Topology Vol. 29, No. 3, pp. 257–315, 1990
Printed in Great Britain.

0040-9393/90 $3.00+.00

POLYNOMIAL INVARIANTS FOR SMOOTH FOUR-MANIFOLDS

S. K. DONALDSON

(*Received in revised form 6 March 1989*)

§I. INTRODUCTION

THE TRADITIONAL methods of geometric topology have not given a clear picture of the classification of smooth 4-manifolds. This gap has been partially bridged by the introduction into 4-manifold theory of methods using Yang–Mills theory (or Gauge theory). Riemannian 4-manifolds carry with them an array of *moduli* spaces, finite dimensional spaces of connections cut out by the first order Yang–Mills equations. These equations depend on the Riemannian geometry of the 4-manifold but at the level of homology we find properties of the moduli spaces which do not change as the metric is varied continuously. Any two Riemannian metrics can be joined by a path so, by default, these properties depend only upon the underlying smooth 4-manifold, and they furnish a mine of potential new differential topological invariants. This point of view was developed in [7], and the invariant defined there did indeed go beyond the classical ones. In fact Friedman and Morgan [13], and Okonek and Van de Ven [29] showed that this invariant could distinguish mutually distinct differentiable structures on an infinite family of homeomorphic 4-manifolds (Dolgachev surfaces).

This paper takes the same ideas further. In [7] a single moduli space was used to define an invariant for manifolds with b_+^2—the rank of a maximal positive subspace for the intersection form—equal to 1. Here we will use infinite families of moduli spaces to define an infinite set of invariants for a simply connected 4-manifold X with b_+^2 odd and strictly greater than 1. These invariants are distinguished elements of the ring $S^*(H^2(X))$ of *polynomials* in the cohomology of the underlying 4-manifold. They can be viewed equivalently as symmetric multilinear functions:

$$q\colon H_2(X) \times \ldots \times H_2(X) \to \mathbf{Z}.$$

We will define a family of such polynomials $q_{k,X}$, where k is a sufficiently large integer, and the degree of $q_{k,X}$ grows linearly with k (see §III below). Certainly one of the most striking facts is that we get infinitely many invariants for a single manifold. Discovering to what extent these are independent (i.e., whether there are universal relations between the invariants) is an interesting target for future research.

The restriction in our work to manifolds with b_+^2 odd is made for a very simple reason. In essence our invariants are defined by the fundamental homology classes $[M]$ of the Yang–Mills moduli spaces within the ambient space $\mathscr{B}_X^*$ of SU(2) connections modulo equivalence (see §II for a summary of notation and basic facts from Yang–Mills Theory). The methods of Atiyah and Bott [1], §I (see also [4], §II) show that the rational cohomology ring of $\mathscr{B}_X^*$ is a polynomial algebra with generators in dimensions 2 and 4. The two-dimensional generators are obtained from the homology of X via a map:

$$\mu\colon H_2(X) \to H^2(\mathscr{B}_X^*).$$

[The additional generator in dimension 4 can be seen as arising in the same fashion from $H_0(X)$.] In particular, all rational cohomology of $\mathscr{B}^*_X$ lies in even dimensions. If our moduli space M is even dimensional we can obtain invariants by, roughly speaking, pairing the fundamental class $[M]$ with the appropriate cohomology of $\mathscr{B}^*_X$, which leads then to a polynomial function on the homology of X. The dimension of the Yang–Mills moduli space of connections on a bundle P over X, with structure group G, is given by a general formula of the shape:

$$\dim M = 4a(G)k - \dim G(1 + b^2_+(X)),$$

where $a(G)$ is an integer depending on G and the integer k is a topological invariant of the G-bundle P. So when $G = \mathrm{SU}(2)$ (or any odd dimensional Lie group) this dimension is even precisely when b^2_+ is odd. The invariant $q_{k,X}$ is associated to an SU(2) bundle with $c_2 = k$.

It would certainly be possible to define other invariants using the techniques we develop in this paper, and also to consider larger classes of 4-manifolds. In one direction, one could look at manifolds which are not simply connected. In another, one could consider the moduli spaces of connections for more general structure groups G (and it is possible that, by considering even dimensional Lie groups, one could obtain results about manifolds with b^2_+ even). A third extension would be to consider invariants defined by other cohomology classes on $\mathscr{B}^*_X$. We have chosen in this paper to stick to the simplest interesting and reasonably general class of manifolds and invariants for which we can give concrete applications. Indeed even within this class the invariants we obtain are still rather mysterious because of the very small number of cases where explicit calculations have been carried out. One such calculation is discussed in §VI. Otherwise, our results are obtained by appeal to two contrasting general properties of the invariants, and the main purpose of this paper is to develop these two properties in detail.

As in [7] the main application of our work is to the topology of complex algebraic surfaces. We shall prove the following general theorem.

THEOREM A. *Let* S *be a smooth, simply connected, complex projective surface. If the oriented* C^∞ *4-manifold underlying* S *can be decomposed as a smooth, oriented connected sum of oriented 4-manifolds* Y_1, Y_2 *then one of* Y_1, Y_2 *has a negative definite intersection form.*

Informally, this says that the positive part of the homology of a simply connected surface is "indivisible" in connected sums. This theorem gives many new examples of homeomorphic 4-manifolds with distinct smooth structures. Consider, for example, a smooth hypersurface S_d in $\mathbb{CP}^3$ of odd degree d. This surface is simply connected, by the Lefschetz theorem, and it has:

$$b^2_+(S_d) = \alpha_d = (1/3)(d-1)(d-2)\ (d-3) + 1$$
$$b^2_-(S_d) = \beta_d = (1/3)(d-1)(2d^2 - 4d + 3).$$

The intersection form is odd so, by Freedman's classification, S_d is homeomorphic to a connected sum of α_d copies of $\mathbb{CP}^2$ (standard orientation) and β_d copies of $\overline{\mathbb{CP}}^2$ ($\mathbb{CP}^2$ with orientation reversed.) But if $d \geqslant 5$, so $\alpha_d > 1$, these two manifolds cannot be diffeomorphic, by Theorem A.

Combining Theorem A with the Theorem of [3] we see that in any smooth connected sum decomposition of a simply connected complex algebraic surface S one of the factors has a diagonalisable intersection form, with an integral basis for the homology consisting of class of self-intersection -1. Conversely, there certainly are surfaces for which such decompositions exist: we can just take S to be a multiple blow-up of another surface.

(Blowing up a point in a complex surface corresponds, differentiably, to taking a connected sum with $\overline{\mathbb{CP}}^2$.) It is tempting to conjecture that this is the only possibility—i.e. that a minimal model in the algebro-geometrical sense is indecomposable in the differential topological sense. (There is an extensive discussion of the topology of complex surfaces, and connected sum decompositions, in the survey article [21].)

As we have stated above, Theorem A follows from two contrasting properties of the Yang–Mills polynomial invariants. These properties are as follows.

THEOREM B. [THEOREM (4.9) IN §IV BELOW]. *Suppose* X *is a simply connected, oriented* 4-*manifold with* b_+^2 *odd and there is an orientation-preserving diffeomorphism between* X_2 *and an oriented connected sum of manifolds* Y_1, Y_2 *where both* $b_+^2(Y_1)$ *and* $b_+^2(Y_2)$ *are strictly positive. Then all the invariants* $q_{k,X}$ *are* 0.

THEOREM C [THEOREM (5.1) IN §V BELOW]. *If* S *is a simply connected complex projective surface and* H *is a hyperplane class in* $H_2(S)$ *then for large enough integers k we have*:

$$q_{k,S}(H, \ldots, H) > 0.$$

(Recall that H is a hyperplane class if it can be realised by a hyperplane section of some embedding of S in projective space.) Theorem A is an immediate consequence of these two results about the polynomial invariants.

The detailed proof of Theorem B is long and technical but, as we explain in §IV(ii), this vanishing phenomenon stems from two simple and general topological mechanisms. The theorem also fits in very tidily with results obtained, more than 20 years ago, by Wall [35, 36]. Wall proved:

(i) If Y is a simply connected, smooth 4-manifold with an indefinite intersection form and Z is the connected sum of Y with $S^2 \times S^2$ then all automorphisms of $H_2(Z)$ which preserve the intersection form can be realised by self-diffeomorphisms of Z.

(ii) If Y' is another manifold which is homotopy equivalent to Y then for large enough r the connected sums $Y \# r(S^2 \times S^2)$, $Y' \# r(S^2 \times S^2)$ are diffeomorphic.

One can use (i) to deduce many special cases of Theorem B, sufficient for most applications, arguing that the zero polynomial is the only one with the required symmetry (compare §VI below). Equally we see that Theorem B is consistent with (ii)—taking connected sums with $S^2 \times S^2$ kills the polynomials and allows distinct manifolds to become "stably" diffeomorphic. (Similar patterns appear for the Dolgachev surfaces, see [13].) On the way to the proof of Theorem B we give an "excision" property of the invariants [Theorem (4.8)] which shows that one cannot hope to use them to detect distinct homotopy spheres, for example, by adding on homology in a connected sum.

Turning now to Theorem C: the proof here is again long and complicated but, as we explain at the beginning of §V, this positivity again stems from simple and general facts about the Yang–Mills moduli spaces when the underlying 4-manifold is a complex surface. In that case the moduli spaces can be identified with moduli spaces of holomorphic bundles, and hence can be studied algebro-geometrically. The technical aspects of the proof hinge on a result about the singularities of moduli spaces of holomorphic bundles [Theorem (5.8)] which may be of interest in its own right. In turn, the proof of this theorem lead us into a digression to discuss facts about holomorphic families of matrices in two variables [Appendix 3 and §V(vi)] which the author has not been able to find in the literature. We close the paper by discussing, in §VI, almost complex structures and diffeomorphisms of K3 surfaces.

§II. REVIEW OF TECHNIQUES AND NOTATIONS

We will make a special definition for the purposes of this paper.

Definition (2.1). A C-manifold is a pair (X, β) where:

(i) X is a smooth, compact, simply connected, oriented 4-manifold with $b_2^+(X) = 2a + 1$, for a strictly positive integer a.

(ii) β is an orientation of a maximal positive subspace $H^2_+ \subset H^2(X;R)$ for the intersection form (so $\dim H^2_+ = b^2_+$).

Note that in part (ii) the choice of subspace H^2_+ is immaterial, since the space of positive subspaces is contractible. A better definition is this: let N be the null cone of the intersection form, minus the origin. N is homotopy equivalent to a product of spheres and β can be viewed as a choice of generator for the relevant part of the homology of N. As we shall see in §VI the orientation data is closely related, at the level of homotopy, to almost complex structures on 4-manifolds. For the remainder of §II and §III, we fix a C-manifold (X, β) (and, throughout the paper, we will often suppress the orientation to simplify our notation).

(i) Connections

In Yang–Mills Theory we consider the space $\mathscr{A}$ of connections on a principal bundle $P \to X$. For the moment we fix attention on SU(2) bundles, determined up to isomorphism by an integer $k = \langle c_2(P), [X] \rangle$. Denote by $\mathscr{B}_k$ (or sometimes $\mathscr{B}_X$, $\mathscr{B}_P$ or $\mathscr{B}_{k,X}$, depending on the context) the quotient $\mathscr{A}/\mathscr{G}$ of $\mathscr{A}$ by the gauge group $\mathscr{G} = \operatorname{Aut} P$. $\mathscr{B}_k$ is the space of gauge equivalence classes of connections. We write $[A]$ for the equivalence class containing a connection A. Let $\mathscr{B}_k^*$ be the open subset of $\mathscr{B}_k$ consisting of orbits of irreducible connections. The trivial bundle, with $k = 0$, is distinguished by the fact that it supports the flat reducible connection θ, otherwise all the reducible connections have holonomy group S^1. When appropriate we consider a set-up based on Sobolev spaces, with connections of class L_1^p and gauge transformations of class L_2^p for some fixed large even integer exponent p (for example $p = 4$). Then the $\mathscr{B}_k^*$ are Banach manifolds (see [10], §3 and Appendix A, for example). Similarly we work with the space $\mathscr{R}$ of Riemannian metrics on X of class C^r, for some fixed r (for example $r = 2$). By a path in the space of metrics we mean a C^1 map from the unit interval to $\mathscr{R}$.

(ii) Moduli spaces and transversality

Let g be a Riemannian metric on X. For each $k \geq 0$ the moduli space $M_k(g) \subset \mathscr{B}_k$ is the set of gauge equivalence classes of anti-self-dual (ASD) connections. These are the connections A for which the self-dual part $F^+(A)$ of the curvature is zero. They are characterised by the equality:

$$\int_X |F(A)|^2 \, d\mu = 8\pi^2 k, \quad \text{if } A \text{ is ASD.} \tag{2.2}$$

It follows that when $k = 0$ the moduli space is the single point $[\theta]$.

For $k > 0$ the moduli spaces are more complicated finite dimensional spaces which we think of as being cut out by the Fredholm ASD equations. Viewed in this way, we attach to each bundle P the "virtual" dimension of the corresponding moduli space. We will sometimes write this $V.d.(M_{k,X})$ to emphasise that it may not coincide with the actual dimension of the moduli space, for some metrics. The main properties of this virtual dimension are summarised in the next proposition.

PROPOSITION (2.3). (a) ([12], §2) *The virtual dimension of the moduli space* $M_k(g)$ is

$$8k-6(1+a) = 8k-3(1+b_+^2(X)).$$

(b) ([12], §3) *There is a second category subset of* $\mathscr{R}$ *(the "generic metrics") consisting of metrics* g *for which all the moduli spaces* $M_k(g)$ $(k > 0)$ *are smooth submanifolds of the* $\mathscr{B}_k^*$, *cut out transversely by the ASD equations and thus of dimension equal to their virtual dimension.*

(c) *Let* g_t *be a path in* R *whose end points* g_0, g_1 *are generic. Then* g_t *can be aproximated arbitrarily closely by a path* g_t^* *having the same end points and also the following transversality property. Let* Z *be the parametrised moduli space:*

$$Z = \{([A], t) \in \mathscr{B}_k^* \times [0, 1] \mid [A] \in M_k(g_t^*)\},$$

then Z *is a smooth manifold-with-boundary,* $\dim Z = \dim M_k(g_0) + 1$ *and the boundary of* Z *is the disjoint union of* $M_k(g_0)$ *and* $M_k(g_1)$.

Only part (c) requires comment. The proof is much the same as Freed and Uhlenbeck's proof of (b). First one shows that we can arrange that all moduli spaces encountered in the path lie in $\mathscr{B}_k^*$. This uses the fact that $b_2^+(X) > 1$, see [4], §6. For fixed k the same is true for nearby paths. Freed and Uhlenbeck's basic result is that for each $k > 0$ the universal moduli space of irreducible solutions, $\mathscr{M}_k \subset \mathscr{B}_k^* \times \mathscr{R}$ say, is a manifold. Our problem is to choose $g^*: [0, 1] \to \mathscr{R}$, close to g and transverse to the projection map from $\mathscr{M}_k$ to $\mathscr{R}$. This can be done using the appropriate version of Sard's theorem ([30], Theorem (3.1)).

This result of Freed and Uhlenbeck is very convenient. It means in particular that for generic g all the moduli spaces $M_k(g)$ for $k > 0$ and with *negative* virtual dimension are empty. Of course when $k = 0$ the trivial solution is immoveable. It is not, however, really essential to our purposes to appeal to Freed and Uhlenbeck's rather difficult result. One can achieve the same ends by making artificial perturbations to the ASD equations [3, 6]. We shall have to do this in §IV of this paper.

(iii) Orientation

The orientation β involved in the definition of a C-manifold serves to orient all the moduli spaces $M_k(g)$. This is done in [6]. We use the "standard orientation" of [6], (3.20), (3.24). In general this depends upon an orientation of $\Lambda^{\max} H^1(X) \otimes \Lambda^{\max}(H^0(X) \oplus H_+^2(X))$. In our case H^1 is zero and we use the orientation $1 \otimes \beta$ of the second term. Changing β to $-\beta$ reverses the orientation of all the moduli spaces.

(iv) Compactness

We now fix a generic metric g on X and write M_k for $M_k(g)$. Uhlenbeck's Theorem [33] yields a natural compactification of M_k ([4], §III):

$$\bar{M}_k = \bar{M}_{k,X}(g) \subset M_k \cup (M_{k-1} \times X) \cup (M_{k-2} \times S^2(X)) \cup \ldots.$$

$\bar{M}_k$ is a compact metrisable space. The key property of its topology is this: if a sequence $[A_\alpha]$ in M_k converges to a limit $([A_\infty]; x_1, \ldots x_n)$ in $M_{k-n} \times S^n(X)$ then, after gauge transformations over the punctured manifold, the A_α converge in C^∞ to A_∞ over $X \backslash \{x_1, \ldots x_n\}$. We will refer to this kind of convergence as *weak convergence*. There is an obvious extension of these ideas to the case when we consider a compact set T of metrics on X and compactify the parametrised moduli space in $\mathscr{B}_k \times T$, by adjoining points of $\mathscr{B}_{k-n} \times T \times S^n(X)$.

(v) Cohomology

The maps $\mu: H_2(X;\mathbf{Z}) \to H^2(\mathscr{B}_k^*;\mathbf{Z})$, mentioned in the Introduction, were defined in [4], §II. We will use the representation of these cohomology classes $\mu(\alpha)$ by codimension 2-submanifolds, much as developed in [4], §III. However, as the referee of the present paper has kindly pointed out to the author, there is a difficulty with the detailed procedure adopted in [4], so, to get around this point, we will now develop a slightly different construction.

Let Σ be an oriented surface embedded in X, and N be an open neighbourhood of Σ in X with smooth boundary such that $H_2(N) = \mathbf{Z}[\Sigma]$. We can form a space $\mathscr{B}_N = \mathscr{A}_N/\mathscr{G}_N$ of L_1^p connections over N modulo gauge equivalence. The open subset $\mathscr{B}_N^*$ of irreducible connections is a Banach manifold (we can use Neumann boundary conditions to construct slices for the action, just as in [33]). The complement $\mathscr{B}_N\backslash\mathscr{B}_N^*$ has components labelled by the degree of the reduction. We let $\mathscr{B}_N^+$ denote the union of $\mathscr{B}_N^*$ and the component of the complement consisting of reductions of degree zero; $\mathscr{B}_N^+$ is again an open subset of $\mathscr{B}_N$.

Over $\mathscr{A}_N$ there is a $\mathscr{G}_N$-equivariant line bundle $\tilde{\mathscr{L}}_N$, which descends to a line bundle $\mathscr{L}_N$ over $\mathscr{B}_N^+$ such that $\mu([\Sigma])$ is the pullback of $c_1(\mathscr{L}_N)$ to $H^2(\mathscr{B}_{k,X}^*)$ (cf. [4], §II). More precisely, let

$$R_N: \mathscr{B}_{k,X} \to \mathscr{B}_N$$

be the continuous map induced by restriction of connections over X to N. Then the restriction of $\mu([\Sigma])$ to $R_N^{-1}(\mathscr{B}_N^+)$ is $R_N^*(c_1(\mathscr{L}_N))$. The annoying complication—and this is the point which was not treated adequately in [4]—is that there are irreducible connections over X which become reducible when restricted to N so we only get a description of $\mu([\Sigma])$ over a subset of $\mathscr{B}_k^*$ by this restriction procedure. This difficulty should certainly not be taken too seriously: for example, if the metric g is real analytic the ASD solutions will be also, and in that case an irreducible ASD connection cannot be reducible on any non-empty open set, so the difficulty does not arise for the ASD connections we are ultimately interested in. It is not too hard to modify the theory so that one always works with real analytic metrics (see (3.6), (3.7) below). However, here we will adopt another approach, appealing to the following result which is proved in Appendix 0.

PROPOSITION (2.4). *For any given* k *and homology class* α *in* $H_2(X;Z)$ *there is a representative* Σ *of* α *and a neighbourhood* N *of* Σ *such that for all* $j \le k$ *the image of* M_j *under the restriction map* R_N *lies in* $\mathscr{B}_N^+$. *Moreover for any finite collection* $\alpha_1, \dots \alpha_d$ *of homology classes we can find surfaces* Σ_ν *and neighbourhoods* $N_\nu (\nu = 1, \dots, d)$ *having the above property and such that all triple intersections* $N_\lambda \cap N_\mu \cap N_\nu$ *(λ, μ, ν distinct) are empty.*

Now, continuing with the notation above, let s be a smooth $\mathscr{G}_N$—equivariant section of $\tilde{\mathscr{L}}_N$ over $\mathscr{A}_N$. The zeros of s define a closed subset V of the quotient $\mathscr{B}_N$, which necessarily contains the non-trivial reductions $\mathscr{B}_N\backslash\mathscr{B}_N^+$ (reflecting the fact that $\mathscr{L}_N$ does not extend over all of $\mathscr{B}_N$). By abuse of notation we will denote also by V the closed subsets $R_N^{-1}(V)$ of $\mathscr{B}_{j,X}$ for all the different j. So we have closed subsets $M_j \cap V$ of the different moduli spaces. If we choose N as in (2.4) then, for $j \le k$, these are the zero sets of smooth sections, which we will denote again by s, of the line bundles $R_N^*(\mathscr{L}_N)^l M_j$. Also, if these sections vanish transversely on the moduli spaces, the zero sets $M_j \cap V$ are smooth, codimension 2, submanifolds Poincaré dual to the cohomology classes $\mu([\Sigma])$. (For $j = 0$ we should take this to mean that the intersection is empty, i.e. that s does not vanish on the trivial flat connection θ.)

The existence of the sections of these line bundles with the desired transversality properties is spelled out in the following proposition.

PROPOSITION (2.5). *If* N *is chosen as in* (2.4) *there is a smooth section* s *of* $\tilde{\mathscr{L}}_N$ *such that for all* $j \le k$ *the induced section of* $R_N^*(\mathscr{L}_N)$ *over* M_j *vanishes transversely on a codimension* 2 *submanifold* $M_j \cap V$. *Moreover, if* $N_1, \ldots N_d$ *is a finite collection of such neighbourhoods, as in* (2.4), *then we can choose sections* $s_1, \ldots s_d$ *with zero sets* $V_1, \ldots, V_d$ *such that all multiple intersections:*

$$M_j \cap V_{v(1)} \cap \ldots \cap V_{v(p)}, \; \{v(1), \ldots, v(p)\} \subset \{1, \ldots, d\}$$

are transverse.

This is proved by a straightforward modification of the argument in [4] [Lemma (3.17)]. The fact that the exponent p defining our Sobolev spaces is an even integer is used here to give the existence of smooth, $\mathscr{G}_N$-invariant partitions of unity over $\mathscr{A}_N$.

The key point about the constructions above is that we get representatives for the classes $\mu(\alpha)$ over all of the moduli spaces simultaneously and these depend only on the restriction of connections to the tubular neighbourhoods N. The precise form in which we shall use this fact is as follows: if a sequence $[A_\alpha]$ lies in $M_k \cap V$ and $[A_\alpha]$ converges weakly to a limit $([A_\infty]; x_1, \ldots, x_n)$ and none of the x_r lie in N then $[A_\infty]$ is in $M_{k-n} \cap V$. This follows immediately from the fact that V is closed and the connections converge in C^∞ over N.

(vi) SO(3) connections

The range of invariants we are able to define is increased substantially by considering SO(3) connections. In a slightly different context the value of these connections has been shown by the work of Fintushel and Stern [9, 10]. In the present paper they will appear in the detailed applications in §VI.

The SO(3) bundles $P \to X$ are classified topologically by $w_2(P)$ in $H^2(X; \mathbf{Z}/2)$ and $p_1(P)$ in $H^4(X; \mathbf{Z})$; these characteristic classes can be specified independently, subject to the condition $w_2^2 = p_1 \bmod 4$. When w_2 is zero the bundle lifts uniquely to an SU(2) bundle with $c_2 = -\frac{1}{4}p_1$, and similarly for the connections on the bundle. In any case the theory goes through much as before, with only two significant changes.

First, if w_2 is non-zero, the trivial connection does not appear in the moduli spaces, whatever the value of p_1. So we can work throughout with irreducible connections. This simplifies many arguments. In fact if w_2 is non-zero the moduli space is empty for all $p_1 \ge 0$, by the SO(3) version of (2.2).

Second, the orientation of the moduli spaces is more complicated. Let w be a class in $H^2(X; \mathbf{Z}/2)$ and divide the integral lifts c of w into equivalence classes by the relation:

$$c \sim c' \text{ if } \{\tfrac{1}{2}(c - c')\} \text{ is even.}$$

When X is spin there is only one equivalence class, otherwise there are two. An orientation of a moduli space of SO(3) connections can be specified by the given orientation β of H_+^2 and a choice of equivalence class $[c]$ of integral lifts of w_2. In the notation of [6], §III(e), we choose a line bundle L over X with $c_1(L)$ representing $[c]$ and fix the orientation $o(L, \mathbb{C}, 1 \otimes \beta)$. Changing β or $[c]$, where possible, changes the orientation of the moduli space.

We will denote by $M_{k,w}$ the moduli space of ASD SO(3) connections on a bundle with $w_2 = w$ and $p_1 = -4k$. So k need not be an integer. The two-dimensional characteristic class w_2 is preserved under weak limits and we have a compactification:

$$\bar{M}_{k,w} \subset M_{k,w} \cup (M_{k-1,w} \times X) \cup (M_{k-2,w} \times S^2(X)) \ldots\ldots$$

We can define a map μ and representatives V much as before; the only additional complication comes from certain divisibility questions. Working over the rationals we can define:

$$\mu(\alpha) = -\tfrac{1}{4}p_1(\mathbf{P})/\alpha$$

Where $\mathbf{P}$ is the universal SO(3) bundle over the product space $\mathscr{B}^* \times X$ [cf. [4], § II (ii)].

§III. DEFINITION OF INVARIANTS

Let (X, β) be a C-manifold, fix $k > 0$ and write the virtual dimension of $M_{k,X}$ as $2d$, so $d = d(a, k) = 4k - 3(1 + a)$. We will define an invariant $q_{k,X}$, a polynomial of degree d in $H^2(X)$, assuming that k is in a "stable range" determined by a. Begin by fixing a generic metric g on X, homology classes $\alpha_1, \ldots \alpha_d$ in $H_2(X)$, neighbourhoods $N_1, \ldots, N_d$ and representatives $V_1, \ldots, V_d$, transverse to all moduli spaces, as in Propositions (2.4) and (2.5).

We will use one elementary argument, in slightly different contexts, many times in this paper. The only substantial ingredient in it is Uhlenbeck's compactness theorem. It is much the same as the argument used in [4], §III (iv).

LEMMA (3.1). *If* $k > (3/2)(1 + a) = (3/4)(1 + b_+^2(X))$ *then the intersection* $V_1 \cap V_2 \cap \ldots \cap V_d \cap M_k$ *is compact.*

Proof. Let $[A_\alpha]$ be an infinite sequence in this intersection. Going to a subsequence we can suppose that $[A_\alpha]$ is weakly convergent to a point $([A_\infty]; x_1, \ldots x_n)$ in the compactified moduli space. We have to show that n is zero, for then $[A_\infty]$ is in the intersection and $[A_\alpha]$ converges strongly to $[A_\infty]$ so any sequence in the intersection has a convergent subsequence and hence the intersection is compact.

Now we know that for each v *either* $[A_\infty]$ lies in V_v *or* some point x_r lies in the closure of N_v. This is an immediate consequence of the way that the V_v were defined (see the remarks at the end of §II (v) above). On the other hand no point x_r can lie in more than two of the $\bar{N}_v$ by the last property in (2.4). So suppose first that $0 < n < k$. Then of the d neighbourhoods N_v at most $2n$ are accounted for by the points x_r. So $[A_\infty]$ lies in the common intersection of at least $d - 2n$ of the V_v with M_{k-n}. By our general position assumptions we must have:

$$\dim M_{k-n} \geq 2(d - 2n).$$

But in fact we know that the dimension of M_{k-n} is equal to the virtual dimension, since we supposed that $k - n > 0$,

$$\dim M_{k-n} = 8(k - n) - 3(1 + b_+^2(X)) = 2d - 8n < 2(d - 2n),$$

and we obtain a contradiction.

Now consider the case when $n = k$, so A_∞ is the trivial flat connection which lies in none of the V_v. We must then have

$$2k \geq d = 4k - 3(1 + a),$$

and this contradicts the hypothesis that $k > (3/2)(1 + a)$. So the only possibility is $n = 0$ and the proof is complete.

From now on, when making arguments like that in the proof above we will just say "by dimension counting". The argument can be viewed as saying that when k is large enough the lower strata $M_{k-n} \times S^n(X)$ $(n > 0)$ in the compactified moduli space have codimension at least 4, so by general position the intersection considered above should lie in the interior or the open stratum. However, the approach we have taken is more elementary and avoids discussing the detailed structure of the compactified moduli space. (This situation, in the "stable range" of large k, is thus simpler than that of [7], where the boundary had codimension 1 and a corresponding correction term was required.)

Now the moduli space M_k has dimension $2d$ and so, since all the intersections are chosen to be transverse, the intersection $M_k \cap V_1 \cap \ldots \cap V_d$ is a finite set of points. To each point we attach a sign ± 1, using the orientations of the moduli space and of the normal bundles to the V_v furnished by the line bundles used in their definition. Then we set:

$$Q_k(V_1, \ldots V_d; g) = \text{total number of points in } M_k \cap V_1 \cap \ldots \cap V_d \quad \text{(counted with signs)}. \tag{3.2}$$

This is all standard differential topology. Another way of expressing the definition is that the V_v or, better, cochains ω_v supported in small neighbourhoods of the V_v, give representatives for the cohomology classes $\mu(\alpha_v)$ in the ambient space $\mathscr{B}_k^*$ such that $\cap$ supp(ω_v) meets M_k in a compact set. Then we get a representative for the cup product $\mu(\alpha_1) \cup \ldots \cup \mu(\alpha_d)$ which restricts to the compactly supported cochains on M_k. We are then evaluating this compactly supported cohomology class on the fundamental class of M_k.

Our next task is to show that the integers defined by (3.2) actually yield differential topological invariants of X. We could formulate this in the same style as the remarks above. One has to show first that any two choices for the compactly supported representatives of $\mu(\alpha_1) \cup \ldots \cup \mu(\alpha_d)$ yield the same class in $H^{2d}_{\text{comp}}(M_k)$. Second, that different moduli spaces $M_k(g_0)$, $M_k(g_1)$ differ by the boundary of a locally finite $(2d+1)$ chain Z over which the cohomology class extends, with compact support. However, we shall continue to formulate the work in differential topological language, considering submanifolds cut out of various moduli spaces.

LEMMA (3.3). *If* $k > \frac{3}{2}(1 + a)$ *then:*

(i) $Q_k(V_1, \ldots V_d; g)$ *depends only on the homology classes* $\alpha_1, \ldots, \alpha_d$ *and defines a symmetric multilinear map:*

$$q_k^{(g)}: H_2(X) \times \ldots H_2(X) \to \mathbf{Z},$$

$$q_k^{(g)}(\alpha_1, \ldots \alpha_d) = Q_k(V_1, \ldots V_d; g).$$

(ii) $q_k^{(g)}$ *is independent of the generic metric* g *on* X.

Proof. (i) It suffices to show that if Σ_1, Σ_1' and ${}_1\Sigma''$ are surfaces in X with $[\Sigma_1] = [\Sigma_1'] + [\Sigma_1'']$ in homology and if V_1, V_1', V_1'' are codimension 2 representatives defined using corresponding neighbourhoods N_1, N_1', N_1'' and if $V_2, \ldots V_d$ are similarly defined using neighbourhoods $N_2, \ldots N_d$ [such that each of the three d-tuples satisfy the genericity conditions of (2.4), (2.5)], then:

$$Q_k(V_1, \ldots V_d; g) = Q_k(V_1', \ldots V_d; g) + Q_k(V_1'', \ldots V_d; g).$$

Choose a 3-chain K in X with boundary $\Sigma_1 - \Sigma_1' - \Sigma_1''$ and not meeting the intersections $N_j \cap N_k$ for $j, k > 1$. Let U be a neighbourhood of K containing N_1, N_1', N_1'', but not meeting any of the other double intersections. The restriction map from $\mathscr{B}_k$ to $\mathscr{B}_{N_1}$ factors through $\mathscr{B}_W$, and similarly for N_1', N_1''. We can regard the three line bundles we have associated to the neighbourhoods as line bundles, $\mathscr{L}$, $\mathscr{L}'$, $\mathscr{L}''$ say over $\mathscr{B}_W^+$. The homology definition of the μ-map ([4], §II) shows that $\mathscr{L}$ is isomorphic to $\mathscr{L}' \otimes \mathscr{L}''$ (since $[\Sigma_1]$

$= [\Sigma_1'] + [\Sigma_1'']$ in $H_2(W)$). Fixing such an isomorphism, we find a smooth section σ of $\pi_1^*(\mathscr{L})$ over $W \times [0, 1]$ equal to s_1 (the section cutting out V_1) over $W \times \{0\}$ and to $s_1' \otimes s_1''$ over $W \times \{1\}$. Let U be the zero set of σ, and choose σ in general position, in the manner of (2.5), relative to all of the intersections of the V_v with all $M_j (v > 1, j \leqslant k)$. (More precisely, of course, we mean the corresponding submanifolds in $M_j \times [0, 1]$.) Then, repeating our dimension counting argument we find that

$$U \cap V_2 \cap \ldots \cap V_d \cap (M_k \times [0, 1])$$

is a compact 1-manifold with boundary. Counting the boundary points, with orientations, we get the desired equality. (Here again we make an abuse of notation by identifying the V_v with their pre-images under the projection map from the $\mathscr{B}_j^* \times [0, 1]$ to $\mathscr{B}_j^*$.)

(ii) Any two metrics on X can be joined by a path (e.g. the linear path). If g_0, g_1 are generic metrics on X we choose a generic path g_t^* between them as in (2.3). Then we have parametrised moduli spaces $Z_j \subset \mathscr{B}_j^* \times [0, 1]$ for $0 < j \leq k$. We can choose the V_v so that all the multiple intersections of the V_v with the Z_j are transverse. This requires extensions of (2.4) and (2.5) to parametrised families which can safely be left to the reader to formulate and prove. Then by the same dimension counting argument (applying Uhlenbeck's Theorem for the compact family of metrics g_t^*) we find that $V_1 \cap \ldots V_d \cap Z_k$ is a compact 1-manifold with boundary, and this gives the desired homology between the two intersections.

We have now achieved the first goal of this paper, the definition of an infinite family of differential topological invariants of C-manifolds. From now on we just write q_k (or $q_{k,X}$ or $q_{k,X,\beta}$ as the context demands) for the invariants $q_k^{(g)}$ of (3.3). To sum up we have the following.

THEOREM (3.4). *Let* (X, β) *be a C-manifold and* k *be an integer bigger than* $(3/4)(1 + \mathrm{b}_2^+(\mathrm{X})) = (3/2)(1 + \mathrm{a})$. *There is a unique polynomial* $\mathrm{q}_{\mathrm{k},\mathrm{X},\beta}$ *in* $\mathrm{H}^2(\mathrm{X})$, *of degree* $\mathrm{d} = 4\mathrm{k} - 3(1 + \mathrm{a})$ *such that if* g *is a generic metric on* X *and* $\{\mathrm{V}_v\}$ *are generic representatives for classes* $\mu(\alpha_v)$, *as in* (2.4), (2.5), *then:*

$$\mathrm{q}_{\mathrm{k},\mathrm{X},\beta}(\alpha_1, \ldots \alpha_\mathrm{d}) = \mathrm{Q}_\mathrm{k}(\alpha_1, \ldots \alpha_\mathrm{d}; \mathrm{g}).$$

If $\mathrm{f}\colon \mathrm{X} \to \mathrm{X}'$ *is an orientation-preserving diffeomorphism and if* $\beta = \mathrm{f}^*(\beta')$ *then* $\mathrm{q}_{\mathrm{k},\mathrm{X},\beta} = \mathrm{f}^*(\mathrm{q}_{\mathrm{k},\mathrm{X}',\beta'})$. *Also, if* $-\beta$ *is the opposite orientation then* $q_{\mathrm{k},\mathrm{X},-\beta} = -\mathrm{q}_{\mathrm{k},\mathrm{X},\beta}$.

Here, of course, we are identifying symmetric multilinear maps on H_2 with polynomials in the dual space H^2.

We can carry all of these constructions over to SO(3) connections on a bundle P over X with $w_2(P) = w$ not zero. To fix signs we have to choose an orientation $[c]$ of w. On the other hand we do not need to impose any restriction on k, since we do not encounter any reducible connections under weak limits. The outcome is summarised by the following theorem.

THEOREM (3.5). *Let* (X, β) *be a C-manifold,* w *a non-zero class in* $\mathrm{H}_2(\mathrm{X}; \mathbf{Z}/2)$, $[\mathrm{c}]$ *an orientation of* w, *and* $\mathrm{k} \in \frac{1}{4}\mathbf{Z}$ *a strictly positive number such that* $4\mathrm{k} = \mathrm{w}^2 (mod\, 4)$. *The anti-self-dual moduli spaces* $\mathrm{M}_{\mathrm{k},\mathrm{w}}$ *define a polynomial* $\mathrm{q}_{\mathrm{k},\mathrm{w},\mathrm{X},\beta,[\mathrm{c}]}$ *in* $\mathrm{H}^2(\mathrm{X}; \mathbf{Z})$, *of degree* $4\mathrm{k} - 3(1 + \mathrm{a})$. *These are natural for orientation-preserving diffeomorphisms of C-manifolds, and changing* $[\mathrm{c}]$ *or* β *changes the sign of the polynomials.*

We finish this section by returning to discuss the genericity requirements imposed on our metrics and representatives V_v. In the definition of the invariants transversality is used to obtain the compactness of the intersection of the V_v with M_k. However, it is possible to compute the invariants using a less restrictive class of metrics and representatives . Let $N_1, \ldots N_d$ be any open sets in X containing surfaces $\Sigma_1, \ldots \Sigma_d$ representing $\alpha_1, \ldots, \alpha_d$, and such that the triple intersections of the N_ν are empty. Let $V_\nu (\nu = 1, \ldots d)$ be closed subsets of $\mathscr{B}_{N_\nu}$ cut out by equivariant sections of the $\tilde{\mathscr{L}}_{N_\nu}$, and by our usual abuse of notation, regard the V_ν as being simultaneously closed subsets of all the $\mathscr{B}_{j,X}$, via the restriction maps. We can also allow the case when the V_ν depend on the restriction of connections to the surfaces Σ_ν. Let g be any Riemannian metric on X. For fixed $k > (3/2)(1 + a)$ we say that $(g; V_1, \ldots, V_d)$ is an *admissible system* if the intersections

$$M_{j,X}(g) \cap V_{\nu(1)} \cap \ldots V_{\nu(p)}$$

are *empty* whenever $j < k$ and $p \geq d - 2(k - j)$ (i.e., when $p \geq \frac{1}{2} V.d.(M_{j,X}) + 2(k - j)$. If $(g; V_1, \ldots V_d)$ is admissible our argument above shows that the intersection $V_1 \cap \ldots \cap V_d \cap M_{k,X}(g)$ is compact. We can then define a numerical intersection number $i(g; V_1, \ldots V_d)$ by making a general perturbation in a neighbourhood of this compact set in $\mathscr{B}_{k,X}$. In the definition of our invariants we have shown that admissible systems exist; however, as we shall now explain *any* admissible system can be used to define the invariant $q_k(\alpha_1, \ldots, \alpha_d)$. This gives much more flexibility, both in explicit calculations with examples and in our proofs in the later parts of this paper.

PROPOSITION (3.6). *For any admissible system* $(g; V_1, \ldots V_d)$ *We have* $i(g, V_1, \ldots V_d) = q_{k,X}(\alpha_1, \ldots \alpha_d)$.

Let $(g, V_1, \ldots, V_d) = (g, \mathbf{V})$ be any system, not necessarily admissible, with the V_ν defined by restriction to open sets N_ν. We will show first that there are admissible systems arbitrarily close to $(g, \mathbf{V})$. In fact we will show that there are systems $(g'; \mathbf{V}')$ close to $(g, \mathbf{V})$ such that the intersection of $M_{j,X}(g')$, for $j \leq k$, with any set of p of the V'_ν is empty if $2p$ exceeds the virtual dimension of the moduli space. Let us call this property $P(k)$. The metric g' can be taken to be real analytic and the V'_ν will be determined by restriction to the same open sets N_ν. The proof is by induction on k. We use two facts: first that for any compact set $K \subset M_{j,X}(g)$ there is a real analytic metric g^*, arbitrarily close to g, and a neighbourhood W of K in $\mathscr{B}^*_{j,X}$ such that $M_{j,X}(g^*) \cap W$ is cut out transversely (and avoids reducibles). This follows from Freed and Uhlenbeck's Theorem (cf. [12], p. 72). Second, for a real analytic base metric the ASD solutions are also real analytic, and hence are either reducible on X or irreducible on any open subset in X. We also make heavy use of the weak compactness theorem.

First note that if $(g, \mathbf{V})$ satisfies property $P(k)$ then so do all sufficiently nearby systems $(g^*, \mathbf{V}^*)$. The proof is by contradiction, using weak compactness and a dimension counting argument. By a similar argument we see that if $(g, \mathbf{V})$ satisfies $P(k)$ then all intersections of $M_{k+1,X}(g)$ with p of the V_ν, with $2p \geq V.d.(M_{k+1,X})$, are compact. We can then perturb the metric slightly to a nearby real analytic metric g^* to make $M_{k+1,X}(g^*)$ transverse in a neighbourhood of these compact sets. The restrictions of connections in $M_{k+1,X}(g^*)$ to the N_ν are irreducible so we can also perturb the V_ν to V^*_ν making all the corresponding intersections in this neighbourhood transverse, i.e. empty. If the perturbations are sufficiently small then $(\mathbf{V}^*, g^*)$ will also satisfy $P(k)$ and, by yet another application of weak compactness, one sees then that the intersection of any p of the V^* with $M_{k+1,X}(g^*)$ is actually empty for $2p > V.d.(M_{k+1,X})$ [i.e., we do not introduce any new intersections "at infinity" in $M_{k+1,X}(g^*)$.] So $(\mathbf{V}^*, g^*)$ satisfies $P(k+1)$ and the inductive step is complete.

We now consider this construction in the particular case when $(\mathbf{V}, g)$ was already admissible. It is easy to see that for $(\mathbf{V}', g')$ as above, sufficiently close to $(\mathbf{V}, g)$, the intersection numbers $i(g; V_1, \ldots V_d)$ and $i(g'; V'_1, \ldots, V'_d)$ are equal. So we can reduce to the case of real analytic metrics and systems satisfying $P(k)$. Suppose that $(g, \mathbf{V})$ and $(h, \mathbf{W})$ are two such systems with $V_v = W_v$ for $v = 2, \ldots d$. We can now combine the arguments used in Lemma (3.2) with the inductive argument over the moduli spaces given above. We start with an arbitrary 1-parameter family of metrics joining g to h, and an arbitrary homotopy between the sections cutting out V_1 and W_1, depending on the restriction of connections to a neighbourhood of a 3-chain in X. This gives us parametrised moduli spaces Z_j. We apply our inductive argument, adapted in the obvious way to fmilies, to find a small perturbation of these paths which gives a compact cobordism between the two intersections, and so deduce that the intersection numbers $i(g; V_1, \ldots V_d)$, $i(h; W_1, \ldots W_d)$ are equal.

Remark (3.7). It is easy to show, using similar arguments to those in (3.6), that for any surfaces Σ_v in X and metric g there are arbitrarily close surfaces Σ'_v and real analytic metric g', together with representatives V_v depending on restriction to the Σ'_v, such that $(g'; V'_1, \ldots V'_d)$ is admissible [or even satisfies property $P(k)$].

§IV. CONNECTED SUMS

(i) Shrinking the neck

In this section we consider a C-manifold X which can be written as a connected sum of oriented manifolds Y_1, Y_2. We let Y be the disjoint union of Y_1 and Y_2. We study the polynomial invariants $q_{k,X}$ by constructing a 1-parameter family of Riemannian metrics g_λ $(0 < \lambda < 1)$ on X. These metrics are all the same outside a connecting tube of diameter $O(\lambda)$. The idea is to study the moduli spaces $M_{k,X}(g_\lambda)$ as $\lambda \to 0$, and to compare them with the moduli spaces for Y. (The latter are, of course, unions of products of the moduli spaces for the components Y_1, Y_2.) In this subsection (i) we will set up notation and definitions, and review the main facts about these degenerating families of metrics which will underlie our detailed arguments.

Fix a metric g on Y. The precise properties required will be spelled out below. We suppose that the injectivity radius of g is greater than 1. For a point y in Y we let $B(y, r)$ denote the image of the closed r-ball under the exponential map at y. If $r < 1$ this is an embedded ball in Y. Fix points y_1, y_2 in Y_1, Y_2 respectively and for $r < 1$ put $Z_i(r) = Y_i \backslash B(y_i, r)$. Let $Z(r)$ be the union of $Z_1(r)$ and $Z_2(r)$. We identify the balls $B(y_i, 1)$ with balls in the tangent spaces, using the exponential maps. Choose an orientation-reversing isometry $\sigma: (TY_1)_{y_1} \to (TY_2)_{y_2}$. Then the map f between punctured tangent spaces given by

$$f(\xi) = (\lambda/|\xi|^2)\sigma(\xi)$$

identifies the annulus in Y_1, centred on y_1, having inner radius λ and outer radius 1, with the corresponding annulus in Y_2. We form the oriented connected sum X as a quotient of $Z(\lambda)$ by this gluing map. (So, strictly, X depends upon λ.) For any $r > \lambda^{\frac{1}{2}}$ we have a natural embedding of $Z(r)$ in X and we will often regard this as an inclusion. So, for example, a submanifold of Y which is contained in $Z(r)$ will also be regarded as a submanifold of X.

We will now define the metric g_λ on X. The key property is that it should agree with g on a large open set, for example on $Z(2\lambda^{\frac{1}{2}})$. Over the "neck" in the connected sum we can define g_λ to be a weighted average of the metrics on the Y_i, compared via the identification map f, and averaged by a cut-off function in the obvious way (cf. [4], §IV (vi)).

Just as in the definition of the polynomial invariants above, the cornerstone of our discussion will be a version of Uhlenbecks compactness theorem. Let P be an SU(2) bundle over X and A_α a sequence of connections on P with A_α ASD relative to the metric g_{λ_α}, for a sequence $\lambda_\alpha \to 0$. Let A' be an ASD connection on a bundle P' over Y, and let (z_i) $(i = 1, \dots, n)$ be a multiset of points in $Y \backslash \{y_1, y_2\}$. We say that $[A_\alpha]$ converges weakly to $([A']; (z_i))$ if, after suitable gauge transformations, defined over punctured 4-manifolds, the A_α converge to A' on $Z(r) \backslash \{z_i\}$ for each $r > 0$, and if the curvature densities $|F(A_\alpha)|^2$ converge on $Z(r)$ to $|F(A)|^2 + 8\pi^2 \Sigma \delta_{z_i}$. Notice that we have $n + c_2(P') \leq c_2(P)$, and equality need not hold since we may "lose" curvature over the neck.

We will also introduce a notion of "strong" convergence in this situation. For a given λ and $\eta > 0$ say that $[A_\alpha]$ is (η, λ) close to $[A']$ if there is a bundle map ϕ: $P|_Z \to P'|_Z$, where $Z = Z((1/100)\lambda^{\frac{1}{4}})$, such that

$$\| A_\alpha - \phi^*(A') \|_{L^p(Z)} < \eta.$$

[The factor (1/100) here is, of course, just a fixed, suitably small, number.] We say that the sequence $[A_\alpha]$ is, *λ_α-strongly convergent* to $[A']$ if for any $\eta > 0$, $[A_\alpha]$ is (η, λ_α) close to A for large α.

The convergence results we need can be summarised by the following proposition.

PROPOSITION (4.1). *If $\lambda_\alpha \to 0$ and A_α is a sequence of g_{λ_α}- ASD connections on a bundle P over X then there is a bundle P′ over Y, a connection A′ on P′ and a multiset $(z_1, \dots, z_n)$ in $Y \backslash \{y_1, y_2\}$, with $c_2(P') + n \leq c_2(P)$, such that a subsequence $[A_{\alpha'}]$ of $[A_\alpha]$ is weakly convergent to $([A'], (z_i))$. Moreover, if $c_2(P') = c_2(P)$ then $A_{\alpha'}$ is $\lambda_{\alpha'}$-strongly convergent to A′.*

The first sentence of this proposition follows easily from Uhlenbeck's Theorem. The second sentence, asserting strong convergence, is more involved. It requires decay estimates for ASD connections over an annulus or, applying a conformal map, a cylinder. These estimates can be proved by adapting the techniques of [3], Appendix or [12], §9.

The other major piece of theory we need is complementary to (4.1); allowing us to go from ASD connections over Y to ASD connections over X. We recall the notion, explained in [4], §III(i) of the "connected sum" of connections. Let P' be a bundle over Y, as above, and ρ: $P'|_{y_1} \to P'|_{y_2}$ be an isomorphism. If A' is an ASD connection on P' we can form a new bundle-with-connection over X by flattening A' in a small neighbourhood of $\{y_1, y_2\}$ and gluing by ρ. Of course this construction depends upon some arbitrary choices of cut-off functions, etc., and the new connections will not be ASD with respect to any metric. However, we can look for *nearby* ASD connections over X, as in [4], §IV. The results we need are summarised in the next two propositions. We consider two situations, first when both of the $b_2^+(Y_i)$ are strictly positive.

We introduce more notation, the moduli space $M_{k,Y}$ has components $M_{(k_1,k_2),Y}$ labelled, in the obvious way, by the disposition of the Chern class. Let $M^*_{k,Y}$ be the union of all the components $M_{(k_1,k_2),Y}$ with k_1 and k_2 both strictly positive. Now if $b_2^+(Y_1)$ and $b_2^+(Y_2)$ are both positive there are, for generic metrics g on Y, no non-trivial reducible ASD connections over Y. The moduli spaces $M^*_{k,Y}$ are smooth manifolds, cut out transversely. Let q: $T_k \to M^*_{k,Y}$ be the fibre bundle, with fibre SO(3), consisting of isomorphism classes of pairs (A', ρ), as above. Now let C be an open set in $M^*_{k,Y}$ with compact closure, and η be a small positive number. The L^p norm induces a metric on $M^*_{k,Y}$. Let D be the 2η-neighbourhood of C. We want to consider smooth maps τ from the open set $q^{-1}(D)$ in T_k to the moduli space $M_{k,X}(g_\lambda)$, which are well approximated by the gluing construction discussed above. So we suppose that $\tau([A'], \rho)$ is (g, λ) close to $[A']$. We can also look at the situation over a common open set $Z(r)$. More precisely, we have two maps from $q^{-1}(D)$ to $\mathscr{B}^*_{Z(r)}$ given by the composites:

$$\begin{array}{ccc} q^{-1}(\mathrm{D}) & \xrightarrow[\tau]{} & M_{k,X} \\ q\downarrow & & \downarrow R_{Z(r)} \\ M^*_{k,Y} & \xrightarrow[R_{Z(r)}]{} & \mathscr{B}^*_{Z(r)} \end{array}$$

PROPOSITION (4.6). *Suppose that* $b_2^+(Y_i)$ *are both strictly positive. Let* C *be a pre-compact open set in* $M_{k,Y}$. *For* $\eta < \eta(C)$ *and* $\lambda < \lambda(C)$ *there is a diffeomorphism* τ_λ *from* $q^{-1}(D) \subset T$ *(where* D *is the* 2η*-neighbourhood of* C, *as above) to an open set* $S_{\lambda,\eta}$ *in* $M_{k,X}(g_\lambda)$ *such that:*

(i) $\tau_\lambda([A', \rho])$ *is* (η, λ) *close to* A';
(ii) $S_{\lambda,\eta}$ *contains all points in* $M_{k,X}(g_\lambda)$ *which are* (η, λ) *close to a point of* C;
(iii) *as* $\lambda \to 0$ *the composites* $R_{Z(r)}\tau_\lambda$ *converge in* C^1 *to* $R_{Z(r)}q$.

In the second case we consider connections which are flat over the part of X corresponding to one factor, say Y_1, but we assume that $b_2^+(Y_1)$ is zero. Thus we are gluing a connection A' over Y which is trivial over Y_1. This means that the gluing parameter ρ can be cancelled by the isotropy group of A': the condition that $b_2^+(Y_1)$ is zero means that there is no obstruction to deforming the glued connection to an ASD connection over X. Precisely, we have the following.

PROPOSITION (4.7). *Suppose that* $b_2^+(Y_1) = 0$. *Let* C *be a pre-compact open set in the moduli space* M_{k,Y_2}. *For* $\eta < \eta(U)$, $\lambda < \lambda(U, \eta)$ *there is a diffeomorphism* τ_λ *from the* 2η*-neighbourhood* V *of* U *to an open set* S_λ *in the moduli space* $M_{k,X}(g_\lambda)$ *such that:*

(i) $\tau_\lambda([A])$ *is* (η, λ) *close to* A' *(the connection over* Y *which is equal to* A *over* Y_2 *and is flat over* Y_1*)*;
(ii) S_λ *contains all points of* $M_{k,X}(g_\lambda)$ *which are* (η, λ) *close to points of* C;
(iii) *as* $\lambda \to 0$ *the composites* $R_{Z(r)}\tau_\lambda$ *converge in* C^1 *to* $R_{Z(r)}$.

[In (4.6, iii) and (4.7, iii) one has to interpret the convergence in suitable way in the unlikely event that a connection in $M_{k,X}(g_\lambda)$ is reducible over $Z(r)$.]

(ii) The main results

In this subsection we first give a corollary [Theorem (4.8)] of the structure theorems above for $M_{k,X}(g_\lambda)$ as $\lambda \to 0$. This result illustrates the main ideas, which will be used again in later arguments, in a simple way. Next we formulate our main theorem, a "vanishing" theorem for the polynomial invariants of connected sums [Theorem (4.9)], and go on to indicate the main lines of the argument used in the proof of this which follows. For simplicity, we concentrate throughout on the SU(2) case, although we make some brief remarks on the corresponding results for SO(3) connections.

THEOREM (4.8). *Suppose* (X, β) *and* (Y_2, γ) *are C-manifolds, and* Y_1 *is a smooth, compact, oriented 4-manifold with* $b_2^+(Y_1) = 0$. *Suppose there is an orientation-preserving diffeomorphism of* X *and the connected sum* $Y_1 \# Y_2$, *under which* β *and* γ *correspond. Then for all* $k > (3/4)(b_2^+(X) + 1)$ *the restriction of* $q_{k,X,\beta}$ *to* $H_2(Y_2) \times \ldots \times H_2(Y_2) \subset H_2(X) \times \ldots \times H_2(X)$ *is equal to* $q_{k,Y_2,\gamma}$.

Proof. Let $\alpha_1, \ldots \alpha_d$ be classes in $H_2(Y_2)$, where $2d = \dim M_{k,X}$. Choose a generic metric g on Y and representatives $V_1, \ldots, V_d$, as in (2.4), (2.5), using neighbourhoods $N_1, \ldots, N_d$ contained in some $Z_2(r)$ and transverse to all the moduli spaces for Y_2. By our usual abuse of notation we consider the V_v as being simultaneously closed subsets of the $\mathscr{B}_{j,Y_2}$ and $\mathscr{B}_{j,X}$. We claim first that when λ is small the intersections:

$$I = M_{k,Y_2} \cap V_1 \cap \ldots \cap V_d, \quad I(\lambda) = M_{k,X}(g_\lambda) \cap V_1 \cap \ldots \cap V_d$$

can be naturally identified as sets. For, in one direction, given a point $[A]$ in I we can construct a map τ_λ from a neighbourhood of $[A]$ to $M_{k,X}(g_\lambda)$. The intersection of the V_v and M_{k,Y_2} is transverse so, for small λ, there will be a unique intersection point (also transverse) in $I(\lambda)$ close to $\tau_\lambda([A])$, by (4.7, iii). In the other direction we need to show that there are no extra points of $I(\lambda)$. By (4.7, ii) it suffices to show that for any η all points of $I(\lambda)$ are (η, λ) close to a point of I, for small enough λ. Suppose the contrary, so there is a sequence $\lambda_\alpha \to 0$ and points $[A_\alpha]$ of $I(\lambda_\alpha)$ which are not (η, λ_α) close to I. Then by (4.1) we can suppose that the sequence is weakly convergent to $([A']; z_1, \ldots z_n)$, with an ASD connection A' on a bundle P' having $c_2(P') = k'$. Let k'_1 and k'_2 be the components of $c_2(P')$ on Y_1, Y_2, so $k' = k'_1 + k'_2$. We know that $k' + n \leq k$ and that $[A']$ lies in V_v for any v such that N_v does not contain one of the points z_i. On the other hand we have that $M_{j,Y} \cap V_{v(1)} \cap \ldots \cap V_{v(p)}$ is empty if $2p > \dim M_{j,Y_2}$. Then the counting argument used in (3.1) shows that n and k' are zero and $k'_2 = k$. Thus the $[A_\alpha]$ converge strongly to $[A']$, by the last part of (4.1), and we obtain the desired contradiction.

Now it is easy to see that the sign with which the points of I and $I(\lambda)$ are counted agree, using the definition in [6] of the standard orientation of the moduli spaces. This completes the proof of the theorem [note that we do not have to worry about the V_v being transverse to all the lower moduli spaces, or some restrictions being reducible on the N_v since the counting argument above shows that the system $(g(\lambda); \mathrm{V})$ is admissible for small λ, cf. (3.6)].

We will now move on to the main topic of this section.

THEOREM (4.9). *Suppose a C-manifold X can be decomposed as a smooth oriented connected sum $X = Y_1 \# Y_2$ with $b_2^+(Y_1)$ and $b_2^+(Y_2)$ both strictly positive. Then all the polynomial invariants $q_{k,X}$ are zero.*

[Here we are considering the invariants derived from SU(2) moduli spaces. In fact in the situation of (4.9) all the SO(3) invariants also vanish. The proof is in some cases much easier than for the SU(2) result and we leave the reader to supply details.]

We shall now discuss the proof of (4.9) in an informal way to motivate the detailed work in the proof in §IV (iii)–(vi) below. Fix attention on a set of classes $\alpha_1, \ldots \alpha_d$ in $H_2(X)$ and let $V_1, \ldots V_d$ be representatives for the corresponding classes, defined by restriction to subsets of some $Z(r)$. We want to study the intersection $I(\lambda)$ of $M_{k,X}(g_\lambda)$ with the V_v as λ tends to 0. So suppose that λ_α is a sequence tending to zero and that $[A_\alpha]$ is a point of $I(\lambda_\alpha)$. By (4.1) we can suppose, taking a subsequence, that the $[A_\alpha]$ are weakly convergent to some $([A']; z_i)$. Let us assume that all intersections of the V_v are transverse to all the moduli manifolds for Y. Then we can apply our familiar dimension counting argument. Let the limit A' have Chern class (k'_1, k'_2) and let there be d_1 of the α_v in $H_2(Y_1)$ and d_2 in $H_2(Y_2)$. Similarly, let n_1 of the points z_i be in Y_1 and n_2 be in Y_2. The important thing to observe first is that one of k'_1, k'_2 is zero. Indeed, if not we have the inequalities:

$$d_i \leq 2n_i + \tfrac{1}{2}\dim M_{k'_i, Y_i} \quad (i = 1, 2)$$
$$n_1 + n_2 + k'_1 + k'_2 \leq k.$$

But $d_1 + d_2 = d = \frac{1}{2}\dim M_{k,X}$ and we get a contradiction, since the index formula gives:

$$\dim M_{k,X} = \dim M_{k'_1,Y_1} + \dim M_{k'_2,Y_2} + 3$$

when $k'_1 + k'_2 = k$. This numerology is clarified by the gluing construction of (4.6). Open sets in the moduli spaces $M_{k,X}(g_\lambda)$ are, for small λ, described as SO(3) bundles over the products of the moduli spaces for the Y_i. The point is that the three dimensions in the fibre are effectively invisible when the connections are restricted to $Z(r)$. If we imagine we are in a situation where the moduli spaces for Y were all compact then we could phrase the argument up to this point in more algebro-topological terms: on the component of $M_{k,X}$ diffeomorphic to an SO(3) bundle T the classes $\mu(\alpha_v)$ lift from the base of the fibration, and so their cup product must vanish on this part of the moduli space, since it again lifts from the base.

We interpose two paranthetic remarks here. First, if one works with SO(3) connections where w_2 has a non-zero component on each factor in the connected sum then one only needs this dimension counting argument to prove the vanishing theorem. Second, it is useful when visualising the asymptotic behaviour of the moduli spaces $M_{k,X}(g_\lambda)$ to know that in the natural "L^2 metric" on the moduli space the fibre of T collapses as λ tends to 0. In fact the diameter of the fibre is $O(\lambda)$.

We now move on to the second stage of the proof of the vanishing theorem. The argument above tells us that for small λ the intersection $I(\lambda)$ can be divided into two parts, consisting of connections which are close in the weak topology to the flat connection over Y_1 and Y_2 respectively (i.e., they are almost flat outside a union of small balls in one of the Y_i.) No point of $I(\lambda)$ can satisfy both conditions, by the dimension counting argument. [since $\lambda' > \frac{3}{2}(1+a)$]. So let us consider the part of the intersection close to the flat connection over Y_1. In terms of our sequence $[A_\alpha]$ above, we assume that k'_1 is zero. Our counting argument still gives us some information. Combining our inequalities we get:

$$4(n_1 + n_2) \le b_2^+(Y_1)$$

$$2n_1 \ge d_1 .$$

When $b_2^+(Y_1)$ is large the range of phenomena that can occur becomes complicated, since we can encounter many "concentrated instantons" in the moduli spaces as $\lambda_\alpha \to 0$. Let us first consider the simplest case when $b_2^+(Y_1) = 1$. In this case we must have $n = 0$, $k_2 = k$ and $d_1 = 0$. So the sequence $[A_\alpha]$ is strongly convergent. To understand the possible limits [A′] we need the extension of (4.7) to the case when the "obstruction space" $H = H^2_+(Y_1)$ is non-zero. The theory involved is developed in [4], §IV(vi). Let $Q \to M_{k,Y_2}$ be the principal SO(3) bundle consisting of isomorphism classes of pairs (A, f) where A is a connection on an SU(2) bundle E over Y_2 and f is a framing of the associated SO(3) bundle at y_2. Let $\Theta \to M_{k,Y_2}$ be the real 3-plane bundle associated to Q. Then on a neighbourhood of a pre-compact subset C of M_{k,Y_2} we can find sections ϕ_λ of $\Theta \otimes H$ and a map τ_λ from C to $\mathscr{B}_X$ such that τ_λ takes the zeros of ϕ_λ to $M_{k,X}(g_\lambda)$, and all the points of $M_{k,X}(g_\lambda)$ which are close to U are contained in the image of τ_λ. Moreover there is a formula for the leading term in ϕ_λ as $\lambda \to 0$. This description becomes especially simple in the case at hand when H is one-dimensional. Let ω be the harmonic form on Y_1 generating H. The curvature of a connection at y_2 gives a section of $\Theta \otimes \Lambda^2_{+,y_2}$. The leading term in ϕ_λ is given by contracting this with the value of ω at y_1 (using the isomorphism σ). Let us call this leading term ϕ. We suppose that the intersections of the V_v with the lower moduli spaces M_{j,Y_2} are transverse. Then a dimension counting argument shows that the intersection $V_1 \cap \ldots \cap V_d \cap M_{k,Y_2}$ is a compact 3-manifold J. (Recall that the "lower strata" have condimension at least 4.) The structure theorem for the moduli spaces now tells us that if the section ϕ vanishes transversely on a finite set of points I in J then τ_λ gives a bijection from I to the relevant portion of the intersection $I(\lambda)$. Thus the total contribution to the intersection $I(\lambda)$, counted algebraically with signs, is the *Euler class* of the bundle Θ over J. But the rational Euler class of three plane bundles is always zero so we deduce that this contribution to $I(\lambda)$ is algebraically zero (although we do not know that the intersection is actually empty when λ is small).

We can say, roughly, that there are two topological mechanisms underlying the vanishing theorem: the SO(3) fibration on one large subset of the moduli space $M_{k,X}(g_\lambda)$ for small λ, and this Euler class argument on the remainder. Both these ideas can be applied rather generally. For example, it is an easy exercise in Chern–Weil theory to prove that if Q is any principal G-bundle, where G is a compact Lie Group, and Ξ is the vector bundle associated to the adjoint representation, then the Euler class of Ξ in rational cohomology is zero. This makes it seem very likely that if (as is certainly possible) one constructs rational Yang–Mills invariants using moduli spaces of G-connections for more general groups G, then all these invariants will vanish in the situation considered in (4.9).

The discussion above for the case when $b_2^+(Y_1) = 1$ gives the main idea behind the proof of (4.9). We will now describe the outlines of the argument in the general case. The important point is that we want to avoid describing our moduli spaces explicitly since, while one knows in principle how to analyse different regions in the $M_{k,X}(g_\lambda)$ asymptotically, the descriptions become extremely complicated. So we proceed as follows. We let U_1 be an open subset in $\mathscr{B}_X$ consisting of connections which are almost flat over $Z_1(\rho)$ minus at most k small balls. (i.e. roughly, a neighbourhood of the flat connection in the weak topology). U_1 will be defined to be $(E_1 \circ R_{Z_1(\rho)})^{-1}([0, \varepsilon])$ where E_1 is a function from $\mathscr{B}_{Z(\rho)}$ to $\mathbf{R}^+$, measuring the "weak distance to the flat connection". Thus the defining criteria depend only on the restriction of connections to $Z(\rho)$. By abuse of notation we write $E_1 \circ R_{Z_1(\rho)}$ simply as E_1 and regard E_1 as being a function on all the spaces $\mathscr{B}_{Z_1(\rho)}$, $\mathscr{B}_{j,X}$, $\mathscr{B}_{j,Y_1}$ simultaneously. Then we have corresponding sets $W_1^j = E_1^{-1}([0, \varepsilon])$ in $\mathscr{B}_{j,Y_1}$. We also have symmetrical objects U_2, E_2 etc. The first part of the argument above shows that the intersection $I(\lambda)$ can be divided into disjoint pieces $I_1(\lambda)$ and $I_2(\lambda)$ in U_1 and U_2 respectively. We want to show that, counted algebraically, the contribution from each of these is zero, so we can fix attention on $I_1(\lambda)$ in U_1.

The argument continues by embedding the ASD equations over U_1 in a larger family of equations. Recall that connections A in U_1 are almost flat over most of $Z_1(\rho)$. We show that when ρ is small one can choose, in a gauge invariant way, sections of the adjoint bundle $\mathfrak{g}_P$ over $Z_1(\rho)$ which are similarly almost covariant constant over most of $Z_1(\rho)$, and which give an approximately flat trivialisation of $\mathfrak{g}_P$ there. We do this in a way that is compatible with weak convergence of connections. These sections define a rank 3 bundle $\Xi \to U_1$; the points of Ξ are isomorphism classes of pairs (A, ξ) where ξ is one of the chosen sections of $\mathfrak{g}_P$ and A is a connection on P. Let ω be a self-dual 2-form supported on $Z_1(\rho)$ obtained by cutting- off a harmonic form on Y_1. We can consider the equation over the total space of Ξ:

$$F^+(A) + \xi . \omega = 0.$$

This is a Fredholm equation and the solution space $L_k \subset \mathscr{B}_{k,X}$ has virtual dimension $2d + 3$. The moduli space $M_{k,X}$ is clearly the intersection of L_k with the zero section in Ξ. This procedure is consistent with that followed above in the case $b_2^+(Y_1) = 1$, since we can then think of L_k as being identified with M_{k,Y_2}, τ_λ as being the projection map π from Ξ to $\mathscr{B}_X$, and ϕ_λ as being the tautological section of $\pi^*(\Xi)$.

We then go on to show that, for suitable choices of parameters, the component of the intersection $J = L_k \cap V_1 \cap \ldots \cap V_d$ which contains $I_1(\lambda)$ is a compact 3-manifold. Once this is done the proof is completed by the same Euler class argument, since $I_1(\lambda) \subset J$ is by definition the zero set of a section of the rank 3 bundle $\pi^*(\Xi)$. There are two main parts to this study of J. On the one hand we look at the subsets $\bar{J}(\varepsilon') = E_1^{-1}[0, \varepsilon'] \cap J$ for $\varepsilon' < \varepsilon$. To study these we want to appeal to the same kind of arguments, mixing transversality with weak compactness, that we use for the ordinary moduli spaces. Building up the analogous results here makes up the bulk of the arduous labour in the proof below. Given the same basic properties as for the ASD connections, one can see that $\bar{J}(\varepsilon)$ should be generically compact by a dimension counting argument.

The other part of the study of J is much more interesting. The discussion so far leads to the point where we see that J is the interior of a compact manifold with boundary $\bar{J} = \bar{J}(\varepsilon)$. We have two important small parameters, λ defining the metric and ε defining the open set U_1 that we are considering. (The parameter ρ is fixed earlier in the argument.) Suppose that, with ε fixed, the boundary of $\bar{J}$ is not empty for arbitrarily small λ. Then we can use weak compactness to get a weak limit $([A', \xi]: z_i)$ of the boundary points. Our function E_1 is continuous in the weak topology on solutions to the generalised equations, so $E_1(A') = \varepsilon$. In particular A' is not flat over Y_1. Then using a dimension counting argument we find that we actually have strong convergence. Now our extended equations $F^+(A) + \xi . \omega = 0$ can be defined equally well on a bundle over the open subsets W_1^i of $\mathscr{B}_{j, Y_1}$, since all our constructions depend only on the restriction of connections to $Z_1(\rho)$. The strong limit A' is made up of an ASD connection over Y_2 and a solution to the extended equations over Y_1. When we take account of the intersection with the V_ν these individual constituents are isolated points. Finally we extend the results of our gluing construction [Theorem (4.6)] to the generalised equations. This shows that A' yields, for small λ, a 3-parameter family of points of $J(\varepsilon'')$, for any $\varepsilon'' > \varepsilon$, parametrised by a copy of SO(3). This forms a complete component of $J(\varepsilon'')$ on which the variation of E_1 tends to zero with λ. It is then clear that the components of J containing $I_1(\lambda)$ do not meet the boundary when λ is small. This argument is really dealing with the intersection of the regimes in the moduli spaces where the two basic mechanisms apply.

We complete this preliminary discussion of the proof of (4.9) with three more remarks. First, to obtain transversality we are forced to introduce yet more general equations, using the perturbations constructed in [6]. These are discussed in §4(v) below. Second, we have to divide the discussion of transversality into two parts: first we study the extended equations over the W_1^i and make these transverse, then analyse the boundary of J as sketched above, and finally fix λ and perturb the equations and representatives V_ν to obtain a compact transverse intersection in U_1. Thirdly, it is helpful in understanding this argument to see precisely why it would fail if $b_2^+(Y_1)$ were zero, contrary to our hypothesis. In that case there would be a 1-parameter family of reducible solutions to the extended equations over W_1^0 having the trivial solution (the flat connection over Y_1) as a boundary point. When we analyse the space J we get a contribution to the boundary of $\bar{J}(\varepsilon)$ coming from these reducible solutions glued to points of M_{k, Y_2}. The isotropy group of these reducible solutions is S^1, and each such pair now furnishes a two-dimensional family, parametrised by $\mathrm{SO}(3)/S^1 = S^2$, of solutions in Ξ. So we can no longer deduce that the components of J containing $I_1(\lambda)$ do not cross $E^{-1}(\varepsilon)$.

(iii) Preliminary steps in the proof of Theorem (4.9)

In this subsection we will first define the functions E_i whose role in the proof has been outlined above and then go on to verify that they have the detailed properties required in our construction of the extended equations in §IV (iv) below. Our definitions will depend on a small parameter ρ, with $2\lambda^{\frac{1}{2}} < \rho < 1$, which will be fixed at the end of §IV(iv). We begin by choosing a system (g, V) for Y, satisfying $P(k)$, with a real analytic metric g and V_v defined by restriction to surfaces Σ_v as in (3.7), with the Σ_v contained in $Z(1)$.

Let $f: \mathbf{R}^+ \to \mathbf{R}^+$ be a smooth, monotone, function with $f(x) = x^2$ for small x and $f(x) = x$ for large x. If A is any SU(2) connection over $Z_1(\rho)$ we put:

$$E_1(A) = \int_{Z_1(\rho)} f(|F(A)|) + |F^+(A)|^p \, d\mu.$$

As we have explained above, we regard E_1 as being simultaneously a function on all the $\mathscr{B}_{j,X}$ and $\mathscr{B}_{j,Y_1}$. The reason for using $f(|F(A)|)$ rather than just $|F(A)|$ in the integrand is that E_1 is then a smooth function. For $\varepsilon > 0$ we let:

$$U_1(\varepsilon) \subset \mathscr{B}_{k,X}, \; W_1^j(\varepsilon) \subset \mathscr{B}_{j,Y_1}$$

be the open sets $E_1^{-1}([0, \varepsilon])$. To simplify the discussion in these preliminary sections we will state our results for $U_1(\varepsilon)$, since the corresponding statements for the other spaces are established in just the same way. Of course we also define E_2, $U_2(\varepsilon)$ in a symmetrical fashion.

The next two lemmas show that the function E_1 does indeed measure the "distance to the flat connection in the weak topology".

LEMMA (4.10). *For fixed ρ and any $\delta > 0$ there is an ε such that for all small λ and any point $[A]$ in $U_1(\varepsilon)$ there can be found a set of points $z_1, \ldots z_n$ in $Z_1(\rho)$, with $n \leq k$, such that if Ω is the complement in $Z_1(2\rho)$ of the δ balls centered on the z_i, we have:*

$$\int_\Omega |F(A)|^p \, d\mu \leq \delta.$$

This is, once again, a corollary of Uhlenbeck's Thoerem. A sequence $[A_\alpha]$ in $\mathscr{B}_{Z(\rho)}$ with $F(A_\alpha)$ bounded in L^2 and $F^+(A_\alpha)$ tending to zero in L^p has a weakly convergent subsequence; i.e. the connections converge in L_1^p on interior domains minus finite sets. Now the function E_1 dominates $\| F^+(A)|_{Z(\rho)} \|_{L^p}^p$ so, arguing by contradiction, if the conclusion of the lemma was false we would find a weakly convergent sequence $[A_\alpha]$ with $E_1(A_\alpha) \to 0$ but for which there is no set of suitable δ balls for any α. However the function E_1 also dominates a multiple of the L^1 norm of the curvature. So the limit A_∞ must have curvature zero, and we obtain a contradiction.

Conversely we have the following.

LEMMA (4.11). *For any ρ and ε we can find $\delta(\rho, \varepsilon) > 0$ such that if $\delta < \delta(\rho, \varepsilon)$ and A is an ASD connection over $Z(\rho)$ such that:*

(i) $\displaystyle\int_{Z_1(\rho)} |F(A)|^2 \, d\mu \leq 16\pi^2 k$

say, and

(ii) $\displaystyle\int_{\Omega'} |F(A)|^p \, d\mu < \delta,$

where Ω' is the complement in $Z_1(\rho)$ of k balls of radius δ, then $E_1(A) \leq \varepsilon$.

Proof. We divide the integral defining $E_1(A)$ into two parts corresponding to Ω' and the small balls. Clearly the contribution from Ω' is $O(\delta^{1/p})$. Suppose $f(x) = x$ for $x \geq c$. Then the contribution from the small balls to $E_1(A)$ is dominated by the L^1 norm of $F(A)$ over the balls plus c times the volume of the balls: now use the inequality,

$$\| F(A) \|_{L^1(B)} \leq \| F(A) \|_{L^2(B)} \cdot \mathrm{Vol}(B)^{\frac{1}{2}},$$

and the fact that the volume of the balls decreases with δ.

We will now set down the main properties of the open sets defined by these functionals. Recall that we have three parameters at our disposal, to wit ε, ρ, λ (with $2\lambda^{\frac{1}{2}} < \rho$). The first property is an auxiliary fact that can be deduced easily from the removeability of singularities theorem in the simple, linear, case of reducible connections and we omit further details of proof.

PROPOSITION (4.12). *For sufficiently small ρ and ε there are no reducible connections in any* $\mathrm{W}_1^j(\varepsilon)$, $\mathrm{W}_2^j(\varepsilon)$ *except for those on the trivial* S^1 *bundle when* $\mathrm{j} = 0$.

The main properties we need are given in the next two propositions.

PROPOSITION (4.13). *For any ρ, λ_0 there is an $\varepsilon_0(\rho, \lambda_0)$ such that*

$$I(\lambda) \cap \mathrm{U}_1(\varepsilon) \cap \mathrm{U}_2(\varepsilon)$$

is empty for all $\lambda \leq \lambda_0$ and $\varepsilon \leq \varepsilon_0(\rho, \lambda_0)$.

PROPOSITION (4.14). *For any fixed ε, ρ there is a $\lambda_1(\varepsilon, \rho)$ such that $\mathrm{I}(\lambda)$ is contained in $\mathrm{U}_1(\varepsilon) \cup \mathrm{U}_2(\varepsilon)$ for $\lambda \leq \lambda_1$.*

The proofs of these two propositions are straightforward applications of the techniques we have used many times already. Beginning with (4.13): if the result were false we could find a sequence $[A_\alpha]$ in $I(\lambda_\alpha) \cap U_1(\varepsilon_\alpha) \cap U_2(\varepsilon_\alpha)$ with $\varepsilon_\alpha \to 0$ and $\lambda_\alpha \in (0, \lambda_0)$. By (4.1) we can suppose that the $[A_\alpha]$ converge to the trivial connection over $Z(\rho)$ minus a finite exceptional set, and in the usual way there are at most k points in the exceptional set. But, just as in the proof of Lemma (3.1), there must be a surface Σ_ν which does not contain any of these points [since $k > 3/2(1 + a)$], and we obtain a contradiction to the fact that the $[A_\alpha]$ lie in all the V_ν.

Turning now to (4.14), if the statement were false there would be an ε, a sequence λ_α tending to zero, and points $[A_\alpha]$ in $I(\lambda_\alpha)$ but not in either $U_1(\varepsilon)$ or $U_2(\varepsilon)$. We can suppose the sequence is weakly convergent to a limit $([A']; z_i)$ and, by Lemma (4.11) above, A' cannot be flat over either component Y_1, Y_2. Then if there are n points z_i and the Chern class of A' has components k_1, k_2 we have:

$$2n + \tfrac{1}{2}\dim M_{k_1, Y_1} + \tfrac{1}{2}\dim M_{k_2, Y_2} \geq d = \tfrac{1}{2}\dim M_{k, X},$$

and

$$n + k_1 + k_2 \leq k,$$

which contradicts the index equality:

$$\dim M_{k_1 + k_2, X} = \dim M_{k_1, Y_1} + \dim M_{k_2, Y_2} + 3.$$

Put together, these two propositions imply that when λ and ε are small we can unambiguously write $I(\lambda)$ as the union of $I_1(\lambda) \subset U_1(\varepsilon)$ and $I_2(\lambda) \subset U_2(\varepsilon)$. We let $i_1(\lambda)$ and $i_2(\lambda)$ be the corresponding algebraic intersection numbers, so:

$$q_{k, X}(\alpha_1, \dots \alpha_d) = i_1(\lambda) + i_2(\lambda).$$

The integers $i_1(\lambda)$, $i_2(\lambda)$ are independent of the parameters λ, ε, ρ, provided these are suitably small (although we do not use this fact in the proof below). We will complete the proof of (4.9) by establishing the following theorem.

THEOREM (4.15). *In the situation above both* $i_1(\lambda)$ *and* $i_2(\lambda)$ *are zero, for suitably small values of the parameters* λ, ε, ρ.

To prove (4.15) we can plainly fix attention from now on one component $i_1(\lambda)$, and simplify our notation by writing E, $U(\varepsilon)$, $W^j(\varepsilon)$ in place of E_1, U_1, W_1^j.

(iv) The extended equations

In this subsection we show how the ASD equations over $U(\varepsilon)$ and $W^j(\varepsilon)$ can be embedded in a larger family of equations. To do this we need to find a procedure for choosing an approximately flat trivialisation for our bundles over large regions in Y_1 where the connections are almost flat. There are various approaches one might take to this; the approach we follow is analytical and is based on Taubes' construction in [31]. This approach has the disadvantage that it requires some rather long and tedious analytical digressions which are certainly not a central feature of the proof. It is possible that a more abstract topological approach would work better, but the author has not yet been able to develop such a procedure in detail. We begin with some preliminary lemmas showing that connections in $U(\varepsilon)$ and $W^j(\varepsilon)$ are almost flat outside a large set whose geometry we can control; this allows us to get around any difficulties with the "diagonals" in the symmetric products appearing in the definitions of weak convergence.

LEMMA (4.16). *There is a number* $N(k)$ *such that for any set* S *of points in* Y_1 *containing at most* $k+1$ *elements we can find a collection of disjoint balls* $B_1, \ldots, B_m$ *in* Y_1 *with* $m \leq k+1$ *such that:*

(i) *for each point* p *in* S *the* 2ρ *ball centered on* p *lies in some* B_i;
(ii) *the radii of the balls* B_i *lie between* $10\,\rho$ *and* $N(k)\,\rho$;
(iii) *The distance between any two different balls* B_i, B_j *is at least* ρ.

The proof is elementary and we leave it as an exercise for the reader. Let us call the complement in Y_1 of a collection of at most $k+1$ balls satisfying the conditions (ii) and (iii) of Lemma (4.16) a "good" domain in Y_1. We then have the following.

COROLLARY (4.17). *For any* δ *and* ρ *we can find an* ε *such that for any point* $[A]$ *in* $U(\varepsilon)$ *or* $W^j(\varepsilon)$ *there is a good domain* D *contained in* $Z_1(\rho)$ *such that:*

$$\int_D |F(A)|^p \, d\mu \leq \delta.$$

This follows immediately from (4.10) and (4.16), where we take the set S to consist of y_1 and the points $z_1, \ldots z_n$ of (4.10).

PROPOSITION (4.18). *There are constants* $C(\rho)$, $\theta(\rho)$, *depending only on* ρ, k *and the Riemannian manifold* Y_1, *such that any* $SU(2)$ *connection* A *over a good domain* D *with* $\int_D |F(A)|^p \, d\mu \leq \theta(\rho)$ *can be represented by a connection matrix* $A^{\#}$ *with* $\int_D |A^{\#}|^2 \, d\mu \leq C(\rho) \left\{ \int_D F(A)|^p \, d\mu \right\}^{2/p}$.

This is a consequence of Uhlenbeck's gauge fixing theorem, proved in [33]. For any simply connected domain Ω the corresponding result holds, but with constants depending upon Ω. It is easy to see, however, that for the good domains D the constants depend only on ρ. For example, we can construct a diffeomorphism ϕ_D from any good domain D to one of $k+1$ standard models in such a way that the derivatives of ϕ_D are bounded by a function of ρ, then the desired uniformity follows by transforming to connections over the standard models.

The conclusion of this discussion is as follows.

PROPOSITION (4.19). *For fixed ρ and any $\eta > 0$ we can find ε such that any connection in $\mathrm{U}(\varepsilon)$ or $\mathrm{W}^j(\varepsilon)$ can be represented over a good domain D by a connection matrix with L^2 norm less than η.*

We can now proceed with the main construction of this subsection. Let $\sigma > 0$ be the first eigenvalue of the Laplacian Δ on the functions on Y_1. Choose d so that the volume of the ball $B(y_1, d)$ is less than $\frac{1}{4}\operatorname{Vol}(Y_1)$, and a positive function R on Y_1, equal to 2σ on $B(y_1, d)$ and supported in $B(y_1, 2d)$. Suppose A is a connection on a bundle P over $Z_1(\rho)$. We define a Dirichlet form Q on the sections of the adjoint bundle $\mathfrak{g}_P$ which vanish on the boundary $\partial Z_1(\rho)$ by:

$$Q(s) = \int_{Z_1(\rho)} |\nabla_A s|^2 + R|s|^2 \, d\mu. \tag{4.20}$$

(The term involving R could be omitted, but we include it to simplify the analysis below.) The associated eigenvalue problem is to find sections s, vanishing on the boundary, and constants τ such that:

$$\Delta_A s + Rs = \tau s. \tag{4.21}$$

We will use the eigenfunctions belonging to low eigenvalues to construct our extensions of the ASD equations. Note first that for an L_1^p connection A the Laplacian $\Delta_A = \nabla_A^* \nabla_A$ makes sense as an operator mapping L_2^p to L^p, and we have smooth families of eigenvalue problems, via restriction, parametrised by $U(\varepsilon)$ and $W^j(\varepsilon)$.

For a connection A as above we denote by Ξ_A the space of sections of $\mathfrak{g}_P$ spanned by eigenfunctions s belonging to eigenvalues τ with $\tau < \frac{1}{2}\sigma$, vanishing on the boundary of $Z_1(\rho)$.

PROPOSITION (4.22). *There is ρ_0, with $\frac{1}{2}\mathrm{d} > \rho_0 > 0$, and a function $\varepsilon(\rho)$ such that if $\rho < \rho_0$, $\varepsilon < \varepsilon(\rho)$ and $[\mathrm{A}]$ is a point in $\mathrm{U}(\varepsilon)$ or $\mathrm{W}(\varepsilon)$ then space Ξ_{A}, [associated to the restriction of A to $\mathrm{Z}_1(\rho)$] is three-dimensional.*

This assertion is not at all surprising: the low eigenvalues correspond to sections of the three-dimensional bundle $\mathfrak{g}_P$ which are approximately covariant constant over most of $Z_1(\rho)$.

We will now prove Proposition (4.21). We first show that the dimension of Ξ_A is at least 3. To do this we use the variational description of the eigenspaces: it suffices to exhibit a three-dimensional space H of sections of $\mathfrak{g}_P$ over $Z(\rho)$ such that $Q(s) < \frac{1}{2}\sigma \|s^2\|_L^2$ for all non-zero s in H. There is a constant C such that for any good domain D we can find a cut-off function β_D supported in D, equal to 1 outside the $\frac{1}{2}\rho$ neighbourhood of the boundary of D and with $\|\nabla\beta_D\|_L^2 \le C_\rho$. Choose $\rho < (1/100)\operatorname{Vol}(Y_1)C^{-1}\sigma$. Then make ε so small that any connection A in $U(\varepsilon)$ is represented over a good domain D by a connection matrix $A^{\#}$ with $\|A^{\#}\|_L^2 < (1/100)\operatorname{Vol}(Y_1)\sigma$. In this trivialisation of P over D we let H be the image of the map:

$$u\colon \mathbf{R}^3 = \mathrm{Lie}(\mathrm{SU}(2)) \to \Omega^0_{Z(\rho)}(\mathfrak{g}_P)$$
$$u(t) = \beta_D t,$$

where we think of a vector t as defining a "constant" section of the bundle in this trivialisation over D, and extend by zero to the rest of $Z(\rho)$. Then we have:

$$Q(u(t)) = \|\nabla_A(\beta_D t)\|^2 + \int_D R|\beta_D t|^2$$
$$= \|(\nabla\beta_D)t + A^{\#}(\beta_D t)\|^2 + \int_D R|\beta_D t|^2.$$

Now $\|\nabla\beta_D t\|^2 \leq C\rho|t|^2 \leq (1/100)\mathrm{Vol}(Y_1)\sigma|t|^2$ and

$$\|A^{\#}(\beta_D t)\|^2 \leq \|A^{\#}\|^2|t|^2 \leq (1/100)\mathrm{Vol}(Y_1)\sigma|t|^2.$$

For the third term, recall that $R \leq 2\sigma$ and R is supported on a set of volume at most $\frac{1}{4}$ $\mathrm{Vol}(Y_1)$. So:

$$\int R|\beta_D t|^2 \leq \tfrac{1}{4}.\mathrm{Vol}(Y_1)\sigma|t|^2.$$

Putting these three inequalities together we get:

$$Q(u(t)) \leq (1/4) + (1/25)\mathrm{Vol}(Y_1)\sigma|t|^2.$$

On the other hand $\|u(t)\|^2$ is at least $|t|^2$ times the volume of the set on which β_D equals 1. Clearly, for ρ sufficiently small (depending only on σ), we can make

$$\|u(t)\|^2 \geq (9/10)\mathrm{Vol}(Y_1)|t|^2,$$

and this shows that our map has the desired properties. (Notice that we can construct a three-dimensional space on which Q is arbitrarily small, for small enough ρ and λ.)

The other half of the proof of (4.21), showing that the dimension of Ξ_A is at most 3, is quite similar. Suppose we have established a uniform bound:

$$\|s\|_{L^\infty} \leq B\|s\|^2_L, \tag{4.23}$$

for eigenfunctions s of (4.19) belonging to eigenvalues $\tau < \frac{1}{2}\sigma$, with a constant B *independent of* ρ. Then, if there were four such linearly independent eigenfunctions we could use the same procedure as above to construct a four-dimensional space H' of sections of the trivial bundle, with fibre R^3, such that for small ε and ρ, $\|\nabla h\|^2 \leq \frac{1}{2}\sigma\|h\|^2$ for all h in H', and this would contradict the defintion of σ.

To obtain the uniform estimate (4.23) we observe that for any such eigenfunction:

$$\Delta(|s|^2) < \langle s, \Delta_A s\rangle \leq (\tfrac{1}{2}\sigma - R)|s|^2.$$

In particular $\Delta(|s|^2) \leqslant 0$ on the annulus $B(y_1, d)\backslash B(y_1, \rho)$ where $R = 2\sigma$. We now use the boundary condition that $s = 0$ on $\partial Z_1(\rho)$ and the maximum principle to deduce that the maximum value of $|s|$ over this annulus is attained on the outer boundary $\partial B(y_1, d)$. Thus it suffices to establish the estimate (4.23) over the fixed interior region $Z_1(d)$, and this follows from routine elliptic theory (integrate to get an L^2_1 bound, apply the Sobolev inequality to deduce an L^4 estimate and then bootstrap using the differential inequality).

From now on we fix ρ to be any value less than the ρ_0 of (4.22) and assume ε, λ are always small enough for the conclusions of (4.12), (4.13), (4.14) and (4.22) to apply.

With this preliminary work completed we can define our extended equations. Let s_0 be a solution of the eigenvalue problem (4.21) belonging to the low eigenvalue, when A is the trivial connection θ. This can be regarded as a function on $Z_1(\rho)$. Choose a self-dual 2-form ω, supported in $Z_1(1)$ and such that $s_0\,\omega$ is not in the image of:

$$d^+ : \Omega^1_{Y_1} \to \Omega^{2+}_{Y_1}.$$

It is easy to see that this is possible, since $b^2_+(Y_1) > 0$ and d^+ is not surjective. [Note that if the term R in (4.21) is omitted then s_0 would be a constant.] Let $[A]$ be a point in $U(\varepsilon)$ and ξ a vector in the three-dimensional space Ξ_A. We can regard $\xi\omega$ as a $\mathfrak{g}_P$-valued self-dual 2-form over X, extending by zero outside $Z_1(1)$. The extended equations for the pair (ξ, A) are then:

$$F^+(A) + \xi\omega = 0. \tag{4.24}$$

We complete this subsection by extending the main properties of the ASD equations to these generalised equations. First, equation (4.24) is completely gauge invariant so we can pass to the corresponding quotient space. This is a space Ξ whose points consist of equivalence classes $[A, \xi]$ of pairs. There is an obvious map $p\colon \Xi \to U(\varepsilon) \subset \mathscr{B}_{k,X}$ sending $[A, \xi]$ to $[A]$. Next, recall that the ASD solutions themselves can be viewed as the zero set of a section of an infinite dimensional bundle, $\mathscr{F}$ say, over $\mathscr{B}_{k,X}$. The fibre of $\mathscr{F}$ over $[A]$ is a copy of $\Omega^+_X(\mathfrak{g}_P)$ (or, more precisely, the L^p completion of this). In a similar way we can regard $F^+(A) + \xi\omega$ as defining a section γ of $p^*(\mathscr{F})$ over Ξ.

LEMMA (4.25). (i) *The map* p: $\Xi \to$ U(ε) *is a smooth orientable* 3-*plane bundle.*
(ii) *The section* γ *of* p$^*(\mathscr{F})$ *over* Ξ *is Fredholm of index* $2d + 3$.

Proof. All these formal properties follow from more or less routine checks. Let $L^p_{2,\partial}$ be the space of L^p_2 sections of $\mathfrak{g}_P$ over $Z_1(\rho)$ vanishing on the boundary. The assignment

$$A \to \Delta_A + V$$

gives a smooth equivariant map from the relevant open subset of $\mathscr{A}$ to the bounded operators mapping $L^p_{2,\partial}$ to L^p. Let $\pi_A\colon L^p \to \Xi_A$ be given by standard L^2 projection. To show that Ξ is a smooth bundle we have to check that π_A varies smoothly with A. For this we use the spectral formula:

$$\pi_A = \int_\Lambda (\Delta_A + R - z)^{-1}\,dz \tag{4.26}$$

where Λ is a contour in the complex plane, running around the interval $[0, \frac{1}{2}\sigma]$ and not meeting the spectrum of $\Delta_A + R$. This condition is open in A and on such open sets the integrand varies smoothly with A. The bundle Ξ is easily seen to be orientable. Our proof of (4.22) gives an explicit class of maps from $\mathbf{R}^3$ to each Ξ_A and, while these depend on some arbitrary choices, any two choices induce the same orientation on Ξ_A.

We can construct local models for the Banach manifold Ξ. We write the points in $U(\varepsilon)$ in a neighbourhood of $[A_0]$ using the standard slice:

$$A_0 + a, \quad d^*_A a = 0.$$

Then we identify the fibres of Ξ in a small neighbourhood with Ξ_{A_0} using the restriction of π_{A_0} to Ξ_{A_0+a}.

Turning to part (ii) of (4.25), we call a section Fredholm if it is represented in local trivialisations by maps with Fredholm derivatives. The index of the section is then the index of these linear maps. In the trivialisation given in (i) the section γ represented by the map Φ where:

$$\Phi(a, \xi) = F^+(A_0) + d_{A_0}^+ a + \tfrac{1}{2}[a, a] + \pi_{A_0 + a}(\xi) \in \Omega_X^{2,+}(\mathfrak{g}_P),$$

for a in Ker d_A^* and ξ in Ξ_{A_0}. Now the derivative $\delta\pi_A$ of π_{A_0+a} with respect to a, evaluated at $a = 0$, is given by differentiating (4.26). This gives:

$$\delta\pi_A = -\int_\Lambda G_z \delta\Delta_A G_z \, dz, \tag{4.27}$$

where G_z is the Greens operator $(\Delta_A + R - z)^{-1}$. We then have:

$$(d\Phi)_0(b, \eta) = \eta\omega + d_{A_0}^+ b - \int_\Lambda G_z(d_{A_0}^*(b(G_z\eta)) + b^*(d_{A_0}G_z\eta))\,\omega. \tag{4.28}$$

The first term is finite rank, the second term is the familiar linearisation of the ASD equations, which we know is Fredholm of index $2d$ on Ker $d_{A_0}^*$. So we have to check that for fixed z in Λ and for fixed η in Ξ_{A_0} the map

$$b \mapsto G_z(d_{A_0}^*(b(\theta)) + b^*(\phi)) \tag{4.29}$$

is compact from L_1^p to L^p, where we have set $\theta = G_z\eta$ and $\phi = d_{A_0}\theta$. But this is true because the map factors through the compact inclusion of L_2^p in L_1^p. Now (4.25, ii) follows from the fact a sum $F + K$ of a Fredholm operator F and a compact operator K is again Fredholm and index $(F + K) = \text{index}(F)$.

We have stated all the above for the bundle Ξ over $U(\varepsilon) \subset \mathscr{B}_{k,X}^*$, obtained from the eigenvalue problem over $Z_1(\rho) \subset X$. Of course we can carry out completely parallel arguments to obtain extended spaces $\Xi^j \to W^j(\varepsilon) \subset \mathscr{B}_{j,Y_1}$. Here we have to note that when $j = 0$ there are reducible connections in $W^j(\varepsilon)$ [cf. (4.12)], so Ξ^0 is not strictly a bundle. However, we will not encumber ourselves with extra notation to handle this, and just assume that all our statements are interpreted in the appropriate way over the reductions in $W^0(\varepsilon)$. In practise this linguistic simplification will not cause any difficulties.

Over Ξ^j we have a Fredholm section, still called γ, of a bundle $p^*(F)$ and the index of the section is clearly three plus the usual formal dimension of the ASD moduli space, i.e. $8j - 3b_2^+(Y_1)$. We let $L_{k,X}$ denote the set of zeros of the extended equations in Ξ and L_{j,Y_1} denote the corresponding zero set in Ξ^j. So we can regard the ordinary moduli spaces $M_{k,X}$, M_{j,Y_1} as being contained in $L_{k,X}$ and L_{j,Y_1} respectively, as the intersections with the zero sections in the 3-plane bundles.

Analytically there is little difference between the extended equations and the ordinary ASD equations. The extended equations enjoy the three key properties explained below in (4.30), (4.31) and (4.32). The main points in the proofs are discussed in Appendix 1; we do not go into great detail here since this would mean repeating a large amount of the theory from the ordinary case, with only minor changes.

(4.30) *Regularity.* Solutions (A, ξ) of the extended equations with A in L_1^p and ξ in L_2^p are smooth in some bundle trivialisation.

(4.31) *Weak compactness and the removeability of singularities.* This means that we can define compactification of our enlarged solution spaces just as in the usual case. The only complication is that $U(\varepsilon)$ and $W^j(\varepsilon)$ are *open* subsets of the spaces of connections, so the solution spaces can fail to be compact for trivial reasons. However if $\varepsilon' < \varepsilon$ and we let

$\bar{L}_{j,Y_1}(\varepsilon') \subset L_{j,Y_1}$ be the closed subset $E^{-1}([0, \varepsilon'])$ there is a compactification:

$$\bar{L}_{k,Y_1}(\varepsilon') \subset \bigcup_j \bar{L}_{j,Y_1}(\varepsilon') \times S^{k-j}(Y_1).$$

Similarly, when we define the lower extended moduli spaces $L_{j,X}$, in the obvious way, we get a compactification of $L_{k,X}$. Moreover, the function E extends *continuously* to the compactified spaces, with $E(A; x_1, \ldots. x_n) = E(A)$. This follows from the principle used in the proof of Lemma (4.11), that given an L^2 bound on $F(A)$ small balls in Y_1 can only give a small contribution to the integral of $f(|F(A)|)$. For the term involving $F^+(A)$ we use the fact that, for solutions of the extended equations, that L^∞ norm of $F^+(A)$ is controlled by that of the low eigenfunctions of Δ_A, which can in turn be uniformly estimated by the L^2 norm of the eigenfunction (4.23). Thus we have a uniform estimate $\|F^+(A)\|_{L^\infty} \le \text{const.}\ |\xi|$ for all solutions (A, ξ) of the generalised equations and this again means that the small balls in Y_1 give only a small contribution to the integral of $|F^+(A)|^p$.

(4.32) *Weak semi-continuity of cohomology.* Just as for the ordinary ASD equations the cokernel of the linearisation of the extended equations at a solution A, ξ is a finite dimensional vector space $H^2_{A,\xi}$, a quotient of the fibre of $p^*(\mathscr{F})$. If $H^2_{A,\xi}$ is zero a neighbourhood of $[A, \xi]$ in the solution space is a manifold of the proper dimension, cut out transversely by the equations. In Appendix 1 we will prove the following semi-continuity property of these spaces, which will be used in the transversality discussion in (v) below.

PROPOSITION (4.33). *Let $[A_\alpha, \xi_\alpha]$ be a sequence of points in L_{j,Y_1} converging weakly to $([A_\infty; \xi_\infty]z_1, \ldots, z_n)$. Suppose i_α are bundle maps over $Y_1 \setminus \{z_1, \ldots, z_n\}$ such that $i_\alpha^*(A_\alpha)$ converges to A_∞, and $i_\alpha^*(\xi_\alpha)$ to ξ_∞, over this punctured manifold. Let V be a space of bundle-valued 2-forms generating $H^1_{A_\infty,\xi_\infty}$. Then for large α, $i_\alpha(V)$ generates $H^2_{A_\alpha,\xi_\alpha}$.*

(v) Transversality

We cannot appeal to Freed and Uhlenbeck's Theorem when using the extended equations so, to achieve transversality, we consider a further family of deformations. We need to construct these deformations over all the W^j so that we preserve, as far as possible, control of solutions under weak limits. To do this we use the method introduced in [6], §II(b). This method does not give all the weak continuity properties one would hope for, but by keeping careful track of the dimensions of the various spaces one shows that this failure does not in the end affect the main argument. The discussion is, unfortunately, rather complicated and on first reading it is best to assume that the extended moduli spaces L_{j,Y_2} are themselves cut out transversely, and proceed directly to (vi) after reading to the beginning of Lemma (4.34) below.

We consider deformations of the extended equations defined by sections σ of the bundle $\mathscr{F}$ over the $W^j(\varepsilon)$. These will be constructed as linear combinations of sections associated to loops in Y_1. Let τ be a section of $\mathscr{F}$, so for each connection A we have an element $\tau(A)$ of $\Omega^+{}_{Y_1}(\mathfrak{g}_P)$ and the map τ commutes with the action of the gauge group. We say that τ is supported on a tubular neighbourhood G of a loop l in Y_1 if:

(a) $\tau(A)$ is supported in G for all connections A;
(*b*) τ depends only on the restriction of connections to G.

We will suppose the neighbourhoods G are chosen to be disjoint from y_1 and all the surfaces Σ_v in Y_1. Also we assume our deformations τ satisfy:

(c) there is a constant $C(G)$ such that $\|\tau(A)\|_{L^\infty} \le C(G)$ for all connections A.

There are of course many ways of constructing such sections; explicit examples are given in [6] where the parallel transport of the connection is used. (It is also possible to arrange that the sections are defined by non-linear pseudo-differential operators of negative order.) For any pair A, ξ, with A irreducible, satisfying the extended equations we can find a finite collection of loops l_i such that for arbitrarily small neighbourhoods G_i of the l_{i_2} there are deformations τ_i supported on G_i so that the $\tau_i(A)$ generate $H_{A,\xi}$ (see Lemma (2.5) of [6]). We want to choose a finite set of this kind which will handle *all* the solutions in $E^{-1}([\frac{1}{2}\varepsilon, \frac{3}{4}\varepsilon])$ simultaneously. First, however, we discuss the reducible solutions of the extended equations.

Recall that we have chosen ε so small that there are no reducible connections at all in W^j, for $0<j\leq k$ [cf. Proposition (4.12)]. In W^0 there certainly are reducible connections, the trivial connection $[\theta]$ and also connections with holonomy S^1. The trivial pair $[\theta, 0]$ satisfies the extended equations, so L_{0,Y_1} contains at least one point. Consider the linearisation of the extended equations about this point:

$$(b, \eta) \to d^+ b + \eta\omega,$$

where $b \in {}_{\mathrm{Ker}\, d^*c}\Omega^1_{Y_1} \otimes R^3$ and $\eta \in \Xi_\theta$. The kernel of this is trivial (by our choice of ω) so we see that *the trivial solution of the extended equations is isolated* (recall the discussion at the end of §IV(ii) above). By making ε small we can suppose that the trivial solution is the only point in L_{0,Y_1}, in particular there are no other reducible solutions of the extended equations. From now on ε can be fixed, and we can simplify our notation still further by writing U and W^j for $U(\varepsilon)$ and $W^j(\varepsilon)$.

We will use the following lemma in the construction of our deformations.

LEMMA (4.34). *There is a finite set of loops* $l_1, \ldots, l_s$ *in* Y_1, *disjoint neighbourhoods* G'_i *of the* l_i, *deformations* τ_i *supported on interior neighbourhoods* $G_i \subset\subset G'_i$ *and a number* $B>0$ *with the following property. Let* $[A, \xi]$ *be a point in* L_{j,Y_1} *with* $\frac{1}{2}\varepsilon \leq E(A) \leq \frac{3}{4}\varepsilon$ *and* I *be a subset of the index set* $\{1, \ldots, s\}$ *containing at most* k-j *elements. Let* N *be the subset of* $\{1, \ldots, s\}$ *consisting of those* i *which are not in* I *and for which:*

$$\int_{G'_i} |F(A)|^p \leq B.$$

Then the sections $\tau_i(A)$ *for* i *in* N *generate* $H^2_{A,\xi}$.

More briefly, this rather complicated condition says that we can remove all the sections τ_i for which the curvature is very large on G'_i, plus any other set of $k-j$ sections, and we still get a generating collection for the cokernel of the linearisation of the extended equations.

We will now prove Lemma (4.34). It is convenient to begin by allowing sections $\tau^*(A)$ which are combinations of δ functions at points of Y_1, rather than smooth sections of $\Omega^+_{Y_1}(\mathfrak{g}_P)$, and which depend on the holonomy around loops in Y_1. These make perfectly good sense in this context since the δ-functions can be paired with the harmonic representative of $H^2_{A,\xi}$. As in [6], for each irreducible solution of the extended equations we can choose a finite set of such τ^* generating the cohomology space. Moreover the loops can, by general position, be taken to avoid y_1, the surfaces Σ_ν, and any other given finite set of loops. Our first goal is to find a finite set of sections τ^*_i $(i=1, \ldots, s)$ of this kind, depending upon disjoint loops, and such that for any solution $[A, \xi] \in L_{j,Y_1}$ with $\frac{1}{2}\varepsilon \leq E(A) \leq \frac{3}{4}\varepsilon$, and any set I of indices containing at most $k-j$ elements, the $\tau^*_i(A)$ for i not in I generate $H^2_{A,\xi}$. This is done by induction on j, using the semi-continuity property (4.32) above. When $j=0$ the condition is vacuous. When $j=1$, the part of L_{1,Y_1} in $E^{-1}[\frac{1}{2}\varepsilon, \frac{3}{4}\varepsilon]$ is compact, by the weak compactness principle

(4.31). Then we can find a finite set of τ^*s to deal with this compact set by an obvious covering argument. The condition on removing sets I of size k is handled by taking $k+1$ parallel copies of each loop. With this set of τ^*s the part of L_{2,Y_1} on which the condition fails is again compact, using the weak compactness and semi-continuity properties. Any sequence in $E^{-1}[\frac{1}{2}\varepsilon, \frac{3}{4}\varepsilon] \subset L_{2,Y_1}$ without strongly convergent subsequences has a subsequence converging to a pair in $L_{1,Y}$ on the complement of a single point in Y_1. This point can lie at one of the loops already chosen, and by the semi-continuity property (4.33) the generating condition is satisfied for all but a finite number of pairs in this subsequence. We now extend our set of τ^*s to deal with the remaining compact part of L_{2,Y_1}, and continue by induction in the same way.

Having found this set $\{\tau_i^*\}$ a similar argument shows that we can mollify the τ_i^* slightly to get sections τ_i supported on tubular neighbourhoods G_i, with $G_i \subset\subset G_i'$, G_i' disjoint, and having the same generating properties as the τ_i. It remains only to choose the constant B so that the conditions of Lemma (4.34) are satisfied. To do this we again use induction on j. For $j=1$ we can immediately find an appropriate constant B_1: we can just put,

$$B_1 = 2 \sup \int_{G_i'} |F(A)|^p$$

where the supremum is taken over all i and all pairs $[A, \xi]$ in the compact set $E^{-1}[\frac{1}{2}\varepsilon, \frac{3}{4}\varepsilon]$ in L_{1,Y_1}. Then with this constant B_1 the set of points in L_{2,Y_1} where the condition fails is compact, since in a weakly convergent sequence only one of the integrals over a G_i' can tend to infinity, and this is compensated for by the fact that we are allowed to remove finite sets of size $(k-1)$ when considering L_{1,Y_1} but only those of size $k-2$ when considering L_{2,Y_1}. We now find a new value $B_2 > B_1$ to deal with this remaining compact part, and continue by induction.

Fix a set $\{\tau_i\}_{i=1}^s$ and a constant B as in (4.34). Let h be a monotone cut-off function, equal to 1 on $[0, B]$ and supported in $[0, 2B]$. For a connection A define:

$$h_i(A) = h\left(\int_{G_i'} |F(A)|^p\right).$$

We are now able to define a family of perturbations—sections of $\mathscr{F}$—p arametrised by $\mathbf{R}^s$. For a vector v in $\mathbf{R}^s$ we let:

$$\sigma_v(A) = \sum_{i=1}^{s} v_i h_i(A) \tau_i(A).$$

In brief, we consider general linear combinations of the τ_i, except that we arrange to ignore those values of i for which the curvature is large over G_i'. The point of the definition is this: suppose A_α converges to A_∞ on the complement of a finite set $\{z_1, \ldots. z_n\}$ in Y_1. Then $\sigma_v(A_\alpha)$ is supported away from the points z_r for large α. Going to a subsequence we can suppose that:

$$\sigma_v(A_\alpha) \to \sum_{i=1}^{s} t_i v_i h_i(A_\infty) \tau_i(A_\infty)$$

where:

$t_i = 0$ if there is a point z_r in the interior of G_i',
$t_i \in [0, 1]$ if there is a point z_r on the boundary of G_i',
$t_i = 1$ if no point z_r lies in the closure of G_i'.

In the light of this we make the following further definition. Let I be a subset of $\{1, \ldots, s\}$ with $|I| \le k$ and $\mathbf{t}: I \to [0, 1] \subset \mathbf{R}$ be a vector (with components t_i, for $i \in I$). We define the "contraction" of a section σ_v by $\mathbf{t}$ to be the section:

$$\mathbf{t}^*\sigma_v = \sum_{i=1}^{s} t_i v_i h_i(A)\tau_i(A),$$

where we set $t_i = 1$ for i not in I.

We now consider the equation:

$$F^+(A) + \xi\omega + \sigma_v = 0$$

in the three variables A, ξ, v. In fact we restrict v to lie in a small ball $B(\delta) \subset \mathbf{R}^s$ of radius δ. Then we can consider $(A, \xi, v) \to F^+(A) + \xi\omega + \sigma_v$ as a Fredholm section ψ of a vector bundle over $\Xi_j \times B(\delta)$. We can make yet another generalisation of the weak compactness principle to these equations, just as in [6]. Similarly, there is an extension of the semi-continuity property (4.32) to this situation. These are used to prove the following.

LEMMA (4.35). *For sufficiently small* δ *the zeros* $([A, \xi], v)$ *of* ψ *in* $\Xi_j \times B(\delta)$ *with* $\frac{1}{2}\varepsilon \leq E(A) \leq \frac{3}{4}\varepsilon$ *are transverse.*

Proof. By construction the derivatives of ψ in the v variable span the cokernel $H^2_{A,\xi}$ for all of these zeros in $\Xi_j \times \{0\}$. If there were a sequence of non-transverse zeros in $E^{-1}([\frac{1}{2}\varepsilon, \frac{3}{4}\varepsilon])$, with vectors $v_\alpha \to 0$, we could take a weakly convergent subsequence and get a contradiction to the semi-continuity property.

We now want to specialise the equation

$$F^+(A) + \xi\omega + \sigma_v = 0 \tag{4.36}$$

to a fixed value of v, obtaining a perturbation of the extended equation in Ξ_j. With this fixed v we can also consider the equations:

$$F^+(A) + \xi\omega + \mathbf{t}^*\sigma_v = 0 \tag{4.37}$$

in $([A, \xi], \mathbf{t})$ for all different subsets I of the index set.

COROLLARY (4.38). *There are arbitrarily small vectors* v *such that for all* j *the zeros* $[A, \xi]$ *of equation* (4.36) *on* Ξ_j *with* $\frac{1}{2}\varepsilon \leq E(A) \leq \frac{3}{4}\varepsilon$ *are all transverse, therefore forming a manifold of dimension* $8j - 3b_2^+(Y_1)$. *Moreover, we can choose* v *so that for each subset* I *of the index set* $\{1, \ldots, s\}$ *the space of solutions* $([A, \xi], \mathbf{t})$ *to equation* (4.37) *forms a manifold of dimension* $8j - 3b_2^+(Y_1) + |I|$.

Proof. The previous Lemma gives a universal zero set Z in the relevant portion of the $\Xi_j \times B(\delta)$. For the first part of the corollary we just have to choose a regular value of the projection from Z to $B(\delta)$. For the second part we choose similarly regular values of the projections to all the coordinate planes in $\mathbf{R}^s$.

Now fix v as in (4.38). We have achieved the goal of this subsection, finding simultaneous perturbations (4.36) of the extended equations on the Ξ_j with transverse zero sets. The extra condition in the second part of (4.38) will enter in an auxiliary way in (vi) below. To complete this subsection we modify (if necessary) our representatives V_ν, for all the Σ_ν in Y_1, so that all their multiple intersections are transverse to all the moduli manifolds of (4.38). To simplify notation we will still call these modified representatives V_ν. We can make the modification in such a way that the regions defining the V_ν are kept disjoint from the regions G_i defining the perturbation σ_v. No difficulties with reducible connections arise (our connections will in fact still be real analytic away from the G_i and the support of the potential function R).

(vi) End of proof

We now move on to complete the proof of (4.15), and hence of the vanishing theorem (4.9). We turn the focus of our attention back to the space $U=U_1(\varepsilon)$ of connections over X which are almost flat over most of $Z_1(\rho)$, and the bundle $\Xi \to U$. With ω and v fixed as in (v) above we can now consider the equation $F^+(A)+\xi\omega+\sigma_v=0$ over Ξ. This makes good sense, for small enough λ, since the second two terms are supported on a common subset $Z_1(r)$ for some r. We denote the solution spaces by $L'_{j,X}(\lambda)$, for $j \le k$. With the V_ν fixed as at the end of (v) we consider the intersection:

$$J(\lambda)=L'_{k,X}(\lambda)\bigcap V_1\bigcap \cdots\cdots \bigcap V_d.$$

Let $I'(\lambda)$ be the intersection of $J(\lambda)$ with the zero section in Ξ; it is obtained from the intersection $I_1(\lambda)$ that we considered in (4.15) by perturbing the ASD equations by the term σ_v. Recall that the virtual dimension of $I'(\lambda)$ is 0 while that of $J(\lambda)$ is 3. Finally let $J^*(\lambda)$ be the union of the path components of $J(\lambda)$ which contain points of $I'(\lambda)$. A simple adaptation of the argument in (4.14) shows that for small enough λ $I'(\lambda)$ is contained in $E^{-1}([0,\frac{1}{4}\varepsilon])$ say. The main step in the argument is then given by the following lemma.

LEMMA (4.39). *For small values of the parameter λ the union of components* $J^*(\lambda)$ *lies in* $E^{-1}([0,\frac{1}{2}\varepsilon])$.

In the proof of this vital lemma we have to appeal to yet another extension of our ordinary theory for the ASD equations to the modified equations we are considering here. In this case we need the analogue of the gluing result stated in (4.6). The technical aspects of this are discussed in Appendix 2.

To prove (4.39) it suffices to show that E does not take on the value $\frac{1}{2}\varepsilon$ on $J^*(\lambda)$ when λ is small. So suppose we have a sequence λ_α tending to zero and points $[A_\alpha, \xi_\alpha]$ in $J(\lambda_\alpha)$ with $E(A_\infty)=\frac{1}{2}\varepsilon$. We apply the analogue of the weak convergence property (4.1) to these solutions of the modified equations (4.36) over X. Thus we can suppose that the pairs $[A_\alpha, \xi_\alpha]$ converge to $[A', \xi']$ on the compliment of a finite set $z_1, \ldots, z_n$ in $Z(r)$. A' satisfies the ASD equation over Y_2 and over Y_1 it satisfies some contraction of the modified equation, i.e.

$$F^+(A')+\xi\omega+\mathbf{t}^*\sigma_v(A)=0$$

with a vector $\mathbf{t}$ defined by a set Q. We now carry out our familiar counting argument, but in a slightly more involved form because of this contraction. The important point to note first is that A' cannot be flat over either component Y_1, Y_2 since E is continuous under weak limits, so $E(A')=\frac{1}{2}\varepsilon$. Thus we can count dimensions without any exceptional cases.

Suppose that of the n points z_r where convergence fails, p points lie on at least one of the surfaces Σ_ν and q points lie in the closure of one of the disjoint neighbourhoods G'_j in Y_1 used to define σ_v. Then if k' is the Chern class of A' we have:

$$p+q+k' \le k.$$

Now the size of the set Q defining the contraction in (4.40) is q. So, by the last part of (4.38) the moduli space of solutions $([A, \xi], t)$ to the equation (4.40) over Y_1 containing $[A', \xi']|_{Y_1}$, has dimension $8k'_1-3b_2^+(Y_1)+q$, where k'_1 is the component of k' over Y_1. Combining with the contribution from Y_2, and using the fact that the intersections with the V_ν are all transverse we get that:

$$d=\tfrac{1}{2}(8k-3(1+b_2^+(X))\leq 2p+\tfrac{1}{2}(8k'-3b_2^+(X)+q),$$

so $k\leq k'+\frac{1}{2}p+\frac{1}{8}q$. So we deduce that $p=q=0$ and in fact the sequence $[A_\alpha, \xi_\alpha]$ is strongly λ_α convergent to $[A', \xi']$.

We now invoke the gluing construction for the modified equations. For small enough λ we can construct from A', ξ' a three-dimensional family S_λ of points in $J(\lambda)$, parametrised by a copy of SO(3). Clearly this forms a complete connected component of $J(\lambda)$. For large α the pairs $[A_\alpha, \xi_\alpha]$ lie in S_λ. But we also have that the variation of E from $\frac{1}{2}\varepsilon$ over S_λ tends to zero with λ [cf. (4.6, iii)]. So plainly the $[A_\alpha, \xi_\alpha]$ cannot be joined by paths in $J(\lambda)$ to $I'(\lambda)\subset E^{-1}[0, \frac{1}{4}\varepsilon]$, for small λ. This completes the proof of (4.39).

We now complete the proof of (4.15). We can fix λ in accordance with the Lemma above and look at the spaces $L_{j,X}$ of solutions to our modified equations. Assume first that all of these are cut out transversely, and similarly for all their intersections with the V_v. Then the familiar argument shows that $E^{-1}[0, \frac{1}{2}\lambda]\subset J(\lambda)$ is a compact subset of a smooth 3-manifold and by Lemma (4.39) the union of components $J^*(\lambda)$ containing all the points of $I'(\lambda)$ is a closed 3-manifold. $I'(\lambda)$ is the set of zeros of a section of a 3-plane bundle over $J^*(\lambda)$ so the algebraic sum $i'(\lambda)$, defined by perturbing to a transverse section, represents the Euler class of this bundle and hence is zero. It is then straightforward to show, arguing as in § III, that $i'(\lambda)$ represents the intersection number $i_1(\lambda)$ that we wished to calculate.

Finally we have to consider the gloomy possibility that the zero sets $L'_{j,X}$, or their intersections with the V_v, are not transverse. In that case we follow the inductive procedure of (v) above to choose further, arbitrarily small, deformations τ supported on neighbourhoods of loops in X such that the solutions of

$$F^+(A)+\xi\omega+\sigma_v(A)+\tau(A)=0$$

are transverse. As $\tau\to 0$ solutions of these equations with $\frac{1}{4}\varepsilon\leq E(A)\leq\frac{3}{4}\varepsilon$ and lying in all the V_v converge strongly to solutions of $F^+(A)+\xi\omega+\sigma_v(A)=0$. So the description of the part of the intersection crossing $E^{+1}\{\frac{1}{2}\varepsilon\}$ is the same, for small τ. Then perturb the V_v slightly to achieve compactness and argue as above.

§V. PROOF OF THEOREM C

(i) Moduli spaces of holomorphic bundles over complex curves and surfaces

Throughout this section we let S be a smooth, compact, simply connected complex algebraic surface. Let H be a hyperplane class in $H_2(S)$, which is dual to the first Chern class of a very ample line bundle over S, which we also denote by H. The hyperplane class is represented by the algebraic curves cut out by holomorphic sections of H, i.e. by the hyperplane sections of S in the corresponding projective embedding. Equally we can consider a Kahler metric ω on S with $[\omega]$ Poincaré dual to H. The Hodge index theorem asserts that:

$$\mathrm{H}^2_+=[\omega]\mathbf{R}\oplus H^{2,0}$$

so $b_2^+(S)=1+2\,p_g(S)$. We suppose the geometric genus p_g is strictly positive. Then the surface gives a C-manifold with the orientation β specified by: $-[\omega]\wedge$ (complex orientation of $H^{2,0}$). (see [6, 7] §(A)). The aim of this section is to prove the following.

THEOREM (5.1). *There is a number* $k(S, H)$ *such that for all* $k \geqslant k(S, H)$ *the polynomial invariants* $q_{k,S}$ *satisfy:*

$$q_{k,S}(H, \ldots\ldots, H) > 0.$$

This will be proved by combining some general theory which gives an algebraic description of the invariants of S with a technical algebro-geometrical theorem on the singularities in the moduli spaces for large values of k. For the general theory we consider the ASD moduli spaces defined using the Kahler metric ω, which will be fixed throughout this section. (This can be thought of as the opposite approach to that taken in §II–IV since the Kahler metrics form a subset of infinite codimension in the space of Riemannian metrics on S.) The basic point which enables us to transfer our discussion to the realm of algebraic geometry is that, in this special situation, the ASD solutions can be identified with certain *holomorphic bundles* over the algebraic surface S. Let us just consider rank 2 bundles.

Definition (5.2). A holomorphic rank 2 vector bundle E over S is H-stable if for every line bundle L over S admitting a non-trivial, holomorphic, bundle map into E we have:

$$c_1(L).H < \tfrac{1}{2}c_1(E).H.$$

Then we have the following.

THEOREM (5.3). *For each* $k > 0$ *there is a natural* (1:1) *correspondence between the gauge equivalence classes of irreducible ASD connections over* S, *on an* $SU(2)$ *bundle with* $c_2 = k$, *and the isomorphism classes of* H-*stable holomorphic bundles* E *over* S *with* $c_1(E) = 0$ *and* $c_2(E) = k$.

This was proved in [5] and more general results in the same direction were obtained by Uhlenbeck and Yau [34]. In outline, one associates to an ASD connection A the (0, 1) part of its covariant derivative, acting on sections of the 2-plane bundle U constructed from the fundamental representation of $SU(2)$. This is a coupled Cauchy–Riemann operator

$$\bar{\partial}_A : \Omega_S^{0,0}(U) \to \Omega_S^{0,1}(U).$$

The ASD condition implies that this operator endows U with a holomorphic structure (local holomorphic sections s of U are those with $\bar{\partial}_A s = 0$). This holomorphic structure is stable if A is irreducible and Theorem (5.3) asserts that, conversely, for any stable bundle we can construct an ASD connection. The reducible ASD solutions correspond to decomposable bundles $E = L \oplus L^{-1}$, with $c_1(L).H = 0$. It is not hard to extend all of our discussion to take account of these reducible connections but, for simplicity of exposition, we will assume that there are in fact no reductions of this kind in the moduli spaces we consider. Then we can say that for $k > 0$ the moduli spaces $M_k = M_{k,S}(\omega)$ are identified with the moduli spaces of stable holomorphic bundles. Theorem (5.3) refers only to an identification at the level of sets but the correspondence does respect, in an obvious sense, the structures on the two spaces. This is discussed in [7], §III (b). In particular our moduli spaces are endowed with the structure of complex analytic spaces, and the possible singularities of these precisely reflect the way that the moduli spaces are cut out by the ASD equations. Our main concern here, however, will be with the structure of the moduli spaces in the large and in particular the fact that, as we shall see below, they are naturally *quasi-projective complex varieties.*

This discussion can be extended to connections for other gauge groups. In the case of SO(3) connections we can get a general correspondence with SO(3, $\mathbb{C}$) holomorphic bundles. If we consider connections on an SO(3) bundle P for which $w_2(P)$ is the reduction of an integral class c of type $(1, 1)$ then there is a rather simpler algebro-geometric description in terms of stable bundles E with $c_1(E)=c$. In this situation an analogue of (5.1) holds for our invariants defined by the SO(3) moduli spaces. However, for brevity we will henceforth confine ourselves to the SU(2) case.

We want to describe now a projective embedding of the moduli spaces M_k which is compatible with the procedure used to define our invariants. For this we begin by discussing moduli spaces of holomorphic bundles over algebraic curves (Riemann surfaces). In our application these curves will be hyperplane sections in S. We use a description due to Gieseker [14, 15] of the moduli spaces of bundles over curves.

(ii) Gieseker's projective embedding

Recall [cf. §II(v), Remark (3.7)] that the polynomial invariants are defined using an equivariant line bundle on the space of connections over a (real) two-dimensional surface. We shall now explain the algebro-geometrical significance of this line bundle. Let C be a compact Riemann surface of genus $g \geqslant 2$ and U a complex 2-plane bundle over C with structure group SU(2). The relation between connections on U and holomorphic structures is very simple (see [1]). Any connection A endows U with a holomorphic structure (there are no "integrability conditions") via it's $\bar{\partial}$-operator, just as in the higher dimensional case above. Conversely, any holomorphic $SL(2, \mathbb{C})$ bundle is isomorphic to one obtained in this way. Let $\mathscr{A}_C$ be the space of connections on U. The familiar action of the gauge group $\mathscr{G}$ on A extends to the "complexified gauge group" $\mathscr{G}^c$ of special linear (but not necessarily unitary) automorphisms of U. The set of isomorphism classes of holomorphic $SL(2, \mathbb{C})$ bundles over C is identified with the space of $\mathscr{G}^c$ orbits in $\mathscr{A}_C$. A holomorphic $SL(2, \mathbb{C})$ bundle E over C is defined to be stable if any proper sub-bundle has negative degree; we then have a set $\mathscr{A}_C^s \subset \mathscr{A}_C$ of stable points and the quotient $\mathscr{A}_C^s/\mathscr{G}^c = M_C$ is naturally a smooth complex manifold—the moduli space of stable $SL(2, \mathbb{C})$ bundles over C.

Now the action of $\mathscr{G}$ on the line bundle $\tilde{\mathscr{L}}$ over $\mathscr{A}_C$ extends to the larger group $\mathscr{G}^c$. To see this recall that $\tilde{\mathscr{L}}$ can be described as the determinant line bundle of the coupled Dirac operator over C ([4], §II). This description depends upon the choice of a spin structure on C which we can regard as the choice of a square root $K_C^{1/2}$ of the canonical line bundle. The coupled Dirac operator can be identified with the $\bar{\partial}$ operator on $U \otimes K_C^{1/2}$, and the action of $\mathscr{G}^c$ on the kernels and cokernels is then immediate. In algebro-geometrical terms we consider the line bundle which attaches to any holomorphic bundle E the line:

$$\chi(E \otimes K_C^{1/2})^{-1} = \Lambda^{\max}(H^0(E \otimes K_C^{1/2}))^* \otimes \Lambda^{\max}(H^1(E \otimes K_C^{1/2})).$$

[Here H^i refers to the usual sheaf cohomology; $\chi(F)$ is the multiplicative Euler characteristic of a holomorphic bundle F.] We therefore obtain a holomorphic bundle $\mathscr{L}$ over the quotient M_C. (In fact, we can regard M_C as a subset of $\mathscr{B}_C^*$, the subset of flat connections, via the theorem of Narasimhan and Seshadri [26].) Holomorphic sections of $\mathscr{L}$ lift to equivariant sections of $\tilde{\mathscr{L}}$ over $\mathscr{A}_C^s$ and these automatically extend to $\mathscr{A}_C$ since the complement has codimension at least 3. Thus a holomorphic section of $\mathscr{L}$ gives a representative V of the kind we considered in §II(v), (3.6) and (3.7). All of our constructions apply equally well if we replace $\mathscr{L}$ by a power $\mathscr{L}^{\otimes n}$, we just multiply our polynomials by a factor. The key result we need is then as follows.

PROPOSITION (5.4) (Gieseker [15]). *$\mathscr{L}$ is an ample line bundle over M_C—the sections of some power $\mathscr{L}^n$ embed M_C as a quasi-projective variety in $\mathbb{CP}^N$.*

We will now review Gieseker's proof of this proposition, since the identification of the ample line bundle with our determinant line bundle is not given explicitly in Gieseker's work.

Given an SL(2, $\mathbb{C}$) bundle E over C we can find a positive divisor $D = p_1 + \ldots + p_d$ (with the p_i distinct, for simplicity) such that:

$$H^1(E \otimes K_C^{1/2} \otimes [D]) = 0.$$

Write $F = E \otimes K_C^{1/2} \otimes [D]$. Then $\chi(F) = \Lambda^{2d}(H^0(F))$, since H^1 is zero and H^0 has dimension $2d$ by Riemann–Roch. The exact sequence:

$$0 \to E \otimes K_C^{1/2} \to F \to \bigotimes_{i=1}^{d} F_{p_i} \to 0$$

induces a natural isomorphism (via the long exact cohomology sequence)

$$\chi(E \otimes K_C^{1/2}) = \chi(F) \otimes \prod_i \Lambda^2 F_{p_i}. \tag{5.5}$$

Gieseker shows first that we can find a divisor D so that the vanishing of H^1, as above, holds for all stable bundles E. One can then shift attention to the corresponding moduli space of bundles F, and recover the determinant line bundle using (5.5). Let T_F be the evaluation map:

$$T_F: \Lambda^2 H^0(F) \to H^0(\Lambda^2 F).$$

For all the bundles we are considering $\Lambda^2 F$ is isomorphic to the fixed line bundle $L = K_C \otimes [2D]$. Let $H = H^0(L)$ and $W = \mathrm{Hom}\,(\Lambda^2 \mathbb{C}^{2d}, H)$. The group SL($2d$, $\mathbb{C}$) acts on W and we can associate to T_F a point $[T_F]$ of the quotient set $\mathbb{P}(W)/\mathrm{SL}(2d, \mathbb{C})$ (i.e., the ambiguity in fixing isomorphisms of W with $\mathbb{C}^{2d}$ and $\Lambda^2 F$ with L is precisely absorbed by the action of $\mathbb{C}^* \times \mathrm{SL}(2d, \mathbb{C})$].

Geometric invariant theory considers exactly these kind of quotients. According to the general theory [15, 25, 27] there is an open set of "stable points" $\mathbb{P}(W)_s \subset \mathbb{P}(W)$, preserved by the group action, such that:

(i) $Q = \mathbb{P}(W)_s/\mathrm{SL}(2d, \mathbb{C})$ is Hausdorff in the quotient topology and hence is naturally a complex manifold.

(ii) The line bundle $\mathcal{O}(d)$ descends to Q. For large enough n the invariant sections of $\mathcal{O}(nd)$ (invariant polynomials on W) embed Q as a quasi-projective subvariety of a projective space $\mathbb{CP}^N$.

Thus the quotient $\mathcal{O}(nd)/\mathrm{SL}(2d, \mathbb{C})$ is an ample line bundle over Q. Gieseker then shows that:

(iii) If E is stable then $[T_F]$ lies in Q.

(iv) The map $i: M_C \to Q$, $i([E]) = [T_F]$ is an embedding, with image a closed algebraic subvariety of Q.

Putting (ii) and (iv) together we see that M_C is displayed as a quasi-projective variety, and $i^*(\mathcal{O}(d)/\mathrm{SL}(2d, \mathbb{C}))$ is an ample line bundle over the moduli space. It is now elementary to identify this line bundle with the determinant line bundle $\mathscr{L}$. Fix isomorphisms of the fibres $\mathscr{L}_{p_i}$ with $\mathbb{C}$. Given a bundle E choose isomorphisms:

$$\phi: H^0(F) \to \mathbb{C}^{2d},$$

$$\psi: \Lambda^2 F \to L.$$

Using these isomorphisms T_F yields a point $S(\phi, \psi)$, say, in W. We can think of $S(\phi, \psi)$ as being contained in a fibre of the tautological line bundle $\mathcal{O}(-1)$ over $\mathbb{P}(W)_s$. If E is stable $S(\phi, \psi)$ is not zero, by (iii), and we can take the $(-d)$ power and descend to the quotient to get a point:

$$B_1(\phi, \psi) = [S(\phi, \psi)^{-d}]$$

in the fibre of $\mathcal{O}(d)/\mathrm{SL}(2d, \mathbb{C})$ over $i(E)$. On the other hand ϕ and ψ define a point:

$$B_2(\phi, \psi) = \det(\phi)^{-1} \otimes \Pi\, \psi_{p_i} \in \Lambda^{2d}(H^0(F)) \otimes \Pi\, \Lambda^2 F^*_{p_i}.$$

Suppose now that we change ϕ, ψ to ϕ', ψ' where $\phi' = \alpha g \phi$, $\psi' = \beta \psi$ with g in $\mathrm{SL}(2d, \mathbb{C})$, α and β in $\mathbb{C}^*$. Then:

$$S(\phi', \psi') = \alpha^2 \beta^{-1} (\Lambda^2 g)^{-1} S(\phi, \psi)$$

so:

$$B_1(\phi', \psi') = \alpha^{-2d} \beta^d B_1(\phi, \psi)$$

whereas:

$$B_2(\phi', \psi') = \det(\alpha g) \beta^{-d} B_1(\phi, \psi) = \alpha^{2d} \beta^{-d} \beta_2(\phi, \psi).$$

Hence there is a natural isomorphism, independent of choices, between $\Lambda^{2d}(H^0(F)) \otimes \Pi\, \Lambda^2 F^*_{p_i}$ and the dual of the fibre of $\mathcal{O}(d)/\mathrm{SL}(2d, \mathbb{C})$ (sending $B_2(\phi, \psi)$ to $B_1(\phi, \psi)$), and in view of (5.5) this shows that the determinant line bundle is the ample line bundle over M_C as required.

(5.6) *Remark.* As we have explained above these sections of $\mathcal{L}^n$ define equivariant sections of $\tilde{\mathcal{L}}^n$ over all of $\mathcal{A}_C$. In fact, the sections are forced to vanish on the points of $\mathcal{A}_C \backslash \mathcal{A}_C^s$ corresponding to bundles which are not "semi-stable", much as, in the unitary set-up, a $\mathcal{G}$ invariant section is forced to vanish at reductions of positive degree. However, we can choose sections which do not vanish on semi-stable bundles, and in particular on the trivial holomorphic bundle, corresponding to the trivial flat connection.

(iii) Restriction to curves: the main argument

Gieseker also applied the techniques sketched above to construct a projectiven embedding of moduli spaces $M^+_{k,S}$ of "Gieseker stable" bundles over the algebraic surface S [16]. The condition of Gieseker stability is weaker than that in Definition (5.2), so these are not the same as our moduli spaces. However, $M_{k,S}$ is a Zariski open subset of $M^+_{k,S}$: in particular $M_{k,S}$ can be viewed as a quasi-projective variety.

The basic fact that we use to prove Theorem (5.1) is that the cohomology class $\mu(H)$ is the first Chern class of an ample line bundle over the moduli space $M_{k,S}$, i.e. a multiple of $\mu(H)$ is pulled back under a projective embedding of $M_{k,S}$. This is *not* the same as Gieseker's projective embedding of $M_{k,S} \subset M^+_{k,S}$, but is defined rather by restriction to curves in S.

We will use a theorem of Mehta and Ramanathan [23]. They prove that if E is an H-stable bundle over S then there is a number $p(E)$ such that the restriction of E is stable for generic curves C in $|pH|$. Clearly (since the moduli spaces have finite type) p can be chosen to depend only on k. Then the finiteness condition also shows that we can choose a set of curves $C_1, \ldots C_N$ in $|pH|$ such that for any stable bundle E with $c_2 \leq k$ the restriction of E to at least one of the C_i is stable. Now recall that our determinant line bundle associated to a curve C is

$$\mathcal{L}_{C,F} = \chi(F \otimes K_C^{1/2})^{-1}$$

for a bundle F over C. Fix a section s of $H^{\otimes p}$ cutting out C in S, and suppose for simplicity that H and K_S are even. Then we have a natural exact sequence:

$$0 \to E \otimes K_S^{1/2} \otimes H^{-p/2} \to E \otimes K_S^{1/2} \otimes H^{p/2} \to E|_C \otimes K_C^{1/2} \to 0.$$

As in (5.5) this induces an isomorphism of Euler characteristics:

$$\chi(E|_C \otimes K_C^{1/2}) = \chi(E \otimes K_S^{1/2} \otimes H^{-p/2})^{-1} \otimes \chi(E \otimes K_S^{1/2} \otimes H^{p/2}),$$

where the right hand side is independent of the particular curve C in $|pH|$. So we get natural holomorphic isomorphisms between the pull backs of all the line bundles $\mathscr{L}_{C_i}$ to $M_{k,S}$. (If K_S is not even we can choose a suitable H for which the combination of square roots used above is well defined.) Thus we can think of a single line bundle $\mathscr{L}$ over $M_{k,S}$ and spaces W_i^* of holomorphic sections of $\mathscr{L}^n$ determined by restriction to C_1. These sections then give a regular map:

$$\alpha: M_{k,S} \to \mathbb{P}(W_1 \oplus \ldots \oplus W_N) = \mathbb{P}.$$

(The sections W_i^* individually give rational maps on $M_{k,S}$; the composite of α with the projections in the projective spaces.) It is easy to show that for large enough p the map α is generically an embedding. However, we may as well then suppose that $p=1$, since the statement of Theorem (5.1) is unaffected by multiplying H by p.

The image of α is a quasi-projective variety in $\mathbb{P}$. As such it has a non-negative degree, deg α $(M_{k,S})$, the number of intersection points of $\alpha(M_{k,S})$ with the appropriate number of generic hyperplanes in $\mathbb{P}$. Moreover, this degree is unchanged if we adjoin an extra curve to our collection $C_1, \ldots C_N$ and is therefore independent of the choice of these curves. The degree is zero only if $M_{k,S}$ is empty.

We will now use two important technical facts about the moduli spaces $M_{k,S}$ when k is large. First, the moduli spaces are not empty. This was proved in the context of ASD connections by Taubes [31]. In the algebro-geometric setting similar results have been obtained by Gieseker [16]. Second, we need a fact whose proof will be the main business of the remainder of this section. Let $\Sigma_k \subset M_{k,S}$ be the algebraic sub-variety representing bundles E with $H^2(End_0 E)$ not equal to 0. Here End_0 denotes the trace-free endomorphisms. Thus, by the deformation theory, the complement of Σ_k in $M_{k,S}$ is a smooth complex manifold of the "correct" dimension.

$$\dim_{\mathbb{C}} M_{k,S} \backslash \Sigma_k = d_k = 4k - 3(1 + p_g(S)). \tag{5.7}$$

And, viewed as part of the ASD moduli space, it is cut out transversely by the ASD equations (see [7] §III(b)). Although it may itself be singular Σ has a complex dimension (largest dimension of a smooth stratum). The result we shall prove in this paper is given below.

THEOREM (5.8). *There are constants* a, b *depending only on* S *and the ray spanned by* H *in* $\mathrm{H}_2(\mathrm{S})$ *such that*:

$$dim_{\mathbb{C}} \Sigma_k \leq \mathrm{a} + \mathrm{b}\mathrm{k}^{1/2} + 3\mathrm{k}.$$

Roughly speaking, this says that the Kahler metrics are not too far from being generic metrics if we take k to be large. We will now go on to complete the proof of Theorem (5.1) (Theorem C of §I) assuming Theorem (5.8).

We claim that for large enough k the degree deg $(\alpha(M_{k,S}))$ is n^d times the integer $q_{k,S}(H, \ldots\ldots, H)$. This plainly implies (5.1) since the moduli spaces are non-empty so the degree is positive. What we have to show then is that the intersection procedures used to calculate the polynomial invariant and the projective degree are compatible. First choose k_0 so large that M_k is non-empty for $k \geqslant k_0$ and

$$4k_0 - 3(1+p_g(S)) > a + bk^{1/2} + 3k.$$

Then $M_k \backslash \Sigma_k$ is non-empty for $k \leq k_0$ (since the dimension of the moduli space in this holomorphic situation must be at least the virtual dimension). We now make a careful choice of the set of curves C_i defining the embedding (although in the end we see that any generic collection would do). As usual, we want to consider all the moduli spaces $M_{j,S}$ for $j \leq k$, as well as $M_{k,S}$. So we can suppose that we have projective embedding, also denoted by α, of these moduli spaces. A vector in W_i^* gives representatives $V_i \cap M_j$ in all the moduli spaces simultaneously [in the notation of §II(v), §III (3.6), (3.7)]. Here we use the obvious fact that the restriction from S to C_i in the holomorphic and differential geometric pictures is compatible. The image $\alpha(V_i \cap M_j)$ is a hyperplane section of α (M_j). Suppose inductively that we have chosen curves $C_1, \ldots C_m$ and elements $w_1, \ldots w_m$ of $W_1^*, \ldots W_m^*$ satisfying the following genericity properties:

(i) $V_{i(1)} \cap \ldots \cap V_{i(p)} \cap M_j$ has complex dimension dim $M_j - p$ for all collections $\{i(1), \ldots i(p)\} \subset \{1, \ldots m\}, j \leq k$.

(ii) $V_1 \cap \ldots \cap V_m$ has complex dimension dim $\Sigma_k - m$.

In condition (i) here we should emphasize that we are referring to the (complex) dimension of M_j as a complex variety, which may be quite different from its complex virtual dimension, for small values of j. To pass from stage m to $m+1$ we look at all the possible intersections made from $V_1, \ldots V_m$. Each has a finite number of components and we choose a point from each component. We can choose the next curve C_{m+1} so that the restriction of each of this finite set of bundles to C_{m+1} is stable. Then the generic element of W_{m+1}^* gives a hyperplane section V_{m+1} with the desired properties. In this way we choose a set of d representatives $V_1, \ldots V_d$ having the general position properties above with $m = d$, depending on curves $C_1, \ldots, C_d$. We now extend this set of curves, if necessary, to get projective embeddings α of our moduli spaces. Let Π_i be the hyperplane in $\mathbb{P}$ corresponding to V_1, so that

$$\Pi_1 \cap \Pi_2 \cap \ldots \cap \Pi_d \cap \alpha(M_k) = \alpha(V_1 \cap \ldots V_d \cap M_{k,S}).$$

We finish the proof by checking three things. First, the system $(g, V_1, \ldots, V_d)$ is admissible, so long as k was chosen large enough. In fact if D is the maximal dimension of the moduli spaces M_j for $j \leq k_0$ it suffices to take k such that:

$$d_k = 4k - 3(1+p_g) > D + 2(k - k_0 + 1).$$

This follows from the genericity condition (i) above on V_i. Thus the intersection $I = V_1 \cap \ldots \cap V_d \cap M_k$ (which is a finite set of points) can be used to calculate $q_{k,S}(H, \ldots, H)$ when we take account of the relevant multiplicities. Second, the multiplicities with which the points of I are counted in the calculation of $q_{k,S}$ agree with those with which the points of $\alpha(I)$ are counted in the intersection $\Pi_1 \cap \ldots \cap \Pi_d \cap \alpha(M_k)$ in $\mathbb{P}$. This is straightforward checking of definitions. Note that no point of I lies in Σ_k so we only have to perturb the V_i to get transverse intersections. Finally, we check that the intersection $\Pi_1 \cap \ldots \cap \Pi_d \cap \alpha(M_k)$ is a valid way of calculating the degree of $\alpha(M_k)$ in $\mathbb{P}$, i.e. that the intersection:

$$\Pi_1 \cap \ldots \cap \Pi_d \cap \{\overline{\alpha(M_k)} \backslash \alpha(M_k)\},$$

is empty, where $\overline{\alpha(M_k)}$ is the ordinary closure in $\mathbb{P}$. This follows from the fact that the system is admissible. If the intersection contained a point p we could find a sequence E_λ of holomorphic bundles over S such that $\alpha(E_\lambda)$ tends to p. Let A_λ be the corresponding sequence of ASD connections over X, we can suppose they converge weakly to $(A_\infty; x_1, \ldots x_r)$. Then if no point x_s lies on a curve C_i we must have that $\alpha(A_\infty)$ lies in Π_i [since the $\alpha(E_\lambda)$ converge to a point in Π_i and the section cutting out V_i is defined by restriction to C_i). Then it would follow, counting points, that the system $(g, V_1, \ldots V_d)$ is not admissible and we have a contradiction. This completes the proof of (5.1), assuming Theorem (5.8). [Notice that this final step in the proof can be interpreted as saying that there is some compatibility between the two compactifications $\bar{M}_k$ and $\overline{\alpha(M_k)}$ of the moduli space. Also, in the argument above we have ignored the fact that the trivial bundle is not stable; however, it is semi-stable and this is all that is needed, by remark (5.6).]

(iv) Differential geometric approach

The ideas we have been developing can be summarised by saying that the moduli spaces $M_{k,X}$ are quite like compact manifolds and, when X is an algebraic surface, $\mu(H)$ is the Chern class of an ample line bundle $\mathscr{L} \to M_k$. One can see a corresponding approach in de Rham cohomology.

Let X be any compact 4-manifold and ρ a closed 2-form on X. Write α for the Poincaré dual of the de Rham class of ρ. Recall that one way of defining the class $\mu(\alpha)$ is:

$$\mu(\alpha) = c_2(\mathbb{E})/\alpha$$

where $\mathbb{E}$ is the "universal" bundle on the product space $\mathscr{B}_X^* \times X$. If we choose a connection $\mathbb{A}$ on $\mathbb{E}$ this gives a definite de Rham representation of $\mu(\alpha)$:

$$\tilde{\Omega}_\alpha = \int_X c_2(\mathbb{A}) \wedge \rho. \tag{5.9}$$

Here we have written $c_2(\mathbb{A})$ for the Chern–Weil 4-form $\frac{1}{8\pi^2} Tr(F(\mathbb{A})^2)$. Suppose we use the same formula (5.9) upstairs on the product $\mathscr{A} \times X$. Then there is a very simple universal connection $\mathbb{A}$, equal to A on the slice $\{A\} \times X$ and flat in the X direction. A 2-form on $\mathscr{A}$ is a skew-symmetric function on tangent vectors (bundle valued 1-forms) a, b. Then an easy calculation shows that:

$$\Omega_\alpha(a, b) = \frac{1}{8\pi^2} \int_X Tr(a \wedge b) \wedge \rho. \tag{5.10}$$

This is a $\mathscr{G}$-equivariant form on $\mathscr{A}$. In general it does not descend to $\mathscr{A}/\mathscr{G}$ because it does not ahilliate the tangent vectors along $\mathscr{G}$-orbits. However, if ρ is *self-dual* and we restrict to *anti-self-dual* connections (5.10) does induce a well-defined 2-form Ω_α on the quotient. For if

$$a = d_A u$$

is a vector along the orbit and b is a tangent vector to the ASD connections, so $d_A^+ b = 0$, we have:

$$\begin{aligned}
\Omega_\alpha(a, b) &= \frac{1}{8\pi^2} \int_X Tr(d_A u \wedge b) \wedge \rho \\
&= \frac{1}{8\pi^2} \int_X d(Tr(u \wedge b)) \wedge \rho - Tr(u \wedge d_A b) \wedge \rho \\
&= \frac{-1}{8\pi^2} \int_X Tr(u \wedge d_A b) \wedge \rho \\
&= 0,
\end{aligned}$$

since the wedge product of self-dual and anti-self-dual forms is zero. Thus $\Omega_\alpha(a, b)$ is independent of the choice of lifts a, b used to represent tangent vectors in the moduli space. It is not hard to show that Ω_α is closed and is a de Rham representative for $\mu(\alpha)$. (This is linked to the ideas of moment maps and the Marsden–Weinstein quotient in symplectic geometry [1]).

Now suppose that X is a complex Kahler surface, with Kahler form ω. The Kahler form is a closed self-dual 2-form so the construction above gives an explicit representative Ω for $\mu(H)=\mu(P.D.(\omega))$. This ties in on the one hand with the complex structure on the moduli space; Ω is a form of type (1, 1) and is the Kahler form for the standard L^2 metric on M:

$$\|a\|^2=\int_X |a|^2 d\mu$$

where a is the horizontal representative with $d_A^* a=0$ [20]. On the other hand if H is an integral class Ω is, up to a multiple, the curvature form of a natural connection on the determinant line bundle $\mathscr{L}$ [8]. So if the moduli space M is non-empty, smooth and compact we can prove the positivity of $\langle \mu(H)^d, M\rangle$ by the "Wirtinger" argument: $\langle \mu(H)^d, M\rangle$ is $(d-1)!$ times the Riemannian volume of M.

To take this approach further one would need, in addition to the discussion of singularities in (vi) below, to understand the behaviour of these 2-forms at the ends of the moduli spaces. This should be possible but we have chosen the algebro-geometric approach in our proof.

(v) Controlling the singularities, determinant zero case

This section and the next contain a proof of Theorem (5.8). The dimension of $H^2(End_0 E)$, as E varies over Σ_k is semi-continuous. So there is a decomposition of Σ_k into sub-varieties $\Sigma^{(i)}$ $(i=1, 2, \ldots.)$ on which

$$\dim H^2(End_0 E)=\dim H^0(End_0 E \otimes K_S)=i.$$

It suffices to give bounds on the dimension of each $\Sigma^{(i)}$. Over a fixed one of these pieces of Σ the vector spaces $H^\circ(End_0 E \otimes K_S)$ fit together into the total space of a vector bundle. Let $\mathscr{H}^{(i)}$ be the complement of the zero section in this bundle. Points of $\mathscr{H}^{(i)}$ correspond to pairs (E, s) where E is stable and s is a non-zero section of $End_0 E \otimes K_S$. The determinant det s is a section of K_S^2 (since E has rank 2). Let $U^{(i)} \subset \mathscr{H}^{(i)}$ be the Zariski open set (possibly empty) on which det s is *not* identically zero. Our proof divides into two parts, establishing the following propositions which together imply Theorem (5.8).

PROPOSITION (5.11). *There are constants* a, b (*independent of* k, i) *such that:*

$$dim(\mathscr{H}^{(i)}\backslash U^{(i)})\leq 3k+a\sqrt{k}+b.$$

PROPOSITION (5.12). *There is a constant* c (*independent of* k, i) *such that:*

$$dim\, U^{(i)}\leq 2k+c.$$

Proof of Proposition (5.11). Let (E, s) represent a point in $\mathscr{H}^{(i)}\backslash U^{(i)}$. In local trivialisations s is represented by a holomorphic, matrix-valued function with trace and determinant identically zero. Over the non-empty, Zariski-open, subset $\Omega \subset S$ where s is non-zero the kernel of:

$$s\colon \mathcal{O}(E)|_\Omega \to \mathcal{O}(E \otimes K_S)|_\Omega$$

is clearly a line sub-bundle of E. It is a basic fact that the sheaf kernel

$$\mathscr{K}=\operatorname{Ker} s\subset\mathcal{O}(E)$$

is a locally free rank 1 sheaf over S. That is, there is a line bundle $L\to S$ and a holomorphic bundle map

$$\alpha\colon L\to E$$

which defines an isomorphism $\mathcal{O}(L)\to\mathscr{K}$. So we get an exact sequence of sheaves:

$$0\to\mathcal{O}(L)\xrightarrow{\alpha}\mathcal{O}(E)\to\mathcal{O}(L^{-1})\otimes\mathscr{I}\to 0 \tag{5.13}$$

where $\mathscr{I}\subset\mathcal{O}_S$ is an ideal sheaf. The quotient $\mathcal{O}_s/\mathscr{I}=\mathcal{O}_z$ is the structure sheaf of a 0-dimensional scheme z, supported on a finite subset $\{x_1,\ldots,x_m\}$ in S. Let z_λ be the part of z supported on x_λ. There is an associated degree or multiplicity, $d_\lambda>0$, which can be defined topologically or algebraically [18]. Topologically it is the multiplicity of the zero x_λ of the section α of $E\otimes L^{-1}$. So:

$$c_2(E\otimes L^{-1})=k+c_1(L)^2=\sum_{\lambda=1}^{m} d_\lambda. \tag{5.14}$$

Algebraically, it is $\dim_{\mathbf{C}}\mathcal{O}_{z_\lambda}$. We can think of z_λ as representing d_λ points of s which have collided at x_λ.

Extensions like (5.13) are very familiar in the theory of 2-plane bundles over surfaces (compare [11, 18, 29]). They are classified by a vector space

$$T=\operatorname{Ext}^1(L^{-1}\otimes\mathscr{I},L)$$

which fits into an exact sequence, relating local and global terms [15, p. 729]

$$H^1(L^2)\to T_1\to\bigoplus_\lambda\operatorname{Ext}^2(\mathcal{O}_{z_\lambda}\otimes L^{-1},L).$$

There is a duality:

$$\operatorname{Ext}^2(\mathcal{O}_{z_\lambda}\otimes L^{-1},L)\cong\mathcal{O}_{z_\lambda}\otimes_{\mathbf{C}}(L^2\otimes K)^*_{x_\lambda}$$

[18, p. 707] so $\dim\operatorname{Ext}^2(\mathcal{O}_{z_\lambda}\otimes L^{-1},L)=d_\lambda$ and

$$\dim T\leq h^1(L^2)+k+c_1(L)^2. \tag{5.15}$$

The pair (E,α) is determined, up to isomorphism, by L, z and the extension class in T. We will now use the fact that L arose as the kernel of s. Clearly

$$s^2\colon E\to E\otimes K_S^2$$

is identically zero. This implies that s has a factorisation:

$$\mathcal{O}(E)\to\mathcal{O}(L^{-1})\xrightarrow{\tilde{s}}\mathcal{O}(L\otimes K_S)\xrightarrow{\alpha\otimes 1}\mathcal{O}(E\otimes K_S)$$

(One sees this immediately away from the zeros of α, the map $\tilde{s}$ extends by Hartog's Theorem).

So there is a non-zero bundle map from L^{-1} to $L\otimes K_S$. Hence:

$$\deg(L^{-1})\equiv c_1(L^{-1})\cdot H\leqslant\deg(L\otimes K_S).$$

on the other hand E is stable and $c_1(E)=0$ so we have:

$$-\frac{1}{2}\deg K_S \leqslant \deg L < 0. \tag{5.16}$$

The Hodge Index Theorem asserts that $\langle H\rangle$ is a maximal positive subspace for the intersection form on $\mathrm{Pic}(S)\otimes\mathbb{R}\subset H^{1,1}(S)$. We use this to get a bound on $h^1(L^2)=\dim H^1(L^2)$. Write

$$c=c_1(L)=rH+c^\perp$$

where $c^\perp$ is orthogonal to H. (5.16) gives a fixed bound on r. Then:

$$\begin{aligned} c^\perp\cdot c_1(K_S) &\le A_0(S)\sqrt{|c^\perp . c^\perp|} \\ &\le A_0(S)(\sqrt{k}) \end{aligned} \tag{5.17}$$

using (5.14) [$A_i(S)$ will be constants depending only on S and the ray of H]. So

$$c_1(L)\cdot c_1(K_S)\le A_0(S)\sqrt{k}+A_1(S).$$

Now Riemann–Roch gives:

$$\begin{aligned} h^1(L^2)= &-2c_1(L)^2+c_1(L)\cdot c_1(K_s) \\ &-(1+p_g(S))+h^0(L^2)+h^0(L^{-2}\otimes K_S). \end{aligned} \tag{5.18}$$

L^2 cannot have a non-trivial section, since its degree is negative, while the section $\tilde{s}$ of $L^2\otimes K_S$ gives an injection:

$$\tilde{s}: H^0(L^{-2}\otimes K_S)\hookrightarrow H^0(K_S^2).$$

So:

$$h^1(L^2)\le A_0(S)\sqrt{k}+A_2(S)-2c_1(L)^2,$$

and

$$\dim T\le k+A_0(S)\sqrt{k}+A_2(S)-c_1(L)^2$$

by (5.15).

Finally, we decompose $\mathscr{H}^{(i)}\setminus U^{(i)}$ into open sets labelled by the different line bundles L (there are only finitely many possibilities satisfying our inequalities). Let J be the $\mathbb{C}^*$ bundle over one of these sets given by a choice of α. For each point j of J we get a scheme $Z(j)$ of fixed total degree $d=\Sigma_{d_\lambda}$. These schemes are parametrised by a Hilbert scheme $S^{[d]}$; a smooth, connected complex manifold of dimension $2d$ [13, 19]. (This is the dimension of the open subset of $S^{[d]}$ parametrising unordered d-tuples of points in S.) The universal property of the Hilbert scheme shows that there is a regular map $J\to S^{[d]}$ sending j to $Z(j)$. By the discussion above the fibres of this map have dimension at most $\dim T$. So:

$$\begin{aligned} \dim J &= 2c_2(E\otimes L^{-1})+\dim T \\ &= 2(k+c_1(L))^2+\dim T \\ &\le 3k+A_0(S)k+A_2(S)+c_1(L)^2. \end{aligned}$$

But $c_1(L)^2$ is bounded above by a fixed constant [combine the Hodge Index Theorem with (5.16)]. So

$$\dim J\le 3k+a\sqrt{k}+b$$

as required.

(vi) Branched covers: proof of (5.12)

(a) It is convenient to arrange the argument by considering first a more general situation. Let $F \to X$ be a 2-plane bundle over surface. Suppose $L \to X$ is a line bundle and that s is a holomorphic section of $End\, F \otimes L$ with trace s identically zero and $\det s = -\Delta$ is a section of L^2 vanishing on an effective divisor $C \in |2L|$. The main idea we need is a spectral representation of F (and s) by a rank 1 sheaf on a double cover of X. This is similar to techniques which are used for studying integrable differential equations—see [17] for example.

Define a subspace $\tilde{X}$ in the total space of L:

$$\tilde{X} = (\lambda \in L \,|\, \det((\lambda - s): F \to F \otimes L) = 0. \tag{5.19}$$

Since $Tr(s) = 0$ the defining equation is: $\lambda^2 = \Delta$. $\tilde{X}$ is the double cover of X, branched along C. We regard it as a complex space, with the obvious structure sheaf:

$$\mathcal{O}_{\tilde{X}} = \mathcal{O}_{(L)} / (\lambda^2 - \Delta) \tag{5.20}$$

[$\mathcal{O}_{(L)}$ the holomorphic functions on the total space of L]. Let $\pi: \tilde{X} \to X$ be the holomorphic projection. We define a sheaf $\mathscr{S}$ over $\tilde{X}$ to be the sheaf associated to a pre-sheaf $\mathscr{S}_0$:

$$\mathscr{S}_0(V) = \{\varphi \in \mathcal{O}_V(\pi^*(F) | (s - \lambda)\varphi = 0 \bmod (\Delta - \lambda^2)\} / (\Delta - \lambda^2) \mathcal{O}_V(\pi^*(F))$$

for V open in L.

It is easy to see that $\mathscr{S}$ is a sheaf of $\mathcal{O}_{\tilde{X}}$ modules. We let $[X]$ denote the line bundle over L, or $\tilde{X}$, associated to the zero section (so sections of $[X]$ are functions with a pole at $\lambda = 0$).

Main Claim. There is a natural isomorphism of sheaves over X:

$$\pi_*(\mathscr{S} \otimes [\mathrm{X}]) \cong \mathcal{O}(\mathrm{F}).$$

Under this isomorphism the bundle map s *is induced from the canonical map*

$$\mathscr{S} \otimes [\mathrm{X}] \to \mathscr{S} \otimes \pi^*(\mathrm{L})$$

(*multiplication by λ, in local coordinates*).

Notice that the claim is rather obviously true away from C: the map s has distinct eigenvalues and the assertion is just that F is the direct sum of eigenspaces.

To prove the claim in general we write down maps $\mathcal{O}(F) \underset{\beta}{\overset{\alpha}{\leftrightarrows}} \pi_*(\mathscr{S} \otimes [X])$ and check that they are mutually inverse.

Definition of α. Let $W \subset X$ be open and define: $\alpha_0: \mathcal{O}_W(F) \to \mathscr{S}_0(\pi^{-1}(W)) \otimes [X]$ by

$$\alpha_0(f) = \left[\frac{1}{\lambda}(s + \lambda)\right].$$

The image lies in $\mathscr{S} \otimes [X]$ because:

$$(s - \lambda)(s + \lambda) f = (s^2 - \lambda^2) f = (\Delta - \lambda^2) f.$$

α_0 induces a sheaf map α.

Definition of β. An element of the stalk of $\pi_*(\mathscr{S}_0 \otimes [X])$ at a point x in X can be represented (not uniquely) by a section φ of $\pi^*(F)[X]$ defined over an open subset $V \subset L$ containing $\pi^{-1}(x)$. We can suppose that, in a local trivialisation of L, V is a product $W \times D$, where D is a domain in "λ-space", and that there is a contour $\Gamma \subset D$ surrounding $\pi^{-1}(W) \cap \tilde{X}$. ($\Gamma$ need not be connected.) Then set:

$$\beta_0(\varphi) = \frac{1}{2\pi i} \int_\Gamma \frac{\lambda\varphi}{\lambda^2 - \Delta^2} d\lambda, \text{ a section of } F \text{ over } W.$$

One checks that this is independent of the choices of Γ and local trivialisation. Also: $\beta_0(\varphi + \psi(\lambda^2 - \Delta)) = \beta_0(\varphi)$ by Cauchy's Theorem. So β_0 induces the desired sheaf map β.

Now $\beta\alpha = 1$ since

$$\int_\Gamma \frac{s+\lambda}{\lambda^2 - \Delta} d\lambda = \int_\Gamma (\lambda - s)^{-1} d\lambda$$
$$= 2\pi i \cdot 1.$$

Similarly $\alpha\beta = 1$. This is clear when $\Delta \neq 0$. At a point of c any element of $\pi_*(\mathscr{S} \otimes [X])$ can be written:

$$\frac{1}{\lambda}(sf + \lambda f),$$

where f is a local section of F over X. (For we can expand in the fibre direction and use the relation $\lambda^2 = \Delta$ on $\tilde{X}$ to write any power series in this form.) So α is surjective, hence $\alpha\beta = 1$ by the result above.

Finally, the relation between $\mathscr{S}$ and the map "multiplication by λ" on $\tilde{X}$ follows tautologically from the identity $(\lambda - s) = 0$ on $\mathscr{S}$. To sum up, we have a complete equivalence.

$$\left\{ \begin{array}{l} \text{2-bundles } F \text{ with } s \in H^0 \\ (End_0\ F \otimes L), \det s = -\Delta \end{array} \right\} \leftrightarrow \left\{ \begin{array}{c} \text{sheaves } \mathscr{S} \text{ on } \tilde{X} \text{ such that } \pi_*(\mathscr{S} \otimes [X]) \\ \text{is a 2-plane bundle.} \end{array} \right\}.$$

For, given the sheaf $\mathscr{S}$, multiplication by λ induces a map s with $s^2 = \lambda^2 = \Delta$ which is thus a trace-free map of $\pi_*(\mathscr{S} \otimes [X])$.

(b) We will now discuss the singularities that may occur in the sheaves $\mathscr{S}$ defined above. The main proofs are set aside in Appendix 3. The discussion is a local one, so we can work with an endomorphism s of the trivial bundle over a neighbourhood Y of the origin in $\mathbb{C}^2$, with branched cover $\pi: \tilde{Y} \to Y$. There is no need to distinguish between $\mathscr{S}$ and $\mathscr{S} \otimes [Y]$.

LEMMA (5.21). *If $s(0) \neq 0$ then $\mathscr{S}$ is locally free in a neighbourhood of $\pi^{-1}(0)$.*

Proof. SL(2, $\mathbb{C}$) act transitively on the cone:

$$\{A \in M_{2\times 2}(\mathbb{C}) \mid Tr A = \det A = 0, A \neq 0\}.$$

Represent s by a matrix valued function and choose local coordinates on this cone near $s(0)$ using conjugates of $s(0)$. The determinant function is regular at $s(0)$ so the implicit function theorem implies that s is conjugate, near 0, to a function:

$$\begin{pmatrix} 0 & 1 \\ \Delta & 0 \end{pmatrix} \qquad (\Delta(z, w) = -\det s).$$

Then $\mathscr{S}$ is freely generated by the element written as $\binom{\lambda}{1}$ in the new trivialisation.

To describe the position when s has isolated zeros we introduce some notation. First, for the application in (iii) below, we need only discuss the cases when the determinant function

(z, w) has one of the forms:

(I) $\Delta(z, w)=z^{2p+1}$
(II) $\Delta(z, w)=z^{2q}$
(III) $\Delta(z, w)=z^{2p+1}w^{2q}$
(IV) $\Delta(z, w)=z^{2p}w^{2q}$.

Let v: $Y^* \to \tilde{Y}$ be the *normalisation* of $\tilde{Y}$. Y^* is smooth and has a single component in cases I and III, and two separate sheets in cases II and IV. Consider the family of endomophisms

$$v^*\pi^*(s-\lambda)$$

over Y^*. They have determinant zero so, as in (v) above, their kernels define a line bundle J over Y^* with a bundle map:

$$i: J \to \mathbb{C}^2.$$

The proof of (5.21) shows that i vanishes only over the zeros of s.

Let $\delta(s)\geq 0$ be the total multiplicity of the zeros of i in Y^*. In cases II and IV the two sheets give separate contributions. Our goal is to relate these multiplicities to the parameters describing the singularities of s.

There is a natural inclusion of the sheaf $\mathscr{S}$ into the sheaf of sections of J. So $\mathscr{S}$ can be regarded, locally, as a subset of $\mathcal{O}_{Y^*}$. To make this concrete, consider case I and introduce coordinates (t, w) on Y^* so the normalisation is represented by the substitution:

$$z=t^2,\ \lambda=t^{2p+1}. \tag{5.22}$$

Then, away from the zeros of s we can choose a generator σ of J induced from a generator of $\mathscr{S}$, and $\mathscr{S}$ is represented by sections $f\cdot\sigma$ of J where $f(t, w)$ satisfies

$$f_{(1)}=f_{(3)}= \dots =f_{(2q-1)}=0. \tag{5.23}$$

Here we write $f_{(j)}$ for the coefficient of t^j in the expansion of f—so $f_{(j)}$ is a holomorphic function of w. In terms of an arbitrary local generator τ of J, $\mathscr{S}$ is represented (away from the zeros of s) by sections $g\tau$ where g satisfies a system of equations:

$$\begin{aligned} g_{(1)}&=\psi_1 g_{(0)}\\ g_{(3)}&=\psi_1 g_{(2)}+\psi_2 g_{(0)}\\ &\ \vdots\\ g_{(2p-1)}&=\psi_1 g_{(2p-2)}+ \dots \psi_p g_{(0)}, \end{aligned} \tag{5.24}$$

for certain holomorphic functions $\psi_i(w)$ depending on τ. This follows immediately from (5.23). The transformation behaviour of the ψ_i under change of trivialisation in J is rather complicated, but it is of course possible to give an invariant description of them.

Singular points, where s vanishes, can be treated in the same way. We have the following.

LEMMA (5.25). (i) *If the matrix function* s *over* Y *has determinant* $-z^{2p+1}$ *and isolated zeros, then given a trivialisation* τ *of* J *elements of* $\mathscr{S}$ *correspond to holomorphic sections* $g(t, w)\cdot\tau$ *of* J *satisfying a system of equations* (5.24), *with meromorphic functions* $\psi_i(w)$ *whose poles lie over the zeros of* s.

(ii) *Conversely, to any system* Ψ *of meromorphic functions* $\psi_i(w)$ *we can associate a matrix function* $s[\Psi]$ *with determinant* $-z^{2p+1}$.

Proof. For (i) write τ as a vector valued function $\begin{pmatrix} a(t, w) \\ b(t, w) \end{pmatrix}$. The condition that $g \cdot \tau$ lies in $\mathscr{S}$ is:

$$(g \cdot \tau)_{(j)} = 0 \qquad j = 1, 3, \ldots, 2p-1.$$

Expanding each component of this gives rise to a system of equations (5.24) with meromorphic coefficients. The two systems are the same, for if they differed over Y they would also differ away from the zeros of s, contradicting the description above. Part (ii) follows from the analysis in Appendix 3.

We now define a number $N(s)$ which can be viewed as the number of parameters in the polar data of the meromorphic functions ψ_i corresponding to a matrix function s. Suppose s_ξ is a holomorphic family of matrix functions over Y, all with determinant $-z^{2p+1}$, parametrised by a point ξ in a polydisc D in $\mathbb{C}^n$, with $s_0 = s$. We can choose a smooth family τ_ξ of local trivialisation of the line bundles J_ξ over Y^* and so describe the associated sheaves by functions $\psi_{i,\xi}$ varying smoothly with ξ. That is:

$$\psi_{i,\xi} = \frac{p_i(\xi, w)}{q_i(\xi, w)}$$

with p_i, q_i holomorphic in both variables. We define an equivalence relation on D by $\xi_1 \sim \xi_2$ if, for all i, $\psi_{i,\xi_1} - \psi_{i,\xi_2}$ is a holomorphic function of w. Notice that this depends on the trivialisations τ_ξ. Clearly the equivalence classes are unions of locally closed subvarieties in D. We let $N(s)$ be the maximum, over all such families and trivialisations, of the codimension of these equivalence classes.

The other cases II, III and IV are similar. In case II Y^* has two sheets Y^*_+, Y^*_-. There are line bundles $J_+ \to Y^*_+$, $J_- \to Y^*_-$ and $\mathscr{S}$ is identified with pairs of sections of J_+, J_- satisfying conditions along the branch locus $z = 0$. In terms of local trivialisations τ_+, τ_- the conditions on $g^+\tau_+, g^-\tau_-$ take the form:

$$\begin{aligned} g^+_{(0)} &= \gamma \cdot g^-_{(0)} \\ g^+_{(1)} &= \gamma(g^-_{(1)} + \varphi_1 g^-_{(0)}) \\ &\vdots \\ g^+_{(q-1)} &= \gamma(g^-_{(q-1)} + \ldots + \varphi_{q-1} g^-_{(0)}), \end{aligned} \tag{5.26}$$

with meromorphic functions $\gamma(w)$, $\varphi_i(w)$. Observe that γ has a simple invariant description as a meromorphic bundle map from $J_-|_{(z=0)}$ to $J_+|_{(z=0)}$. We define $N(s)$ in a similar way to that above, using an equivalence relation $\xi_1 \sim \xi_2$ if $\varphi_{i,\xi_1} - \varphi_{i,\xi_2}$ is holomorphic and the divisors of poles and zeros of $\gamma_{\xi_1}, \gamma_{\xi_2}$ are equal.

In case III we take coordinates t, w on the normalisation Y^* such that ν is represented by the substitutions

$$z = t^2, \quad \lambda = t^{2p+1} w^q.$$

There are two sets of identification data:

$$\begin{aligned} g_{(1)} &= \psi_1 g_{(0)} \\ &\vdots \\ g_{(2p-1)} &= \psi_1 g_{(2p-1)} + \ldots + \psi_p g_{(0)} \end{aligned} \tag{5.27A}$$

where $\psi_1 = \psi_i(w)$ and $g_{(i)}$ is the coefficient of t^i,

$$g(t, 0) = \gamma g(-t, 0) \tag{5.27B}$$

$$\vdots$$

$$g_{(q-1)}(t, 0) = \gamma[g_{(q-1)}(-t, 0) \ldots + \chi_{q-1} g(-t, 0)]$$

where $\chi_i = \chi_i(t)$ and $g_{(i)}$ is the coefficient of w^i. We define $N(s)$, recording residues of the ψ's and the χ's and the divisor of γ, in the same way. Case IV is similarly a combination of two copies of case II.

The result we shall need—similar to the duality formula in the proof of Proposition (5.11)—is the following.

LEMMA (5.28). *For each of the cases I to IV,* $N(s) \leqslant \delta(s)$.

This is proved in Appendix 3.

(iii) We will now apply the discussion of (ii) to bound the number of free parameters describing the sheaves $\mathcal{S}$ of (i)—and so complete the proof of (5.12). Let us first suppose that X is a compact algebraic surface, $H_X \to X$ is ample, $L \to X$ a fixed holomorphic line bundle and $C \in |2L|$ a curve with singular points of the types considered in (ii) above. (So C_{red} has only ordinary double points and components of C with odd multiplicity do not meet each other.) Let $M(C, k)$ be the moduli space whose points represent pairs (F, s) where F is H_X stable, $\Lambda^2 F = \mathcal{O}$, $c_2(F) = k$, $s \in H^0(End_0\, F \otimes L)$ has determinantal divisor C and isolated zeros.

We claim that there is a constant $A(X, H_X, C)$ such that:

$$\dim M_{k,C} \leq 2k + A. \tag{5.29}$$

To see this form the branched double cover $\pi\colon \tilde{X} \to X$ as in (i) and its normalisation $\nu\colon X^* \to \tilde{X}$. X^* is smooth and may be disconnected. Let $\sigma\colon X^* \to X^*$ be the covering involution. The pre-image of C in X^* can be written $D \cup E$ where $\sigma = 1$ on D—the pre-image of the components of C with odd multiplicity—and σ restricts to a non-trivial involution of E—the pre-image of the even components. Notice that the reduced curve D_{red} is smooth and σ induces an involution of the normalisation E^* of E_{red}.

Now since X^* is smooth there is an eigenspace line bundle $J \to X^*$, as in (ii), and a bundle map $i\colon J \to \nu^*\pi^*(F)$. If δ denotes the total multiplicity of the zeros of i we have:

$$2k = c_2(\nu^*\pi^*(F)) = \delta - c_1(J)^2. \tag{5.30}$$

Our first task is to find a bound on $c_1(J)^2$ and we use separate arguments for the cases when X^* is connected or disconnected. In either case we have the standard formula:

$$0 = c_1(\nu^*\pi^*(F)) = c_1(J) + c_1(\sigma^* J) + 2\nu^*\pi^*(C). \tag{5.31}$$

Also, in both cases $\tilde{H} = \nu^*\pi^*(H_X)$ is ample on X^*. Suppose X^* is connected; by (5.31) the degree

$$c_1(K) \cdot c_1(\tilde{H}) = c_1(\sigma^*(K)) \cdot c_1(\tilde{H})$$

is fixed by H_X, C. Then, as in §V above, the Hodge Index Theorem implies that there is a fixed upper bound:

$$c_1(J)^2 \leq A_0(H_X, C). \tag{5.32}$$

If X^* is disconnected the components X^*_+, X^*_- are each copies of X and so $\nu^*\pi^*(F)$ is $\tilde{H}$-stable on each. So if J_+, J_- are the restrictions of J to X^*_+, X^*_- both of the degrees:

$$c_1(J_+)\cdot c_1(\tilde{H}),\ c_1(J_-)\cdot c_1(\tilde{H})$$

are negative. But their sum is fixed by (5.31) so we get a bound on $|c_1(J_\pm)\cdot c_1(\tilde{H})|$ and so again on $c_1(J)^2$.

Now the line bundles J define a holomorphic map:

$$M(C,k)\to Jac(\tilde{X}).$$

Clearly it suffices to bound the dimension of the fibres of this map. But any polydisc in such a fibre defines locally on X a family of endomorphisms s_ξ of the kind considered in (ii) above. So by (5.28) the codimension of the subsets on which the poles of the φ, ψ data and divisor of γ correspond is at most $\delta\le 2k+A_0$ [by (5.30), (5.32)]. Finally then, we consider a subset V of $M(C,k)$ on which the bundles k, the polar data of the φ's and ψ's and the divisor of γ agree. We want to bound dim V.

Consider the ψ-data along the smooth curve D_{red} in X^*. It is easy to see that if D_0 is a component of D_{red} and $\{\psi_i\}$, $\{\hat{\psi}_i\}$ are two sets of ψ data along $D_0(1\le i\le\rho(D_0))$ such that $\psi_i=\hat{\psi}_i$ for $i<\lambda$, then $\psi_\lambda-\hat{\psi}_\lambda$ is naturally a holomorphic section of the line bundle $N_{D_0}^{-\lambda}$ (where N_{D_0} is the normal bundle of D_0 in X^*).

The dimension of this space of holomorphic sections is fixed by C so, by considering the ψ_i in turn, we find a subset in V of bounded codimension on which all the ψ data are equal. Similarly, for the γ's and φ's. (The difference between the φ's is analysed in terms of holomorphic sections of the pull back of the normal bundle to the smooth curve E*.) But any points in $M(C,k)$ for which all this data agree give rise to isomorphic sheaves $\mathcal{S}$ on $\tilde{X}$ and so are equal by the discussion of (i) above. This completes the verification of (5.29).

Completion of proof of (5.12). Recall that $U^{(i)}$ represents pairs (E,s) where s is a section of $End_0E\otimes K_S$ with non-trivial determinant and $h^0(End_0E\otimes K_S)=i$. We will reduce to the situation considered above in four steps.

First decompose $U^{(i)}$ into the fibres of the determinant map

$$\det: U^{(i)}\to|2K_S|.$$

It suffices to bound the dimensions of the fibres of this pair (E,s) with a fixed determinant curve C_0. The constructions below depend on C_0 but it will be clear that the constants we get are uniform, since we can stratify $|2K_S|$ into a finite number of pieces on each of which the singularities of C_0 have the same numerical invariants.

Second, blow up S at the singularities of C_0 to get a new surface S_1 and map $\tau_1: S_1\to S$ such that the reduced curve of $C_1=\tau_1^{-1}(C_0)$ has only ordinary double points. One can show that $\tau_1^*(E)$ is stable on S_1 for a suitable hyperplane class (compare [11]).

Third, blow up S_1 at points where two branches of C_1, both having odd multiplicity, cross. If C_1 has a local model

$$z^{2p+1}w^{2q+1}=0$$

near such point in S_1 the singularities of the new determinant locus C_2 in the blow up S_2 of S_1 have the forms:

$$z^{2p+1}w^{2(p+q+1)}=0,\ w^{2q+1}z^{2(p+q+1)}=0.$$

So the singularities of C_2 are of the kind considered above, and if $T_2: S_2\to S_1$ is the blow up map, $\tau_2^*\,\tau_1^*(E)$ is stable on S_2 for a suitable hyperplane class.

Finally, if $\tau_2^* \tau_1^*(s)$ vanishes on a curve T in S_2 we consider the induced section s' of $End_0(\tau_2^* \tau_1^*(E)) \otimes \tau_2^* \tau_1^*(K_S) \otimes [T]^{-1}$. Only finitely many bundles $[T]$ can occur and we can stratify our moduli space into pieces on which $[T]$ is fixed. This twist will change the determinant locus by $2T$ so will not introduce any new odd/odd crossings. Thus, we are in a position to apply (5.29) to (F, s') with $F=\tau_2^* \tau_1^*(E)$ and $L=\tau_2^* \tau_1^* K_S \otimes [T]^{-1}$. But (F, s') determines (E, s) when restricted to the complement of the finite set of points in S which have been blown up to obtain S_2. By Hartog's Theorem they determine (E, s) over all of S (an isomorphism between two pairs will automatically extend over the exceptional points). So the deformations of (E, s), on a given strata, map injectively to those of (F, s') and we can deduce (5.12) from (5.29).

§VI. K3 SURFACES

A K3 surface is a compact simply connected complex surface S with trivial canonical bundle. Information from complex geometry can be used to deduce differential topological properties of the underlying 4-manifolds; for example, it is known [2, Ch. 8] that all such surfaces are diffeomorphic. We will use our invariants to make two observations of this kind.

(i) *Diffeomorphisms*

The existence of a large moduli space of complex K3 surfaces implies that there are many self-diffeomorphisms of S. More precisely we consider the isometry group O_Q of the intersection form Q on the integral homology $H_2(S)$ and the homomorphism

$$\rho\text{:Diff}(S) \to O_Q.$$

O_Q contains an index 2 subgroup O_Q^+ consisting of transformations which preserve the orientation of the "positive part" of H_2 (cf. §II). Since -1 does not lie in O_Q^+ there is a splitting:

$$O_Q = Q_Q^+ \times \{\pm 1\}.$$

Then we have the known result given below.

PROPOSITION (6.1) [22]. *The image of* ρ *contains* O_Q^+.

This fact can be understood via the description of the moduli space of Kahler–Einstein metrics on S.

Any Riemannian metric g determines a positive subspace $\mathscr{H}^2_+(g) \subset H^2$ represented by g self-dual harmonic forms. The classification theorem for Kahler–Einstein metrics asserts that all positive subspaces arise in this way from Kahler–Einstein metrics g and conversely if g_1, g_2 are Kahler–Einstein they are isometric if and only if there is an element φ of O_Q with

$$\varphi \mathscr{H}^2_+(g_1) = \mathscr{H}^2_+(g_2) \quad [2, \text{Ch. } 8].$$

Thus the moduli space of Kahler–Einstein metrics is G/O_Q where G is the open subset of the Grassmannian consisting of positive 3-planes. Now -1 acts trivially on G and O_Q^+ acts freely on the complement of an isolated set of points. Thus we obtain a self diffeomorphism of S for each transformation O_Q^+ through the monodromy of the universal family over G/O_Q.

We will now deduce a converse to (6.1).

PROPOSITION (6.2). *The image of* ρ *is contained in* O_Q^+.

Proof. To prove this in a concrete way, without appeal to all of §V, we consider an SO(3) bundle P over S with $k=3/2$ (i.e. $p_1=6$ and $w_2^2=2 \bmod 4$). Then the moduli space of ASD connections on P is zero-dimensional. Choose the standard orientation β of H^2_+ defined by a complex structure. S is spin so this orients the moduli space and we obtain from (3.5) an integer valued invariant

$$q=q_{3/2,\,w,\,S,\,\beta}\in\mathbb{Z}.$$

Notice here that the class $w=w_2(P)$ we choose is not important since the group O_Q^+ acts transitively on elements $w\in H^2(S;\ \mathbb{Z}/2)$ with $w^2=2 \bmod 4$, and these are realised by automorphisms of S [Proposition (6.1)].

Suppose that there is a self-diffeomorphism $f\colon S\to S$ such that f^* does not lie in O_Q^+. By (6.1) we can suppose $f^*=-1$ so f^* preserves $w_2(P)$. But $f^*(\beta)=-\beta$ so, by naturality, $q=-q$ and thus the invariant q must be zero.

We claim that in fact $q=1$. Fix a complex model for a K3 surface as double cover of $\mathbb{CP}^2$ branched over a smooth curve of degree 6; $\pi\colon S\to\mathbb{CP}^2$. [18, p. 593]. Let P be the SO(3) bundle defined by the complex 2-plane bundle $\pi^*(T\mathbb{CP}^2)$. This has the correct characteristic classes. According to §V we need to find the moduli space of stable holomorphic bundles on S topologically equivalent to $\pi^*(T\mathbb{CP}^2)$. We fix the usual hyperplane class $\pi^*(\mathcal{O}(1))$ on S.

In one direction, this moduli space contains *at most* one point. This was proved by Mukai [24]; his argument is so simple that we repeat it here. If E, F are two stable bundles with this topological type the Riemann–Roch formula implies that at least one of

$$H^0(\mathrm{Hom}(E,F)),\ H^2(\mathrm{Hom}(E,F)$$

is non-zero. But the latter is dual to $H^0(\mathrm{Hom}(F,E))$ since K_S is trivial. Thus we can suppose there is a non-zero homomorphism from E to F. Now it is a general fact that a non-zero homomorphism between topologically equivalent stable bundles is an isomorphism [28], so E, F are isomorphic.

On the other hand we show that $E=\pi^*(T\mathbb{CP}^2)$ is stable, so the moduli space contains at least one point. Furthermore, $H^2(End_0E)=H^0(End_0E)^*$ is zero so this point is smooth and counts with multiplicity one. The stability of E follows from the well known fact that $T\mathbb{CP}^2$ is stable [28]. Indeed, it is true in general that if $p\colon Y\to X$ is a branched cover and V is H_X-stable over X then $p^*(V)$ is $p^*(H_X)$-stable over Y. This can be proved by a simple argument, directly from the definition of stability.

This completes the proof of (6.2). In line with the discussion above, we can give an equivalent formulation in terms of metrics.

COROLLARY (6.3). *Any differentiable fibration $\Sigma\to S^1$ with a fibre K3 surface can be given a Riemannian metric which is Kahler–Einstein along the fibres.*

(ii) Almost complex structures

Recall that an almost complex structure on an oriented Riemannian 4-manifold X corresponds to a section of the sphere bundle Z of $\Lambda^2_{+,X}$ (the "twistor space"). Let A_X denote the set of homotopy classes of almost complex structures, so we have a map:

$$c_1\colon A_X\to H^2(X;\mathbb{Z}).$$

The image of this consists of classes c such that $c=w_2(X) \bmod 2$ and $c^2=3\tau(X)+2\chi(X)$ [12, p. 130]. The first Chern class does not completely determine the homotopy class of an almost complex structue. We define

$$p\colon A_X\to A_X$$

with $p^2 = id.$ and $c_1 \circ p = c_1$, as follows. For $\sigma \in A_X$, $p(\sigma)$ agrees with σ outside a small ball in X and over this ball the two compare by the non-zero element of

$$[S^4, \text{fibre of } Z] = [S^4, S^2] \cong \mathbf{Z}/2.$$

It is easy to see that there are at most these two homotopy classes; σ, $p(\sigma)$ lying over a given class in $H^2(X; \mathbf{Z})$.

We have used in §V the fact that a complex Kahler structure on a 4-manifold singles out a preferred orientation β. The same is true for almost complex structures σ, because given such a structure we can deform the operator $[-d^* + d^+]$ to an operator commuting with the actions of σ on Λ^1, $\Lambda^0 \oplus \Lambda^2_+$ (see [6], Prop. (3.25)). So for σ in A_X we have a preferred orientation $\beta(\sigma)$ of $H^2_+(X)$.

Lemma (6.4). $\beta(\sigma) = -\beta(\mathrm{p}(\sigma))$.

Proof. By an excision argument (cf. [6]) it suffices to verify this in any one case. Consider the 4-manifold

$$Y = \mathbb{CP}^2 \# \mathbb{CP}^2 \# \mathbb{CP}^2$$

and let σ be an almost complex structure with $c_1(\sigma) = (3, 3, 1)$ in the standard basis. Let f be a diffeomorphism of Y which acts on H^2 by interchanging the first two factors. Then f^* reverses the orientation of H^2_+ so $f^*(\sigma)$ is not equal to σ, but f^* preserves $c_1(\sigma)$ so we must have $f^*(\sigma) = p(\sigma)$. Thus the orientations $\beta(\sigma)$ and $\beta p(\sigma)$ are opposite.

Corollary (6.5). *There is a homotopy class of almost complex structures with* $\mathrm{c}_1 = 0$ *on the* $K3$ *surface* S *which does not contain any integrable representative.*

Proof. Let σ be the standard integrable almost complex structure. Then $p(\sigma)$ cannot be integrable, for if it were it would define a new complex K3 surface. There would then be a self-diffeomorphism g of S with $g^*(\sigma) = p(\sigma)$. But by Lemma (6.4) g would reverse the orientation of H^2_+, in contradiction to (6.2).

Acknowledgements—The author is grateful to A. Beauville, D. Gieseker, F. Kirwan, P. Kronheimer, J. Morgan and A. Todorov for invaluable help, in various different ways, in the production of this paper.

REFERENCES

1. M. F. Atiyah and R. Bott: The Yang–Mills equations on a Riemann surface. *Phil. Trans. R. Soc. Lond. Ser. A* **308** (1982), 524–615.
2. W. Barth, C. Peters and A. Van de Ven: *Compact Complex Surfaces.* Springer-Verlag, Berlin (1984).
3. S. K. Donaldson An application of gauge theory to 4-dimensional topology. *J. Diff. Geom.* **18** (1983), 279–315.
4. S. K. Donaldson: Connections, cohomology and the intersection forms of 4-manifolds. *J. Diff. Geom.* **24** (1986), 275–341.
5. S. K. Donaldson: Anti-self-dual Yang–Mills connections over a complex algebraic surface and stable vector bundles. *Proc. Lond. Math. Soc.* **50** (1985), 1–26.
6. S. K. Donaldson: The orientation of Yang–Mills moduli spaces and four manifold topology. *J. Diff. Geom.* **26** (1987), 397–428.
7. S. K. Donaldson: Irrationality and the h-cobordism conjecture. *J. Diff. Geom.* **26** (1987), 141–168.
8. S. K. Donaldson: Infinite determinants stable bundles and curvature. *Duke Math. J.* **54** (1987), 231–247.
9. R. Fintushel and R. Stern: SO(3) connections and the topology of four-manifolds. *J. Diff. Geom.* **20** (1984), 523–539.
10. R. Fintushel and R. Stern: Definite four manifolds. *J. Diff. Geom.* **28** (1988), 133–141.
11. J. Fogarty: Algebraic families on an algebraic surface. *Am. J. Math.* **90** (1968), 511–520.
12. D. S. Freed and K. K. Uhlenbeck: *Instantons and four-manifolds.* M. S. R. I. publications, Springer, New York (1984).

13. R. FRIEDMAN and J. W. MORGAN: On the diffeomorphism type of certain elliptic surfaces I, II. *J. Diff. Geom.* **27** (1988), 297–369, 371–398.
14. D. GIESEKER: On the moduli of vector bundles on an algebraic surface. *Ann. Math.* **106** (1977), 45–60.
15. D. GIESEKER: Geometric invariant theory and applications to moduli problems. In *Lecture Notes in Mathematics* 996 (Edited by F. GHERARDELLI). Springer, Berlin (1982).
16. D. GIESEKER: A construction of stable bundles on an algebraic surface. *J. Diff. Geom.* **27** (1988), 137–154.
17. P. A. GRIFFITHS: Linearising flows and cohomological interpretation of Lax equations. *Am. J. Math.* **107** (1985), 1445–1484.
18. P. GRIFFITHS and J. HARRIS: *Principles of Algebraic Geometry.* John Wiley (1978).
19. A. GROTHENDICK: Sem. Bourbaki, Exposé 221, Paris (1960).
20. M. ITOH: The moduli space of Yang–Mills connections over a Kähler surface is a complex manifold. *Osaka J. Math.* **22** (1985), 845–862.
21. R. MANDELBAUM: Four manifold topology an introduction. *Bull. Am. Math. Soc.* **12** (1980), New Series, 1–160.
22. T. MATUMOTO: *On Diffeomorphisms of a K3 Surface.* Hiroshima University Reprint (1985).
23. V. B. MEHTA and A. RAMANATHAN: Restriction of stable sheaves and representations of the fundamental group. *Invent. Math.* **77** (1984), 163–172.
24. S. MUKAI: Moduli of bundles on a K3 surface in: Vector bundles on algebraic varieties, Proc Tata Institute Colloquium. Oxford U.P. 1987.
25. D. MUMFORD and J. FOGARTY: *Geometric Invariant Theory.* 2nd edn. Springer, Berlin (1982).
26. M. S. NARASIMHAN and C. S. SESHADRI: Stable and unitary vector bundles on compact Riemann surfaces. *Ann. Math* **82** (1965), 540–567.
27. P. E. NEWSTEAD: *Introduction to Moduli Problems and Orbit Spaces.* Tata Inst. for Fundamental Research, Bombay (1978).
28. C. OKONEK, M. SCHNEIDER and H. SPINDLER: *Vector Bundles on Complex Projective Spaces.* Birkhaüser, Boston (1980).
29. C. OKONEK and A. VAN DE VEN: Stable bundles and differentiable structures on certain elliptic surfaces. *Invent. Math.* **86** (1986), 357–370.
30. S. SMALE: An infinite dimensional version of Sard's Theorem. *Am. J. Math.* **87** (1968), 861–866.
31. C. H. TAUBES: Self-dual connections on manifolds with indefinite intersection matrix. *J. Diff. Geom.* **19** (1984), 517–560.
32. C. H. TAUBES: Self-dual connections on non-self-dual four-manifolds. *J. Diff. Geom.* **17** (1982), 139–170.
33. K. K. UHLENBECK: Connections with L^p bounds on curvature. *Comm. Math. Phys.* **83** (1982), 31–42.
34. K. K. UHLENBECK and S.-T. YAU: The existence of Hermitian Yang–Mills connections on stable bundles over Kähler manifolds. *Comm. Pure Appl. Math.* **39-S** (1986), 257–293.
35. C. T. C. WALL: On simply connected 4-manifolds. *J. Lond. Math. Soc.* **39** (1964), 141–149.
36. C. T. C. WALL: Diffeomorphism of 4-manifolds. *J. Lond. Math. Soc.* **39** (1964), 131–140.

The Mathematical Institute, 24–29 St. Giles,
Oxford OX1 3LB, U.K.

APPENDIX 0

Proof of (2.4). If N is any open set in X and A an SU(2) connection over X we call A *good* on N if the restriction of A to N is either irreducible or a reducible connection induced from a connection on a topologically trivial S^1 bundle over N. Let $\alpha_1, \ldots\ldots \alpha_n$ be homology classes, as in (2.4), and choose representative surfaces $\Sigma_1, \ldots\ldots \Sigma_n$ in general position. Let t and r be small parameters with $t<r$ and define open sets $N_v(t, r)$ $(v=1, \ldots\ldots n)$ as follows. $N_v(t, r)$ is the union of the tubular neighbourhood about Σ_v of thickness t and the collection of r-balls centered on the intersection points $\Sigma_v \cap \Sigma_\mu$ (μ not equal to v). Let $Y(t, r)$ be the interior of the complement in X of $\bigcup_v N_v(t, r)$.

LEMMA A0. *For any given k there is an r_0 such that for all $r<r_0$ and t less than some $t(r)$, and for all points* [A] *in M_j, $j \leq k$, we have: either the restriction of A to $Y(t, r)$ is irreducible or A is good on each $N_v(t)$.*

Proof of A0. Suppose the contrary, so with r fixed we can find $t_i \to 0$ and points $[A_i]$ in some M_j for fixed $j<k$ such that A_i is not good on $N_v(t_i, r)$ for some fixed v and A_i is reducible on $Y(t_i, r)$. By the compactness theorem we can suppose that A_i is weakly convergent to $([A_\infty]; x_1, \ldots\ldots x_r)$ say. Plainly A_∞ is reducible over the complement of the finite set of r-balls centered on the intersection points of the surfaces Σ_μ for μ not equal to v. Moreover, the reduction is topologically non-trivial over any

surface homologous to Σ_v. Now denote this limiting connection by $A_\infty(r)$. If the statement of the lemma does not hold we can find a sequence of values $r_i \to 0$, such that the $A_\infty(r_i)$ are reducible outside r_i-balls with these fixed centres, with a non-trivial reduction. Applying the compactness theorem again we can suppose the $A_\infty(r_i)$ converge weakly to a limit $(A; y_1, \ldots y_b)$. Then A is reducible over all of X and the reduction is topologically non-trivial, which gives the desired contradiction.

Given this lemma the proof of (2.4) is completed by an argument very like that in Lemma (4.34) (see also [10]). Fix t and r satisfying the conditions of the above lemma. We want to choose disjoint one-dimensional sets $K_1, \ldots\ldots K_n$ (each a wedge of circles) in $Y(t, r)$ such that for any point $[A]$ in a moduli space $M_j(j \le k)$ and for each v we have:

either A is good on $N_v(t, r)$
or the holonomy of A around K_v is irreducible.

We can construct these K_v inductively, using the compactness theorem, just as in (4.34). Finally choose disjoint neighbourhoods P_v of the K_v in $Y(t, r)$ and let N_v be union of P_v and $N_v(t, r)$. These sets then have the properties asserted in (2.4.).

APPENDIX 1

(a) We will prove Proposition (4.32) on the weak semi-continuity of the spaces $H^2_{A,\xi}$. First, let us consider the compactness property (4.29). Suppose $([A_\alpha], \xi_\alpha)$ is an infinite sequence in L_{j,Y_1}-equivalence classes of solutions of $F_+(A_\alpha) + \xi_\alpha \cdot \omega = 0$. Then the condition that $[A_\alpha]$ lie in $W_1^j(\varepsilon)$ clearly implies that $\|\xi_\alpha\|_{L^2}$ is bounded [and in fact $O(\varepsilon^{1/p})$]. The uniform bound (4.33) gives:

$$\sup_{Y_1} |\xi_\alpha| \le \text{const.} \|\xi_\alpha\|_{L^2}.$$

So $\|F_+(A_\alpha)\|_{L^\infty}$ is bounded and Uhlenbeck's Theorem shows that there is a subsequence converging weakly in all $L^q_{1,\text{loc}}$ over $Y_1 \backslash \{p_1, \ldots, p_l\}$ to a limit A_∞. Similarly, we have local L^2_1 bounds on the ξ_α from the Dirichlet forms so we can suppose the ξ_α converge weakly in $L^2_{1,\text{loc}}$ to a ξ_∞ over $Y_1 \backslash \{p_1, \ldots, p_l\}$. We could go on to show that the convergence is C^∞. More important we check that (A_∞, ξ_∞) extends over the punctures. A variant of Uhlenbecks Removable Singularities Theorem shows that any $L^q_{1,\text{loc}}$ connection $(q \gg 0)$ on a punctured manifold with F having finite L^2-norm and F_+ having finite L^∞ norm extends over the punctures to an L^q_1 connection (in some gauge). We leave the reader to adapt the proofs in [12] or [3], Appendix. We do the (easier) linear problem of extending ξ_∞ here since it is relevant to the proof of (4.32) below.

Clearly ξ_∞ is a linear combination of solutions s of the eigenvalue equation:

$$\Delta_A s + (R - \mu)s = 0 \tag{A1}$$

$(A = A_\infty)$ on the punctured manifold, for eigenvalues $\mu \le \frac{1}{2}\sigma_0$. We know that $\sup |s|$, $\int |\nabla_A s|^2$ are finite. We need to show that $s \in L^2(Y_1)$ is a weak solution of (A1), that is, if β_r is a cut-off vanishing on small r-balls about the punctures, and ψ is a smooth test function then:

$$\int_{Y_1} |(\Delta_A + R - \mu)(\beta_r s) \cdot \psi$$

tends to zero with r. This follows if

$$I(r) = \int_{Y_1} \psi \cdot (\Delta\beta_r \cdot s + \nabla\beta_r \nabla_A s)$$

tends to zero. But $\|\nabla\beta_r\|_{L^2}$ and $\|\Delta\beta_r\|_{L^{3/2}}$ are each $O(r)$ so we obtain $I(r) = O(r)$ by Holder's inequality, using the fact that $\|\nabla_A s\|_{L^2}$, $\|s\|_{L^4}$ are finite.

With this preparation we can go on to the proof of (4.32). Observe first that we can represent the cokernel $H^2_{A,\xi}$ of the operator $d\Phi$ by the "harmonic space" $\mathscr{H}^2_{A,\xi}$; the zeros of the formal adjoint. A short calculation using (4.28) shows that this operator has $\Omega^1(\mathfrak{g}_P)$ component:

$$(d\Phi)^*(\sigma) = d_A^* \sigma + \int_Y [d_A \theta_z, \Gamma_z] - [\theta_z, d_A \Gamma_z]\, dz \tag{A2}$$

where

$$\text{(i)}\ \theta_Z=(\Delta_A+R-z)^{-1} \tag{A3}$$
$$\text{(ii)}\ \Gamma_z=(\Delta_A+R-z)^{-1}(\omega\cdot\sigma)$$

and $\sigma\in\Omega^2_+(g_P)$. We want to show first that there is a uniform bound: $(d\Phi)^*\sigma=0\Rightarrow\|\nabla_{A_0}\sigma\|_{L^2}\leq C\|\sigma\|_{L^2}$ with C independent of $[A]\in W^i_1(\varepsilon)$. The proof is based on Taubes' argument in [32].

Suppose $(d\Phi)^*\sigma=0$ and $\|\sigma\|_{L^2}=1$. Apply the operator d_A^+ to (A2) to get:

$$0=(d_A^+\,d_A^*\sigma)+\int_\gamma d_A^+\,([d_A\theta_Z,\Gamma_Z]-[\theta_Z,d_A\Gamma_Z])dz. \tag{A5}$$

We can write the first term, by a Weitzenbock formula:

$$d_A^+d_A^*\sigma=\frac{1}{2}\nabla_A^*\nabla_A\sigma+\{K,\sigma\}+\{F_A^+,\sigma\}$$

[12] where the $\{,\}$ are algebraic bilinear combinations and K is the curvature of the base manifold. Since we have a fixed bound on $\|F_A^+\|$ (for $[A]$ in $W^i_1(\varepsilon)$) we can deduce, using the Sobolev embedding theorem, that:

$$\int|\nabla_A\sigma|^2\leq\text{const.}\left(\int\sigma d_A^+\,d_A^*\sigma+1\right).$$

So consider the integral:

$$\left|\int_{Y_1}\sigma\cdot d_A^+([d_A\theta_z,\Gamma_z]-[\theta_z,d_A\Gamma_z])\,d\mu\right|$$

for fixed z in γ. We rewrite it as:

$$\left|\int_{Y_1}\sigma\cdot d_A^+(2[d_A\theta_z,\Gamma_z]-d_A[\theta_z,\Gamma_z])\,d\mu\right|$$
$$=\left|2\int_{Y_1}d_A^*\sigma[d_A\theta_z,\Gamma_z]-\sigma(F_A^+,[\theta_z,\Gamma_z])\,d\mu\right|$$
$$\leq\text{const.}(\|\nabla_A\sigma\|_{L^2}\|\nabla_A\theta_z\|_{L^2}\,\Gamma_Z\|_{L^\infty}+\|\sigma\|_{L_4}\|\theta_z\|_{L^\infty}\|F_A^+\|_{L^2}\|\Gamma_z\|_{L^4}.$$

Now (A3, i, ii) imply:

$$\|\theta_z\|_{L^2}\leq\text{const.}\|\xi\|_{L^2}$$
$$\|\Gamma_z\|_{L^2}\leq\text{const.}\|\sigma\|_{L^2}$$

with constants independent of z and A, since the contour γ is a definite distance from Spec $(\Delta+R)$. Integrating the equation we get:

$$\|\nabla_A\theta_z\|_{L^2}\leq\text{const.}\|\xi\|_{L^2}$$
$$\|\nabla_A\Gamma_z\|_{L^2}\leq\text{const.}\|\sigma\|_{L^2}.$$

We can also apply the argument of [32] to (A3, i) to deduce $\|\theta_z\|_{L^\infty}\leq\text{const.}\|\xi\|_{L^2}$. Putting this together and using the Sobolev inequality $\|\sigma\|_{L^4}\leq\text{const.}\ \|\nabla\sigma\|_{L^2}$ gives:

$$\|\nabla_A\sigma\|^2_{L^2}\leq\text{const.}\left(\int_\gamma\left|\int_{Y_1}\sigma d_A^+(d_A\theta_z,\Gamma_z)-\theta_z d_A\Gamma_z)\right|+1\right)$$
$$\leq\text{const.}\,(\|\nabla_A\sigma\|_{L^2}+1)$$

and hence a uniform bound on $\|\nabla_A\sigma\|_{L^2}$.

Now suppose that $([A_\alpha],\zeta_\alpha)$ is a sequence converging over the complement of $\{p_1,\ldots p_l\}$ to $([A_\infty],\zeta_\infty)$ as in (4.32). If $\sigma_\alpha\in H^2_{A_\alpha,\zeta_\alpha}$ with $\|\sigma_\alpha\|_{L^2}=1$ we can use the elliptic system (A3) to find a subsequence converging to a limit σ_∞ on the punctured manifold. The uniform bound on $\|\sigma_\alpha\|_{L^4}\leq\text{const.}\ \|\nabla_{A_\alpha}\sigma_\alpha\|_{L^2}$ established above shows that $\|\sigma_\infty\|_{L^2}=1$. Then we repeat the argument above, used to show that ζ_∞ extends over the puncture, twice, to show that σ_∞ does. For each z in γ we have sections $(\theta_z)_\alpha$:

$$(\Delta_{A_\alpha}+R-z)(\theta_z)_\alpha=\zeta_\alpha.$$

We deduce that there is a limiting $(\theta_Z)_\infty$ as before. Then we regard (A5) and (A3, ii) as linear elliptic systems for the σ_α, $(\Gamma_Z)_\alpha$ with convergent coefficients. The same discussion serves again to show that σ_∞ extends over the punctures.

Now we prove (4.32) by contradiction. Suppose, on the contrary, that the $i_\alpha(L)$ do not generate $H^2_{A_\alpha, \zeta_\alpha}$ so there is a sequence of normalised harmonic forms σ_α with σ_α orthogonal to $i_\alpha(L)$. Then the limit σ_∞ above would be orthogonal to L, contradicting the assumption that L generates $H^2_{A_\infty, \zeta_\infty}$.

APPENDIX 2

In our proof of (4.9) we did not use Proposition (4.7) directly but rather its analogue for the deformed equations $F_+ + \xi\cdot\omega + \sigma_\mu = 0$. The extension needs only very general remarks. Indeed, we can consider any equations

$$F_+(A) + \psi_i(\xi, A) = 0 \tag{A6}$$

over Y_i where ξ takes values in a finite dimensional bundle over the space of connections. We suppose:

(i) ψ_i depends only on the restriction of (A, ξ) to a compact set $K_i \subset Y_i \backslash \{y_i\}$
(ii) For every (A, ξ), $\psi_i(A, \xi)$ is supported in K_i.
(iii) ψ_i are smooth perturbations of the ASD equations represented by non-linear maps whose partial derivatives are bounded operators $L^p_1 \to L^p$.

Our perturbations fit into this class.

We assert then that for such equations the analogue of (4.7) holds—when λ is small the solutions of:

$$F_+(A) + \psi_1(A, \xi) + \psi_2(A, \xi) \tag{A7}$$

over X can be analysed in terms of solutions over the constituent manifolds. To see this we recall the basic feature of the "alternating method" used in [4]. This method constructs solutions of (A7) over X by an iterative scheme. At each stage an equation of type (A6) over one of the Y_i with a source term is solved. More precisely, it suffices to study equations

$$F_+(A_1 + a + b) + \psi_1(A_1 + a, \xi_1 + \eta + \tau) = \sigma \tag{A8}$$

where (i) (A_1, ξ_1) is a solution of (A6) over Y_1 with $H^2_{A_1, \xi_1} = 0$.
(ii) a is a section of $T^*_{Y_1} \otimes \mathfrak{g}_{P_1}$ defined on the complement of a small ball containing y_1.
(iii) σ is supported in a small annulus surrounding y_1.

We need to show first that if a is small in $L^{2p}(Y_1) \cap L^p_1(K_1)$, σ is small in L^p and η is small, then there is a solution (b, τ) to (A8) with b small in L^p_1 and τ small. Second we need to know that for y near y_1 $|b(y)|$ is bounded by a multiple of

$$\int_{Y_1} \frac{1}{|y-z|^3} |\sigma(z)| \mathrm{d}\mu_z.$$

The existence of an L^p_1 solution follows immediately from the inverse function theorem. Just as in [4], (4.22) we deduce the required bound on $|b(y)|$ from the linearised equation:

$$\mathrm{d}^+_{A_1}\beta + [a, \beta] + K(\beta, v) = \sigma$$

say, where K is the Frechet derivative of ψ. Our bound on β comes from:

$$\beta = L(\sigma - [a, \beta] + K(\beta, v)),$$

where L is a right inverse for $d^+_{A_1}$. The new term is:

$$L(K(\beta, v)).$$

But $K(\beta, v)$ is supported in K_1, by Hypothesis (ii), so for y near y_1

$$|L(K(\beta, v))(y) \leq \text{const.} \|K(\beta, v)\|_{L^1} \leq \text{const.} \|\sigma\|_{L^p},$$

since the kernel of L is smooth away from the diagonal. With this elementary discussion to augment [4], §IV the whole proof goes over to the generalised equations.

APPENDIX 3

Proof of Lemma (5.28)—*Case I*: $\Delta = z^{2p+1}$. We give a partial classification of matrix functions over Y with the given determinant, beginning from the system of equations (5.24)

$$g_{(1)} = \psi_1 g_{(0)}$$
$$g_{(3)} = \psi_1 g_{(2)} + \psi_2 g_{(0)}$$
$$\vdots$$
$$g_{(2p-1)} = \psi_1 g_{(2p-2)} + \dots \psi_p g_{(0)}.$$

We define generators of the solution space, over $\mathcal{O}_Y$ (the functions of t^2, w) inductively. Suppose that at state i we have found generators h, k for the solutions of the first $(i-1)$ equations. Write $g = ah + bk$ with unknowns $a(t^2, w)$, $b(t^2, w)$. The ith equation takes the form:

$$a_{(0)}(h_{(2i-1)} - [\psi_1 h_{(2i-2)} + \dots + \psi_i h_{(0)}])$$
$$= b_{(0)}[\psi_1 g_{(2i-2)} + \dots + \psi_i g_{(0)}],$$

or, more briefly,

$$a_{(0)}(w)A(w) = b_0(w)B(w)$$

say. The reader can check that if A and B are both identically zero the matrix $s(z, w)$ that will result is divisible by z, contrary to our assumptions. Two cases then arise—if the zero of B at $w=0$ is of higher order than that of A (or pole of lower order) we let $C(w)$ be the holomorphic function $B(w)/A(w)$ and the solutions are generated by the pairs:

$$a = C(w) \qquad b = 1$$
$$a = t^2 \qquad b = 0.$$

If A has a zero of higher order we put $\hat{C}(W) = A(w)/B(w)$ and find generators.

$$a = -1 \qquad b = \hat{C}(w)$$
$$a = 0 \qquad b = t^2$$

(If the zeros have the same order we may choose either form). In this way we find generators H, K for the whole system in the form:

$$\begin{pmatrix} H \\ K \end{pmatrix} = N_p \dots N_1 \begin{pmatrix} t \\ 1 \end{pmatrix}$$

where each matrix N_i is either

$$\begin{pmatrix} t^2 & 0 \\ C_i(w) & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 0 & t^2 \\ -1 & \hat{C}_i(w) \end{pmatrix}.$$

Now

$$t\begin{pmatrix} t \\ 1 \end{pmatrix} = \begin{pmatrix} t^2 \\ t \end{pmatrix} = U\begin{pmatrix} t \\ 1 \end{pmatrix}$$

where $U = \begin{pmatrix} 0 & t^2 \\ 1 & 0 \end{pmatrix}$. Also, if adN_i is the adjugate matrix,

$$ad\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

then $N_i\, adN_i = adN_i N_i = \det N_i = t^2$. So:

$$t^{2p+1}\begin{pmatrix} H \\ K \end{pmatrix} = t^{2p} N_p \dots N_1 U \begin{pmatrix} t \\ 1 \end{pmatrix}$$
$$= N_p \dots N_1\, U adN_1 \dots adN_p \begin{pmatrix} H \\ K \end{pmatrix}.$$

Thus, relative to the basis H, K for the two-dimensional solution space, the matrix of s is:

$$s(t^2, w)=[N_p \dots N_1 U\, adN_1 \dots adN_p]^T. \tag{A6}$$

(The reader will be able to show in this way that, for example, any matrix with determinant $-z^3$ and isolated zeros is conjugate to one of the form

$$s(z, w)=\begin{pmatrix} C(w)z & z^2 \\ z-C(w)^2 & -C(w)z \end{pmatrix}.$$

To find the multiplicity $\delta(s)$ in this framework observe that:

$$(U^T-t)\begin{pmatrix} 1 \\ t \end{pmatrix}=0.$$

So:

$$(s-t^{2p+1})[adN_1 \dots adN_p]^T\begin{pmatrix} 1 \\ t \end{pmatrix}=0$$

and the bundle map $i{:}J\to\mathbf{C}^2$ is represented by

$$[adN_1 \dots adN_p]^T\begin{pmatrix} 1 \\ t \end{pmatrix}.$$

δ is the sum of the multiplicities of the zeros of this vector function over Y^*.

Let us call a set of ψ_i's *elementary* at $w=w_0$ of level $j(1\leqslant j\leqslant p)$ if for all $i\neq j$, $C_i(w)$ is holomorphic and non-vanishing at w_0 and $\hat{C}_j(w)$ has a simple zero there. So for $i\neq j$ we can use the C description. We say that the ψ_i are elementary on Y^* if they are elementary at all the singular points. We will first verify Lemma (5.28) for the matrix functions associated to such elementary ψ_i's. This is plainly an open condition and we shall then be able to deduce the general case by a deformation argument.

We assume that there is just one singular point at $z=w=0$ (the general case follows immediately from this) and the ψ_i are elementary of level j. First note that for $i<j$ we can suppose (choosing a suitable trivialisation τ) that $C_i(w)=0$. Indeed the matrix function we obtain by taking just the first $i-1$ equations is of the form:

$$\left[\begin{pmatrix} t^{2j-2} & 0 \\ Q(w,t) & 1 \end{pmatrix}\begin{pmatrix} 0 & t^2 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -Q(w,t) & t^{2j-2} \end{pmatrix}\right]^T$$

which does not vanish at $w=0$. So by Lemma (5.21) it can be placed in standard form by conjugation. Thus at the beginning of stage j we can suppose our generators are t^{2j-1}, 1. If $\hat{C}_j=k\cdot w[k(w)$ holomorphic, non-vanishing] the next pair is t^2, $k\cdot w-t^{2j-1}$ and the final basis elements are:

$$\begin{pmatrix} H \\ K \end{pmatrix}=\begin{pmatrix} t^{2(p-j)} & 0 \\ P(w,t) & 1 \end{pmatrix}\begin{pmatrix} t^2 \\ k\cdot w-t^{2j-1} \end{pmatrix}$$

where $P(w,t)=\sum_{i=j+1}^{p} C_i(w)t^{2(i-j-1)}$.

So:

$$K=kw-t^{2j-1}+\sum_{i=j+1}^{p} c_i(w)t^{2(i-j)}$$

and the condition that K satisfies the equations becomes:

$$\begin{aligned} -1&=K_{(2j-1)}=\psi_j\cdot K_{(0)}=\psi_j k\cdot w \\ 0&=\psi_\lambda\cdot k+\psi_{\lambda-1}\cdot C_{j+1}+\psi_j C_\lambda \\ &(j<\lambda<p). \end{aligned}$$

Hence $\psi_j=1/k\cdot w$ has a simple pole and for $\lambda>j$:

$$\psi_\lambda=(1/k\cdot w)[\psi_{\lambda-1}C_{j+1}\dots +\psi_j\cdot C_\lambda].$$

So ψ_λ is determined by the C_i, ψ_i for $i<\lambda$—hence by the ψ_i for $i<\lambda$ up to the addition of a term $\frac{1}{kw}\psi_j C_\lambda$ having at worst a double pole. So the polar parts of all the ψ_i are determined by a total of

$2(p-j+1)$ parameters (including one parameter for the position of the singularity, normalised here to $w=0$).

On the other hand, for this singularity the bundle map i is given by:

$$ad\left[\begin{pmatrix} t^{2(p-j)} & 0 \\ P & 1 \end{pmatrix}\begin{pmatrix} 0 & t^2 \\ -1 & \hat{C}_j \end{pmatrix}\begin{pmatrix} t^{2j-2} & 0 \\ 0 & 1 \end{pmatrix}\right]^T\begin{pmatrix} 1 \\ t \end{pmatrix}$$
$$=\begin{pmatrix} k\cdot w+t^2P(t,w)+t^{2j-1} \\ -t^{2(p-j+1)} \end{pmatrix}$$

which has a zero of multiplicity $\delta=2(p-j+1)$. This verifies (5.28) in Case I for elementary systems.

For the general assertion we appeal, twice, to a semi-continuity property. Suppose s_ξ, τ_ξ are holomorphic families parametrised by ξ in D, as described in §V(f, ii). A little thought shows that the equivalence relation $\sim$ is closed. If sequences ξ_α, η_α tend to limits ξ_∞, η_∞ in D and $\xi_\alpha\sim\eta_\alpha$ then $\xi_\infty\sim\eta_\infty$. Thus the equivalence classes are in fact Zariski closed subsets.

Suppose that, contrary to Lemma (5.28), an equivalence class V has codimension greater than δ. Choosing a transverse slice we may suppose V is a single point ξ_0 and $\dim D>\delta$. We claim then that for ξ sufficiently close to ξ_0 the equivalence class V_ξ of ξ is also zero-dimensional. For if not V_ξ meets the boundary of D and, by letting ξ tend to ξ_0 and taking a convergent subsequence of these points on the boundary, we would find another point in V.

Now for each ξ in D we analyse the solutions of the system of equations defined by the $\psi_{i,\xi}$ in the manner above. Thus we write D as the union of 2^p disjoint subsets according to the stages in the process at which we are forced to use the $\hat{C}$ as opposed to the C description. Each is a constructible set so one of them must have an interior whose closure contains ξ_0. Combining this with the observation of the previous paragraph we may suppose D chosen so that only one of these subsets is non-empty. Then we may write the matrix functions s_ξ, after conjugation in the form (A6) with each matrix $N_{i,\xi}$ holomorphic in ξ.

Finally deform the polynomials $C_{i,\xi}$ $\hat{C}_{i,\xi}$ defining the $N_{i,\xi}$ with a new parameter ε to get families $C_{i,\xi,\varepsilon}$, $\hat{C}_{i,\xi,\varepsilon}$ and matrices $s_{\xi,\varepsilon}$. We can clearly do this in such a way that, for $\varepsilon\neq 0$, $s_{\xi_0,\varepsilon}$ is elementary. So for fixed $\varepsilon\neq 0$ the equivalence class of ξ_0 in the ξ variable has dimension bigger than 0. Repeating the limiting argument above we get a contradiction to the hypothesis that V is a single point. This completes the proof of (5.28) in Case I.

Case II: $\Delta=z^{2q}$. We analyse the system of equations (5.26):

$$g^+_{(0)}=\gamma g^-_{(0)}$$
$$g^+_{(0)}=\gamma(g^-_{(1)}+\varphi_1 g^-_{(0)})$$
$$\vdots$$
$$g^+_{(q-1)}=\gamma(g^-_{(q-1)}+\;\dots\;+\varphi_{q-1}g^-_{(0)})$$

in just the same way. Generators for the solutions are H, K where:

$$\begin{pmatrix} H \\ K \end{pmatrix}=M_{q-1}\;\dots\;M_0\begin{pmatrix} h \\ k \end{pmatrix}$$

with each

$$M_i=\begin{pmatrix} z & 0 \\ D_i(w) & 1 \end{pmatrix}\text{ or }\begin{pmatrix} 0 & z \\ -1 & \hat{D}_i(w) \end{pmatrix}$$

and

$$h^+=1,\qquad k^+=0$$
$$h^-=0,\qquad k^-=1.$$

The D_i, $\hat{D}_i$ are holomorphic functions, determined by $\gamma\varphi_j$ for $j<i$ in the manner above. In this basis $S(\varphi_i,\gamma)$ has matrix

$$[M_{q-1}\;\dots\;M_0 V\,adM_0\;\dots\;adM_{q-1}]^T$$

where $V=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$. The inclusions of J_+, J_- are represented by

$$ad[M_q \dots M_0]^T\begin{pmatrix}1\\0\end{pmatrix}$$

$$ad[M_q \dots M_0]^T\begin{pmatrix}0\\1\end{pmatrix}$$

respectively. We define the elementary singularities in the same way and repeating the argument above shows that it suffices to verify the inequality for these.

The analyses of the elementary singularities of level $j=0$ and $j>0$ are a little different. First suppose that $j>0$ so γ is holomorphic and non-vanishing. We may then take $\gamma=1$ and $\varphi_i=0$ for $i<j$ so at the beginning of stage j our generators are

$$\begin{aligned} h' &= z^{j+1} \text{ on } Y^*_+\\ &=0 \text{ on } Y^*_-\\ k' &=1 \text{ on } Y^*_+ \text{ and } Y^*_-. \end{aligned}$$

If $\hat{D}_j=k\cdot w$, the final basis elements are

$$\begin{pmatrix}H\\K\end{pmatrix}=\begin{pmatrix}z^{q-j}\\ kw-z^{j+1}+ZP(z,w)\end{pmatrix} \text{ on } Y^*_+$$

$$=\begin{pmatrix}z^{q-j}\\ kw+zP(z,w)\end{pmatrix} \qquad \text{on } Y^*_-$$

with $P(z,w)=\sum_{i=j+1}^{q-1} D_1(w)\, z^{i-j}$. The φ's are recovered from the D's from the equations:

$$\begin{aligned} 1&=\varphi_j\cdot k(w)\\ 0&=\varphi_\lambda k(w)+\dots\dots \varphi_j D_\lambda \end{aligned}$$

for $j<\lambda\leqslant q-1$. So again φ_λ is determined from the previous φ's up to a term with at worst a double pole and there are in total $2(q-j)$ parameters involved in the polar data.

On the other hand

$$ad(M_q \dots M_0)^T=\begin{pmatrix}k(w)+z^p & z^j-k-z^p\\ -z^{q-j} & z^{q-j}\end{pmatrix}$$

and the multiplicity of the zero is $q-j$ on each sheet, so $\delta=2(q-j)$.

If $j=0$ there are two cases to consider—either γ has a simple zero or simple pole. In the first we can choose the D_i description at each stage so

$$M_{q-1}\dots\dots M_0=\begin{pmatrix}z^q & 0\\ P(z,w) & 1\end{pmatrix}$$

with $P(z,w)=\Sigma D_{i-1}(w)z^i$. Then

$$K^+=P(z,w) \qquad K^-=1$$

and $K^+_{(i)}=\gamma\varphi_i K^-_{(0)}$ so all the φ_i have at worst simple poles. Thus the number of polar parameters is q. On the other hand the inclusion maps are given by:

$$\begin{pmatrix}1\\0\end{pmatrix} \text{ on } Y^*_+,\ \begin{pmatrix}-p\\ z^q\end{pmatrix} \text{ on } Y^*_-.$$

There is no zero on Y^*_+ and a zero of multiplicity q on Y^*_- [since $D_0(w)$ has a simple zero]. So δ is also equal to q. The case when γ has a pole is similar. (Indeed it is clear that this is obtained from the first case by interchanging the sheets).

Case III. $\Delta = z^{2p+1}w^{2q}$.

The two systems of equations (5.27A) and (5.27B) are handled in the same way. Considering (A) first then (B) we find generators:

$$\begin{pmatrix} H \\ K \end{pmatrix} = M_{q-1} \ldots . M_0 N_p \ldots . N_1 \begin{pmatrix} t \\ 1 \end{pmatrix}$$

where

$$N_i = \begin{pmatrix} t^2 & 0 \\ C_i(w) & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 0 & t^2 \\ -1 & C_i \end{pmatrix}$$

and

$$M_i = \begin{pmatrix} 1 & D_i(t^2) \\ 0 & w \end{pmatrix} \text{ or } \begin{pmatrix} D_i(t^2) & -1 \\ w & 0 \end{pmatrix}.$$

Repeating the continuity argument given in Case I it suffices to show that there is a small deformation through matrices with the same determinant to a matrix function which does not vanish at $z = w = 0$. We do this by deforming to the situation when C_i, D_i are holomorphic and $D_0(0) \neq 0$. As in Case I we may then suppose that the C_i are all 0. Our deformed matrix is then:

$$\begin{pmatrix} w^q P(z, w) & w^{2q} \\ z^{2p+1} - P(z, w)^2 & -w^q P(z, w) \end{pmatrix}$$

where $P(z, w) = \sum_{i=0}^{q-1} D_i(z)\, w^i$, which does not vanish at the origin. Case IV is similar.