Y_2 is obtained from Y_1 by doing a 0-**surgery along a framed knot** K. This means that we choose an identification[10] of a neighborhood $N(K)$ with $S^1 \times D^2$, attach one part $\partial D^2 \times D^2$ of the boundary of the 4-ball $D^2 \times D^2$ to Y_1 via the obvious map

$$\psi : \partial D^2 \times D^2 \to S^1 \times D^2 \equiv N(K) \subset Y_1,$$

and then define Y_2 to be a smoothed out version of the union

$$Y_2 = \left(Y_1 \setminus \mathrm{int}\, N(K)\right) \sqcup_\psi (D^2 \times S^1),$$

where ψ identifies the boundary torus $S^1 \times S^1$ in $D^2 \times S^1$ to $\partial N(K)$. Note that the 4-manifold $W = \left([0,1] \times Y_1\right) \sqcup_\psi (D^2 \times D^2)$ is a cobordism from Y_1 to Y_2 obtained by adding a 2-handle to $[0,1] \times Y_1$ along $N(K) \subset \{1\} \times Y_1$.

---

[10] This is called a **framing** of the knot. It corresponds to choosing a pair of linearly independent vector fields along K that trivialize its normal bundle. Note that any knot in S^3 has a canonical framing; because $H_2(S^3) = 0$, K bounds an embedded surface S in S^3, and one can choose the first vector field to be tangent to S.

Given a knot K in Y_1, one can always choose a Heegaard diagram (Σ, α, β) for Y_1 so that K lies in the surface $\Sigma - \beta_2 - \cdots - \beta_g$ (and is given the obvious framing) and intersects β_1 once transversally. Pushing K into U_β, one sees that this is equivalent to requiring that K be disjoint from the discs D_j with boundary β_j for $j > 1$ and meet D_1 transversally in a single point. Hence one can construct a suitable diagram by starting with a neighborhood $N(K)$ of the knot and then adding 1-handles to obtain U_β. Since doing 0-surgery along K adds a disc with boundary K, it is easy to check that

$$Y_2 = Y_{\alpha, \gamma}, \quad \gamma := \{K, \beta_2, \ldots, \beta_g\}.$$

Further $Y_{\beta, \gamma}$ is a connected sum $\#(S^2 \times S^1)$ of copies of $S^2 \times S^1$, and so is standard. Pairing the map f^∞ in equation (2.4) with a canonical element in $HF^\infty(Y_{\beta, \gamma})$, one obtains a map

$$f^\infty : HF^\infty(Y_1; \mathfrak{s}_1) \to HF^\infty(Y_2; \mathfrak{s}_2)$$

for suitable $\mathfrak{s}_i$.

This construction can be extended to any cobordism.

Lemma 2.3. *Suppose that X is an oriented connected cobordism from Y_1 to Y_2, where each Y_i is an oriented connected 3-manifold. Then, for each Spin^c structure $\mathfrak{s}$ on X there is a natural induced map*

$$F_{X, \mathfrak{s}}^\infty : HF^\infty(Y_1, \mathfrak{s}_1) \to HF^\infty(Y_2, \mathfrak{s}_2),$$

where $\mathfrak{s}_i$ is the restriction of $\mathfrak{s}$ to Y_i.

There are corresponding maps for the other groups, $HF^\pm, \widehat{HF}$ and so on, that are well defined up to an overall choice of sign. All of them have the obvious functorial properties, behaving well for example under compositions of cobordisms. Another important property is that the image of the induced map

$$F_{X, \mathfrak{s}}^- : HF^-(Y_1, \mathfrak{s}_1) \to HF^-(Y_2, \mathfrak{s}_2)$$

is contained in HF_{red} if $b_2^+(X) \geq 1$. (This is the first appearance so far of this condition[11] on b_2^+ which is so ubiquitous in Seiberg–Witten theory.) In other words, im $F_{X, \mathfrak{s}}^-$ is contained in the image of the boundary map $\delta : HF^+ \to HF^-$ in the long exact sequence (2.2).

These functorial properties are the basis for most applications and calculations; see for example Némethi [21] and Owens–Strle [23].

A 4-manifold invariant: We now define an invariant $\Phi_{X, \mathfrak{s}}$ of a closed connected 4-manifold X with Spin^c-structure $\mathfrak{s}$. Conjecturally it agrees with the Seiberg–Witten invariant. Its construction illustrates the use of the different groups HF.

Suppose that X is a closed connected 4-manifold with $b_2^+(X) > 1$ and Spin^c-structure $\mathfrak{s}$. (For example, any symplectic 4-manifold has a canonical Spin^c-structure.) An **admissible cut** of X is a decomposition of X into two pieces X_1, X_2, each with $b_2^+(X_i) \geq 1$, along a 3-manifold $Y := X_1 \cap X_2$. We assume also that the restriction map

$$H^2(X) \to H^2(X_1) \oplus H^2(X_2)$$

---

[11] Given a connected, oriented 4-manifold X, $b_2^+(X)$ is the number of positive squares in the diagonalization of the cup product pairing on $H^2(X, \partial X)$. The relevant fact here is that when $b_2^+(X) \geq 1$ there is a closed surface C in X with self-intersection $C \cdot C \geq 0$.

is injective. Delete small 4-balls from the interior of each piece X_i and consider them as giving cobordisms from S^3 to Y. Denote $\theta := [\mathbf{x}, -1] \in HF^-(S^3)$ (cf. Example 2.2). Then consider

$$\Phi_{X,\mathfrak{s}} := \left\langle \delta^{-1} \circ F^-_{X_1,\mathfrak{s}_1} \theta, F^-_{X_2,\mathfrak{s}_2} \theta \right\rangle \in \mathbb{Z},$$

where we use the pairing (2.3) on $HF_{\mathrm{red}}(Y, \mathfrak{s})$. This element turns out to be independent of choices and nonzero for symplectic manifolds. (In this case it can be calculated using a decomposition of X coming from a Donaldson–Lefschetz pencil.) Hence in any admissible cut of a symplectic manifold, $HF_{\mathrm{red}}(Y)$ must be nonzero. Ozsváth–Szabó conclude in [25] that:

Proposition 2.4. *A connected closed symplectic 4-manifold X has no admissible cut $X = X_1 \cup X_2$ such that $Y := X_1 \cap X_2$ has $HF_{\mathrm{red}}(Y, \mathfrak{s}) = 0$ for all $\mathfrak{s}$.*

The first proof of this was in the case $Y = S^3$ and is due to Taubes [34]. He combined the well known fact that gauge theoretic invariants vanish on connected sums together with his proof that the Seiberg–Witten invariants do not vanish on symplectic 4-manifolds.

Rational homology 3-spheres Y for which HF^+ has no torsion and where $HF_{\mathrm{red}}(Y, \mathfrak{s}) = 0$ for all $\mathfrak{s}$ are called **L-spaces** in [28, §3.4]. All lens spaces are L-spaces, but not all Brieskorn homology spheres are: $\Sigma(2, 3, 5)$ is an L-space, but $\Sigma(2, 3, 7)$ is not (see Rustamov [31]). The class of L-spaces is not yet fully understood, but it has interesting geometric properties. For example it follows from the above proposition that L-spaces do not support any taut foliations, i.e. foliations in which the leaves are minimal surfaces for some Riemannian metric on Y, for if Y supports such a foliation, then results of Eliashberg–Thurston [7], Eliashberg [6], Giroux [13] and Etnyre [8] about contact structures allow one to construct a symplectic manifold X that has an admissible cut with $Y = X_1 \cap X_2$.

As a final corollary, we point out that similar arguments imply that Heegaard Floer theory can **detect the unknot** in S^3. This means the following. Suppose that K is a knot in S^3 and denote by $S^3_0(K)$ the result of doing 0-surgery along K with the canonical framing described above.

Corollary 2.5. *If $HF(S^3_0(K)) = HF(S^3_0(\mathrm{unknot}))$, then K is the unknot.*

Sketch of proof. Let $Y = S^3_0(K)$. Suppose that $Y \neq S^3_0(\mathrm{unknot}) = S^2 \times S^1$. By a deep result of Gabai [12], Y admits a taut foliation. As above, this implies that $HF_{\mathrm{red}}(Y, \mathfrak{s}) \neq 0$ for some $\mathfrak{s}$. But $HF_{\mathrm{red}}(S^2 \times S^1, \mathfrak{s})$ is always 0. $\qquad\square$

Acknowledgements

About the author

Dusa McDuff is a professor at Stony Brook University. She is a member of the U.S. National Academy of Sciences and a Fellow of the Royal Society of London.

References

[1] R. Bott, Morse theory indomitable, *Publ. I.H.E.S.* **68** (1988), 99–114. MR1001450 (90f:58027)

[2] R. T. Cohen, Morse theory, graphs and string topology, GT/0411272.

[3] C. Conley, *Isolated invariant sets and the Morse index*, CBMS Notes, **38**, American Mathematical Society, Providence, RI (1978). MR0511133 (80c:58009)

[4] N. Dunfield, S. Gotov and J. Rasmussen, The superpotential for knot homologies, GT/0505662.

[5] E. Eftekhary, Filtration of Heegaard Floer homology and gluing formulas, GT/0410356.

[6] Y. M. Eliashberg, A few remarks about symplectic filling, SG/0311459, *Geom. and Top.* **8** (2004), 277–93. MR2023279 (2005a:57022b)

[7] Y. M. Eliashberg and W. Thurston, *Confoliations*, Univ. Lecture Series, no. 13, AMS (1998). MR1483314 (98m:53042)

[8] J. Etnyre, On symplectic fillings, SG/0312091, *Alg. and Geom. Top.* **4** (2004), 73–80. MR2023278 (2005a:57022a)

[9] A. Floer, An instanton invariant for 3-manifolds, *Comm. Math. Phys.* **118** (1988), 215–240. MR0956166 (89k:57028)

[10] A. Floer, Morse theory for Lagrangian intersections, *Journal of Differential Geometry* **28** (1988), 513–47. MR0965228 (90f:58058)

[11] A. Floer, Symplectic fixed points and holomorphic spheres, *Communications in Mathematical Physics* **120** (1989), 575–611. MR0987770 (90e:58047)

[12] D. Gabai, Foliations and the topology of 3-manifolds. III, *J. Diff. Geom.* **26** (1987), 479–536. MR0910018 (89a:57014b)

[13] E. Giroux, Geométrie de contact: de la dimension trois vers les dimensions supérieures, *Proc. ICM-Beijing* **2** (2002), 405–414, Higher Ed. Press, Beijing, 2002. MR1957051 (2004c:53144)

[14] M. Gromov, Pseudo holomorphic curves in symplectic manifolds, *Inventiones Mathematicae* **82** (1985), 307–47.

[15] M. Hutchings and Y.-J. Lee, Circle-valued Morse theory and Reidemeister torsion, *Geom. and Top.* **3** (1999), 369–96. MR1716272 (2000h:57063)

[16] M. Khovanov, A categorification of the Jones polynomial, *Duke Math. J.* **101** (2000), 359–426. MR1740682 (2002j:57025)

[17] Yi-Jen Lee, Heegaard splittings and Seiberg–Witten monopoles, GT/0409536.

[18] C. Livingston and S. Naik, Ozsváth–Szabó and Rasmussen invariants of doubled knots, GT/0505361.

[19] C. Manolescu, Nilpotent slices, Hilbert schemes and the Jones polynomial, SG/0411015.

[20] J. Milnor, *Morse Theory*, Annals of Math. Studies #51, Princeton Univ. Press, 1963. MR0163331 (29:634)

[21] A. Némethi, On the Heegaard Floer homology of $S^3_{-p/q}(K)$, GT/0410570.

[22] S. Novikov, Multivalued functions and functionals, an analog of the Morse theory, *Soviet Math. Dokl.* **24** (1981), 222–226. MR0630459 (83a:58025)

[23] B. Owens and S. Strle, Rational homology spheres and four-ball genus, GT/0308073.

[24] P. Ozsváth and Z. Szabó, Holomorphic discs and three-manifold invariants for closed 3-manifolds, SG/0101206, *Ann. Math.* **159** (2004), 1027–1158. MR2113019

[25] P. Ozsváth and Z. Szabó, Holomorphic triangle invariants and the topology of symplectic four manifolds, SG/0210127, *Duke Math. J.* **121** (2004), 1–34. MR2031164 (2004m:57059)

[26] P. Ozsváth and Z. Szabó, Holomorphic discs and knot invariants, *Adv. Math* **186** (2004), 58–116. MR2065507 (2005e:57044)

[27] P. Ozsváth and Z. Szabó, Knots with unknotting number one and Heegaard Floer homology, GT/0401426.

[28] P. Ozsváth and Z. Szabó, Heegaard diagrams and holomorphic discs, GT/0403029, *Different faces of geometry*, Int. Math. Ser. (N.Y.), Kluwer/Plenum, New York (2004), 301–348. MR2102999 (2005g:57057)

[29] O. Plamenevskaya, Transverse knots and Khovanov homology, GT/0412184.

[30] J.A. Rasmussen, Khovanov homology and the slice genus, GT/0402131.

[31] R. Rustamov, On Plumbed L-spaces, GT/0505349.

[32] P. Seidel, Fukaya categories and deformations, SG/0206155.

[33] P. Seidel and I. Smith, A link invariant from the symplectic geometry of nilpotent slices, SG/0405089.

[34] C. H. Taubes, The Seiberg–Witten invariants and symplectic forms, *Math. Res. Letters* **1** (1994), 809–822. MR1306023 (95j:57039)

[35] E. Witten, Supersymmetry and Morse theory, *J. Diff. Geo.* **17** (1982), 661–692. MR0683171 (84b:58111)

Department of Mathematics, Stony Brook University, Stony Brook, New York 11794-3651

E-mail address: dusa@math.sunysb.edu

URL: http://www.math.sunysb.edu/~dusa