RESEARCH REPORTS

THE SEIBERG-WITTEN EQUATIONS
AND 4-MANIFOLD TOPOLOGY

S. K. DONALDSON

Since 1982 the use of gauge theory, in the shape of the Yang-Mills instanton equations, has permeated research in 4-manifold topology. At first this use of differential geometry and differential equations had an unexpected and unorthodox flavour, but over the years the ideas have become more familiar; a body of techniques has built up through the efforts of many mathematicians, producing results which have uncovered some of the mysteries of 4-manifold theory, and leading to substantial internal conundrums within the field itself. In the last three months of 1994 a remarkable thing happened: this research area was turned on its head by the introduction of a new kind of differential-geometric equation by Seiberg and Witten: in the space of a few weeks long-standing problems were solved, new and unexpected results were found, along with simpler new proofs of existing ones, and new vistas for research opened up. This article is a report on some of these developments, which are due to various mathematicians, notably Kronheimer, Mrowka, Morgan , Stern and Taubes, building on the seminal work of Seiberg [S] and Seiberg and Witten [SW]. It is written as an attempt to take stock of the progress stemming from this initial period of intense activity. The time period being comparatively short, it is hard to give complete references for some of the new material, and perhaps also to attribute some of the advances precisely. The author is grateful to a number of mathematicians, but most particularly to Peter Kronheimer, for explaining these new developments as they unfolded.

1. The Seiberg-Witten equations

The equations introduced by Witten in [W3], following up work of Seiberg and Witten [SW], involve two entities, a $U(1)$ connection and a "spinor" field. Thus a main pre-requisite for their study is a knowledge of spinors on 4-manifolds. More precisely, the most relevant notion is that of a $Spin^c$ structure. Recall that if X is an oriented Riemannian 4-manifold, a spin structure on X is a lift of the structure group of the tangent bundle from $SO(4)$ to its double cover $Spin(4)$. The exceptional isomorphism $Spin(4) = SU(2) \times SU(2)$ means that this can be given a more concrete description in terms of vector bundles. Giving a spin structure

is the same as giving a pair of complex 2-plane bundles $S^+, S^- \to X$, each with structure group $SU(2)$ and related to the tangent bundle by a structure map $c : TX \to Hom(S^+, S^-)$. The basic algebraic model is

$$(x_0, x_1, x_2, x_3) \mapsto \begin{pmatrix} x_0 + ix_1 & x_2 + ix_3 \\ -x_2 + ix_3 & x_0 - ix_1 \end{pmatrix}.$$

Now the map $e \wedge f \mapsto c(e)^* c(f) - c(f)^* c(e)$ induces a map ρ from the self-dual 2-forms Λ^+ to $Hom(S^+, S^+)$, which corresponds to the standard isomorphism between the Lie algebras of $SU(2)$ and $SO(3)$.

The map c is the symbol of the Dirac operator $D : \Gamma(S^+) \to \Gamma(S^-)$, and one of the most fruitful calculations in differential geometry leads to the Lichnerowicz-Weitzenbock formula for the Dirac operator:

$$(1) \qquad\qquad\qquad D^* D\psi = \nabla^* \nabla \psi + \tfrac{1}{4} R\psi.$$

Here ∇ is the covariant derivative on spinors, induced by the Levi- Civita connection, and R is the scalar curvature, which acts in equation (1) by scalar multiplication at each point. If we have an additional auxiliary bundle $E \to X$, with a Hermitian metric and connection, we may consider spinors with values in E— sections of $S^\pm \otimes E$. The Dirac operator on these coupled spinors satisfies

$$D^* D\psi = \nabla^* \nabla \psi + \tfrac{1}{4} R\psi - F_E^+(\psi)$$

where F_E^+ is the self-dual part $\tfrac{1}{2}(F_E + *F_E)$ of the curvature of E. Here the self-dual forms act on spinors in the way described above. Now a spin structure may not exist globally—the Stiefel-Whitney class $w_2(X) \in H^2(X; \mathbf{Z}/2)$ is the obstruction—but a variant, a $Spin^c$ structure, always does. A $Spin^c$ structure is given by a pair of vector bundles $W^\pm$ over X with an isomorphism $\Lambda^2 W^+ = \Lambda^2 W^- = L$ say, such that locally $W^\pm = S^\pm \otimes L^{1/2}$, where $L^{1/2}$ is a local square root of L: $L^{1/2} \otimes L^{1/2} = L$. An old result of Hirzebruch and Hopf assures the existence of $Spin^c$ structures on any oriented 4-manifold; up to an action of the finite group $H^1(X; \mathbf{Z}/2)$ they are classified by the lifts of $w_2(X)$ to $H^2(X; \mathbf{Z})$, the first Chern class of the line bundle L. A connection on L gives a Dirac operator $D : \Gamma(W^+) \to \Gamma(W^-)$, which is locally just the same as the Dirac operator on $L^{1/2}$-valued spinors. In particular we get the Lichnerowicz formula:

$$(2) \qquad\qquad D^* D\psi = \nabla^* \nabla \psi + \tfrac{1}{4} R\psi - \tfrac{1}{2} F_L^+(\psi),$$

where the factor of $1/2$ comes from the square root of L. Note that $Hom(W^+, W^+) \cong Hom(S^+, S^+)$.

We now come to the Seiberg- Witten equations for a 4-manifold X with $Spin^c$ structure $W^\pm$. These are equations for a pair (A, ψ) where

   1. A is a unitary connection on $L = \Lambda^2 W^\pm$,

   2. ψ is a section of W^+.

If ξ and η are in W^+, we write $\xi\eta^*$ for the endomorphism $\theta \mapsto \langle\theta, \eta\rangle\xi$ of W^+. The trace-free part of this endomorphism lies in the image of the map ρ, and we write $\tau(\xi, \eta)$ for the corresponding element of $\Lambda^+ \otimes \mathbf{C}$. So τ is a sesquilinear map $\tau : W^+ \times W^+ \to \Lambda^+ \otimes \mathbf{C}$. The Seiberg-Witten equations are:

$$D_A \psi = 0,$$
$$F_A^+ = -\tau(\psi, \psi).$$

The sign of the quadratic term $\tau(\psi, \psi)$ is crucial. One sees this in a calculation, using the Lichnerowicz formula, which, in various guises, underpins the whole theory. Suppose (A, ψ) is a solution of the Seiberg- Witten equations over a compact 4-manifold X. We have

$$0 = D_A^* D_A \psi = \nabla_A^* \nabla_A \psi + \tfrac{1}{2} F_A^+(\psi) + \tfrac{1}{4} R\psi.$$

Taking the L^2 inner product with ψ, we get

$$\int_X |\nabla_A \psi|^2 + \tfrac{1}{2}(F_A^+(\psi), \psi) + \tfrac{1}{4}R|\psi|^2 d\mu = 0.$$

Now using the second equation, $(F_A^+(\psi), \psi) = -(\tau(\psi, \psi)(\psi), \psi)$, and one readily sees that $\tau(\psi, \psi)(\psi) = \tfrac{1}{2}|\psi|^2\psi$. This just corresponds to the matrix calculation, where ψ is represented as the column vector $\begin{pmatrix} a \\ b \end{pmatrix}$:

$$\begin{pmatrix} \tfrac{1}{2}(|a|^2 - |b|^2) & a\bar{b} \\ b\bar{a} & \tfrac{1}{2}(|b|^2 - |a|^2) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \tfrac{1}{2}(|a|^2 + |b|^2) \begin{pmatrix} a \\ b \end{pmatrix}.$$

So the term $(F_A^+(\psi), \psi)$ is $-\tfrac{1}{2}|\psi|^4$, and we have

$$\int_X |\nabla_A \psi|^2 + \tfrac{1}{4}|\psi|^4 d\mu = -\int_X \tfrac{1}{4}R|\psi|^2 d\mu.$$

One immediate consequence of this calculation occurs if the scalar curvature R of the base manifold is everywhere non-negative. Then, considering the sign of the terms in the equation, we deduce that the only solutions to the Seiberg- Witten equation occur when $\psi = 0$ and $F_A^+ = 0$, i.e. from the solutions of the $U(1)$ instanton equation. This application of the Lichnerowicz formula is, of course, highly reminiscent of the way that the Dirac equation has been applied in Riemannian geometry by Lichnerowicz, Hitchin, Gromov and Lawson and others (cf. [GL]), but with the additional ingredient of the quadratic term coupling the connection and the spinor. Without any assumptions on the scalar curvature, we can still obtain some information. Let $-c \leq 0$ be the minimum of the scalar curvature R over X. Then

$$\int_X |\nabla_A \psi|^2 + \tfrac{1}{4}|\psi|^4 d\mu \leq \tfrac{1}{4}c \int_X |\psi|^2 d\mu \leq \tfrac{1}{4}c \mathrm{Vol}(X)^{1/2} \left(\int_X |\psi|^4 d\mu \right)^{1/2},$$

using the Cauchy-Schwarz inequality. In particular the integral of $|\psi|^4$ is bounded by $+c^2 \mathrm{Vol}(X)$, and substituting into the equations we get a bound on the L^2 norm of F_A^+ in terms of the geometry of the base manifold. We may compare this with the topological invariant $c_1(L)^2 \in H^4(X; \mathbf{Z}) = \mathbf{Z}$ which is represented via the Chern-Weil theory as

$$c_1(L)^2 = \frac{1}{4\pi^2} \int_X F_A \wedge F_A = \frac{1}{4\pi^2} \int_X |F_A^+|^2 - |F_A^-|^2 d\mu,$$

so, if a solution exists, we get an upper bound on $c_1(L)^2$.

A similar calculation to the one above gives a variational description of the Seiberg-Witten equations. For any pair (A, ψ) consider the integral

$$(3) \qquad E(A, \psi) = \int_X |\nabla_A \psi|^2 + \tfrac{1}{2}|F_A|^2 + \tfrac{1}{8}(|\psi|^2 + R)^2 d\mu.$$

The coupling terms that appear here are of the standard kind considered in the Mathematical Physics literature [JT]. If one uses the Weitzenbock formula to express the term involving $\nabla_A \psi$ in terms of $D_A \psi$, together with the integral representation of $c_1(L)$, one can write:

$$(4) \qquad E(A, \psi) = \int_X |D_A \psi|^2 + |F_A^+ + \tau(\psi, \psi)|^2 d\mu + E_0,$$

where E_0 depends only on X and L, in fact

$$E_0 = \int_X \frac{R^2}{8} d\mu + 2\pi^2 c_1(L)^2.$$

So we see that the Seiberg-Witten solutions are the absolute minima of the functional E, on the given bundle, much as the Yang-Mills instantons minimise the ordinary Yang-Mills functional.

2. Seiberg-Witten invariants

We begin by recalling some general points about "topological invariants" defined by solutions of partial differential equations. In differential topology one is familar with many contexts in which the solutions of an "equation" $f(x) = y$ are, at the level of homology, unchanged by continuous variations of parameters. For example f might be a map $f : P \to Q$ between compact oriented manifolds, then the homology class in $H_*(P)$ of $f^{-1}(y)$, for generic y in Q, is a homotopy invariant of f—just the Poincaré dual of the pull-back of the fundamental cohomology class of Q. Or f might be a section of an oriented vector bundle $V \to P$, and $y = 0$, so the solutions are the zero set of the section which, assuming transversality, give a submanifold representing the Poincaré dual of the Euler class of V. Now if we have a family of partial differential equations, depending on continuous parameters, we may hope to find similar invariants from the homology class of the solution space. This can be developed abstractly in the framework of differential topology in Banach manifolds. In any case, the main ideas are now well-known. The key points one needs to establish in order to find invariants analogous to the finite dimensional case are:

1. The maps involved should be *Fredholm* maps, which in practise means that the linearisation of the equations about a solution should be represented by linear *elliptic* differential equations, say over a compact manifold. The index of the linearised equation gives the "expected dimension" of the solutions space. One needs to see that after suitable generic perturbations of the equations the solutions spaces do have this dimension.

2. One needs to establish the *compactness* of the space of solutions, or some weaker analogue of this.

3. One needs to establish orientability, analogous to the finite dimensional case; otherwise one only gets invariants modulo 2. This can be set up in terms of the index theory of families of operators.

In the cases arising from gauge theory—with P.D.E.'s for a connection and other structures on a bundle $P \to X$—there is another general feature. The equations are invariant under the action of the gauge group of bundle automorphisms, and one studies spaces of solutions modulo this action. Difficulties can occur because of *reducible solutions*, with a non-trivial stabiliser in the gauge group, and

4. one must not encounter reducible solutions in generic 1-parameter families of equations.

The 4-manifold invariants obtained from the instanton moduli spaces fit into this framework. One studies the instanton equation $F_A^+ = 0$ for a connection on, say, an $SU(2)$ bundle $E \to X$ with Chern class $k > 0$. The linearisation of the equation, modulo the action of the gauge group, is represented by the operator

$$d_A^* + d_A^+ : \Omega^1(\text{ad } E) \to \Omega^0(\text{ad } E) \oplus \Omega_+^2(\text{ad } E),$$

which is elliptic with index $8k - 3(1 - b_1 + b_+)$, where b^+ is the dimension of a positive subspace for the intersection form (the dimension of the space H^+ of self-dual harmonic forms). For generic metrics on X the moduli space M_k of solutions modulo gauge equivalence is a manifold of this dimension, away from reducible solutions. If $b^+ > 0$, there are no reducible solutions for generic metrics, and if $b^+ > 1$, there are none in generic 1-parameter families. So we suppose that $b^+ > 1$. The moduli spaces can be oriented by a choice of orientation of the line:

(5) $$\det(H^1(X)) \otimes \det(H^+(X)),$$

where $\det(\ )$ denotes the highest exterior power. The moduli spaces are not compact—a phenomenon related to the conformal invariance of the equation— but there is a well-understood compactification which allows one to define an appropriate fundamental class $[M_k]$ which pairs with suitable cohomology classes. These cohomology classes are given by maps:

$$\mu_i : H_i(X) \to H^{4-i}(M_k).$$

Let us assume that $b^1(X) - b^+(X)$ is *odd*. Then the moduli spaces are even-dimensional and yield invariants

(6) $$q_{p,r}(\alpha) = < \mu_2(\alpha)^p u^r, [M_k] >,$$

where $\alpha \in H_2(X)$, $u = \mu_0(\text{point}) \in H^4(M_k)$, and we arrange the indexing so that $2p + 4r = 2d$ is the dimension of the moduli space. These are a collection of polynomial functions on the homology group $H^2(X; \mathbf{R})$. They are defined using a choice of Riemannian metric on X, but the essential point is that in the end one gets differential topological invariants, independent of metrics. We refer to the literature for further details, including some technicalities we have glossed over in the sketch above [DK].

With this background in place, we return to the Seiberg-Witten equations. It is not hard to check the essential points listed above, and so to see that these also define differential-topological invariants of the underlying 4-manifold. Indeed, the theory is significantly simpler than for the instanton equations. To check the Fredholm property we can ignore the quadratic term $\tau(\psi, \psi)$, since this does not

affect the symbol (leading term) of the linearisation. At the level of the symbol, the linearisation is given by the sum of the linearisation of the $U(1)$ instanton equation, which modulo gauge is represented by the operator $d^* + d^+$ acting on ordinary forms, and the Dirac operator D_A. Each of these is elliptic and the index of the problem is

$$i(L) = \mathrm{ind}(d^* \oplus d^+) + \mathrm{ind}_{\mathbf{R}}(D_A) = (b^1 - 1 - b^+) + \tfrac{1}{4}(c_1(L)^2 - \tau(X))$$

(7) $$\qquad = \tfrac{1}{4}(c_1(L)^2 - (2\chi(X) + 3\tau(X))),$$

where the calculation for $\mathrm{ind}(D_A)$ uses a simple case of the Atiyah-Singer Index Theorem (χ, τ denote the Euler characteristic and signature). The number $i(L)$ is the expected dimension of the Seiberg-Witten solution space. There is a straight-forward class of perturbations of the Seiberg-Witten equations: for any self-dual 2-form θ on X we replace the second equation by

$$F^+(A) = \tau(\psi, \psi) + \theta.$$

Then one can find arbitrarily small θ such that the moduli space of solutions to the perturbed equation is a manifold of dimension $i(L)$, away from reducibles [KM3],[W3]. (The discussion here is much simpler than for the $SU(2)$ instanton equations, because the curvature of a line bundle is an ordinary 2-form, rather than a bundle-valued 2-form.)

The next issue is compactness. Unlike the instanton case, the Seiberg-Witten moduli spaces are compact, without qualification. This follows from *a priori* estimates on the solutions. These can be obtained from energy estimates, using integration by parts as in the previous section, or, more directly by the maximum principle applied to second-order equations [KM3],[W3]. If (A, ψ) is a Seiberg-Witten solution then, as we have seen:

$$\nabla_A^* \nabla_A \psi = -\tau(\psi, \psi)\psi - R\psi.$$

This implies that

$$2\Delta(|\psi|^2) \leq c|\psi|^2 - |\psi|^4,$$

where $-c$ is the lower bound of the scalar curvature as before. At a point where $|\psi|$ is maximal, $\Delta|\psi|^2 \geq 0$, so we get

$$\max|\psi| \leq \sqrt{c}.$$

Substituting back into the equations we get an L^∞ bound on $F^+(A)$, and elliptic theory gives an L^p bound on the full curvature $F(A)$, for all p. Compactness follows easily from this. (By contrast, there is an *a priori* bound on the L^2 norm of the curvature of an instanton, but in general not on the L^p norm for $p > 2$.)

The remaining issues are reducibles and orientations. If a non-trivial gauge transformation $g \in \mathrm{Aut}(L)$ fixes a pair (A, ψ), then ψ must be zero, and $g \in U(1)$ a constant scalar. Thus the only reducible Seiberg-Witten solutions are the self-dual $U(1)$ connections, and as we have mentioned these do not occur in generic r-dimensional families of metrics on X, so long as $b^+(X) > r$. Thus if $b^+ > 1$, reducibles do not interfere with the definition of invariants. Considering orientations:

an orientation of the moduli space is furnished by an orientation of the "determinant line" of the relevant index bundle over the space $\mathcal{C}^*$ of all irreducible pairs (A, ψ) modulo gauge transformation. As in the calculation of the index $i(L)$, this line splits into the tensor product of a piece coming from the $d^* + d^+$-operator on ordinary forms and a piece from the Dirac operator. The first piece is independent of the connection and is oriented by an orientation of the determinant line

$$\det \operatorname{ind}(d^* + d^+) = \det H^1 \otimes \det H^+,$$

and the second carries a canonical orientation coming from the fact that the spinors are complex vector bundles and the Dirac operator is complex-linear. Thus the data required to orient the Seiberg-Witten moduli spaces is the same as for the instanton spaces.

To sum up, if $b^+ > 1$ and we fix an orientation of the line (5), the homology class $[M_{SW}]$ of the Seiberg-Witten moduli space (perhaps after a small perturbation of the equations) gives an element of $H_i(\mathcal{C}^*)$, where i is given by (7), independent of the metric on X. Now there is a natural class $h \in H^2(\mathcal{C}^*)$: if x_0 is a base-point in X, evaluation at X_0 gives a homomorphism $\operatorname{Aut}(L) \to U(1)$, which induces a principle $U(1)$ bundle over the quotient space $\mathcal{C}^* = \{(A, \psi) : \psi \neq 0\}/\operatorname{Aut}(L)$. The class h is the first Chern class of this $U(1)$ bundle. (If $H^1(X) \neq 0$, there are other cohomology classes on $\mathcal{C}^*$ which one could also try to exploit.) Let us suppose that $b_1 - b^+$ is *odd*: this implies that $i(L)$ is even, $i(L) = 2s$ say, and finally we obtain a numerical invariant:

$$(8) \qquad\qquad n_L = <h^s, [M_{SW}]> .$$

Of course the line bundle L is determined by its first Chern class, so we can equally regard these as invariants of a 4-manifold X with a chosen class in $H^2(X; \mathbf{Z})$. If we replace the line bundle L by its inverse L^{-1}, the moduli space is essentially unchanged—one just reverses the complex structures on $W^\pm$—and this gives $n_{L^{-1}} = \pm n_L$.

There are only a finite number of line bundles L for which $n_L \neq 0$. Indeed for a fixed, generic, metric on X there are only a finite number of line bundles with non-empty moduli spaces. The integral formulae of the previous section give an upper bound on the self-dual part of the harmonic form representing $c_1(L)$. On the other hand the condition that the index $i(L)$ be positive gives a lower bound on $c_1(L)^2$, and hence an upper bound on the anti-self-dual part of the harmonic representative.

In the case when $b^+ = 1$ the picture is more complicated. Just as for the instanton theory, one gets invariants which depend upon a choice of "chamber" in the cohomology of X—a Riemannian metric determines a chamber via its self-dual harmonic form, and the "invariants" change as one passes between the chambers due to the occurrence of reducible solutions: despite this complication useful information can still be obtained as we shall describe in Section 6.

3. Background: basic classes and the main conjecture

We will now review some of the background which led Seiberg and Witten to introduce these invariants. Let α be a 2-dimensional homology class in a 4-manifold

X. One can always find a smooth embedded oriented surface Σ in X which represents α, and the intrinsic topology of such a surface is of course completely determined by its genus $g(\Sigma)$. By adding on small handles one can find representatives of arbitrarily large genus, but it is an interesting problem to find the *minimal* genus of a representative in a given homology class. A famous conjecture, frequently ascribed to Thom, was that the complex curves in $\mathbf{CP}^2$ minimised the genus in their homology classes. In a lecture at the International Congress in 1986 [K], Kronheimer set out a programme to attack these problems using gauge theory. More precisely he studied instanton solutions over $X \setminus \Sigma$ with a singularity along Σ. This programme was carried through in the following year by Kronheimer and Mrowka. The first result they obtained was that if $b^+(X) > 1$, if the instanton invariants of X are not all trivial, and if $\alpha.\alpha \geq 0$, then

$$(9) \qquad\qquad 2g(\Sigma) - 2 \geq \alpha.\alpha.$$

Recall here that if X is a complex surface and $\Sigma \subset X$ is a complex curve, then one has an adjunction formula:

$$(10) \qquad\qquad 2g(\Sigma) - 2 = \alpha.\alpha + K_X.\alpha,$$

where we use the notation K_X to denote both the canonical line bundle of a complex surface and its first Chern class $-c_1(TX)$. Comparing the formulae, this result of Kronheimer and Mrowka gave a proof of an analogue of the Thom conjecture for homology classes in a $K3$ surface, where $K_X = 0$. (It does not immediately give information about the original Thom problem, since $b^+(\mathbf{CP}^2) = 1$.)

In subsequent work, using these singular instantons, Kronheimer and Mrowka were led to a far-reaching structure theorem for the instanton invariants of simply-connected 4-manifolds. Recall that these are a collection of polynomial functions $q_{p,r}$ on $H_2(X)$. Kronheimer and Mrowka introduced the notion of a 4-manifold of *simple type*. By definition, this condition holds if the polynomials are related by

$$(11) \qquad\qquad q_{p,r+2} = 4q_{p,r}.$$

Effectively this means that there is little extra information captured by the 4-dimensional class u over the instanton moduli spaces. They showed that this condition holds for very many 4-manifolds (for example, all complex surfaces obtained as complete intersections in $\mathbf{CP}^n$, with equations of reasonably high degree), and there are no examples of simply-connected manifolds which are known not to be of simple type. If X has simple type, the instanton invariants can be encoded in a single formal power series:

$$\mathcal{D}(\alpha) = \sum_{p=0}^{\infty} \frac{1}{p!} \tilde{q}_p(\alpha),$$

where $\tilde{q}_p$ is the polynomial of degree p equal to $q_{p,0}$ if $p = b^+ + 1 \mod 2$, and to $\frac{1}{2} q_{p,1}$ if $p = b^+ \mod 2$. Kronheimer and Mrowka's structure theorem, for manifolds of simple type, asserts that there are a finite number of classes—called *basic* classes—$\kappa_i \in H^2(X)$, and co-efficients $a_i \in \mathbf{Q}$ such that:

$$(12) \qquad\qquad \mathcal{D}(\alpha) = e^{\alpha.\alpha/2} \sum_i a_i e^{\kappa_i.\alpha}.$$

Thus the information from the instanton invariants boils down to the basic classes in $H^2(X)$, and the co-efficients a_i. Moreover they proved that (9) could be refined to the lower bound:

$$(13) \qquad\qquad 2g(\Sigma) - 2 \geq \alpha.\alpha + \max_i |\kappa_i.\alpha|.$$

For many complex surfaces it could be shown that K_X is a basic class, so comparison with the adjunction formula yields further analogies of the Thom conjecture.

There have been a number of further developments along these lines. Fintushel and Stern derived the structure theorem for many 4-manifolds by more elementary, but indirect arguments, using "generalised connected sum decompositions" $X = X_1 \cup_Y X_2$ of a 4-manifold split into two by a 3-manifold Y. One basic question is to find formulae which relate invariants of a manifold X with those of the connected sum $X \sharp \overline{CP}^2$. A direct attack on this leads to difficult calculations involving the details of the compactification of the instanton moduli spaces. For manifolds of simple type Kronheimer and Mrowka used their structure theory to find a simple formula involving trigonometric functions. By an indirect argument, Fintushel and Stern showed [FS] that, even if X is not of simple type, there is an explicit formula involving elliptic functions. Further work of Kronheimer and Mrowka suggests a strategy for proving that any simply connected 4-manifold has "finite type". If T is the operator on doubly- indexed collections of numbers $(q_{p,r})$ which maps $(q_{p,r})$ to $q'_{p,r+2} = q_{p,r+2} - 2q_{p,r}$, then a 4-manifold has finite type if its instanton invariants are annihilated by T^n for some n: thus the manifold is of simple type if one can take $n = 1$. Kronheimer and Mrowka have found the outlines of a general structure theorem expressing the instanton invariants by a more complicated expression, involving elliptic functions, but still with the fundamental data comprising preferred cohomology classes in $H^2(X)$ and numerical co-efficients.

We now turn from this differential-geometric and topological work to Mathematical Physics and Quantum Field theory. Sadly, the author's lack of knowledge in these areas will make our account rather superficial. Soon after the introduction of the instanton invariants, Witten [W1] showed that they could be interpreted as functional integrals, integrating over the space of connections and certain auxiliary fields. In fact the invariants are the natural expectation values in a variant of "$N = 2$ Supersymmetric Yang-Mills theory". Atiyah and Jeffrey, and others, elucidated the formulae in Witten's paper, putting them into a general differential-geometric framework of a Chern-Weil theory for the Euler class of a bundle with a section [AJ]. On the one hand this Quantum Field Theory interpretation is subject to the fundamental difficulties involved in the rigorous mathematical definitions of functional integrals. On the other hand it was one of the examples which lead to the idea of a "Topological Quantum Field Theory", which Witten developed in another direction in his renowned work on knot and 3-manifold invariants. In 1993 Witten extended this Quantum Field Theory approach to 4-manifold invariants, beginning with the case of a $K3$ surface. These 4- manifolds admit hyperkahler structures which give extra symmetries, and Witten explained how Quantum Field Theory ideas could be used, in principle, to derive all the invariants from this fact. This result agrees with that obtained from Kronheimer and Mrowka's structure theorem (and earlier explicit calculations of Friedman and Morgan and O'Grady): for a $K3$ surface the only basic class is 0, and one gets

$$\mathcal{D}_{K3}(\alpha) = e^{\alpha.\alpha/2}.$$

Witten went on to consider the case of a general Kahler surface X [W2]. If $b^+(X) > 1$, there is a non-trivial holomorphic 2-form η on X, which gives a kind of hyperkahler structure away from its zero set—a union of complex curves in X. Witten argued that the calculation could be localised around this zero set and obtained the Kronheimer and Mrowka formula, with the additional prediction that the basic classes could be obtained from the components of the zero sets of generic holomorphic forms. This latter prediction, while it was in line with many known examples, went definitely beyond the results obtained by geometers at that time. If one takes a complex surface S such as a suitable ramified cover of $\mathbf{CP}^1 \times \mathbf{CP}^1$ (which will be of basic type), the zero set of a generic holomorphic form is connected, a curve representing the canonical class K_S, so Witten's prediction asserts that the only basic classes are $\pm K_S$. On the other hand there are many other potential basic classes—lifted up from $H^2(\mathbf{CP}^1 \times \mathbf{CP}^1) = \mathbf{Z}^2$— which were not ruled out by known mathematical constraints, such as symmetry and the Hodge decomposition.

This story came to a head in the latter part of 1994 with the work of Witten and Seiberg on invariants for general 4-manifolds. This used advances in understanding of $N = 2$ supersymmetric Yang-Mills theory which we will scarcely attempt to describe, except to paraphrase some parts of [W3]. The Quantum Field Theory analysis of the problem involves a scale parameter t (which re-scales the metric on the 4-manifold) and a complex parameter u, related to the auxiliary fields appearing in the functional integral. The invariants can be calculated for any t—the geometers interpretation of the invariants in terms of instanton moduli spaces arises from the limit $t \to 0$, where attention concentrates in the region of $u = 0$, and the auxiliary fields drop out. In the limit when $t \to \infty$ the situation is described by an analytic function τ of the u variable, with some singularities. The essential ingredient in the analysis is that if $b^+(X) > 1$, this function τ can be shown to be *modular* with respect to the action of $SL(2, \mathbf{Z})$. The nub of this modularity is a symmetry $u \to u^{-1}$ which corresponds to a duality between electricity and magnetism proposed in 1977 by Olive and Montonen. (The modular property leads to elliptic curves and elliptic functions, which it is natural to hope will tie in with the elliptic functions found by Fintushel and Stern in their connected sum formula.) The modular property constrains the function τ so much that it is entirely determined by its Taylor expansion about the special points $u = \pm 1$, and the evaluation of these co-efficients leads to the Seiberg- Witten invariants. In sum, Seiberg and Witten's arguments predicted that the instanton invariants of a 4-manifold with $b^+ > 1$ can be expressed in terms of the Seiberg-Witten invariants. For manifolds of simple type their precise prediction can be written:

$$(14) \qquad \mathcal{D}_X(\alpha) = 2^{m(X)} e^{\alpha.\alpha/2} (\sum_{L \in \Lambda} n_L e^{c_1(L).\alpha}),$$

where $m(X)$ is the elementary topological invariant $m(X) = 2 + \frac{1}{4}(7\chi(X) + 11\tau(X))$ and Λ is the set of isomorphism classes of line bundles L with $c_1(L)^2 = 2\chi + 3\tau$. In other words:

1. the basic classes of Kronheimer and Mrowka all satisfy $\kappa_i^2 = 2\chi + 3\tau$ (this possibility had been suggested by Kronheimer and Mrowka),
2. the basic classes are just the classes with this square which are the first Chern classes of line bundles L with $n_L \neq 0$ (note that the condition on c_1^2 means

    exactly that these are the line bundles with 0-dimensional Seiberg-Witten moduli spaces),

3. the co-efficients a_i are just the Seiberg-Witten invariants, up to an overall factor $2^{m(X)}$.

This is a wonderful, "magic", formula (which is expected to generalise to manifolds not of simple type, using the Seiberg-Witten invariants coming from higher dimensional moduli spaces). Each side of the formula has a rigorous mathematical definition, yet the arguments used to produce it seem to lie way beyond the borders of rigorous mathematical understanding, as for other celebrated predictive formulae, such as those arising in "mirror symmmetry". From the point of view of a pragmatic 4-manifold topologist the formula relating the two theories is perhaps not essential, since, as we shall describe in the remainder of this survey, having once got the Seiberg-Witten invariants, one can prove many things about 4-manifolds perfectly rigorously using them, ignoring the relationship with the older instanton theory. However it is clearly of fundamental importance to put the formula on a sound footing. On the one hand there is overwhelming circumstantial evidence that the formula is true, based on known properties of the two kinds of invariants. Beyond this, V. Pidstragach proposed (in lectures in December 1994), and to a large extent carried through, a geometrical approach which seems very likely to give a rigorous proof of at least a substantial part of the formula.

Pidstragach's idea is a very natural one and fits in with a number of other lines of work in gauge theory. He considers a $U(2)$ bundle $E \to X$. There is a straightforward generalisation of the Seiberg- Witten equations to a pair of coupled equations for a connection on E and a coupled spinor, a section Ψ of $E \otimes W^+$. These constrain the trace-free part of the self-dual curvature, essentially the curvature of the associated $PU(2) = SO(3)$ connection; the central component of the connection is fixed arbitrarily. Then one gets a moduli space M of solutions to these $U(2)$ Seiberg-Witten equations modulo $SU(2)$ gauge transformations. The moduli space M contains the ordinary instanton moduli space (when $\Psi = 0$) and Seiberg-Witten moduli spaces for certain line bundles corresponding to the reducible connections on E. These are the fixed points of a circle action on M, induced by scalar multiplication on E. Otherwise said, if one divides by a larger gauge group of $U(2)$ automorphisms, these are the two kinds of reducible solutions to the $U(2)$ equations. Either way, standard arguments can be used to relate the topological data on the instanton moduli spaces which goes into one system of invariants, to data over the Seiberg-Witten moduli spaces which goes into the other. The only complication— and it is a serious one—arises from the lack of compactness, which introduces terms from the compactifications of the instanton moduli spaces.

With the benefit of hindsight, and Pidstragach's idea, one can trace antecedents of these new developments in other lines of work by geometers on moduli problems. On the one hand there is the work of Pidstragach and Tyurin on "Spin^c-invariants" of 4-manifolds [PT], which involves a kind of uncoupled version of the Seiberg-Witten equations. On the other hand there is a considerable body of work on analogies in 2-dimensional gauge theory. The simplest example is the "vortex equation", for a connection on a bundle E over a Riemann surface together with a section of E; another example is the equation introduced by Hitchin [H] which involves a holomorphic form with values in the bundle of endomorphisms of E (the article [BDGW] contains a survey of some of these developments). One may

compare Pidstragach's approach to the work of Thaddeus [Th], which relates the moduli spaces of flat connections over a Riemann surface to the moduli spaces of abelian vortices. It seems likely that, in the case of a Kahler surface, there should be a translation of Pidstragach's argument into purely algebro-geometric terms, along the lines of Thaddeus' method. In this circle of ideas, it is natural to consider variants of the Seiberg- Witten equation over 4-manifolds, in parallel with the 2-dimensional case: one such variant has already been considered by Vafa and Witten in [VW], leading to intriguing relations with modular forms.

4. KAHLER AND SYMPLECTIC MANIFOLDS

In [W3] Witten showed that the Seiberg-Witten invariants of a Kahler surface X can be completely described in terms of the complex geometry of the surface. The starting point for this is the fact that a Kahler surface has a preferred Spin^c-structure with

$$(15) \qquad W^+ = \Lambda^0 \oplus \Lambda^{02} \ , \ W^- = \Lambda^{01}.$$

The decomposition of W^+ in this setting corresponds to the decomposition of the self-dual 2-forms:

$$(16) \qquad \Lambda^+ = \mathbf{R}.\omega \oplus \Lambda^{02},$$

where ω is the Kahler form. The line bundle determined by this preferred Spin^c-structure is the dual of the canonical bundle $K_X = \Lambda^2 T^*_{\mathbf{C}} X$. This satisfies $K_X.K_X = 2\chi + 3\tau$, the condition for zero-dimensional Seiberg-Witten moduli spaces. In fact for any closed 4-manifold the necessary and sufficient conditions for a class $c \in H^2(X)$ to be the first Chern class of an *almost-complex* structure on X are just that $c = w_2(X)$ modulo 2, and that $c.c = 2\chi + 3\tau$. (As a matter of notation, we use the symbol K_X to denote both the canonical line bundle and its first Chern class.)

The Dirac operator in the Kahler case, induced from the Levi-Civita connection, is

$$D = \sqrt{2}(\overline{\partial} + \overline{\partial}^*) : \Lambda^0 \oplus \Lambda^{02} \to \Lambda^{01}.$$

Consider first the Seiberg-Witten equations corresponding to the preferred Spin^c-structure. If Γ is the connection on K^*_X induced by the Levi-Civita connection, any other connection on K^*_X may be written $\Gamma + B$ where B can be regarded as a connection on the trivial bundle. So we may take the unknowns in the Seiberg-Witten equation to comprise
  1. a connection B on the trivial line bundle Λ^0,
  2. a section α of Λ^0, i.e. a function on X,
  3. a section β of Λ^{02}—a $(0,2)$ form over X.
Under the decomposition (16) the equations become :

$$\overline{\partial}_B \alpha = -\overline{\partial}^*_B \beta,$$
$$F_B^{02} = \overline{\alpha}\beta,$$
$$iF_B^{11}.\omega = (|\beta|^2 - |\alpha|^2) - iF_\Gamma^{11}.\omega.$$

The operators $\overline{\partial}_B, \overline{\partial}_B^*$ appearing in the first equation are obtained in the standard way by coupling to the connection B. Now if (B, α, β) is a solution of these equations, we apply $\overline{\partial}_B$ to the first equation to get

$$\overline{\partial}_B \overline{\partial}_B^* \beta = -\overline{\partial}_B \overline{\partial}_B \alpha = -F_B^{02} \alpha,$$

using the definition of the curvature as the square of the coupled exterior derivative. Substituting into the second equation we get

$$\overline{\partial}_B \overline{\partial}_B^* \beta + |\alpha|^2 \beta = 0.$$

Taking the L^2 inner product with β:

$$\int_X |\overline{\partial}^* \beta|^2 + |\alpha^2||\beta|^2 d\mu = 0.$$

This is essentially a refinement of the basic argument in Section 1, taking account of the splitting of the spin space. It follows that either β is everywhere 0, or α and $\overline{\partial}_B^* \beta$ are zero, (if α or β vanish on an open set they are identically zero, by standard results on unique continuation for elliptic PDE). In any case the product $\overline{\alpha}\beta$, and hence F_B^{02}, is zero, so the connection B defines a holomorphic structure on the trivial line bundle. Now recall that the Seiberg-Witten invariants are only defined, in a straightforward way, if $b^+(X) > 1$. When X is Kahler this means that there is a non-trivial section of the canonical bundle and hence the *degree* $c_1(K_X)).[\omega]$ is non-negative. By the Chern-Weil construction, this is

$$c_1(K_X).\omega = \frac{-i}{2\pi} \int_X (F_\Gamma.\omega) d\mu \geq 0.$$

Since F_B is exact the integral of $F_B.\omega$ over X is zero and we get

$$\int_X |\alpha|^2 - |\beta|^2 \geq 0,$$

so it must be β which vanishes, and the holomorphic section α gives a *holomorphic* trivialisation of the holomorphic structure defined by B. Conversely, the compatible connections on the trivial holomorphic line bundle are parametrised by positive functions e^f on X: the norm of the trivialising section. To solve the Seiberg-Witten equations one has to solve the non-linear equation:

$$(17) \qquad\qquad \Delta f + e^{2f} = \sigma,$$

where $\sigma = -iF_\Gamma.\omega$, which is actually minus one eighth of the scalar curvature of the Riemannian metric. Standard arguments show that, for any σ whose integral is non-negative, equation (17) has a unique solution f. Thus we conclude that there is just one solution to the Seiberg-Witten equations, and in turn the Seiberg-Witten invariant n_K is ± 1. (Equation (17) arises in the construction of constant-curvature metrics on Riemann surfaces and the discussion here can be seen as a generalisation of the uniformisation theorem to higher dimensions.) Another way of going about things is to change the third equation to

$$iF_B^{11}.\Omega = |\beta|^2 - |\alpha|^2 + 1,$$

which has an obvious solution with B the trivial connection, $\beta = 0$ and $\alpha = 1$. This is not really different from the previous method, since the proof that the perturbation does not affect the count of solutions is essentially the same as the "continuity method" proof of the existence of solutions to (17).

For a general line bundle L, Witten's analysis of the equations proceeds in a similar way. We write $L = \xi^2 \otimes K_X^*$ so the data can be taken to be a connection B on the line bundle ξ, a section α of ξ and a $(0,2)$-form β with values in ξ. The argument above shows that if a solution exists, ξ has a holomorphic structure and that one of α or β vanishes identically, which one depending on the sign of the degree $c_1(L).\omega$. For the right choice of sign one gets a moduli space which can be identified with the union over holomorphic structures on ξ of the projectivisations of the spaces of holomorphic sections—that is, with the space of complex curves in X representing $c_1(\xi)$. The dimension of this space may not agree with the dimension predicted by the index formula: the zero set is not "regular". To compute the invariant Witten introduces an artful choice of perturbation of the second equation. For a generic holomorphic 2-form η over X he considers the equation

$$F_B^{0,2} = \overline{\alpha}\beta - \overline{\eta}.$$

We may assume that a solution to the unperturbed equations exists, so the first Chern class of ξ has type $(1,1)$ in the Hodge decomposition of the cohomology of X. Thus

$$\int_X F_B \wedge \eta = \int_X F_B^{02} \wedge \eta = 0.$$

Now the perturbed equations give:

$$\overline{\partial}_B \overline{\partial}_B^* \beta + |\alpha|^2 \beta = \alpha \overline{\eta}.$$

Taking the L^2 inner product with β, using the previous identity and substituting in the second equation again one obtains that

$$\int_X |\overline{\partial}_B^* \beta|^2 + |\alpha\overline{\beta} - \eta|^2 d\mu = 0,$$

so $\overline{\partial}_B^* \beta = \overline{\partial}_B \alpha = 0$ and $F_B^{0,2} = 0$, $\alpha\overline{\beta} = \overline{\eta}$. This means that B defines a holomorphic structure on ξ, that α is a holomorphic section of ξ and $\overline{\beta}$ is a holomorphic section of $K_X \otimes \xi^*$. Conversely, given this data, one can solve the remaining equation to define the connection B, as before. This gives a completely algebro-geometric description of the solutions to the perturbed equation. In terms of divisors on X, if the zero set of η is a divisor $\sum r_i C_i$, the line bundles L with non-zero Seiberg-Witten invariants just correspond to the sums $\sum s_i C_i$, with $0 \le s_i \le r_i$. Witten also gives a formula for the sign with which each of these perturbed solutions should be counted. Notice that the solutions to the perturbed equation are always isolated, so the invariants defined by higher dimensional moduli spaces are all zero. Thus, assuming the general conjectures in Section 3, any Kahler surface is of simple type.

There are essentially two examples when one gets Seiberg-Witten 2 basic classes other than the canonical class. The first occurs if X is the blow-up of a surface X_0 with exceptional divisor E: then the canonical divisor of X is the sum of E and the proper transform of the canonical divisor on X_0. The general description then

tells us that the basic classes on X are just the classes $\kappa_i \pm E$, where κ_i are the basic classes on X_0 (using the canonical injection from $H^2(X_0)$ into $H^2(X)$). The second occurs, when X is an elliptic fibration, where the canonical divisor is a sum of fibres (including possible multiple fibres); in this case the basic classes are all rational multiples of the canonical class, in agreement with the known calculations for the instanton invariants.

Symplectic 4-manifolds. In two papers [T1], [T2] which appeared towards the end of 1994, Taubes showed that large parts of the discussion above can be adapted to the case of *symplectic* 4-manifolds. This leads to some of the most striking new results proved with the Seiberg-Witten invariants. In March 1995 Taubes announced further results [T3], which tie the theory up with work on *pseudo-holomorphic curves* in symplectic manifolds: the techniques Taubes introduced in this later work also allow substantial simplifications in the proofs of his earlier results. Let X be a compact 4-manifold with a symplectic form ω. Then one can choose a compatible almost-complex structure and Hermitian metric, for which the symplectic form is self-dual. The choice of almost-complex structure is unique up to homotopy, in particular the cohomology class $K_X = -c_1(TX)$ depends only on ω; as we have mentioned above K_X has 0-dimensional Seiberg-Witten moduli spaces. Taubes' first collection of results, for manifolds with $b^+ > 1$, are

1. The Seiberg-Witten invariant of K_X is ± 1.
2. If $[\omega]$ is the cohomology class of the symplectic form, then $K_X.[\omega] \geq 0$ and any other class κ with non-zero Seiberg-Witten invariant satisfies

$$|\kappa.[\omega]| \leq K_X.[\omega],$$

with equality if and only if $\kappa = \pm K_X$.

In the Kahler case the first result is just the first result of Witten discussed above: the second follows from Witten's analysis and the fact that if a holomorphic line bundle L has a non-trivial holomomorphic section, then $c_1(L).[\omega] > 0$.

Taubes' work utilises the differential geometry of almost-complex manifolds. On any almost-complex manifold one can decompose the differential forms into bi-type and define operators $\partial, \overline{\partial}$ by taking components of the exterior derivative. However the composite $\overline{\partial}^2$ is not zero in general; instead one has for a function f,

$$(18) \qquad \overline{\partial}^2 f = N \circ \partial f,$$

where N is an algebraic operator given by the Nijenhuis tensor $N \in \mathrm{Hom}(\Lambda^{1,0}, \Lambda^{0,2})$, which vanishes if and only if the almost-complex structure is integrable. On the symplectic manifold X with a choice of almost-complex structure the identification of the canonical Spin^c-structure with the forms goes through as before. Moreover the identity $D = \sqrt{2}(\overline{\partial} + \overline{\partial}^*)$ holds. Thus, after perturbing the equations by a suitable self-dual 2-form we may take the equations corresponding to the canonical Spin^c structure to be

$$\overline{\partial}_B \alpha = -\overline{\partial}_B^* \beta,$$
$$F_B^{02} = \overline{\alpha}\beta,$$
$$iF_B^{11}.\omega = (|\beta|^2 - |\alpha|^2 + 1).$$

If we proceed as before, we would now get:

$$\overline{\partial}_B\overline{\partial}_B^*\beta = -\overline{\partial}_B\overline{\partial}_B\alpha = -|\alpha|^2\beta + N\circ\partial_B\alpha,$$

and the last term spoils the argument.

Taubes' first approach was to modify the equations by a counter term, considering the family of deformations of the second equation

$$F_B^{0,2} = \overline{\alpha}\beta - \frac{r\overline{\alpha}}{1+r|\alpha|^2}N\circ\partial\alpha,$$

where r is a large parameter. He showed that this is an admissible perturbation and that for large r the only solution is the obvious one. This required detailed estimates. In his later work Taubes introduced a different kind of perturbation, for which the corresponding proof is very simple. This second perturbation is the family of equations:

$$\overline{\partial}_B\alpha = \overline{\partial}^*\beta$$
$$F_B^{02} = \overline{\alpha}\beta$$
$$iF_B^{11}.\omega = (|\beta|^2 - |\alpha|^2 + \rho^2),$$

where ρ is a real parameter which is again made large. One now uses a Weitzenbock formula for the $\overline{\partial}$-operator, in this almost-complex case. The formula is just the same as in the more familiar Kahler situation:

$$\nabla_B^*\nabla_B f = 2\overline{\partial}_B^*\overline{\partial}_B f + i(F_B^{1,1}.\omega)f.$$

Applying this to α and using the three equations gives

$$\int_X |\nabla_B\alpha|^2 d\mu = \int_X 2(\overline{\partial}_B^*\overline{\partial}_B\alpha,\alpha) - (|\alpha|^2 - |\beta|^2 - \rho^2)|\alpha|^2 d\mu$$
$$= \int_X -2(\overline{\partial}_B^*\overline{\partial}_B^*\beta,\alpha) - (|\alpha|^2 - |\beta|^2 - \rho^2)|\alpha|^2 d\mu$$
$$= \int_X -2(\beta,\overline{\partial}_B\overline{\partial}_B\alpha) - (|\alpha|^2 - |\beta|^2 - \rho^2)|\alpha|^2 d\mu$$
$$= -2\int_X (\beta, N\circ\partial_B\alpha)d\mu - 2\int_X |\alpha|^2|\beta|^2 - \int_X (|\alpha|^2 - |\beta|^2 - \rho^2)|\alpha|^2 d\mu$$
$$= -2\int_X (\beta, N\circ\partial_B\alpha)d\mu - \int_X |\alpha|^2|\beta|^2 d\mu$$
$$\quad - \int_X (|\alpha|^2 - \rho^2)^2 d\mu - \rho^2\int_X |\alpha|^2 - \rho^2 d\mu.$$

Now comes the crucial step: the integral of $F_B^{11}.\omega$ over X, with the respect to the volume element, is the same as the integral of the 4-form $F_B\wedge\omega$, and this *vanishes*

since the curvature F_B represents the first Chern class of the trivial line bundle and ω is closed. So, using the third equation,

$$\int_X |\alpha|^2 - \rho^2 d\mu = \int_X |\beta|^2 d\mu$$

and one gets

$$\int_X |\nabla_B \alpha|^2 + |\alpha|^2 |\beta|^2 + (|\alpha|^2 - \rho^2)^2 + \rho^2 |\beta|^2 d\mu = -2 \int_X (\beta, N \circ \partial_B \alpha) d\mu,$$

where the terms on the left-hand side are all non-negative. This manipulation, which is related to Witten's argument for the perturbation of the $(0, 2)$ equation, is another variant of the basic integration-by-parts of Section 1. To apply the formula: the right-hand side above is bounded by a constant multiple of $\|\beta\| \|\nabla_B \alpha\|$ and hence by

$$\tfrac{1}{2} \|\nabla_B \alpha\|^2 + C^2 \|\beta\|^2,$$

for some C, depending only on the geometry of the base manifold. Thus:

$$\|\nabla_B \alpha\|^2 + \rho^2 \|\beta\|^2 + \| |\alpha|^2 - \rho^2 \|^2 \leq \tfrac{1}{2} \|\nabla_B \alpha\|^2 + C \|\beta\|^2$$

and it follows that once $\rho^2 > 2C$ the only solution is the obvious one with B the trivial connection, $\alpha = \rho$ a constant, and $\beta = 0$. This essentially proves Taubes' result (1) and the same argument gives (2).

Taubes' later work, the details of which have not appeared at the time of writing, establishes a far-reaching connection between the Seiberg-Witten theory and the "Gromov invariants" defined by counting pseudo-holomorphic curves. He shows that the Seiberg-Witten basic classes have the form $\pm(2\xi - K_X)$ where the Poincaré dual of ξ, a 2-dimensional homology class in X, has non-zero Gromov invariant: in particular the homology class is represented by a pseudo-holomorphic curve. Moreover the Gromov invariant—the "number" of pseudo-holomorphic curves in this homology class—equals the Seiberg- Witten invariant. Roughly speaking, Taubes shows that the zero set of α, where α arises from a solution of a deformed equation as above, but for a general line bundle ξ, is very close to a pseudo-holomorphic curve when ρ is large. (Taubes' results tie in with the Kahler case, since the fixed components of the canonical system are the only curves on a Kahler surface which are "regular" solutions of the Cauchy-Riemann equations, in the sense of Gromov's theory [D4].)

5. Gluing theory

In this section we will discuss techniques for the calculation of Seiberg-Witten invariants on a 4-manifold $X = X_1 \cup_Y X_2$ split into two pieces by a 3-manifold Y. For the instanton invariants there has been much work on this question, using "Floer homology" groups, or variants thereof. The simplest case is that of a connected sum, $X = \overline{X_1} \sharp X_2$, say, where Y is the 3- sphere. This was studied by Witten in [W3], and other authors. The discussion follows closely the corresponding one for the instanton equations and leads to a vanishing theorem: all the Seiberg-Witten invariants of X vanish if $b^+(X_1), b^+(X_2)$ are both strictly positive. One

can consider a family of metrics on the connected sum in which the "neck" is pinched down to zero radius. This can be done in such a way that the scalar curvature is bounded below in the family. The *a priori* estimates for Seiberg-Witten solutions and a simple removable singularities argument show that a sequence of solutions for these metrics corresponding to a Spin^c structure $c \in H^2(X)$ has a subsequence converging away from the neck to solutions over the closed manifolds $\overline{X_1}, \overline{X_2}$ with Spin^c structures c_1, c_2 such that $c = c_1 + c_2$. The index formula gives $i(c) = i(c_1) + i(c_2) - 1$ so if $i(c)$ is zero, one of $i(c_1), i(c_2)$ is negative so there are no irreducible solutions for generic metrics. One deduces that the Seiberg-Witten invariants defined by 0-dimensional moduli spaces over X vanish, provided there are no reducible solutions. The argument extends easily to the higher dimensional moduli spaces. Geometrically, the extra term "1" in the index formula arises from a gluing parameter; solutions over $\overline{X_1}, \overline{X_2}$ can be glued together to give a solution over the connected sum, but there is a $U(1)$ choice in the identification of the bundles over the neck. If one of the summands, say $\overline{X_2}$, has a negative definite intersection form, the invariants for the connected sum can be obtained by analysing the solutions obtained by gluing reducible solutions over $\overline{X_2}$. For example if $\overline{X_2}$ is a homotopy $\overline{\mathbf{CP}}^2$, the Seiberg-Witten basic classes for the connected sum have the form $\kappa_i \pm E$, where κ_i runs over the Seiberg-Witten basic classes for $\overline{X_1}$ and E is the generator of $H^2(\overline{X_2})$—in agreement with the discussion for blow-ups in the case of complex Kahler surfaces.

For more general decompositions $X = X_1 \cup_Y X_2$ it is natural to use a different kind of metric-degeneration, in which the neck region is stretched into a long tube $Y \times [-L, L]$, isometrically embedded in X. This leads to a theory which runs parallel to the Floer theory for the instanton equations. Recall that over a closed 4-manifold the Seiberg-Witten solutions minimise the functional

$$\int_X |\nabla_A \psi|^2 + \tfrac{1}{8}||\psi|^2 + R|^2 + \tfrac{1}{2}|F_A|^2 d\mu.$$

This involves two applications of Stokes' theorem. Consider now a 4-manifold Z with boundary a 3-manifold Y. A Spin^c-structure on Z induces one on Y. Write a, ϕ for the boundary values of A, ψ. The corresponding calculation for this case leads to an extra boundary integral

$$(19) \qquad\qquad J_Y(a, \phi) = \int_Y (\delta_a \phi, \phi) d\mu + \tfrac{1}{2} CS(a),$$

where δ_A is the Dirac operator over Y and CS denotes the Chern-Simons functional. In the case where the relevant bundle is trivial over Y and a can be regarded as a connection 1-form, this is

$$CS(a) = \int_Y a \wedge da.$$

In general the Chern-Simons functional is only defined modulo $4\pi^2$. The space $\mathcal{C}$ of irreducible pairs (a, ϕ), where a is a connection on a $U(1)$ bundle L over Y, has the homotopy type of the product of $\mathbf{CP}^\infty$ and the torus $H^1(Y; \mathbf{R})/H^1(Y; \mathbf{Z})$. The Chern-Simons functional gives a map from $\mathcal{C}$ to S^1 whose homotopy class

naturally corresponds to the Poincaré dual of $c_1(L)$ under the isomorphism $[\mathcal{C}, S^1] = H_1(Y)/\text{Torsion}$. Now, following Floer, the solutions of the Seiberg-Witten equation over the tube $Y \times \mathbf{R}$ can be interpreted as the gradient flow lines of the functional J_Y on $\mathcal{C}$. In particular the translation-invariant solutions correspond to the critical points of J_Y which are the solutions of the equations

$$(20) \qquad \begin{aligned} \delta_a \psi &= 0, \\ F_a &= -\tau(\psi, \psi) \end{aligned}$$

over Y. Here the quadratic map τ is constructed in just the same way as in the 4-dimensional case. Floer's construction gives a chain complex built from the irreducible solutions of this equation (perhaps after a small perturbation) and the Seiberg-Witten solutions over the tube, which define the boundary map in the chain complex. This gives Seiberg-Witten Floer homology and cohomology groups $HFSW_*(Y), HFSW^*(Y)$ which are interchanged by switching the orientation of Y. Then in favourable cases the Seiberg-Witten invariants of a manifold $X = X_1 \cup_Y X_2$ are obtained from relative invariants of X_1, X_2 with values in $HFSW_*(Y), HFSW^*(Y)$ under a dual pairing $HFSW^*(Y) \otimes HFSW_*(Y) \to \mathbf{Z}$. For a general theory one must also consider the reducible critical points on $\mathcal{C}$, in the manner of the equivariant Floer theory of Austin and Braam [AB]. If the line bundle L is non-trivial, there are no reducible solutions: in the case of the trivial bundle the reducible solutions lead to the new phenomenon that the Seiberg-Witten Floer groups are not entirely metric-independent: they can change as one passes through the (codimension 1) space of metrics where the ordinary Dirac operator has a non-trivial kernel. For the instanton equations the corresponding critical points which form the basis of Floer's theory are just the flat connections over Y, which can be described in a very concrete way in terms of the fundamental group. By contrast, the Seiberg-Witten theory leads to the system of equations (20), which are hard to interpret. Two cases in which something can be said are

　　1. the 3-manifold Y admits a metric of positive scalar curvature;

　　2. $Y = \Sigma \times S^1$ where Σ is a surface.

In the first case a variant of the 4-manifold vanishing theorem [W3] shows that the only Seiberg-Witten solutions are reducible (with $\phi = 0$ and a a flat $U(1)$ connection), so the Seiberg-Witten-Floer groups are trivial. In the second case we assume that the line bundle L over $\Sigma \times S^1$ is pulled back from a line bundle over Σ, of degree d say. The elements of the "Floer theory" in this case have been analysed by Kronheimer and Mrowka [KM3], and in [MST]. The key observation is that the solutions to equation (20) over $S^1 \times \Sigma$ are invariant under rotations of the circle. In general if W is any 3-manifold and L is a line bundle over $S^1 \times W$ whose cohomology class has no component in $H^1(S^1) \otimes H^1(W) \subset H^2(S^1 \times W)$, then any Seiberg-Witten solution on the line bundle L over $S^1 \times W$ is invariant under S^1. This follows from the gradient-flow interpretation of the equations and the description of the homotopy class of the Chern-Simons functional (just as the gradient flow of a genuine real-valued function can have no non- constant closed orbits). This is then applied to the manifold $W = \Sigma \times S^1$, multiplying by an additional copy of S^1, to get the assertion stated. Now the equations for S^1-invariant solutions of (20) just reduce to the vortex equations over Σ, mentioned in Section 3. By the symmetry $L \leftrightarrow L^{-1}$ we may suppose that the degree d of the line bundle L over Σ is positive.

The relevant vortex equation is then for
   1. a connection A on L,
   2. a section ψ of $K_\Sigma^{1/2} \otimes L^{-1/2}$,
the equations being that ψ is a holomorphic section—$\overline{\partial}_A \psi = 0$ and

$$iF_A = -|\psi|^2.$$

The condition that ψ is holomorphic implies that the degree of $K_\Sigma^{1/2} \otimes L^{-1/2}$ is positive; so solutions exist only if

$$2g - 2 \geq d.$$

If equality holds, there is a unique trivial solution; if $r = (2g-2) - d > 0$, the space of solutions is identified, via the zero-set of the section, with the symmetric product $s^r(\Sigma)$. This is the basic "correspondence theorem", well-known in the theory of the vortex equation [BDPW]. In turn the Floer groups, for line bundles pulled back from Σ, are the ordinary cohomology groups of the symmetric products of Σ.

6. Applications of the Seiberg-Witten equations

In this Section we will report on some results about 4-manifolds proved using the Seiberg-Witten equations. These are logically quite independent of the relations with the instanton theory, although naturally enough there are many similarities in the methods and results. The striking features are that, in the case of results which can be obtained by both theories, many of the proofs become much simpler and there are new results which seem quite out of reach in the older theory.

Distinguishing 4-manifolds; connections with Riemannian geometry. The most straightforward application of the Seiberg-Witten invariants is to distinguish differentiable 4-manifolds within the same homeomorphism type. Myriads of examples could be given: the simplest being to show that connected sums $X_{p,q}$, say, of p copies of $\mathbf{CP}^2$ and q copies of $\overline{\mathbf{CP}}^2, q > 1$, for which the Seiberg-Witten invariants vanish, are not diffeomorphic to Kahler surfaces (or any other manifolds with non-zero Seiberg-Witten invariants). Of course many of these results have already been obtained using the older theory. One gets stronger and novel results by using the vanishing theorem for manifolds of positive scalar curvature. No manifold with $b^+ > 1$ and non- zero Seiberg-Witten invariants can admit a metric of positive scalar curvature. By contrast the connected sums $X_{p,q}$ do admit such metrics [GL]. This shows that the classification of 4-manifolds with metrics of positive scalar curvature is radically different from the higher dimensional case, in which the only obstruction comes from cobordism theory and characteristic classes. In a similar vein, the vanishing theorem for the solutions of equation (20) over positive scalar curvature 3-manifolds leads to the conclusion that no such 3-manifold can be embedded in a 4-manifold with non-zero Seiberg- Witten invariants, in such a way that $b^+ > 0$ on each side. This is a new kind of obstruction to the existence of positive scalar curvature metrics on 3-manifolds. For the case of the Poincaré sphere, and other constant- curvature 3-manifolds, the result was proved earlier by Austin [A], using the ordinary Floer theory, but the argument in that case was less direct, involving a study of the interaction between the reducible and irreducible flat connections.

Other links with Riemannian geometry have been found by Le Brun [L]. By combining a sharper form of the vanishing theorem with the Hitchin-Thorpe formulae for characteristic numbers of a 4-manifold, Le Brun has shown that any Einstein 4-manifold with non-zero Seiberg-Witten invariants satisfies the inequality $\chi - 3\tau \geq 0$. For complex surfaces this reduces to the celebrated Miyaoka-Yau inequality $c_1^2 \leq 3c_2$: for general 4-manifolds it provides a new restriction for Einstein metrics. In the case of equality, LeBrun proves that the only Einstein metrics on compact complex space-forms $\mathbf{CH}^2/\Gamma$ are the canonical ones.

Kahler and symplectic manifolds. In an influential survey paper [FM], written soon after the introduction of the instanton invariants, Friedman and Morgan set down a collection of conjectures about the differential topology of complex surfaces. Much progress was made on many of these in the following years, by work of various mathematicians. One of the corollaries of Witten's description [W3] of the Seiberg-Witten invariants for complex surfaces is the final solution of one of the principle conjectures: *the canonical class of a minimal surface of general type with $p_g > 0$ is, up to sign, a differentiable invariant* (i.e. preserved by any diffeomorphism of the surface). This follows from the fact that, using results on canonical systems, $\pm K_X$ is the only Seiberg-Witten basic class in this case. Another new result which follows from the Seiberg-Witten theory is the fact that a minimal Kahler surface is smoothly indecomposable: in any connected-sum decomposition one factor is a homotopy 4-sphere. Similarly the Seiberg-Witten invariants have been used by Kronheimer to give a simpler proof of one of the other main conjectures: *the Kodaira dimension of a Kahler surface is a differentiable invariant* (which had been finally established shortly before by Friedman and Qin [FQ]). These applications essentially round off earlier work, and the main conclusion to be drawn from the Seiberg-Witten theory, in the Kahler case, is perhaps a rather negative one. Essentially the only information carried by the invariants is the canonical class. Of course, assuming the relations with the instanton theory described in Section 3, the same is true for the instanton invariants. Thus one has no potential tools at the moment to attack one of the other main questions; whether there are diffeomorphic surfaces which are not "deformation equivalent" (many candidates—multiple branched covers of $\mathbf{CP}^1 \times \mathbf{CP}^1$—have been described by Catanese and Manetti [M]).

Some of the most far-reaching applications of the Seiberg-Witten theory are to symplectic 4-manifolds, through the work of Taubes. Kahler surfaces furnish a basic stock of examples of symplectic 4-manifolds; on the other hand one may consider the class of compact 4-manifolds X equipped with

1. a class $[\omega] \in H^2(X; \mathbf{R})$ with $[\omega]^2 > 0$,
2. an almost-complex structure, giving a "canonical class" $K = -c_1(TX)$.

These are the elementary necessary conditions for the existence of a symplectic structure. This class is much larger than the examples obtained from Kahler surfaces: as we have mentioned above, a cohomology class K arises from an almost-complex structure if and only if $K^2 = 3\tau + 2\chi$, and it follows that any 4-manifold with indefinite intersection form and with $b_1 - b^+$ odd can be equipped with the above structure. Almost 20 years ago Thurston pointed out that not all symplectic 4-manifolds are Kahler. Thurston's example is not simply-connected and for simply-connected manifolds one could, until recently, say very little: both of the extreme possibilities—that all simply-connected symplectic 4-manifolds are Kahler or that any triple $(X, [\omega], K)$ can arise from a symplectic structure—were open.

The first possibilty was ruled out by Gompf [Go] who constructed non-Kahler symplectic manifolds, homotopy equivalent to $K3$ surfaces but distinguished by instanton invariants. The second possibility is now ruled out by the work of Taubes. For example the connected sum $X_{p,q}$ with $p, q > 1$ and p odd, admits suitable classes $[\omega], K$, but cannot be given a symplectic structure since all Seiberg-Witten invariants are 0, in contradiction to Taubes' first result. Again, myriads of other examples can be given: even for an underlying differentiable manifold X which does admit a symplectic structure one can still find many pairs $K, [\omega] \in H^2(X)$ which are *not* realised by a symplectic structure. Thus symplectic 4-manifolds occupy a truly intermediate position. Their Seiberg-Witten invariants share many of the main properties of Kahler surfaces but with scope for much more diversity; whereas the Seiberg-Witten invariants for "most" surfaces boil down just to the canonical class, there are examples due to Stipsicz [St] and Morgan and Szabo of symplectic 4-manifolds with many different basic classes.

Taubes also considers symplectic 4-manifolds with $b^+ = 1$, where the Seiberg-Witten invariants involve chamber structures. In particular he settles two outstanding questions about the complex projective plane:

1. There is no symplectic structure on $\mathbf{CP}^2$ with $K = \lambda[\omega]$ for $\lambda > 0$.
2. Any symplectic structure on $\mathbf{CP}^2$ with $K = \lambda[\omega]$, $\lambda \leq 0$ is diffeomorphic to a multiple of the standard structure.

The proof of the second result uses a theorem of Gromov [Gr], which reduces the problem to that of finding a pseudo-holomorphic curve representing the generator of $H^2(\mathbf{CP}^2)$. Together, these results of Taubes assert that there is essentially only one symplectic structure on the differentiable manifold $\mathbf{CP}^2$.

Adjunction inequalities and the Thom conjecture. As we described in Section 3, one of the main goals which drove the work of Kronheimer and Mrowka was the study of the minimal genus of surfaces within a given homology class in a 4-manifold, and in particular the conjecture that complex curves should minimise the genus. This conjecture has now been proved in considerable generality using the Seiberg-Witten invariants; in particular including the original "Thom conjecture" for curves in $\mathbf{CP}^2$. Moreover the proofs are very much simpler than the older arguments using the instanton invariants. The primary result is the adjunction inequality for the Seiberg-Witten basic classes. If X is a 4-manifold with $b^+ > 1$ and $\Sigma \subset X$ an embedded, oriented surface with $\Sigma.\Sigma \geq 0$, then for any basic class $\kappa \in H^2(X)$:

$$2g - 2 \geq \Sigma.\Sigma + \kappa.\Sigma,$$

where g is the genus of Σ. To prove this consider first the case when $\Sigma.\Sigma = 0$. Then the boundary of a tubular neighbourhood of Σ is the 3-manifold $S^1 \times \Sigma$. According to the gluing theory outlined in Section 5, the Seiberg-Witten invariant of X and a class $\kappa \in H^2(X)$ can be computed from relative invariants taking values in a Floer group obtained from the solution of the three-dimensional Seiberg-Witten equation over $S^1 \times \Sigma$. These reduce to the vortex equations over Σ, as described in Section 5 above, with a line bundle of degree $d = \kappa.\Sigma$. If $\kappa.\Sigma \geq 2g - 2$, there are no solutions to the vortex equations so the Seiberg-Witten invariant of X must vanish and κ cannot be a basic class. In general, if $\Sigma.\Sigma = n > 0$, one can reduce to the case of 0 self-intersection by "blowing up". Let $\tilde{X}$ be the connected sum of X with n copies of $\overline{\mathbf{CP}}^2$, where one thinks of the connected sums as being made at n points of Σ.

Let $\tilde{\Sigma} \subset \tilde{X}$ be the surface obtained by taking a "connected sum" with copies of a projective line in $\overline{\mathbf{CP}}^2$—the "proper transform" of Σ in the complex case. Then $\tilde{\Sigma}$ has self-intersection 0 and, by the connected sum formula for basic classes, for each basic class κ on X there is a basic class $\tilde{\kappa}$ on $\tilde{X}$ with $\tilde{\kappa}.\tilde{\Sigma} = \kappa.\Sigma - n$.

If X is a Kahler surface with $b^+ > 1$, then the canonical class K_X is a basic class. On the other hand any complex curve C satisfies the classical adjunction equation

$$2g(C) - 2 = C.C + K_X.C,$$

so minimises the genus in its homology class. (By Taubes' work, the same holds for pseudo-holomorphic curves in symplectic 4-manifolds.) The proof of the original Thom conjecture for $\mathbf{CP}^2$ uses an extension of this argument taking into account the chamber structure of the Seiberg-Witten invariants in this case [KM3],[MST].

Constraints on intersection forms. The first applications of instantons to 4-manifold topology were not through the construction of invariants but, in the complementary direction, to show that certain intersection forms could not be realised by smooth 4-manifolds. Specifically [D1]:
1. non-standard definite forms can never be realised,
2. the even forms

$$\lambda E_8 + \mu \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

    for $\mu = 1, 2$ cannot be realised by 4-manifolds with no 2-torsion in the first homology.

The well-known conjecture in this direction—the "11/8-conjecture"—predicts that for spin manifolds one always has $\mu \geq (3/2)\lambda$, the connected sum of copies of the $K3$ surface being the extremal case. The results above confirm the conjecture for $\lambda \leq 2$, in the absence of 2-torsion.

These results have been re-proved and extended by Kronheimer and others using the Seiberg-Witten invariants. The simplest argument is to rule out even, negative-definite forms for a simply connected manifold. If X is such a manifold, it admits a spin structure and Kronheimer considers the corresponding moduli space of Seiberg-Witten solutions (i.e. for the trivial line bundle ξ). This has dimension $-2\tau(X) - 1$, where τ is the signature, which is minus the betti number $b_2(X)$ by hypothesis. There is one reducible solution corresponding to the trivial flat connection, and the structure of the moduli space around this point is modelled on a cone over a complex-projective space $\mathbf{CP}^m$, with $m = b_2 - 1$, assuming $b_2 > 1$. The cohomology class h gives an extension of the generator of $H^2(\mathbf{CP}^2)$ over the moduli space, and one gets a contradiction by considering the evaluation of h^m on the boundary of a truncated moduli space. The general scheme of this argument—considering the links of reducible solutions in the moduli space—is similar to those using the instanton moduli spaces. The argument extends to odd, definite forms Q provided one can find a *characteristic vector* c for Q with $-c^2 < \text{rank}(Q)$, and this has recently been shown to be true, for any non-standard form, by N. Elkies [E]. For even, indefinite forms, Kronheimer reproved the older results, with a slight improvement, dropping the hypothesis on 2-torsion. Kronheimer's proofs involve considering the links of the reducible solution in various cobordism theories, but the method seems to run into fundamental obstacles for larger values of μ. However, just at the time of writing, M. Furuta announced the proof of a "10/8" bound:

for any spin manifold $\mu \leq \lambda$. Furuta's proof uses the same general scheme and sophisticated algebro-topological calculations in the relevant cobordism theory.

Interesting generalisations of these results can be obtained by considering 4-manifolds with boundary. In the instanton theory one gets results which relate the intersection form to the Floer homology of the boundary. In particular if a homology 3-sphere Y bounds a 4-manifold with a non-standard negative definite intersection form, then the instanton Floer homology of Y is non-trivial. (Recent work of Froyshov uses these techniques to obtain more concrete results for specific 3-manifolds.) One can also use the Seiberg-Witten equations in the same way, but a new feature enters from the Atiyah-Patodi-Singer η-*invariant* of the boundary. For example the Poincaré sphere P bounds a manifold Z with the non-standard negative definite form E_8 but has trivial Seiberg-Witten-Floer groups, as we have seen. The reason one does not get a contradiction by applying Kronheimer's argument to Z (made into a complete manifold by attaching a semi-infinite tube) is that the index of the Dirac operator over Z is

$$\mathrm{ind}(D_Z) = (1/8)\sigma(Z) + \rho(P, g)$$

where $\rho(P, g))$ is the difference of the η and Chern-Simons invariants of P, for a metric g on P. The reduction of ρ modulo 2 is the Rohlin invariant. For the Poincaré sphere the Rohlin invariant is 1, so one always gets a discrepancy in the index calculation, compared with the case of a closed 4-manifold.

7. Concluding remarks

The Seiberg-Witten equations have led to astounding advances in 4-manifold theory. To some extent they may well have brought the study of the gauge theory invariants to a fairly complete form, resolving many of the main problems that drove research in this area in the last ten years. Perhaps the most exciting challenge for mathematicians is to come to grips with the Quantum Field Theory ideas which led to these new advances—in parallel with other celebrated developments such as mirror symmetry, 3- manifold invariants, conformal field theory etc.—and to understand in a rigorous way the intricate structures discovered by Seiberg and Witten. At the same time there are notable detailed questions which are left open at present. One is the question of whether all simply-connected manifolds are of simple type. A more wide-ranging problem is to understand the structure of the invariants, and the relation between the instanton and Seiberg-Witten theories, for 4-manifolds with $b^+ = 1$. Just as this article was going to print, L. Gottsche [Got] announced the solution of many of the long-standing problems involving the instanton theory, in this area, including complete formulae for the invariants of $\mathbf{CP}^2$, extending earlier calculations of Kotschick and Lisca [KL] and others. Gottsche's formulae involve modular forms, in at least general agreement with the picture predicted by Quantum Field Theory. One direction which has not yet been explored much is the study of invariants of families of 4-manifolds. By considering an r-dimensional family of equations, of either kind, one should get invariants which are roughly speaking cohomology classes in $H^r(B \, \mathrm{Diff} \, (X))$ where $B \, \mathrm{Diff} \, (X)$ is the classifying space of the diffeomeorphism group of a 4-manifold X [D4]. Then the same issues which complicate the story for ordinary invariants when $b^+ = 1$ should arise, for any X, once $r \geq b^+ - 1$. In another direction one may consider

4-manifolds which are not smooth. The instanton theory can be extended to the class of *quasi-conformal* 4-manifolds (where the co-ordinate change maps are only quasi-conformal, not necessarily smooth)[DS]. The Seiberg-Witten theory makes essential use of spinors and the Dirac equation, and it is known in other problems that these depend essentially upon a smooth (or at least C^1) structure [Su]. At the moment there are no examples known of quasi-conformal 4-manifolds which cannot be given a smooth structure, but it is possible that such examples exist and display new phenomena.

Finally we turn to questions of 4-manifold theory *per se*. Despite the great progress of the last ten years, many fundamental questions seem quite out of reach. One set of problems is to distinguish the various examples of 4-manifolds, such as the Catenese-Manetti surfaces, with the same gauge theory invariants—or prove that they are diffeomeorphic. (Until mid-1994 one might reasonably have hoped that these examples would already be distinguished by the instanton invariants.) Another is to prove or disprove the 11/8 conjecture. Still more ambitiously, one can dream of finding systematic *positive* techniques, say for constructing diffeomorphisms between manifolds, or finding embedded surfaces of a given genus, which would complement the obstructions and invariants which have been found. There are, perhaps, glimmers of hope in this direction in the case of symplectic 4-manifolds. In general, all these recent advances have given tremendous new insights into 4-dimensional geometry and topology but it is quite unclear how far they will go towards a complete understanding of the subject.

References

[AJ]   M. F. Atiyah and L. C. Jeffrey, *Topological Lagrangians and cohomology*, Jour. Geometry and Physics **7** (1990), 119–136. MR **92f:**58025

[A]   D. Austin, *Equivariant Floer Theory for binary polyhedral spaces*, Preprint.

[AB]   D. Austin and P. Braam, *Equivariant Floer cohomology*, Topology (To appear).

[BDGW] S. Bradlow, G. Daskalopoulos, O. Garcia-Prada, R. Wentworth, *Stable augmented bundles over Riemann surfaces*, Vector bundles in algebraic geometry (Ed. Hitchin, Newstead, Oxbury), Cambridge UP, 1995.

[D1]   S. Donaldson, *Connections, cohomology and the intersection forms of four manifolds*, Jour. Differential Geometry **24** (1986), 275–341. MR **88g:**57033

[D2]   S. Donaldson, *Irrationality and the h-cobordism conjecture*, Jour. Differential Geometry **26** (1987), 141-168. MR **88j:**57035

[D3]   S. Donaldson, *Polynomial invariants for smooth four-manifolds*, Topology **29** (1990), 257-315. MR **92a:**57035

[D4]   S. Donaldson, *Yang-Mills invariants of four-manifolds*, Geometry of low-dimensional manifolds, Vol I (Ed. Donaldson and Thomas), Cambridge UP, 1990.

[DK]   S. Donaldson and P. Kronheimer, *The geometry of four-manifolds*, Oxford UP, 1990. MR **92a:**57036

[DS]   S. Donaldson and D. Sullivan, *Quasiconformal four-manifolds*, Acta Math. **163** (1989), 181-252. MR **91d:**57012

[E]   N. Elkies, *A characterisation of the $\mathbf{Z}^n$-lattice*, Math. Res. Letters (To appear).

[FS]   R. Fintushel and R. Stern, *The blowup formula for Donaldson invariants*, Preprint (1994).

[FM]   R. Friedman and J. Morgan, *Algebraic surfaces and four-manifolds: some conjectures and speculations*, Bull. Amer. Math. Soc. **18** (1988), 1-19. MR **89g:**57046

[FQ]   R. Friedman and Z. Qin, *The smooth invariance of the Kodaira dimension of a complex surface*, Math. Res. Letters **1** (1994), 369-376. CMP 95:04

[Go]   R. Gompf, *A new construction of symplectic manifolds*, Preprint (1994).

[Got]   L. Gottsche, *Modular forms and Donaldson invariants for 4-manifolds with $b_+ = 1$*, Preprint (1995).

[Gr]   M. Gromov, *Pseudo-holomorphic curves in symplectic manifolds*, Inventiones Math. **82** (1985), 305-347. MR **87j:**53053

[GL]   M. Gromov and H. Lawson, *The classification of simply-connected manifolds of positive scalar curvature* **111** (1980), 423-434. MR **81h:**53036

[H]   N. Hitchin, *The self-duality equations on a Riemann surface*, Proc. Lond. Math. Soc. **55** (1987), 59-126. MR **89a:**32021

[JT]   A. Jaffe and C. Taubes, *Vortices and monopoles*, Birkhauser, 1980. MR **82m:**81051

[K]   P. Kronheimer, *Embedded surfaces in 4-manifolds*, Proc. Int. Congress Mathematicians, Kyoto 1990, 1990, pp. 529-539. MR **93d:**57066

[KL]   D. Kotschick and P.Lisca, *Instanton invariants of* $\mathbf{CP}^2$ *via topology*, preprint.

[KM1]   P. Kronheimer and T. Mrowka, *Gauge theory for embedded surfaces*, Topology **32** (1993), 773-826. MR **94k:**57048

[KM2]   P. Kronheimer and T. Mrowka, *Recurrence relations and asymptotics for four-manifold invariants*, Bull. Amer. Math. Soc. **30** (1994), 215–221. MR **94k:**57046

[KM3]   P. Kronheimer and T. Mrowka, *The genus of embedded surfaces in the projective plane*, Math. Res. Letters **1** (1994), 797-808. CMP 95:05

[L]   C. LeBrun, *Einstein metrics and Mostow rigidity*, Math. Res. Letters **2** (1995), 1-8. CMP 95:07

[M]   M. Manetti, *On some components of moduli spaces of surfaces of general type*, Compositio Math. **92** (1994), 285-297. CMP 94:15

[MST]   J. Morgan, Z. Szabo and C. Taubes, *The generalised Thom conjecture*, In preparation.

[PT]   V. Pidstragach and A. Tyurin, *Invariants of the smooth structure of an algebraic surface arising from the Dirac operator (English translation)*, Russian Math. Surveys (1994).

[S]   N. Seiberg, Phys.Lett. B **318** (1993), 469. MR **93m:**14036

[SW]   N. Seiberg and E. Witten, *Monopoles, duality and chiral symmetry breaking in* $N = 2$ *supersymmetric QCD*, Nucl. Phys. B **431** (1994), 581-640. CMP 95:05

[St]   A. Stipsicz, *Donaldson invariants of certain symplectic 4-manifolds*, Crelles Journal (To appear).

[Su]   D. Sullivan, *Exterior d, the local index and smoothability*, IHES preprint (1995).

[T1]   C. Taubes, *The Seiberg-Witten invariants and symplectic forms*, Math. Res. Letters **1** (1994), 809-822. CMP 95:05

[T2]   C. Taubes, *More constraints on symplectic forms from Seiberg-Witten equations*, Math. Res. Letters **2** (1995), 9-14. CMP 95:07

[T3]   C. Taubes, *The Seiberg-Witten and the Gromov invariants*, Harvard Preprint (1995).

[Th]   M. Thaddeus, *Stable pairs, linear systems and the Verlinde formula*, Inventiones Math. **117** (1994), 317-353. MR **95e:**14006

[VW]   C. Vafa and E. Witten, *A strong coupling test of S-duality*, Nucl. Phys. B (To appear). CMP 95:04

[W1]   E. Witten, *Topological Quantum Field Theory*, Commun. Math. Phys. **117** (1988), 353-386. MR **89m:**57037

[W2]   E. Witten, *Supersymmetric Yang-Mills theory on a four-manifold*, J. Math. Phys. **35** (1994). CMP 95:01

[W3]   E. Witten, *Monopoles and 4-manifolds*, Math. Res. Letters **1** (1994), 769-796. CMP 95:05

THE MATHEMATCAL INSTITUTE, OXFORD UNIVERSITY, 24-29 ST. GILES, OXFORD OX1 3LB, ENGLAND

E-mail address: `donaldson@vax.ox.ac.uk`