http://dx.doi.org/10.1090/pspum/048/974342

Proceedings of Symposia in Pure Mathematics
Volume 48 (1988)

New Invariants of 3- and 4-Dimensional Manifolds

MICHAEL ATIYAH

1. Introduction. Hermann Weyl was probably the most influential mathematician of the twentieth century. The topics he chose to study, the lines he initiated and his general outlook have proved remarkably fruitful and have underpinned much of the development of the past fifty years. Weyl saw mathematics, and to some extent theoretical physics, as an organic whole and not as a collection of special subjects. The relation between geometry and physics was perhaps his central interest, but this led him deeply into the theory of Lie groups and differential equations. He would, I am sure, have been delighted in the resurgence of interest in this whole area which we are now witnessing. The geometry and physics of gauge theories, in its modern form, is one of the most exciting developments of our time and it rests ultimately on Weyl's pioneering ideas.

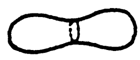
In this talk I am going to describe some of the most recent, and still incomplete, developments involving the application of ideas from the physics of gauge theories to the study of manifolds in 3 and 4 dimensions. The main results are due to S. K. Donaldson who initiated the whole programme a few years ago, but important contributions are being made by C. Taubes and A. Floer. Moreover mathematicians have learnt a great deal about the geometrical interpretation of physical ideas from E. Witten. His paper [**14**] on supersymmetry and Morse theory has been very influential in a number of ways.

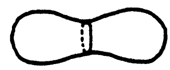
I should emphasize that many of the results I shall describe have not yet been fully written up, but the general picture seems to be fairly clear and the ideas are so beautiful and simple that they deserve a nontechnical presentation. Detailed treatments will hopefully be provided in due course by Donaldson, Floer, and Taubes. I am grateful to all of them for explaining their ideas to me at this early stage.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 81E13, 57R55 ;Secondary 58E05, 58F05.

Let me now outline in broad terms the picture I want to describe. It may be illustrated by the following scheme.

Dimension 2		Dimension 3		Dimension 4
π_1		π_1		$(\pi_1 = 0)$
Aut π_1		Homology spheres $(H_1 = 0)$		Differentiable structures
	$\leftarrow$	Casson invariant Floer homology	$\leftarrow$	Donaldson invariants
X	$\leftarrow$	Y	$\leftarrow$	Z


X


Y

In each dimension I have indicated the object of interest and the invariants used to study them. Thus in dimensions 2 and 3 the fundamental group π_1 is our main concern (and homology spheres concentrate on the nonabelian part), while in dimension 4, even when we restrict to simply connected manifolds, there are many different differentiable structures on the same topological manifold. These are detected by Donaldson's invariants [**5**], while in 3 dimensions there is the integer invariant recently introduced by Casson [**3**] and, as I shall explain, refined by Floer to give a homology theory. The horizontal arrows and the diagrams at the bottom are meant to indicate that the invariants in n dimensions (for $n = 4, 3$) can be related to those of an $(n-1)$-dimensional submanifold when we 'cut it in half'.

2. The Casson invariant. Let Y be an oriented homology 3-sphere, i.e., with $H_1(Y) = 0$. Then the Casson invariant $\lambda(Y)$ is roughly defined by

$$\lambda(Y) = \tfrac{1}{2}\{\text{number of irreducible representations } \pi_1(Y) \to \mathrm{SU}(2)\}.$$

Here, of course, we identify conjugate representations, and since $H_1(Y)$, the abelianization of $\pi_1(Y)$, is zero the only reducible representation in $\mathrm{SU}(2)$ is the trivial one, so irreducible = nontrivial.

The main problem is to give a proper way of counting the number of representations, so as to get a well-defined (and finite) integer. The most natural way has been developed by Taubes and involves identifying representations $\pi_1(Y) \to \mathrm{SU}(2)$ with *flat connections* on Y. We proceed as follows. Let

$$\mathscr{A} = \{\text{space of SU(2)-connections for the trivial bundle over } Y\},$$
$$\mathscr{G} = \{\text{group of gauge transformations, i.e. maps } Y \to \mathrm{SU}(2)\},$$
$$\mathscr{C} = \mathscr{A}/\mathscr{G}.$$

Then, except for reducible connections, $\mathscr{C}$ is an infinite-dimensional manifold. Moreover $A \to F_A$ (the curvature of A) defines a natural 1-form F on $\mathscr{C}$. To see this note that tangents to $\mathscr{A}$ are 1-forms on Y with values in the Lie algebra of

SU(2), which pair naturally (since $\dim Y = 3$) with Lie algebra–valued 2-forms such as the curvature. The Bianchi identity asserts that the $\mathscr{G}$-invariant 1-form on $\mathscr{A}$ defined by the curvature descends to a 1-form on $\mathscr{C}$.

The zeros of F, i.e. flat connections, correspond naturally to representations $\pi_1(Y) \to \mathrm{SU}(2)$. Thus the "number of irreducible representations" appears as the "number of zeros of F" on the nonsingular part of $\mathscr{C}$. To make sense of this number one must, as in finite dimensions, consider perturbations of the 1-form to get simple zeros and then count them up with appropriate signs determined by orientations. Here, in an infinite-dimensional setting one must use appropriate Fredholm perturbations, and this has been carried out and justified by Taubes. The determination of signs is a subtle story and I shall return to this later in §3.

REMARKS. (1) With this definition it is not clear why $\lambda(Y)$ (rather than $2\lambda(Y)$) is an integer. In fact no very satisfactory explanation appears to be known at present.

(2) In principle SU(2) could by replaced here by SU(n), but then more care would need to be taken with reducible representations. So far this has not been fully investigated.

(3) The Casson invariant $\lambda(Y)$ has a more computational definition (due to Casson) which will be described in §5.

(4) The Casson invariant is quite a powerful invariant and was used to settle an outstanding problem on 3-manifolds.

3. Morse theory. The 1-form F on $\mathscr{C}$ given by the curvature turns out to be *closed*. It should therefore locally be the differential of a function on $\mathscr{C}$. In fact there is a well-known function $f\colon \mathscr{C} \to \mathbf{R}/\mathbf{Z}$ introduced by Chern and Simons and an elementary calculation shows that

$$F = 4\pi^2\, df.$$

For convenience I recall the definition of f. Given a connection A for the trivial bundle $Y \times \mathrm{SU}(2)$, let A_0 be the trivial or product connection and put $A_t = (1-t)A + tA_0$ for $0 \le t \le 1$. This is a path of connections on Y or equivalently a connection on $Y \times I$, where I is the unit interval. Now define

$$f(A) = \frac{1}{8\pi^2}\int_{Y\times I} \operatorname{Tr} F^2$$

where F is the connection on $Y \times I$. Note that this integral on a closed 4-manifold gives the second Chern class and so is an integer. For similar reasons $f(A)$ is invariant under the connected component $\mathscr{G}_0$ of the gauge group $\mathscr{G}$, but changes by integers under the full group (note that $\mathscr{G}/\mathscr{G}_0 \cong \mathbf{Z}$). Thus, as a function on $\mathscr{C}$, f is well-defined provided we take its values in $\mathbf{R}/\mathbf{Z}$.

Now in finite dimensions if a 1-form is the differential of a real-valued function then the zeros of the 1-form become critical values of the function. The number of zeros is the Euler characteristic of the manifold, but the Morse theory of the function gives us more precise information related to the homology of the manifold. Traditionally this relation is given by the *Morse inequalities* between

Betti numbers and numbers of critical points of different types. However, as Witten suggested in [**15**], one can do better. It is possible to use the critical point information to *construct* the homology of the manifold. Witten's idea is so beautiful and so important that I will now review it, before passing on to explain how Floer adapts it to the infinite-dimensional case of our manifold $\mathscr{C}$ of connections.

Assume therefore that we have a compact manifold M and a real-valued function f with only nondegenerate critical points. At each such critical point P the Hessian $H_P(f)$ is a nondegenerate quadratic form on the tangent space at P. It has a "type" (n_P^+, n_P^-), the integers n^+ and n^- being the number of $+$ and $-$ entries in a diagonalization. Of course $n_P^+ + n_P^- = \dim M$ is independent of P. As a first approximation to the homology of M Witten forms the chain groups C_q having one generator for each critical point P with $n_P^- = q$. The next and crucial step is to introduce a boundary operator $\partial\colon C_q \to C_{q-1}$. This will be given by a matrix with one entry for each pair of critical points (P, Q) with $n_P^- = q$, $n_Q^- = q-1$. To define ∂ we first choose a generic metric on M and introduce the corresponding (descending) gradient flow of f. We then look at trajectories of this flow that start at P and end at Q. The number of such trajectories is *finite* and, counted with an appropriate sign, this gives the (P, Q) entry of ∂. One then verifies that $\partial^2 = 0$, so that we can define the *homology groups* of the complex C_*. This can be shown to be independent of the metric and finally identified with the homology of M.

REMARKS. (1) For Witten the homology of M is the Hodge–de Rham homology, represented by harmonic forms or, using f, as the zero eigenforms of the modified Laplacian Δ_{tf} where d is replaced by $e^{-tf}\,de^{tf}$. In Δ_{tf} we have a potential term $t^2|\operatorname{grad} f|^2$ and so, for $t \to \infty$, the zero eigenforms are concentrated near the critical points of f. These are the classical ground states. However, the eigenform belonging to P has an exponentially small correction due to Q which is approximately computed by using the trajectories of grad f from P to Q. This is *quantum mechanical tunnelling*, which describes the probability of the transition $P \to Q$, which is forbidden classically, but can occur quantum mechanically. Thus the boundary operator ∂ of Witten's chain complex C_* is to be interpreted in terms of such tunnelling. This remark will acquire even more significance in the Floer theory we shall discuss next.

(2) For any pair of critical points (P, Q) with $n_P^- - n_Q^- = r$ say, the trajectories of grad f from P to Q form (for generic metrics) an $(r-1)$-dimensional family. They are the intersections of the *unstable* manifold of P (with dimension n_P^-) and the *stable* manifold of Q (with dimension $n_Q^+ = \dim M - n_Q^- = \dim M - n_P^- + r$). This intersection has dimension r and is made up of an $(r-1)$-dimensional family of curves.

4. Floer homology. After our finite-dimensional digression in §3 on Witten's version of Morse theory we return to the space $\mathscr{C}$ of (classes of) connections on our homology 3-sphere Y and the Chern-Simons function $f\colon \mathscr{C} \to \mathbf{R}/\mathbf{Z}$.

To develop the corresponding Morse theory for this situation we have two main problems:

(1) f takes values in $\mathbf{R}/\mathbf{Z}$, not in $\mathbf{R}$.

(2) The Hessian of f at a critical point has both Morse indices n^+ and n^- *infinite*.

The first problem is not too serious and can be dealt with by passing to the infinite cyclic covering $\mathscr{C}_0 = \mathscr{A}/\mathscr{G}_0$. The second problem is more fundamental and presents essentially new features. To understand this let us recall that for the classical Morse theory of geodesics in Riemannian manifolds the Hessian (viewed as operator, rather than quadratic form, by using the metric) is of *Laplace type*; i.e., it is a second-order elliptic operator and hence is bounded below, so that n^- is always finite. In our present situation the Hessian is of *Dirac type*, i.e., it is of first order. In fact, this Hessian is essentially the operator $*d$ on $\Omega^1/d\Omega^0$, suitably extended to Lie algebra–valued forms.

The way around our difficulty is to observe that the important quantity in the Morse theory is not the Morse index n_P^-, but the *relative* Morse indices $n_{P,Q}^- = n_P^- - n_Q^-$ for pairs of critical points P, Q. Although $n_P^- = \infty$ for all P we can make sense of the differences $n_{P,Q}^-$, i.e, we can define relative Morse indices. This is done as follows.

First, using a fixed metric on Y we can extend the Hessians at the critical points P to a continuous family of self-adjoint (Dirac type) operators H_C for all $C \in \mathscr{C}$. In particular, for any continuous path $\alpha(t)$ from P to Q we get a 1-parameter family H_t of self-adjoint operators connecting H_P to H_Q. In this situation there is a standard integer invariant, the *spectral flow*, that can be defined [**2**]. It describes the net number of negative eigenvalues of H_P that cross over and end up as positive eigenvalues of H_Q. This is clearly a regularization of the formal quantity $n_P^- - n_Q^-$. It is a topological invariant and so depends only on the homotopy class of the path α from P to Q. If $\mathscr{C}$ (or better the irreducible part of $\mathscr{C}$) were simply connected α would be unique up to homotopy and so the relative Morse index $n_{P,Q}^-$ would be well-defined. Since $\mathscr{C}$ is not simply connected, but has the infinite cyclic covering $\mathscr{C}_0$, it follows that $n_{P,Q}^-$ is defined modulo the spectral flow round a generating closed loop in $\mathscr{C}$. But such a spectral flow can be computed from the index theorem on $Y \times S^1$ [**2**] and one finds the answer 8. This is essentially the same index calculation which determines the dimension of the instanton moduli space, and which figures prominently in Donaldson's work.

It is now clear how one should proceed, at least formally. Assume first that all nontrivial critical points of f (i.e., all irreducible representations $\pi_1(Y) \to \mathrm{SU}(2)$) are nondegenerate: if not we would have to make a Fredholm perturbation as in Taubes' approach to the Casson invariant. Now we follow Witten and define chain groups C_* indexed modulo 8. Finally, using trajectories of grad f from P to Q we define a boundary operator ∂, prove $\partial^2 = 0$ and derive homology groups H_*. Finally one should prove these homology groups are independent of the various choices (perturbations) made.

The groups obtained in this way have a mode 8 grading but no obvious start of the grading, i.e., no obvious H_0. In fact with more care, comparing with the trivial representation, one can fix the grading by identifying H_0. Thus we get groups labelled by the integers modulo 8.

REMARK. The problem of signs involved in the definition of ∂ is essentially the same as the problem of giving signs to the zeros of grad f in Taubes' approach. Also the labelling of the dimensions modulo 8 is a refinement of the modulo 2 labelling that determines the sign of the Casson invariant.

The most important ingredient in this whole procedure is the definition of the boundary ∂, using trajectories of grad f. Now, such a trajectory is explicitly a solution of the differential equation

$$\frac{dA}{dt} = - * F_A \tag{4.1}$$

(since grad $f = *F_A$). This equation, interpreted on the infinite cylinder $Y \times R$, is just the anti-self-duality equation which defines instantons. The boundary conditions we impose are that for $t \to -\infty$ the connection converges to the flat connection corresponding to P, while $t \to +\infty$ corresponds similarly to Q. Recalling Witten's interpretation of ∂ as a tunnelling effect, we see that we are using instantons to tunnel from the ground state, or vacuum, of one flat connection on Y to the ground state of another flat connection. This is precisely the way physicists use instantons and was their original motivation. For this reason, Witten even used the word 'instanton' in the classical Morse theory picture for the trajectories connecting consecutive critical points.

Of course the analytical justification of the formal procedure just outlined requires much careful analysis. However, the essential ingredients centre round the analytical properties of instantons, in particular compactness questions and Fredholm perturbation theory. These properties are by now well understood as a result of the basic work of Uhlenbeck, Taubes, and Donaldson. The conclusion is that we have new invariants defined for homology 3-spheres Y in the form of homology groups, denoted $HF_q(Y)$, indexed by $q \in \mathbf{Z}_8$.

Actually, just as in finite dimensions, reversing the sign of f interchanges n^+ and n^- and corresponds to Poincaré duality. Therefore we should really distinguish between two dual homology groups $HF^+(Y)$ and $HF^-(Y)$. They depend on the orientation of Y and switch if we reverse the orientation.

Of course, from the way we derived these homology groups, the Casson invariant (up to a factor 2) is just the corresponding Euler characteristic:

$$2\lambda(Y) = \sum_{q=0}^{7} (-1)^q \dim HF_q^+(Y).$$

In this sense the groups HF represent a refinement of the Casson invariant, and therefore should be very interesting invariants of Y.

REMARKS. (1) In finite dimensions the Witten-Morse approach leads to the usual homology of the underlying manifold. In the infinite-dimensional case of our space $\mathscr{C}$ the homology groups HF that we obtained are not related to the ordinary homology groups of $\mathscr{C}$. I shall explain later that they should be viewed as "middle-dimensional" homology groups.

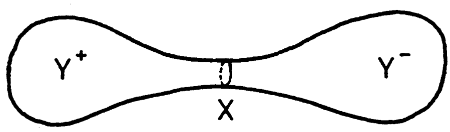
(2) The letter F in HF can stand for Floer who introduced these groups [**8**]. It can also appropriately stand for Fredholm, Fermi or Fock. This will become clear in a later section.

5. Relation with 2 dimensions. I will now indicate how the Casson invariant, and, more generally, the Floer homology groups, can be calculated when our homology 3-sphere Y is presented in terms of a *Heegaard splitting.*

A Heegaard splitting of Y is a decomposition into 2 halves $Y^{\pm}$, where

$$\partial Y^+ = X = -\partial Y^-$$

with X a compact Riemann surface of genus g say, and $Y^{\pm}$ are handlebodies of genus g (connected sums of g solid tori).



In a suitable basis $\pi_1(X)$ is generated by $A_1, \ldots, A_g$, $B_1, \ldots, B_g$ with the single relation

$$\Pi[A_i, B_i] = 1,$$

$\pi_1(X) \to \pi_1(Y^+)$ sends each $B_i \to 1$ and the images of the A_i freely generate $\pi_1(Y^+)$. For $\pi_1(Y^-)$ a similar description holds relative to a different basis—i.e., after applying an automorphism of $\pi_1(X)$.

Casson's procedure [**3**] to define his invariant $\lambda(Y)$ is to count the "number of representations of $\pi_1(Y)$" using the diagram

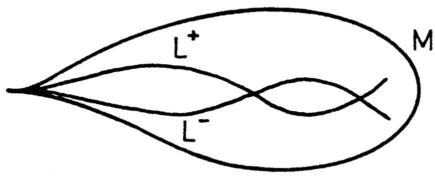
$$\begin{array}{ccc} \pi_1(X) & \longrightarrow & \pi_1(Y^+) \\ \downarrow & & \downarrow \\ \pi_1(Y^-) & \longrightarrow & \pi_1(Y) \end{array}$$

This shows that a representation of $\pi_1(Y)$ is the same as a pair of representations of $\pi_1(Y^+)$, $\pi_1(Y^-)$ which agree when pulled back to $\pi_1(X)$.

Now the (classes of) representations $\pi_1(X) \to \mathrm{SU}(2)$ form a moduli space M which has been extensively studied in algebraic geometry [**1**, **12**, **13**]. In particular, after removing the reducible representations, it is a manifold of dimension $6g - 6$. If X is given a complex structure then M becomes naturally a complex (algebraic) variety and a metric on X induces a Kähler metric on M (outside its singularities). The representations of $Y^{\pm}$ give subspaces $L^{\pm} \subset M$ of dimension $3g - 3$. We can therefore define the Casson invariant $\lambda(Y)$ by

$$2\lambda(Y) = L^+ \cap L^-,$$

where we count the intersections in M (away from the singularities of M). We are now dealing with ordinary homology and usual intersection theory (except that care has to be taken over the singular points of M).



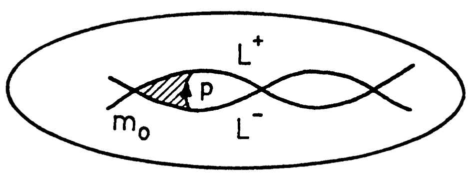
REMARKS. (1) If this approach is taken as a definition of $\lambda(Y)$ then one must show independence of the Heegaard splitting.

(2) Casson uses this approach to produce an effective algorithm for computing $\lambda(M)$, when M is described in terms of surgery on links in S^3 (another way of presenting homology 3-spheres).

Since the Floer homology groups are a refinement of the Casson invariant, it is now reasonable to ask if there is a way of computing $HF(Y)$ using the Heegaard splitting and the Riemann surface X. I shall outline an approach to this problem, which has yet to be fully worked out.

Since M is a Kähler manifold (with singularities) it is in particular symplectic. In fact, as shown in [**1**], the symplectic structure is canonical and independent of the metric on X. Moreover $L^{\pm}$ are *Lagrangian* submanifolds, i.e. submanifolds of middle dimension on which the symplectic 2-form ω of M is identically zero. Now Floer [**6**, **7**] has studied, in general, the problem of intersections of Lagrangian submanifolds of compact symplectic manifolds and, for this purpose, has developed a homology theory. From an analytical point of view this is very similar to the theory leading to the HF groups described in §4. When applied to the particular case of $L^{\pm}$ in M above it is highly plausible that it should coincide with the theory of §4, as I shall indicate later. So first let me outline Floer's "symplectic Morse theory".

We start from any compact symplectic manifold M and two (connected) Lagrangian submanifolds L^+ and L^-. Consider the space Q of paths in M starting on L^- and ending on L^+. Assume for simplicity that $L^+ \cap L^-$ is not empty (otherwise the theory will be trivial) and choose a base point $m_0 \in L^+ \cap L^-$. Define a function $f(p)$ on Q as the area (integral of the symplectic 2-form ω) of a strip obtained by deforming the path p to the constant path m_0.



Since L^+ and L^- are Lagrangian, and ω is closed, this area is unchanged under continuous variations of the strip (with p fixed). However topologically inequivalent strips will differ in area by a "period" of ω. If for simplicity we assume L^+ and L^- simply connected these periods are just the usual periods of ω over 2-cycles in M. If $b_2(M) = 1$ there is just one period which by rescaling ω we can take to be 1. This means f becomes a function

$$f: Q \to \mathbf{R}/\mathbf{Z}.$$

The critical points of f are easily seen to be the constant paths corresponding to the points of intersection of L^+ and L^-. The Hessian is again of Dirac type and one can define a relative Morse index as in §4. This turns out to be well-defined modulo $2N$, where $c_1(M) = N[\omega]$, $c_1(M)$ being the first Chern class of M (note that symplectic manifolds have Chern classes) and $[\omega]$ is the class of ω in $H^2(M)$.

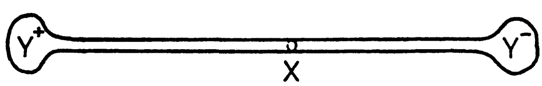
The trajectories of grad f correspond to *holomorphic* strips (with boundaries in $L^\pm$) in the sense of Gromov [**10**]. If M is actually complex Kähler then these are just holomorphic strips in the usual sense.

In this way, following Witten as in §4, Floer defines homology groups graded by Z_{2N} as intrinsic invariants of (M, L^+, L^-).

If now we take (M, L^+, L^-) to be the moduli spaces arising from a Heegaard splitting of a homology 3-sphere Y it is then reasonable to conjecture that the groups defined in the symplectic context (with care taken of the singularities of M) coincide with the groups $HF(Y)$ of §4.

Note that in both cases the representations $\pi_1(Y) \to \mathrm{SU}(2)$ give the generators of the chain group (provided these representations are nondegenerate). One has then to compare the relative Morse indices and the boundary operator ∂.

Geometrically, a path on M, i.e. a 1-parameter family of flat connections on the Riemann surface X, can be viewed as a connection on the cylinder $X \times R$. Moreover the boundary conditions (corresponding to L^+ and L^-) imply that, asymptotically as $t \to \pm\infty$, the connection extends (as a flat connection) over $Y^\pm$, thus giving essentially a connection on Y. In this way the symplectic theory for paths in M should be related to a limiting case of the Floer theory for the space $\mathscr{C}$ of connections on Y. Note that the limit is one in which Y is stretched out along its "neck", so that the two ends get further and further apart.



6. Donaldson invariants. Donaldson [**5**] has introduced certain invariants for smooth 4-manifolds which appear to be extremely powerful in distinguishing different differentiable structures. These invariants are defined in the following context. Let Z be an oriented simply connected differentiable 4-manifold and let b_2^+ and b_2^- be the number of + and − terms in a diagonalization of the quadratic (intersection) form on $H_2(Z)$. We assume b_2^+ odd and > 1. Note that, for a complex algebraic surface, we have the theorem of Hodge:

$$b_2^+ = 1 + 2p_g,$$

where p_g is the geometric genus (number of independent holomorphic 2-forms). Thus b_2^+ is odd and > 1 when $p_g \neq 0$.

The Donaldson invariants are a sequence of integer polynomials ϕ_k on $H_2(Z)$. Here $k > k_0$ and the degree of ϕ_k is

$$d(k) = 4k - 3(b_2^+ + 1)/2. \tag{6.1}$$

Two key results of Donaldson [5] indicate the power of these invariants:

THEOREM 1. *If* $Z = Z_1 \# Z_2$ *is a connected sum with* $b_2^+(Z_i) \neq 0$ *for* $i = 1, 2$ *then* $\phi_k(Z) \equiv 0$ *for all* k.

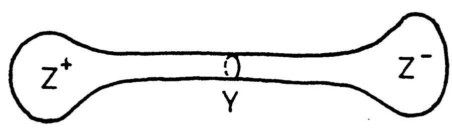
THEOREM 2. *If* Z *is algebraic, then for* $k > k_1(Z)$, $\phi_k(Z) \not\equiv 0$.

These theorems together show that, viewed as smooth manifolds, algebraic surfaces are essentially indecomposable. Note however that blowing-up points always leads to a decomposition in which one factor has $b_2^+ = 0$, $b_2^- = 1$.

Donaldson's invariants are defined using instantons and, in general, are impossible to compute directly. However, for algebraic surfaces another theorem of Donaldson [4] implies that his invariants can be calculated algebraically and this in particular leads to Theorem 2.

For nonalgebraic indecomposable 4-manifolds, the question of computing the Donaldson invariants is therefore an interesting and important one. In particular suppose the quadratic form A of Z is a direct sum $A = A_1 \oplus A_2$ with $b_2^+(A_i) \neq 0$. If $\phi_k(Z) \not\equiv 0$ for all k then from Theorem 1 we know that we cannot decompose Z as a connected sum with Z_i having quadratic form A_i. However, it is known [9] that one can always decompose Z along a *homology* 3*-sphere* Y, inducing the algebraic decomposition $A = A_1 \oplus A_2$ on homology.

(6.2)



This shows that the indecomposability of Z as a usual connected sum (along a genuine 3-sphere), measured by the nonvanishing of the Donaldson invariants, is somehow reflected in the nontriviality of the homology 3-sphere Y.

My aim is now to explain (in the next section) how Donaldson relates his invariants for Z to the Floer homology groups of Y. For this we shall first have to recall briefly the definition of the Donaldson invariants ϕ_k.

We fix a Riemannian metric on Z, a positive integer k and look at the moduli space $M_k(Z)$ of k-instantons on Z, i.e., solutions of the anti-self-duality equations $*F = -F$ with $c_2 = k$. For generic metrics this is a manifold of dimension $2d(k)$, where $d(k)$ is defined by (6.1). If $d(k) = 0$ then M_k is a finite set of points and, when suitably counted, this is Donaldson's invariant. For $d(k) > 0$ we fix $d(k)$ spherical cycles $\alpha_i \colon S^2 \to Z$ with homology classes $[\alpha_1], \ldots, [\alpha_d]$. Each such cycle defines a codimension 2 submanifold A_i of M_k, consisting of connections on Z which pull back via α_i to *special* connections on S^2. A special connection is one which defines a *nontrivial* holomorphic bundle (note: in dimension 2 a unitary connection defines a holomorphic structure). We now consider the intersection number

$$A_1 \cap A_2 \cap \cdots \cap A_d$$

as a function of $\alpha_1, \ldots, \alpha_d$. It depends only on the classes $[\alpha_1], \ldots, [\alpha_d]$ and its values defines $\phi_k([\alpha_1], \ldots, [\alpha_d])$ as a symmetric d-linear function on $H_2(Z)$.

Because M_k is not compact, care has to be taken "at ∞" and this is where the restriction $k > k_0$ enters. Essentially M_k can be compactified and $k > k_0$ ensures that ∂M_k has codimension ≥ 2, which is enough to define intersection numbers (via the fundamental class of M_k).

7. Relation to Floer homology. I now consider a decomposition of the 4-manifold Z along a homology 3-sphere Y as in Figure (6.2). The idea is to study the instanton equations on Z by considering them separately on Z^+ and Z^- and then matching boundary values along Y.

A much simpler prototype problem may help to illustrate the ideas. Consider the 2-sphere S^2 cut in half along the equator. To construct holomorphic functions on S^2 we consider holomorphic functions in each hemisphere and compare their boundary values. This gives the usual Hardy spaces of Fourier series

$$H^+ = \left\{ \sum_{n \geq 0} a_n z^n \right\}, \qquad H^- = \left\{ \sum_{n \leq 0} a_n z^n \right\},$$

and the intersection $H^+ \cap H^-$ gives of course just the constant functions (the only global holomorphic functions on S^2).

This example was both lower-dimensional and linear. We can keep to dimension 2 but make the problem *nonlinear* by looking at the construction of holomorphic maps

$$S^2 \to P,$$

where P is some complex manifold (e.g. projective space). Again, cutting S^2 into half along the equator, a holomorphic map in either hemisphere is determined by its restriction to S^1 which is a point of the free loop space LP. We thus get H^+ and H^- subspaces of LP, but this time these are not linear. Nevertheless the global holomorphic maps are still given by $H^+ \cap H^-$. If we linearize this problem we recover our earlier one, so that $H^{\pm}$ should be seen as infinite-dimensional manifolds of approximately "half" the dimension of LP.

After this 2-dimensional digression let us return to the 4-dimensional situation. Since the instanton equations (4.1) are (like Cauchy-Riemann) of first order, a solution is determined by the appropriate boundary data, which in this case is just a connection on Y (up to equivalence). Thus we should look in the space $\mathscr{C}(Y)$ at the two spaces $\Sigma^{\pm}$ consisting of boundary values of solutions of the instanton equations in $Z^{\pm}$ respectively. Their intersection gives global solutions on Z.

Suppose for simplicity that $d(k) = 0$, so that the Donaldson invariant is just an integer and describes simply the algebraic number of k-instantons on Z. This number should then be computed as an intersection number of Σ^+ and Σ^- in

$\mathscr{C}(Y)$. For this we need to have an appropriate homology theory for $\mathscr{C}(Y)$ in which Σ^+ and Σ^- represent cycles. The Floer homology groups $HF^+(Y)$ and $HF^-(Y)$ provide just such a framework, as I shall try to explain.

Let us revert briefly to the finite-dimensional Morse theory. There if we have a geometric cycle α and we want to associate to it a cycle in Witten's complex, we push it along the gradient flow and see which critical points it "hangs" on. If β is a cycle of complementary dimension and we want to compute the intersection number $\alpha \cdot \beta$ we should deform β along the *ascending* gradient flow of f and find which critical points it "hangs" on. The intersection number $\alpha \cdot \beta$ is now reduced to local calculations near the critical points.

For our infinite-dimensional manifold $\mathscr{C}$ the gradient flows are only defined for appropriately restricted initial data, since we have to solve the heat equation for an operator which is bounded neither above nor below. However, the cycles Σ^+ and Σ^- provide suitable data for the two opposite flows. In this way Donaldson assigns classes $[\Sigma^+] \in HF^+(Y)$ and $[\Sigma^-] \in HF^-(Y)$ whose pairing, under the natural (Poincaré duality) map $HF^+(Y) \otimes HF^-(Y) \to \mathbf{Z}$, gives the Donaldson invariant $\phi_k(Z)$.

The general case when $d(k) > 0$ is essentially similar. Given classes $[\alpha_i] \in H_2(Z)$ for $i = 1, 2, \dots, d$, we choose an integer r with $0 \le r \le d$, and representative spherical cycles

$$\begin{aligned} \alpha_i &: S^2 \to Z^+, \qquad i = 1, \dots, r, \\ \alpha_j &: S^2 \to Z^-, \qquad j = r+1, \dots, d. \end{aligned}$$

The boundary values of instantons on Z^+ which are special on $\alpha_1, \dots, \alpha_r$ define a "cycle" $\Sigma^+(\alpha_1, \dots, \alpha_r) \subset \mathscr{C}$ and hence a Floer homology class $[\Sigma^+(\alpha_1, \dots, \alpha_r)] \in HF^+(Y)$. Similarly we define $[\Sigma^-(\alpha_{r+1}, \dots, \alpha_d)] \in HF^-(Y)$. Pairing these together then defines a homomorphism of symmetric products:

$$S^r(H_2(Z^+)) \otimes S^{d-r}(H_2(Z^-)) \to \mathbf{Z}. \tag{7.1}$$

Since

$$\begin{aligned} H_2(Z) &= H_2(Z^+) \oplus H_2(Z^-), \\ S^d(Z) &= \sum_{r=0}^{d} S^r(H_2(Z^+)) \otimes S^{d-r}(H_2(Z^-)), \end{aligned}$$

summing (7.1) over r gives a homomorphism

$$S^d(H_2(Z)) \to \mathbf{Z}$$

and Donaldson proves that this is his invariant ϕ_k.

REMARKS. (1) Formalizing the construction described above, we see that, when $Y = \partial Z^+$ with Y a homology 3-sphere, there is a sequence of polynomials ϕ_r on $H_2(Z^+)$ with values in $HF^+(Y)$. These may be viewed as a generalization, or a relative version, of the Donaldson polynomials for closed manifolds.

(2) When Y is the standard 3-sphere so that Z is just the usual connected sum, then $HF^+(Y) = 0$ and Donaldson's Theorem 1 follows as immediate corollary

of the general procedure for computing $\phi_k(Z)$ just outlined. Conversely, if we know that Z is indecomposable, e.g. if Z is an algebraic surface, it follows that $HF^+(Y) \neq 0$. Thus Donaldson's Theorem 2 implies the nontriviality of the Floer homology groups for many homology 3-spheres Σ, namely those occurring in a pseudodecomposition of algebraic surfaces.

8. Concluding remarks. Let me return now to the Floer homology groups $HF^{\pm}(Y)$ of a homology 3-sphere Y, and make some general heuristic comments.

For a finite-dimensional manifold there are various ways of defining the homology groups depending on the choice of chain complex. In particular we have:

(1) the Witten complex of a Morse function,

(2) the classical chain complex of geometrically defined cycles (of various kinds),

(3) the de Rham complex,

(4) the Hodge theory.

Witten's argument using Δ_{tf} with $t \to \infty$ relates (1) and (4), while pushing along the gradient flow of f relates (1) and (2). The relation of (3) to (4) is by standard elliptic theory. Definition (1) is the most "finite", using least analysis, while (4) uses the most.

For our infinite-dimensional manifold $\mathscr{C}$ of connections on Y (or better, for the infinite cyclic covering $\mathscr{C}_0$) the Floer definition follows (1), using the Chern-Simons function. Donaldson's work essentially uses (2), since the "cycles" $\Sigma^{\pm} \subset \mathscr{C}$ occur as boundary values of instantons.

Both the Floer and Donaldson approaches show clearly that the homology in question occurs in the "middle dimensions" of the infinite-dimensional manifold $\mathscr{C}$. We could attempt to give a rigorous definition of such middle-dimensional cycles (at least if they are smooth) on the following lines. As already pointed out, a metric on Y enables us to define a continuous family of self-adjoint operators H_C acting on the tangent space to $\mathscr{C}$ at the point C. These are of Dirac type and their spectral decomposition defines subspaces T^+ and T^- of the tangent bundle of $\mathscr{C}$. These are not quite continuous because there are finite jumps in dimension whenever C is in the exceptional set $\mathscr{C}'$ where H_C has a zero eigenvalue. Now given a submanifold $K \subset \mathscr{C}$ we can define it to be a *positive* cycle if, for each $C \in \mathscr{C}$, the tangent space TK is close, in an appropriate sense, to T^+. Close should mean that the projection $TK \to T^+$ is *Fredholm* while the projection $TK \to T^-$ is *compact.* Modulo an appropriate equivalence relation, these should define the *positive* homology groups. Similarly, replacing T^+ by T^-, we would have negative cycles leading to *negative* homology. The index of the Fredholm operator $TK \to T^+$ of a positive cycle would define its (renormalized) dimension, an integer (+ or −). The jumps in T^+ mean that this dimension would depend on the component of $\mathscr{C} - \mathscr{C}'$ in which one was working. The Donaldson cycles $\Sigma^{\pm}$ should fit into this framework.

A more elementary and obvious case when such "middle-dimensional" homology groups enter is for a product $\mathscr{S}^+ \times \mathscr{S}^-$. Each factor has ordinary homology (represented by finite-dimensional cycles) and dually ordinary cohomology represented by finite-codimensional cycles. In the product we then have four types of cycle:

$$\text{finite} \times \text{finite}, \qquad \text{cofinite} \times \text{cofinite},$$
$$\text{cofinite} \times \text{finite}, \qquad \text{finite} \times \text{cofinite}.$$

The first two give ordinary homology and cohomology respectively. The other two are quite different from these and give in an obvious sense "middle-dimensional" homology, one "positive" and one "negative".

Our space $\mathscr{C}$ is not globally a product but only infinitesimally (because its tangent bundle decomposes). Thus its middle-dimensional homology cannot be reduced to ordinary homology and cohomology by a factorization. Thus the Floer homology groups are, from a topological point of view, something essentially new.

Let us now go on to consider the appropriate de Rham theory for $\mathscr{C}$. We clearly want a de Rham complex Ω^+ on the following lines. If e_n ($n \in \mathbf{Z}$) is an orthonormal base for T_C given by the spectrum of H_C, so that T_C^+ is spanned by $n \geq 0$, consider the "volume element" of T_C^+:

$$\omega = e_0 \wedge e_1 \wedge \cdots$$

and also those infinite wedge products which differ from ω by only finitely many terms. Dualizing and taking linear combinations should define the "positive" differential forms at C. Similarly there would be Ω^- and these should lead to the correct de Rham theories.

The semi-infinite volume element ω is of course familiar to all physicists as the vacuum vector of a Fermionic Fock space. Differential forms in Ω^+ are therefore fields of a Fermionic quantum field theory. If we go further and introduce the Laplacian Δ_f^+ (where f is the Chern-Simons function) then this should be the Hamiltonian of the quantum field theory, and the harmonic forms are therefore the ground states of the QFT. Thus, purely formally (and ignoring the trivial ground state), we conclude that the *Floer homology groups $HF^+(Y)$ are the ground states of the QFT with Hamiltonian Δ_f^+.*

All of this discussion is essentially implicit in Witten's original paper [**14**]. Certainly Witten was well aware that QFT was concerned with middle-dimensional homology. Of course our discussion here is completely formal since QFT in dimension $3 + 1$ does not yet exist as a proper mathematical theory and in particular there is no rigorous definition of the Hamiltonian Δ_f^+. However, if we considered instead Floer's symplectic theory for paths in a symplectic manifold then QFT in $1+1$ dimensions is in much better shape and is intensively investigated in connection with strings. The connection between Floer homology and QFT should therefore be taken more seriously in this context.

Finally let me list a few of the major problems that are still outstanding in this area.

(1) Prove that the two definitions of $HF(Y)$ agree (one using $\mathscr{C}(Y)$ and the other via a Heegaard splitting). NOTE. In view of the physical observations

above, this problem suggests a comparison between two quantum field theories, one in $1+1$ dimensions and the other in $3+1$ dimensions.

(2) Find an algorithm to compute $HF(Y)$ which generalizes Casson's algorithm for his invariant.

(3) Find a method to compute the Donaldson invariants $S^*(H_2(Z^+)) \to HF^+(Y)$ when $Y = \partial Z^+$.

More speculatively, I would like to end with

(4) Find a connection with the link invariants of Vaughan Jones [**11**].

As circumstantial evidence that this is reasonable I will list some properties shared by Floer homology and the Jones polynomial.

(i) Both are subtle 3-dimensional invariants.

(ii) They are sensitive to orientation of 3-space (unlike the Alexander polynomial).

(iii) They depend on Lie groups: SU(2) in the first instance but capable of generalization.

(iv) There are 2-dimensional schemes for computing these 3-dimensional invariants.

(v) Whereas the variable in the Alexander polynomial corresponds to $\pi_1(S^1)$, the variable in the Jones polynomial appears to be related to $\pi_3(S^3)$, the origin of "instanton numbers".

(vi) Both have deep connections with physics, specifically quantum field theory (and statistical mechanics).

References

1. M. F. Atiyah and R. Bott, *The Yang-Mills equations on Riemann surfaces*, Philos. Trans. Roy. Soc. London A **308** (1982), 523–615.
2. M. F. Atiyah, V. K. Patodi and I. M. Singer, *Spectral asymmetry and Riemannian geometry*. III, Math. Proc. Cambridge Philos. Soc. **79** (1976), 71–99.
3. A. Casson, *An invariant for homology* 3-*spheres*, Lectures at MSRI, Berkely, 1985.
4. S. K. Donaldson, *Anti-self-dual Yang-Mills connections over complex algebraic surfaces and stable vector bundles*, Proc. London Math. Soc. **50** (1985), 1–26.
5. ——, *Geometry of* 4-*manifolds*, Proc. Internat. Congress of Mathematicians (Berkeley, 1986), Amer. Math. Soc., Providence, R.I., 1987.
6. A. Floer, *A relative Morse index for the symplectic action*, Courant Institute preprint, 1987.
7. ——, *Morse theory for Lagrangian intersections*, Courant Institute preprint, 1987.
8. ——, in preparation.
9. M. Freedman and L. Taylor, Λ-*splitting* 4-*manifolds*, Topology **16** (1977), 181–184.
10. M. Gromov, *Pseudoholomorphic curves in symplectic manifolds*, Invent. Math. **82** (1985), 307–347.
11. V. F. R. Jones, *A polynomial invariant for knots via von Neumann algebras*, Bull. Amer. Soc. (N.S.) **12** (1985), 103–111.
12. M. S. Narasimhan and C. S. Seshadri, *Stable and unitary vector bundles on a compact Riemann surface*, Ann. of Math. (2) **82** (1965), 540–567.
13. P. E. Newstead, *Topological properties of some spaces of stable bundles*, Topology **6** (1967), 241–262.
14. E. Witten, *Supersymmetry and Morse theory*, J. Differential Geom. **17** (1982), 661–692.

Mathematical Institute, Oxford University, England