J. DIFFERENTIAL GEOMETRY
24 (1986) 275–341

CONNECTIONS, COHOMOLOGY AND THE INTERSECTION FORMS OF 4-MANIFOLDS

S. K. DONALDSON

I. Introduction

The Yang-Mills gauge fields, which were first introduced in Mathematical Physics, can be used to obtain strong results about the differential topology of 4-manifolds. In a previous paper [12] simply connected manifolds with definite quadratic intersection form were studied through the associated moduli spaces of "instanton" solutions to the Yang-Mills differential equations. Here we shall extend these methods to discuss the existence of smooth, simply connected, 4-manifolds with certain indefinite intersection forms. A companion article [14] will discuss non-simply-connected manifolds (about which results have recently been obtained by Fintushel and Stern [16]) and we shall develop here a number of techniques to be used in that article and also in other applications [13].

If X is a closed oriented 4-manifold, then the intersection of 2-cycles defines a unimodular bilinear form on the free group: $H_2(X;\mathbf{Z})/\text{Torsion}$. Changing the orientation of the 4-manifold reverses the sign of the form and we shall adopt here the opposite convention to [12], that is, eventually we consider "anti-self-dual" connections—this fits in better with the conventional orientation of complex surfaces. With this convention the Theorem of [12] becomes:

(1.1) Theorem A. *If X^4 is smooth, compact, and simply connected and with negative intersection form ($\alpha \cdot \alpha \leqslant 0$ for all α in H_2), then the form is equivalent over the integers for the standard example*:

$$(-1) \oplus (-1) \oplus \cdots \oplus (-1).$$

Of course, the point of this is that many nonstandard definite forms exists; for example, the positive definite root matrix E_8 and its multiples $\pm n \cdot E_8$.

The nonsingular forms over the real numbers are classified by rank and the number b^+ of positive eigenvalues in a diagonalization. According to the Hasse-Minkowski classification [23] the only other invariant for indefinite

Received July 25, 1985 and, in revised form, January 22, 1986.

unimodular forms, over the integers, is the division into odd and even types. These forms all appear in the two families:

odd: $n \cdot (1) + m(-1)$.

even: $\pm nE_8 + m\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$.

Each of the odd indefinite forms can be realized as the intersection form of a suitable connected sum of copies of the complex projective plane, so the interesting problems regarding existence of manifolds concern the even forms. For simply connected manifolds this even condition on the intersection form is equivalent to the presence of a spin structure.

Thus we consider the family of forms:

$$n(-E_8) + m\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad (n \geqslant 0).$$

The hyperbolic form $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ is realized by the manifold $S^2 \times S^2$ and the next simplest known example, beyond connected sums of $S^2 \times S^2$'s, is the intersection matrix $-2E_8 + 3\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ of a complex $K3$ surface K. All the smooth, simply connected, spin 4-manifolds that are known are homotopy equivalent (and even homeomorphic) to connected sums of these basic examples. It is a well-known general conjecture that, indeed, no other homotopy types are realized. The results of this paper go a small way in the direction dealing with the cases when the number b^+ of positive parts is 1 or 2. We shall establish:

(1.2) Theorem B. *If X^4 is a smooth, simply connected, spin 4-manifold whose intersection form has one positive part, then the form is equivalent over the integers to $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$.*

(1.3) Theorem C. *If the intersection form of such a manifold has two positive parts, then it is equivalent over the integers to $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) \oplus \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$.*

(1.4) Remarks. (i) That these manifolds are homotopy equivalent to the basic examples $S^2 \times S^2$ and $S^2 \times S^2 \# S^2 \times S^2$ follows from the Theorem of Milnor-Whitehead. By the classification of Freeman they are homeomorphic. Of course, such nonexistence results would not be true for topological rather than smooth manifolds.

(ii) By taking sums with $S^2 \times S^2$'s one immediately deduces Theorem B, and some cases of Theorem A, from Theorem C. However, it is more instructive to see the proofs separately.

(iii) We can deduce from Theorem C:

(1.5) Corollary. *The $K3$ surface K is (smoothly) indecomposable; it cannot be expressed as a connected sum $K = K' \# S^2 \times S^2$.*

There is, however, a rather simpler and more direct proof of this corollary which will appear elsewhere [15].

While the main ingredients of the proofs of Theorems B and C below are the same as Theorem A in [12] there are also some essential differences which are easiest to understand by starting with a different proof of Theorem A. In the proof of [12] a 5-dimensional moduli space M of (anti) self-dual connections provided a cobordism between a definite manifold X^4 and a number of copies of $\mathbb{CP}^2$ and the result was deduced from the cobordism invariance of signature. However, as we shall explain in §III(ii) below it is also possible to compute the intersection form of X directly from this moduli space, using certain cohomology classes in M. Restrict, now, to spin manifolds and even forms. Then Theorem A can be summarized as

$$(1.6) \qquad \left\{ \begin{array}{c} X \text{ spin, simply-connected} \\ \text{definite form} \end{array} \right\} \Rightarrow \left(H_2(X) = 0 \right)$$

which is rather trivially equivalent to

$$(1.7) \quad (\alpha) \quad \left\{ \begin{array}{c} X \text{ spin, simply-connected} \\ \text{definite form} \end{array} \right\} \Rightarrow \begin{array}{c} (\alpha_1 \cdot \alpha_2 = 0 \bmod 2 \text{ for} \\ \text{all } \alpha_1, \alpha_2 \in H_2(X; \mathbb{Z})) . \end{array}$$

This latter formulation is the one we shall generalize to indefinite forms, starting from the following simple algebraic observation.

If $(\cdot)$ is an even, unimodular form on a lattice L of rank r, then we may associate to its mod 2 reduction on $L \otimes \mathbb{Z}/2$ a "symplectic" form:

$$\omega \in \Lambda^2(L^* \otimes \mathbb{Z}/2), \qquad \omega(\alpha_1, \alpha_2) = \alpha_1 \cdot \alpha_2 \bmod 2,$$

using the fact that $(\alpha \cdot \alpha) = 0 \bmod 2$. Just as for symplectic forms over the reals, the condition that $(\cdot)$ is unimodular implies that r is even, $r = 2p$, and the exterior power

$$\omega^p \in \Lambda^r(L^* \otimes \mathbb{Z}/2)$$

is nonzero. So, if $\omega^d = 0$, then $d > p$. But, again just as for real forms, the powers of ω may be regarded as multilinear functions on copies of $L \otimes \mathbb{Z}/2$. Thus we have a string of implications:

$$r = 0 \Leftrightarrow \{\alpha_1 \cdot \alpha_2 = 0 \bmod 2 \text{ for all } \alpha_1, \alpha_2 \text{ in } L\},$$

$$r \leqslant 2 \Leftrightarrow \left\{ \begin{array}{c} Q_4(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \equiv (\alpha_1 \cdot \alpha_2)(\alpha_3 \cdot \alpha_4) + (\alpha_2 \cdot \alpha_3)(\alpha_2 \cdot \alpha_4) \\ +(\alpha_1 \cdot \alpha_4)(\alpha_2 \cdot \alpha_3) = 0 \bmod 2 \text{ for all } \alpha_1, \alpha_2, \alpha_3, \alpha_4 \text{ in } L \end{array} \right\},$$

$$r \leqslant 4 \Leftrightarrow \left\{ \begin{array}{l} Q_6(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) \equiv \Sigma(\alpha_{i_1} \cdot \alpha_{i_2})(\alpha_{i_3} \cdot \alpha_{i_4})(\alpha_{i_5} \cdot \alpha_{i_6}) \\ = 0 \bmod 2 \text{ for all } \alpha_1, \alpha_2, \cdots, \alpha_6 \text{ in } L \end{array} \right\},$$

with similar implications for higher ranks.

By the classification, the forms $\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)\oplus\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$ appearing in Theorems B and C are characterized exactly as the even forms with ranks 2 and 4 respectively. So to prove these theorems it suffices to show:

$$(1.8)\quad (\beta)\quad \left\{\begin{matrix} X \text{ spin, simply-connected,} \\ b_2^+(X)=1 \end{matrix}\right\} \Rightarrow \left(\begin{matrix} Q_4(\alpha_1,\alpha_2,\alpha_3,\alpha_4)=0 \bmod 2 \text{ for all} \\ \alpha_1,\alpha_2,\alpha_3,\alpha_4 \text{ in } H_2(X;\mathbf{Z}) \end{matrix}\right),$$

$$(1.9)\quad (\gamma)\quad \left\{\begin{matrix} X \text{ spin, simply-connected,} \\ b_2^+(X)=2 \end{matrix}\right\} \Rightarrow \left(\begin{matrix} Q_6(\alpha_1,\alpha_2,\cdots,\alpha_6)=0 \bmod 2 \text{ for all} \\ \alpha_1,\cdots,\alpha_6 \text{ in } H_2(X;\mathbf{Z}) \end{matrix}\right),$$

which clearly generalize the basic case (α). We could also formulate corresponding statements for larger values of the number b^+ of positive parts of the form, but these would be incorrect as the $K3$ surface, with $b^+=3$, shows.

When the bilinear form $(\cdot)$ is the intersection form of a manifold the multi-linear expressions $Q_{2d}(\alpha_1,\cdots,\alpha_d)$ have a simple geometric meaning derived from the definition of the intersection product. In a 4-manifold the integral homology classes can be represented by oriented surfaces $\Sigma_1,\cdots,\Sigma_d$ placed in general position. Then $Q_{2d}(\alpha_1,\cdots,\alpha_d)$ is equal (modulo 2) to the number of configurations of unordered d-tuples $X_1,\cdots,X_d$ of points in X such that each surface Σ_i contains a point X_j. It is in this way that the expressions Q_{2d} will emerge (in §III(iv)) from the study of Yang-Mills moduli spaces.

In [12] a 4-manifold with definite intersection form was manifested explicitly in its moduli space of gauge fields in the form of concentrated or particle-like connections constructed by Taubes, and this construction lay at the heart of the proof. Similarly, to study indefinite forms we make use of further work of Taubes [25] in which more general "multi-instanton" solutions to the differential equations are constructed corresponding to a configuration of points or particles in the 4-manifold. In §§II and III we will find a way to associate to each homology class in a 4-manifold X a cohomology class in any family of connections over X. In as much as we can think of connections as generalizing the points in the manifold (or particle-like connections) this construction with homology generalizes Poincaré duality. When we consider a cup product of these cohomology classes—represented by intersections within the space of connections—the expressions Q_{2d} will appear as the number of boundary components of multi-instanton solutions. We will thus use homology and cohomology inside families of connections over X to make the deductions (β) and (γ) about the homology of X. This shift from cobordism to homology marks the main difference in the proof here from that of [12]. In a similar spirit, homology and cohomology enters into the work of Fintushel and Stern.

For the proof we also need a more subtle topological construction exploiting the spin condition on the 4-manifold. While the concentrated connections modelled on a single point are described completely by this point and a scale size, in the general case there are extra parameters making up a "link" in the moduli space. Homologically, these parameters are detected by torsion classes developed in §§II and III(v) using the index of the Dirac family. At the geometric level the most arduous part of the work is giving a precise description of these links in §§IV, V, and VI. Their structure is determined by the harmonic forms on the 4-manifold through constraints that were, again, introduced in the work of Taubes. In §§IV, V, and VI we work through a general theory describing the ends of moduli spaces, culminating in Theorem (5.5). The main motivation is to mimic ideas of algebraic geometry and extend the Kuranishi deformation theory to "ideal" ASD connections, with formal point singularities (see §III(ii)). Finally we explain in an Appendix the differences that emerge between the gauge fields over manifolds with $b_2^+ < 3$ (when we can deduce our topological results) and those with $b_2^+ \geqslant 3$ (when these methods do not yet yield any information).

II. Homotopy and connections

Notation. In this section X is a compact connected, oriented 4-manifold; $P \to X$ a principal $SU(2)$ bundle; and E, $\mathfrak{g}_P$ the vector bundles associated to P by, respectively, the fundamental representation on $\mathbf{C}^2$ and the adjoint representation.

II(i). We shall construct cohomology classes in the parameter spaces of families of connections over X. Ultimately these will be the moduli spaces of anti-self-dual (ASD) Yang-Mills connections but our topological discussion will apply to very general families and can be expressed in terms of the homotopy type of the infinite-dimensional space

$$\mathscr{B} = \mathscr{A}/\mathscr{G} \qquad (= \mathscr{B}_{P,X} \text{ or } \mathscr{B}_X) \tag{2.1}$$

of equivalence classes of connections on P. This is formed by dividing the affine space $\mathscr{A}$ of connections by the gauge group $\mathscr{G}$ of automorphisms of P (see [17], [5]). In fact it is easiest to start with the slightly larger space

$$\tilde{\mathscr{B}} = \mathscr{A}/\mathscr{G}_0, \tag{2.2}$$

where $\mathscr{G}_0 \triangleleft \mathscr{G}$ is the subgroup of automorphisms which fix the fiber P_{x_0} over a base point x_0 in X. This smaller group $\mathscr{G}_0$ acts *freely* on $\mathscr{A}$. Any automorphism g of P fixes a connection $\mathscr{A}$ precisely when it commutes with the holonomy maps of $\mathscr{A}$ along paths in X, so if g is the identity on one fiber it will be so on all. An equivalent description is

$$(2.3)\qquad \tilde{\mathscr{B}} = \mathscr{A} \times_{\mathscr{G}} P_{x_0}$$

and the points of $\tilde{\mathscr{B}}$ represent isomorphism classes of connections on "based" bundles, having a preferred identification of P_{x_0} with $SU(2)$. If $\mathscr{G}$ and $\mathscr{A}$ are given suitable Sobolev structures (allowing connections which are not C^∞), $\tilde{\mathscr{B}}$ can be topologized as a Banach manifold and the (weak) homotopy type will not depend upon the precise Sobolev structure which is chosen (cf. [5], [17]).

There is a standard way of representing the space $\tilde{\mathscr{B}}$, up to homotopy, as a function space; generalizing the representation of bundles by homotopy classes of maps into the classifying space $BSU(2) \cong \mathbb{HP}^\infty$ of $SU(2)$.

(2.4) Lemma (*cf.* [5, *Proposition* 2.4]). *There is a weak homotopy equivalence*

$$\tilde{\mathscr{B}} = \mathrm{Maps}_P(X, BSU(2)),$$

where Maps_P *denotes the spaces of based maps in the homotopy class corresponding to the bundle* P.

Proof. This can be understood by introducing families of (based) connections over X parametrized by an auxiliary compact space T.

The homotopy classes of (unbased) maps

$$T \to \mathrm{Maps}_P(X, BSU(2))$$

are in (1-1) correspondence with those of maps

$$T \times X \to BSU(2)$$

which collapse $T \times \{x_0\}$ to the basepoint in $BSU(2)$ and induce the bundle P on each slice $X \times \{t\}$. These correspond, in turn, to the isomorphism classes of $SU(2)$ bundles $\mathbb{P}$ over $X \times T$ trivialized over $T \times \{x_0\}$ and with $\mathbb{P}|_{\{t\}\times X} \cong P$.

Just as any bundle over a manifold admits a connection, so the bundle $\mathbb{P}$ admits a "partial connection" $\mathbf{A}$ in the X direction. For example, if T is a sphere, we can choose a connection on $\mathbb{P}$ over the maniold $X \times T$. Restricting to slices, and using the trivialization over $T \times \{x_0\}$, $\mathbf{A}$ defines a map

$$f_{\mathbf{A}}: T \to \tilde{\mathscr{B}}.$$

If $\mathbf{A}_1$, $\mathbf{A}_2$ are two choices of the partial connection on $\mathbb{P}$, the linear family $s\mathbf{A}_1 + (1-s)\mathbf{A}_2$ induces a homotopy $f_{\mathbf{A}_1} \simeq f_{\mathbf{A}_2}$.

Conversely, since $\mathscr{G}_0$ acts freely on $\mathscr{A}$ there is a "universal" bundle $\tilde{\mathbb{P}} = \mathscr{A} \times_{\mathscr{G}_0} P$ over $\tilde{\mathscr{B}} \times X$ (see [10]), trivialized over $\tilde{\mathscr{B}} \times \{x_0\}$. To any map $f\colon T \to \tilde{\mathscr{B}}$ there corresponds a bundle over $T \times X$ by pull back, and homotopic maps induce isomorphic bundles.

We will not make any real use of this description of the homotopy type of the space of connections $\tilde{\mathscr{B}}$. It has been included to emphasize that the discussion of the cohomology of $\tilde{\mathscr{B}}$ depends, at bottom, only on the homotopy type of the 4-manifold X. It does not make any essential use of the smooth structure (although we are free to do this in our calculations). We will base our constructions directly on the existence of the universal $SU(2)$ bundle $\tilde{\mathbb{P}}$ over $\tilde{\mathscr{B}} \times X$, the reader should compare [5] and [8]. The key to the paper is the following definition:

(2.5) Definition. *Denote by*

$$\tilde{\mu}\colon H_2(X;\mathbb{Z}) \to H^2(\tilde{\mathscr{B}}_X;\mathbb{Z})$$

the slant product $\tilde{\mu}(\alpha) = c_2(\tilde{\mathbb{P}})/\alpha$, *where* $c_2(\tilde{\mathbb{P}})$ *is the usual second Chern class in* $H^4(\tilde{\mathscr{B}}_X \times X;\mathbb{Z})$

The polynomial identities (α), (β), (γ) of the introduction will be derived from this map $\tilde{\mu}$. It depends only on the 2-skeleton of the manifold X. If we represent (as we always can) a 2-dimensional homology class by an embedded surface $\Sigma \hookrightarrow X$, then $\tilde{\mu}[\Sigma]$ factors through the obvious map—induced by restriction of bundles and connections—

$$\tilde{r}_\Sigma\colon \tilde{\mathscr{B}}_{P,X} \to \tilde{\mathscr{B}}_\Sigma. \tag{2.6}$$

(There is no difficulty in choosing Sobolev structures so that this map is a smooth map of infinite-dimensional manifolds.) Notice that while there are topologically distinct bundles P over X-determined by the characteristic class $c_2(P) \in H^2(X;\mathbb{Z}) \cong \mathbb{Z}$—this distinction is lost when the bundles are restricted to surfaces.

The universal bundle $\tilde{\mathbb{P}}$ defines classes in K-theory as well as ordinary cohomology and these can naturally be studied using elliptic operators on X. Indeed elliptic operators may be used to represent cycles in K-homology [4]. Here it would be more strictly correct to formulate the work in terms of compact families of connections, as in the proof of Lemma (2.4), but to save notation we will ignore this point.

The fundamental representation of $SU(2)$ associates to $\tilde{\mathbb{P}}$ a complex vector bundle $\tilde{\mathbb{E}}$ over $\tilde{\mathscr{B}}_X \times X$. If D is an elliptic operator on X, there is a K-theory index of the family of operators formed by coupling D to the partial connection on $\tilde{\mathbb{E}}$:

$$\operatorname{ind}(D_{\tilde{\mathbb{E}}}) \in K(\tilde{\mathscr{B}}_X) \tag{2.7}$$

which is the K-theory slant product $[\tilde{\mathbb{E}}]/[D]$. The slant products in (rational) cohomology and K-theory are related by the Atiyah-Singer index theorem for families. In particular if Z is some spin^c manifold with a (twisted) Dirac operator D, defined by a lifting $f \in H^2(Z;\mathbb{Z})$ of $w_2(Z)$, and if we have a family $\mathbb{E}$ of connection over a product $T \times Z$:

(2.8) Atiyah-Singer index formula [9, I, IV], [10].

$$\operatorname{ch}(\operatorname{ind} D_{\mathbb{E}}) = \left(\hat{A}(Z)e^{f/2}\operatorname{ch}(\mathbb{E})\right)/[Z].$$

We will use this twice—first to give another description of the map $\tilde{\mu}$. If Σ is an embedded surface in X, then, since $w_2(\Sigma) = 0$, there is a Dirac operator ∂_Σ over Σ (with $f = 0$) which is selfadjoint and so has *numerical* index 0. The class

$$\operatorname{ind}(\partial_\Sigma) \in K(\tilde{\mathscr{B}}_\Sigma) \tag{2.9}$$

pulls back to $K(\tilde{\mathscr{B}}_X)$ by r_Σ^*. Any complex virtual bundle defines a first Chern class in H^2. More geometrically a virtual bundle defines a complex line bundle, "the determinant bundle"

$$\det([V] - [W]) = (\Lambda^{\dim V} V) \otimes_{\mathbf{C}} (\Lambda^{\dim W} W)^*. \tag{2.10}$$

Thus to every surface Σ in X we get a complex line bundle

$$\mathscr{L}_\Sigma = \left(\det\operatorname{ind}\{\partial_\Sigma\}\right)^{-1} \tag{2.11}$$

over $\tilde{\mathscr{B}}$; pulled back from the space $\tilde{\mathscr{B}}_\Sigma$ of connections over Σ. (Moreover there is no difficulty in defining these determinant line bundles over noncompact families, unlike the virtual index bundles.)

(2.12) Lemma. $\tilde{\mu}([\Sigma]) = c_1(\mathscr{L}_\Sigma)$.

Proof. Over the rationals the characteristic class $-c_1(\mathscr{L}_\Sigma) = \operatorname{ch}_1(\operatorname{index}\partial_\Sigma)$ can be calculated by the Atiyah-Singer index theorem for families (cf. [5, p. 582])

$$\begin{aligned}\operatorname{ch}_1(\operatorname{index}\partial) &= \operatorname{ch}_2(\tilde{\mathbb{E}})/[\Sigma] = \tfrac{1}{2}\left(c_i^2 - 2c_2\right)(\tilde{\mathbb{E}})/[\Sigma] \\ &= -c_2(\tilde{\mathbb{E}})/[\Sigma] = -\tilde{\mu}[\Sigma].\end{aligned}$$

The result is true over the integers since the space $\tilde{\mathscr{B}}_\Sigma$ is torsion free [5, p. 542].

Next we define mod 2 cohomology classes using indices of operators which make essential use of a spin structure on the manifold. If X is a spin 4-manifold there are spin bundles V^+, V^- corresponding to the fundamental representation of the two factors: Spin (4) $\cong SU(2) \times SU(2)$. The Dirac operator D^* interchanges the bundles:

$$D^*\colon \Gamma(V^-) \to \Gamma(\tilde{V}^+). \tag{2.13}$$

For each connection A on P we construct an extended Dirac operator:

$$D_A^* = \Gamma(V^- \otimes_{\mathbf{C}} E) \to \Gamma(V^+ \otimes_{\mathbf{C}} E).$$

Since the group $SU(2)$ is isomorphic to Sp(1), each of the vector bundles V^-, V^+, E is a quaternionic bundle. Hence the tensor products have real structures, compatible with the Dirac operator, and the kernel and cokernel of the operator are naturally real vector spaces. Thus the index of any family of these operators gives a real virtual bundle

$$\operatorname{ind}(D_A^*) \in KO(\tilde{\mathscr{B}}). \tag{2.14}$$

The Stiefel-Whitney characteristic classes extend to such virtual bundles.

(2.15) Definition. *If X is a spin 4-manifold, define cohomology classes as follows*:

$$\tilde{u}_i = w_i(\operatorname{ind} D_A^*) \in H^i(\tilde{\mathscr{B}}; \mathbb{Z}/2).$$

(2.16) Discussion. Suppose that the 4-manifold X is simply connected. Then it has the homotopy type of a cofibration

$$\bigvee S^2 \to X \to S^4,$$

where the wedge of S^2's represent a basis for $H_2(X)$. Using Lemma (2.4), just as in [5, Proposition 2.10], the space of connections is, at the level of homotopy, fibered:

$$\tilde{\mathscr{B}}_{S^4} \to \tilde{\mathscr{B}}_X \to \prod \tilde{\mathscr{B}}_{S^2}. \tag{2.17}$$

And, as in Lemma (2.4),

$$\tilde{\mathscr{B}}_{S^2} \simeq \operatorname{Maps}(S^2, BSU(2)) \simeq \Omega(SU(2)) = \Omega S^3,$$
$$\mathscr{B}_{S^4} \simeq \operatorname{Maps}(S^4, BSU(2)) \simeq \Omega^3 S^3.$$

The rational cohomology of ΩS^3 is generated by an element in H^2, which corresponds exactly to $\tilde{\mu}[S^2]$. Thus the map $\tilde{\mu}$ when extended to define

$$\tilde{\mu}\colon \text{Polynomial algebra on } H_2(X) \to H^*(\tilde{\mathscr{B}}_X)$$

captures all the (rational) cohomology of the base Prod $\tilde{\mathscr{B}}_{S^2}$ of the fibration.

The classes $\tilde{u}_i$ have more to do with the fiber $\Omega^3 S^3$—corresponding to $SU(2)$ connections over the 4-sphere (cf. [8]). If in place of $SU(2) \cong \operatorname{Sp}(1)$ we consider connections for arbitrarily large symplectic groups $\operatorname{Sp}(l)$, then according to [3], [8] the classifying map

$$\lim \Omega^3 \operatorname{Sp}(l) \to BO$$

of the Dirac index bundle induces the Bott periodicity isomorphism $\Omega^2 \mathrm{Sp} \simeq BO$. Since the mod 2 cohomology of BO is generated by the Stiefel-Whitney classes, the classes u_i capture all the mod 2 cohomology of $\Omega^3 S^3$ which is stable with respect to the size of the (symplectic) gauge group. The point of the spin condition on the 4-manifold is that in this case all such cohomology extends from the fiber of (2.17) over the total space. Conversely, if X is not spin it is easy to see that $\tilde{u}_1$ does not extend—indeed in the spectral sequence

$$d_2(\tilde{u}_1) = \tilde{\mu}_{(2)}(w_2(X)),$$

where $\tilde{\mu}_{(2)}$ is the mod 2 version of (2.5).

II(ii). Stabilizers and universal families. The homotopy of the true space of connections which we need to use is a little more complicated than that of the space $\tilde{\mathscr{B}}$ of based connections. First, just as we have associated cohomology classes to the 2-cells (giving the map $\tilde{\mu}$) and the 4-cell (giving the torsion classes $\tilde{u}_i$), so when we forget the trivialization of the bundle over the base point we obtain more structure from this 0-cell in X. And just as the 2-cells and 4-cell lead to a potentially nontrivial fibration (2.17) so the 0-cell interferes with the previous constructions. Second we meet the new feature that the isotropy in the gauge group varies between one connection and another and we have to take special account of the "reducible" connections. It is at this point that we diverge a little from Atiyah and Bott who used instead $\mathscr{G}$-equivariant cohomology to study connections over Riemann surfaces.

To each connection A on P we associate the stabilizer $\Gamma_A \subset \mathscr{G}$ under the action (2.1). It is a compact Lie group which is identified with a subgroup of $SU(2) \cong \operatorname{Aut} P_{x_0}$ by a choice of base point x_0. Γ_A is the centralizer of the holonomy subgroup of the connection and always contains the center $\{\pm 1\}$ of $SU(2)$. Denote by $\tilde{\mathscr{B}}^* \subset \tilde{\mathscr{B}}$ the subset representing connections for which $\Gamma_A = \{\pm 1\}$. $\tilde{\mathscr{B}}^*$ is open in $\tilde{\mathscr{B}}$ and has complement of infinite codimension so the homotopy type is unaffected by the removal of the *reducible connections* $\tilde{\mathscr{B}} \setminus \tilde{\mathscr{B}}^*$, where $\Gamma_A \neq \{\pm 1\}$.

Dividing by the remaining part $\mathscr{G}/\mathscr{G}_0 \cong SU(2)$ of the gauge group gives a principal fibration:

$$(SU(2)/\pm 1) = SO(3) \to \tilde{\mathscr{B}}^* \to \tilde{\mathscr{B}} \tag{2.18}$$

with base $\mathscr{B}^* \subset \mathscr{B}$ the equivalence classes of irreducible connections. Equivalently a fibration (in homotopy):

$$\tilde{\mathscr{B}}^* \to \mathscr{B}^* \to BSO(3) \quad (\text{cf. } (2.17)). \tag{2.19}$$

The next three propositions bear on the problem of pushing cohomology from $\tilde{\mathscr{B}}^*$ down to $\mathscr{B}^*$, starting with the "universal" problem of pushing down the bundle $\tilde{\mathbb{P}}$ on $\tilde{\mathscr{B}}^* \times X$. We can in any event push down the $SO(3)$ bundle $\mathfrak{g}_{\tilde{\mathbb{P}}}$ associated to $\tilde{\mathbb{P}}$, and working over the rationals this would suffice, but we need to keep track of torsion. Recall that the adjoint representation defines a homomorphism $U(2) \to SO(3)$.

(2.20) Proposition. *If $c_2(P)$ is odd there is a lifting of the structure group of the universal $SO(3)$ bundle over $\mathscr{B}^* \times X$ to give a $U(2)$ bundle $\mathbb{P}$. If in addition the 4-manifold X is spin, then we can take $\mathbb{P}$ to be an $SU(2)$ bundle.*

Proof. Consider the $\mathbb{C}^2$ vector bundle $\tilde{\mathbb{E}}$ over $\tilde{\mathscr{B}}^* \times X$ associated to $\tilde{\mathbb{P}}$. The element -1 in the gauge group $\mathscr{G}$ (or $\mathscr{G}/\mathscr{G}_0$) fixes all connections but acts as (-1) on the fibers of $\tilde{\mathbb{E}}$, so the vector bundle does not descend directly. However, any 4-manifold admits a spin^c structure [19], and fixing one of these gives a Dirac operator

$$D^*\colon \Gamma(W_-) \to \Gamma(W_+)$$

between the twisted spin representations W_-, W_+. As in (2.8) the numerical index of this operator when coupled to a connection A on P is:

$$\begin{aligned} \operatorname{ind} D_A^* &= c_2(P) + \left((\tau - f^2)/4\right) \in \mathbb{Z} \\ &= c_2(P) - 2 \operatorname{ind} D. \end{aligned} \tag{2.21}$$

Now form the corresponding virtual bundle over the space $\tilde{\mathscr{B}}^*$ and determinant line bundle as in (2.10):

$$\operatorname{Det}\left(\operatorname{ind} D_A^*\right) = \left\{ \Lambda^{\dim\ker D_A^*}\left(\operatorname{Ker} D_A^*\right)\right\} \otimes \left\{ \Lambda^{\dim\ker D_A}\left(\operatorname{Ker} D_A\right)\right\}^*.$$

Clearly the scalar α in $\mathbb{C}^*$ acts on $\tilde{\mathbb{E}}$ and the induced action of Det(ind D_A^*) is by $\alpha^{\operatorname{ind} D_A^*}$. From (2.21), the parity of $\operatorname{ind} D_A^*$ is the same as that of $c_2(P)$, so when $c_2(P)$ is odd, -1 acts as -1 on the line bundle Det(Ind D_A^*) over $\tilde{\mathscr{B}}^*$. Pull this back to get a line bundle π_1^*(Det(ind D_A^*)) over $\tilde{\mathscr{B}}^* \times X$: then -1 acts trivially on $\tilde{\mathbb{E}}' = \pi_1^*(\operatorname{Det}(\operatorname{ind} D_A^*)) \otimes_{\mathbb{C}} \tilde{\mathbb{E}}$ over $\tilde{\mathscr{B}}^* \times X$ and we may descend the $U(2)$ bundle $\tilde{\mathbb{E}}'$ to $\mathscr{B}^* \times X$. If X is spin, then, as in (2.14), we can form the *real* line bundle det(ind D_A^*) and set

$$\tilde{\mathbb{E}}' = \pi_1^*\left(\operatorname{Det}\left(\operatorname{ind} D_A^*\right)\right) \otimes_{\mathbb{R}} \tilde{\mathbb{E}} \tag{2.22}$$

in which case $\tilde{\mathbb{E}}'$ descends as an $SU(2)$ bundle.

(2.23) Remark. The existence of these bundles is equivalent to, respectively, the lifting of the second Stiefel-Whitney class of the fibration (2.19) to the integers and to the vanishing of this class. On the other hand the existence of a spin^c or spin structure on X is equivalent to corresponding properties of $w_2(TX)$. It is not hard to understand this duality in the framework of the duality correspondence of §III below. The obstruction $w_2 \in H^2(\mathscr{B}^*, \mathbb{Z}/2)$ coming from the base point fibration (2.19) plays an important part in the arguments of Fintushel and Stern [16].

(2.24) Definition. *If $c_2(P)$ is odd and X is spin set, then*

$$u_i = w_i\left(\operatorname{Ind} D_{\mathbb{E}}^*\right) \in H^i\left(\mathscr{B}^*; \mathbb{Z}/2\right),$$

where $D_{\mathbb{E}}^*$ is the Dirac family on the $SU(2)$ bundle $\mathbb{E}$ over $\mathscr{B}^*$.

If $c_2(P)$ is even we do not get a universal bundle over $\mathscr{B}^*$ so cannot define classes u_i in this way; except for the special case of the one-dimensional class. Since the stabilizer (-1) acts trivially on the real line bundle defined by D_A we can make the following

(2.25) Definition. *If $c_2(P)$ is even and X is spin, let*

$$u_1 \in H^1(\mathscr{B}^*; \mathbf{Z}/2)$$

be the first Stiefel-Whitney class of the real line bundle pushed down from $\det \operatorname{ind} D_A$ *over $\tilde{\mathscr{B}}^*$.*

In the same way, we can push down the two-dimensional cohomology classes associated to the 2-skeleton of X.

(2.26) Proposition. (i) *For each surface $\Sigma \subset X$ the complex line bundle $\mathscr{L}_\Sigma$ descends from $\tilde{\mathscr{B}}^*_X$ to $\mathscr{B}^*$.*

(ii) *The map $\tilde{\mu}$: $H_2(X; \mathbf{Z}) \to H^2(\tilde{\mathscr{B}}^*; \mathbf{Z})$ factors through a linear map*:

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

Proof. (i) Just as in the proof of Proposition (2.20) the line bundle Λ_Σ descends since $(-1) \in \mathscr{G}$ acts as

$$(-1)^{\operatorname{ind} \partial_\Sigma} = (-1)^0 = 1$$

on the fibers of $\mathscr{L}_\Sigma$.

(ii) Since $\tilde{\mu}([\Sigma]) = c_1(\mathscr{L}_\Sigma)$ we can define classes $\mu(\Sigma)$ geometrically using the first Chern classes of the line bundles in (i). The lift $\mu(\Sigma)$ of $\tilde{\mu}(\Sigma)$ to $H^2(\mathscr{B}^*; \mathbf{Z})$ is unique, as one sees from the spectral sequence of (2.19), so μ defines a linear map on homology.

(2.27) Remark. When $c_2(P)$ is odd the map μ can also be defined by using the second Chern class of the $U(2)$ bundle $\mathbb{P}$ over $\mathscr{B}^* \times X$. The fact that the numerical index of ∂_Σ is zero implies that on $\tilde{\mathscr{B}}$ the line bundles

$$\det \operatorname{ind}(\partial_{\Sigma, \hat{\mathbf{E}}}), \quad \det \operatorname{ind}(\partial_{\Sigma, \hat{\mathbf{E}}'})$$

are canonically isomorphic.

Finally we calculate the obstruction to extending the cohomology classes $\mu(\alpha)$ in $H^2(\mathscr{B}^*; \mathbf{Z})$ to the space $\mathscr{B}$ of all equivalence classes of connections. The only important case is when the isotropy group Γ_A is S^1. As explained in [12], [17] this is the only group which can occur when P is not topologically trivial; and the connection A is then induced from an S^1 connection. In terms of vector bundles this corresponds to a splitting $E = L \oplus L^{-1}$ so $c_2(E) = -c_1(L)^2$. Similarly it is explained in those references that while the space $\mathscr{B}^*$ of irreducible connections is an infinite-dimensional manifold, the reducible connections sit as singular points in $\mathscr{B}$. A neighborhood of the reducible connection A is modelled on a quotient $H_{\mathbf{R}} \times H_{\mathbf{C}}/\Gamma_A$, where $H_{\mathbf{R}}$, $H_{\mathbf{C}}$ are respectively real and complex Hibert spaces (making up the kernel of the "gauge fixing operator" d_A^*, see §IV) and corresponding to the bundle decomposition

$$\Omega^1(\mathfrak{g}_P) \cong \Omega^1(\mathbb{R} \oplus L^2).$$

So $H_{\mathbf{C}} \subset \Omega^1(L^2)$ and the action of Γ_A is of weight 2; $e^{i\theta} \in \Gamma_A$ acts as $e^{2i\theta}$ on $H_{\mathbf{C}}$. If Σ is a surface in X the determinant line bundle $\mathscr{L}_\Sigma$ restricts to a Γ_A-equivariant line bundle on $H_{\mathbf{C}}$. The boundary of a small neighborhood of A in $H_{\mathbf{C}}/\Gamma_A$ is a copy of $\mathbf{CP}^\infty$ with canonical cohomology generator h. To determine the restriction of $\mu(\Sigma) = c_1(\mathscr{L})$ to $\mathbf{CP}^\infty$ is the same as knowing the action of the stabilizer Γ_A on the fiber of $\mathscr{L}_\Sigma$ over the origin. (The Hopf bundle on $\mathbf{CP}^\infty$ is induced from an equivariant line bundle l on $H_{\mathbf{C}}$ on which Γ_A acts with weight 2. Since (-1) acts trivially on $\mathscr{L}_\Sigma$ the weight of the action of Γ_A on $\mathscr{L}_\Sigma$ is even, $2w$ say. Then the compact group Γ_A acts trivially on the fiber of $\mathscr{L}_\Sigma \otimes L^{-w}$ over 0—so by averaging we can choose an equivariant trivialization and on the quotient $c_1(\mathscr{L}_\Sigma \otimes l^{-w}) = 0$. Thus $\mu(\Sigma) = w \cdot h \in H^2(\mathbf{CP}^\infty; \mathbf{Z})$.)

(2.28) Lemma. *The cohomology class $\mu(\alpha)$ restricts to $-\langle c_1(L), \alpha\rangle \cdot h$ on the copy of $\mathbf{CP}^\infty$ surrounding a reduction $E \cong L \oplus L^{-1}$.*

Proof. The decomposition of E, restricted to a surface Σ, induces a decomposition

$$\mathscr{L}_\Sigma = \left(\det \operatorname{ind}(\bar\partial_{\Sigma,E})\right)^{-1} = \left(\det \operatorname{ind} \bar\partial_{\Sigma,L}\right)^{-1}\left(\det \operatorname{ind} \bar\partial_{\Sigma,L^{-1}}\right)^{-1}.$$

By definition the element $e^{i\theta} \in \Gamma_A$ acts as $e^{i\theta}$ on L and $e^{-i\theta}$ on L^{-1}. So Γ_A acts with weight

$$\operatorname{ind} \bar\partial_{\Sigma,L} - \operatorname{ind} \bar\partial_{\Sigma,L^{-1}} \in Z$$

on $\mathscr{L}_\Sigma$ (cf. (2.20)). By the index theorem (2.20) on surfaces,

$$\operatorname{ind}(\bar\partial_{\Sigma,L}) = \langle c_1(L), [\Sigma]\rangle, \qquad \operatorname{ind}(\bar\partial_{\Sigma,L^{-1}}) = -\langle c_1(L), [\Sigma]\rangle$$

so the action on $\mathscr{L}_\Sigma$ has weight $2w = -2\langle c_1(L), \alpha\rangle$.

(2.29) Note. The choice $L^{\pm 1}$ of line bundle in the splitting determines the identification $\Gamma_A \cong S^1$ and so the sign of h. So the complete formula is independent of choice.

III. Topology of the ends

In [24], [25] C. H. Taubes has introduced "approximate" solutions of the ASD Yang-Mills equations over any 4-manifold which are modelled on copies of the basic instanton solution over $\mathbb{R}^4$ or S^4. Here we will explain that this construction is closely related to duality in topology and we will go on to reduce the proofs of Theorems B and C to technical assertions about the "ends" of moduli spaces.

III(i). Suppose that a manifold Z is expressed as a union of open sets $Z = Z_0 \cup Z_1$ with $\pi_1(Z_0 \cap Z_1) = \{1\}$ and that A_0 and A_1 are G-connections on bundles P_0, P_1 over Z_0, Z_1 which are both flat on the overlap $Z_0 \cap Z_1$. If z is a base point in $Z_0 \cap Z_1$, then a G-isomorphism

$$\rho\colon (P_0)_z \to (P_1)_z \tag{3.1}$$

defines, by parallel transport, an isomorphism $P_0|_{Z_0 \cap Z_1} \cong P_1|_{Z_0 \cap Z_1}$. Using this gluing map to construct a bundle $P_0 \cup_\rho P_1$ over Z there is an obvious way to define a connection, which we denote

$$A_0 \#_\rho A_1 \tag{3.2}$$

over Z; which restricts to A_i over Z_i. In general the parameter ρ will be effective; it is easy to see that the equivalence classes of connections constructed in this way are in (1-1) correspondence with cosets:

$$G/\Gamma_{A_1} \times \Gamma_{A_0}. \tag{3.3}$$

Here the groups $\Gamma_{A_i} \subset \operatorname{Aut} P_i$ are the holonomy centralizers, as in §II. Clearly the same construction can be made with connections over a number of open sets Z_i, given appropriate identification maps.

Take a standard model for $S^4 = \mathbb{R}^4 \cup \{\infty\}$ with symmetry group $SO(4)$ and let J be a connection on the negative spin bundle $V_- \to S^4$ which is:

(a) flat in a small neighborhood of ∞,

(b) preserved by the $SO(4)$ action on S^4 and V_-.

(In particular J can be a connection close to the basic ASD connection on V_-, distorted to achieve (a).)

Let $S \to X$ be the 4-sphere bundle over a (Riemannian) 4-manifold X obtained by adjoining a section at infinity to the tangent bundle $S = TX \cup X_\infty$.

Then the $SO(4)$-invariance means that on the spin bundle of each fiber $S_x \cong S^4$ there is a way to define a connection isomorphic to J. Furthermore, by using a spin^c structure on X we can define a bundle G over S which is canonically identified with the spin bundle in the direction of the fibers of $S \to X$. For a spin^c structure on X is defined by a pair of Hermitian 2-plane bundles W_-, W_+; an isomorphism $\Lambda^2 W_- \cong \Lambda^2 W_+$ and a Clifford multiplication:

$$\begin{aligned} \alpha\colon TX &\to \operatorname{Hom}(W_-, W_+), \\ \xi &\to \alpha_\xi. \end{aligned} \tag{3.4}$$

When $|\xi|^2 = 1$ the map α_ξ preserves the isomorphism on Λ^2 and the metrics. We define G by clutching W_- on the zero section $X_0 \subset S$ to W_+ on the section at ∞, X_∞, by α over the equatorial 3-spheres. Then G carries a partial connection isomorphic to J on each 4-sphere fiber.

Suppose that A_0 is a connection on some $SU(2)$ bundle E_0 over X (we adopt the language of vector bundles here). Then, following Taubes, the exponential map identifies a small ball M_x in the tangent space $(TX)_x$ with a small ball N_x in X. If we fix a homothety m_λ of the fiber $(TX)_x$ mapping the neighborhood of ∞ where J is flat onto the complement of M_x, then via the exponential identification $M_x \simeq N_x$ we can use J to define a connection $J_{\lambda,x}$ over N_x, flat on a neighborhood of ∂N_x.

We can easily fix a standard procedure—using an exponential gauge and a bump function—for modifying A_0 to get a connection A_0' flat over N_x. Then the construction above will "glue in" the connection $J_{\lambda,x}$ given an identification map

$$\rho\colon (E_0)_x \cong (W_+)_x$$

to form $A_0' \#_\rho J_{\lambda,x}$. As in (3.3) the gluing parameter ρ is, in general, determined up to sign by the connection that is constructed. So the gauge invariant description of the data required is an isomorphism

$$(\pm\rho)\colon (\mathfrak{g}_{E_0})_x \cong (\Lambda^2_{+,X})_x \tag{3.5}$$

of $SO(3)$ bundles.

However, when A_0 (hence A_0') is reducible, with extra symmetries Γ_{A_0}, we get a correspondingly smaller family of connections. In particular if $A_0 = \theta$, the trivial flat connection, the variable ρ can be cancelled by $\Gamma_\theta \cong SU(2)$. Fixing the scale size λ Taubes construction gives a map

$$\tau\colon X \to \mathscr{B}_X', \quad x \mapsto \theta \# J_{\lambda,x}. \tag{3.6}$$

This is a family of connections over X (with $c_2 = 1$) parametrized by X itself and concentrated on the diagonal in $X \times X$. We can write down a bundle $\mathbb{E}_T$ over $X \times X$ which carries this family, just as in §II(ii). Let

$$\pi_1\colon X \times X \to X$$

be projection onto the first factor, the "parameter factor." As x varies, the 4-balls $\{x\} \times N_x \cong M_x$ sweep out a neighborhood N of the diagonal Δ in $X \times X$. The bundle $(m_\lambda^{-1})^* G$ on N is identified with $\pi_1^* W_+$ on the boundary of N, and the $U(2)$ bundle $\mathbb{E}_T$ is formed by clutching with this natural identification.

At the bundle level this construction is much the same as that of the K-theory fundamental class of the diagonal $\Delta \subset X \times X$ [6].

In the latter case the retraction $S \rightleftarrows X_\infty$ defines a direct sum decomposition

$$K(S) = K(X_\infty) \oplus K(S, X_\infty).$$

The Thom class (in K-theory) of the tangent bundle is given in this decomposition by

$$G - \pi_1^*(W^+) \in K(S)$$

which maps to 0 in $K(X_\infty)$. So the K-theory diagonal class is

$$\mathbb{E}_T - \pi_1^*(W^+) \in K(X \times X). \tag{3.7}$$

In the same way the second Chern class of this Taubes family is related to the diagonal class in ordinary cohomology,

(3.8) Lemma. *The composite map*

$$H_2(X; \mathbb{Z}) \xrightarrow{\mu} (\mathscr{B}^*; \mathbb{Z}) \xrightarrow{\tau^*} H^2(X; \mathbb{Z})$$

is the Poincaré duality isomorphism.

Proof. To calculate $\tau^*\mu$ we have to know the cohomology class $c_2(\mathbb{E}_T) \in H^4(X \times X; \mathbb{Z})$ (by Remark (2.27)). Consider

$$H^4(X \times X, X \times X \setminus \Delta) \to H^2(X \times X) \to H^4(X \times X \setminus \Delta),$$

where the first group is infinite cyclic, with generator the Thom class mapping to the diagonal class $\hat{\Delta}$ in $H^4(X \times X)$.

Now $c_2(\mathbb{E}_T) - \pi_1^*(c_2(W_+))$ maps to zero in $H^4(X \times X \setminus \Delta)$, because $\mathbb{E}_T$ and $\pi_1^*(W_+)$ are isomorphic on the complement of the diagonal. So

$$c_2(\mathbb{E}_T) = \pi_1^* c_2(W_+) + n\hat{\Delta}$$

for some integer n, and restricting to a fiber $\{X\} \times X$ shows that $n = c_2(V_-, S^4) = 1$. So $\tau^*\mu(\alpha) = c_2(\mathbb{E}_T)/\alpha = (\pi^* c_2(W_+) + \hat{\Delta})/\alpha = \hat{\Delta}/\alpha$ which is the Poincaré dual of α.

(3.9) Remark. It is clear from this description that a similar "Taubes map" at the level of connections is defined for any even-dimensional spin^c manifold, just as in K-theory. In the case of 2-manifolds this gives, in effect, the classical Abel map of a Riemann surface into its Jacobian.

III(ii). At this stage we can give a slightly different proof of Theorem A, for definite manifolds, using homology. For, as explained in [12], [17], there is a 5-dimensional moduli space M of anti-self-dual connections on a bundle with $c_2 = 1$ associated to any (negative) definite simply connected manifold X^4. It can be supposed to be an orientable manifold except for some singular points labelled by the reductions of the bundle, so by $\pm e$, where $e \in H^2(X; \mathbb{Z})$,

$e^2 = -1$. Chopping off these singular points gives boundary components $\mathbb{CP}_e^2$ sitting in the standard way inside the $\mathbb{CP}_e^\infty$ of Lemma (2.27). We can find a compact manifold-with-boundary $\hat{M}$ inside M which, in addition to the $\mathbb{CP}^2$'s, has a boundary component diffeomorphic to X itself. The boundary map

$$\tau\colon X \to \hat{M} \subset \mathscr{B}^*$$

is clearly homotopic to the map discussed above, since the connections in the boundary were constructed as deformations of the approximate solutions.

Thus the image $\tau[X]$ is homologous in the space $\mathscr{B}^*$ to the projective spaces $\mathbb{CP}_e^2$ surrounding the reductions in M. So for any homology classes α_1, α_2 in $H_2(X; \mathbb{Z})$

$$\begin{aligned}
\alpha_1 \cdot \alpha_2 &= (\mathrm{P.D.}\,\alpha_1) \cup (\mathrm{P.D.}\,\alpha_2)[X] \\
&= \big(\tau^*\mu(\alpha_1) \cup \tau^*\mu(\alpha_2)\big)[X] \quad \text{(by Lemma (3.8))} \\
&= \big(\mu(\alpha_1) \cup \mu(\alpha_2)\big)[\tau X] \\
&= \tfrac{1}{2} \sum_{e^2=-1} \mu(\alpha_1) \cup \mu(\alpha_2)\, \big[\mathbb{CP}_e^2\big] \quad \text{(using the homology } \hat{M}\text{)} \\
&= \tfrac{1}{2} \sum_{e^2=-1} (\pm)\langle \alpha_1, e\rangle \langle \alpha_2, e\rangle \quad \text{(using Lemma (2.28))}.
\end{aligned}$$

At the last stage we have to introduce an unknown sign $(\pm)$ because we do not know how the orientations of M at different points compare. (This will be the main problem considered in [14].) But this equation is plainly enough to tell us that the negative definite intersection form of X is standard. If X is spin, so the form is even, no reductions can appear and using mod 2 homology we get implication (α) of §I.

If the homology classes α_1, α_2 are represented by surfaces Σ_1, Σ_2 in X, then the cohomology classes $\mu(\alpha_1)$, $\mu(\alpha_2)$, restricted to the finite-dimensional part $\hat{M}$ of $\mathscr{B}^*$, are represented by codimension 2 submanifolds V_{Σ_1}, V_{Σ_2}—zero sets of sections of the line bundles $\mathscr{L}_{\Sigma_1}$, $\mathscr{L}_{\Sigma_2}$. The duality relationship (3.8) means that on the boundary $\tau X \subset \hat{M}_0$ we can choose the sections so that

$$V_{\Sigma_i} \cap \tau X = \tau \Sigma_i. \tag{3.10}$$

So a cochain representative for $\mu(\alpha_1) \cup \mu(\alpha_2)$ is given by the 1-dimensional submanifold

$$\hat{N} = V_{\Sigma_1} \cap V_{\Sigma_2} \subset \hat{M} \tag{3.11}$$

whose boundary consists of the intersection points $\Sigma_1 \cap \Sigma_2$ and contributions from the internal projective spaces.

For the proofs of the new Theorems B and C in this paper we shall use the relative homology classes carried by more general moduli spaces. The "ends" of these may be extremely complicated and vary in an essential way with the choice of the Riemannian metric on the 4-manifold. We will be able to extract information from these spaces by defining particular representations—like the V_Σ above—for the cohomology classes $\mu(\Sigma)$ in which the *support* of the cochain is minimized. This will come from a more concrete form, Proposition (3.20) below, of the duality principle Lemma (3.8). First we will recall what can be said about the ends of the moduli spaces in a very general way, using Uhlenbeck's compactness theorem.

III(iii). Denote by M_k the moduli space of ASD connections on a bundle with $c_2 = k \geqslant 0$ over a compact Riemannian 4-manifold X. It is no more difficult to define a compactification $\overline{M}_k$ of this space than to take the corresponding step in the case $k = 1$ considered above. This exploits the basic property that for an ASD connection the Chern-Weil integrand $\mathrm{Tr}(F^2)$ (which represents $8\pi^2 c_2$) is identified with the Yang-Mills action density $|F|^2$.

Define an "ideal ASD connection" of Chern class k to consist of a pair $(A; x_1, \cdots, x_l)$, where A is an ASD connection with $c_2 = k - l$ and $(x_1, \cdots, x_l)$ is an *unordered* l-tuple of points of X, not necessarily distinct. Associate to such a pair an action density given by the measure

$$|F_A|^2 + 8\pi^2 \sum_{\alpha=1}^{l} \delta_{x_\alpha}. \tag{3.12}$$

and say that two ideal ASD connections are equivalent if the connections are gauge equivalent in the usual sense and the action densities agree. Thus the equivalence classes of ideal ASD connections are parametrized by a union of products

$$M_{k-l} \times S^l(X).$$

Put a topology on this union by saying that a sequence of ideal ASD connections converges to a limit $([A], x_1, \cdots, x_l)$ if

(3.13) (i) The action densities converge as measures on X, and

(ii) On each precompact open set in $X \setminus \{x_1, \cdots, x_l\}$ the connections converge to A in C^∞ after a suitable sequence of gauge transformations.

This criterion for convergence passes to subsequences and gives unique limits, so does define a topology. Arguing exactly as in [12], [17], Uhlenbeck's theorem shows that the space of ideal ASD connections with $c_2 = k$ is sequentially compact. Let $\overline{M}_k$ be the closure of M_k in this space of idealized connections. It is not hard to define a metric on $\overline{M}_k$, extending a standard metric on M_k, hence $\overline{M}_k$ is compact and Hausdorff; but we will not discuss this here. Indeed we will really only use the topology of $\overline{M}_k$ as a convenient language. Note that, by definition, $\overline{M}_k$ is "stratified" by subsets,

$$(3.14) \qquad \overline{M}_{k,l} = \overline{M}_k \cap \left(M_{k-l} \times S^l(X)\right), \qquad 0 \leqslant l \leqslant k;$$

that the topology induced on each stratum is the usual one, and that the unions $\overline{M}_{k,l} \cup \overline{M}_{k,l+1} \cup \cdots \cup \overline{M}_{k,k}$ are closed.

The diagonals in the symmetric products make it tricky to describe a base of neighborhoods in $\overline{M}_k$ of an ideal connection $(A; x_1, \cdots, x_l)$. Suppose however that the points $x_1, \cdots, x_j$ $(j \leqslant l)$ each have multiplicity 1 in the l-tuple $(x_1, \cdots, x_l)$ and that $\varepsilon, r > 0$. Then there is a neighborhood W of $(A; x_1, \cdots, x_l)$ in $\overline{M}_k$ on which there are well defined local "center" and "scale" maps

$$(3.15) \qquad W \xrightarrow{p_\alpha} \Omega_\alpha \subseteq X, \; W \xrightarrow{\lambda} [0, \bar{\lambda}) \subseteq \mathbb{R}^+$$

for $\alpha = 1, \cdots, j$; $x_\alpha \in \Omega_\alpha$; Ω_α disjoint. These can be defined exactly as in [12, §III.3], [17, §8]. Moreover we can suppose that every connection A' in W can be put in a gauge such that on the complement of the r-balls about the points $p_\alpha(A')$ $(\alpha = 1, \cdots, j)$ and x_β $(\beta = j+1, \cdots, k)$ in X it is within ε of A in some suitable fixed Sobolev norm. More generally, if U is a precompact subset of M_{k-l} there is a neighborhood W of $U \times (x_1, \cdots, x_l)$ on which local centers and scales are defined and in which any connection is within ε of a connection in U away from the $p_\alpha(A')$, x_β.

For the rest of this section assume that the manifold X is simply connected and has indefinite intersection form, as in Theorems B and C. In §VI it is explained that this means we may suppose each M_k $(k > 0)$ is a smooth manifold containing no reducible connections and that M_0 is a point.

Let Σ be a smooth surface in X. It is conceivable that an irreducible ASD connection over X may restrict to a reducible connection over Σ. But for each such connection we can certainly find a finite set of loops in X on which the connection restricts irreducibly. It follows easily from the weak compactness principle that, given k, we can choose a finite set of loops achieving this condition for all connections in all M_l $(0 < l \leqslant k)$. By adjoining thin, null homologous, 2-tori containing these loops, we can without loss suppose the surface Σ chosen—in a given homology class—so that the moduli spaces M_l $(0 < l < 0)$ map into the infinite-dimensional manifold $\mathscr{B}^*_\Sigma$ under restriction. The trivial connection M_0 maps to a nonmanifold point of $\mathscr{B}_\Sigma$ but as in the proof of (2.27) the line bundle $\mathscr{L}_\Sigma$ extends over this "degree 0" reduction; it makes sense to talk of a smooth section there, by working equivariantly in $\tilde{\mathscr{B}}_\Sigma$.

(3.16) Lemma. (i) *There is a smooth section s of the line bundle $\mathscr{L}_\Sigma$ over $\mathscr{B}_\Sigma \setminus \{$nonzero degree reductions$\}$ such that for every $l \leqslant k$ the pulled-back section of $r^*(\mathscr{L}_\Sigma)$ over the moduli manifold M_l vanishes transversely on a codimension 2 manifold $V_\Sigma \cap M_l$.*

(ii) *If* $\Sigma_1, \cdots, \Sigma_d$ *are surfaces in* X, *we may choose sections* s_i *of the bundles* $\mathscr{L}_{\Sigma_i}$, *as in* (i), *such that all the multiple zeros*

$$V_{\Sigma_{i_1}} \cap V_{\Sigma_{i_2}} \cap \cdots \cap V_{\Sigma_{i_m}} \cap M_l$$

are cut out transversely.

Proof. (i) The point here is that the section s "depends" only on the restriction of the connections to Σ, but we will apply transversality on the finite-dimensional moduli manifolds.

First, we can suppose that the space $\mathscr{B}_\Sigma$ is formed of L_1^2 connections (which carry the topology [5]) and so is modelled on Hilbert spaces. This means that by ([22], II,§3, Theorem 2 and Corollary) it is possible to find a countable collection $\{s_\alpha\}$ of sections of $\mathscr{L}_\Sigma$ with locally finite supports such that at each point some section s_α is nonzero.

Now consider the set of sections $s = \sum x_\alpha \cdot s_\alpha$ parametrized by an infinite vector (x_α) in l^∞ (say). Every point in the union of the restrictions $\bigcup_l r_\Sigma(M_l)$ has a neighborhood $J \subset \mathscr{B}_\Sigma$ meeting the support of only finitely many s_α and with one section, s_1 say, nonzero in J.

Using s_1 to trivialize $\mathscr{L}_\Sigma$ over J we may locally represent a section s pulled back to the moduli manifolds, as

$$\coprod_l (r_\Sigma(J) \subset M_l) \to J \underset{s/s_1}{\to} \mathbf{C}.$$

By Sard's Theorem the set of regular values of this composite is dense in $\mathbf{C}$. Equivalently the set of η_1 in $\mathbf{C}$, such that

$$s + \eta_1 s_1 = \sum x_\alpha s_\alpha + \eta_1 s_1$$

vanishes transversely when pulled back to $\coprod r_\Sigma(J)$, is dense, and a fortiori the set of vectors (x_α) in l^∞ which define such sections over J is dense. Since only finitely many of the s_α have supports meeting J it is clearly also an open subset of l^∞. Now cover the moduli manifolds by the pull-backs of countably many such neighborhoods and apply the Baire category theorem in l^∞ to find the required section s. (ii) is proved similarly.

We will next relate the various submanifolds $V_\Sigma \cap M_l$ of the moduli spaces with the points of the 4-manifold X itself which appear in the compactification. This uses a homotopy lifting property for concentrated connections together with the following lemma.

(3.17) Lemma. *Suppose* (ν, ν') *is a pair of spaces with* $H^0(\nu, \nu') = H^1(\nu, \nu') = 0$, $H^2(\nu, \nu') = \mathbf{Z}\alpha$ *and* p_2: $V_2 \to \nu$ *is a space over* ν *containing a subspace* V_1. *Let* $p_1 = p_2|V_1$ *and* $V_i' = p_i^{-1}(\nu')$. *Suppose that the map* p_2 *has the following lifting property: For any map* f *of a simplex into* V_1 *and homotopy of the* *composite $p_2 \circ f$ in ν there is a lifting to a homotopy of f in the larger space V_2. Similarly, suppose points of the fiber $p_1^{-1}(x)$ may be joined by paths in P_2^{-1}. Then any class in $H^2(V_1, V_1')$ which extends to $H^2(V_2, V_2')$ is a multiple of $p_1^*(\alpha)$.*

The proof (omitted) is a simple adaptation of the standard case when $V_1 = V_2$. In our application ν will be an open tubular neighborhood of a surface Σ in X and $\nu' \subset \nu$ the complement of a smaller closed tubular neighborhood. The spaces V_1, V_2 are sets of concentrated connections, and V_2 should be thought of as a slight thickening of V_1.

In detail, suppose that s is a section of the bundle $\mathscr{L}_\Sigma$ as in (3.16) and U a precompact open subset of the connections $\mathscr{B}_{2\nu}$ over the twice-sized neighborhood 2ν, such that s does not vanish on the closure of $\bar{\nu}$ (i.e., restricting connections to Σ). Suppose U is connected and contains the product connection θ. Since U is precompact there is a constant C such that over each small r-ball in 2ν any connection in U may be represented by a connection matrix A with

$$\|A\|_{C^0} \leqslant C \cdot r, \qquad \|A\|_{C^1} \leqslant C.$$

Let $\varepsilon \cdot r$ be small positive numbers and $N > 1$; define a set $V_{N,r,\varepsilon}$ of pairs $([A], x)$ in $\mathscr{B}_{2\nu} \times \nu$ by the following three conditions.

(i) The restriction of $[A]$ to $\backslash(\Sigma \cap (B_r(x))$ is within ε of the restriction of some connection in U (in a suitable gauge, and relative to the L^2 norm over Σ).

(ii) Over the annulus defined by radii $(r/2, r)$ centered on x, $[A]$ is represented by a connection matrix with

$$\|A\|_{C^0} \leqslant N \cdot C \cdot r, \qquad \|A\|_{C^1} \leqslant N \cdot C.$$

This condition (ii) means that if r is small relative to $C \cdot N$, then there is a well-defined "relative c_2" of the connection A over $B_r(x) \subset 2\nu$. That is we use the given gauge to homotope to the flat connection over the annulus and then take the relative Chern class of the deformed connection.

(iii) The relative c_2 of the connection over $B_r(x)$ is 1.

Let p: $V_{N,r,\varepsilon} \to \nu$ be the projection of $([A], x)$ to x, and $V'_{N,r,\varepsilon} = p^{-1}(\nu')$. If $r < r(\nu)$ and $\varepsilon < \varepsilon(U)$ the section s of $\mathscr{L}_\Sigma$ does not vanish on $V'_{N,r,\varepsilon}$ and the pair $(\mathscr{L}_\Sigma, s)$ defines an element of $H^2(V_{N,r,\varepsilon}, V'_{N,r,\varepsilon}) \cong$ line bundles over $V_{N,r,\varepsilon}$ trivialized over $V'_{N,r,\varepsilon}$.

(3.18) Proposition. *If $\varepsilon < \varepsilon(U, N)$ and $r < r(\nu, \varepsilon, N)$, then the class defined by $(\mathscr{L}_\Sigma, s)$ is the lift by p of the fundamental class in $H^2(\nu, \nu')$.*

Proof. We show that for these parameters the spaces $V_{N,r,\varepsilon/2} \subset V_{2N+3,r,\varepsilon}$ satisfy the hypothesis on $V_1 \cap V_2$ of Lemma (3.18). This uses a basic fact about gauge fixing of families. For any $\eta > 1$ and compact family S of gauge equivalence classes of connections over a standard annulus, sufficiently close to the flat connection θ, we can choose "representatives" $A_s \in \mathscr{A}/SU(2) \cong \Gamma_\theta$, continuously so that

$$\|A_s\|_{C^1} \leqslant \eta \cdot \inf_{g \in \mathscr{G}} \|g(A_s)\|_{C^1}.$$

The construction is a simple patching over the space S. Scaling the annulus to radius r this means that, using property (ii), if $([A], x)$ varies in a compact subset of $V_{N,r,\varepsilon/2}$ we can continuously choose equivalence classes of connections A^* over 2ν which extend $A|_{2\nu \setminus B_r(x)}$ and representable over B_r by small matrices:

$$\|A^*|_{B_r}\|_{C^0} \leqslant (2N+1) \cdot Cr, \qquad \|A^*|_{B_r}\|_{C^1} < (2N+1) \cdot C, \quad \text{say}$$

(to do this multiply in the chosen gauge by a standard bump function).

If r is small, relative to ε and N, $A^*|_{\Sigma_1}$ will still be within 2ε of a connection in U when restricted to Σ (in L^2 norm). Suppose x_t is a path in ν starting at $x_0 = x$ (all parametrized by a map f of a simplex) and use the Riemannian parallel transport to identify the balls encountered along x_t by maps x_t: $B_r(x_t) \to B_r(x_0)$. Similarly use the parallel transport defined by A^*—first along x_t and then radially in the balls, so that $A|_{B_r(x_t)}$ and $x_t^*(A^*|_{B_r(x_0)})$ may be regarded in a gauge invariant way as a pair of connections on the same bundle. Finally piece these together by a partition of unity to define a connection A_t equal to A^* outside $B_r(x_t)$. We can do this in such a way that A_t lies in $V_{2N+3,r,\varepsilon}$, and the procedure defines the path lifting of homotopies assumed in Lemma (3.17). The verification of fiber connectivity is similar, using condition (iii) that the relative Chern class is 1.

Thus Lemma (3.17) implies that the class defined by $(\mathscr{L}_\Sigma, s)$ is some multiple of $p^*[\nu, \nu']$. To check that this multiple is 1 use the Taubes construction to define a section τ: $\nu \to V_1$, exploiting the fact that $[\theta] \in U$. This extends to a map τ: $Y \to \mathscr{B}_Y \times Y$, where Y is the standard 2-sphere bundle over Σ containing ν as an open subset. We know by Lemma (3.8) that $\tau^*(c_1(\mathscr{L}_\Sigma)) \in H^2(Y)$ is the class dual to Σ. But since the section s does not vanish on $\tau(Y \setminus \nu)$ the corresponding relative class is equally the relative fundamental class of Σ; thus completing the proof.

Now let W be a neighborhood of $U \times (x_1, \cdots, x_l)$ in $\overline{M}_k$, as before, where U is precompact open in M_{k-l}. Suppose s is a section of a line bundle $\mathscr{L}_\Sigma$, as in (3.16), not vanishing on U. Abusing notation, we can extend the restrictions of U to a tubular neighborhood of Σ to a connected set containing the product connection and for codimension reasons we can suppose s does not vanish on this larger set. If none of the points x_α lie on Σ then it is clear that W can be chosen to avoid the zeros of S. If just one, say x_1, of the points lies on Σ (and has multiplicity 1) and $p\colon W \to \Sigma_1 \subset X$ is the local center, then for $\bar{\lambda}$ sufficiently small the section s defines a class in $H^2(W, W')$: $W' = p^{-1}(\Omega_1 \cap \nu')$.

(3.19) Corollary. *The class in $H^2(W, W')$ defined by s is $p^*[\Omega_1, \Omega_1 \cap \nu']$.*

This follows from the fact that by taking $\bar{\lambda}$ small we can map W into any given $V_{N,r,\varepsilon}$ by $A \to (A|_{2\nu}, p(A))$.

III(iv). Now we are able to explain the generalizations of the proof described above for the definite manifolds of Theorem A to Theorems B and C. Recall first that on a simply connected 4-manifold X the associated $SU(2)$ moduli spaces have "virtual" dimension given by the Atiyah-Hitchin-Singer formula

$$\dim M_k = 8k - 3\left(1 + b_2^+(X)\right), \qquad k \geqslant 0, \tag{3.20}$$

and, as mentioned above, for indefinite manifolds we can assume that M_k is a smooth manifold of this dimension, once $k > 0$.

Theorem B. $(b^+ = 1)$.

We use the 10-dimensional moduli space M_2. The natural compact closure $\overline{M}$ of this will involve contributions from M_1 and M_0, M_1 is 2-dimensional and since X is simply connected, $M_0 = \{\theta\}$ is a point.

Choose four surfaces $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ in general position in X and transverse sections s_i of the line bundles $\mathscr{L}_{\Sigma_i}$ as in Lemma (3.17).

Thus

$$N = M_2 \cap V_{\Sigma_1} \cap V_{\Sigma_2} \cap V_{\Sigma_3} \cap V_{\Sigma_4}$$

is a smooth manifold of dimension 2.

What are the ends of N? By dimension count every intersection

$$V_{\Sigma_i} \cap V_{\Sigma_j} \cap M_1 \qquad (i \neq j)$$

is empty. This means that no point $([A], x) \in M_1 \times X$ can be in the closure of N since x lies on at most two of the surfaces. Similarly N avoids the points $([A]; x, x)$ in $M_0 \times S^2(X)$. The ends of N are covered by open subsets

$$N_{(xy)} = N \cap W_{(xy)}$$

where $x \in \Sigma_i \cap 2\Sigma_j$, $y \in \Sigma_k \cap \Sigma_l$ ((i, j, k, l) a permutation of $(1,2,3,4)$), and $W_{x,y}$ is a neighborhood of $([\theta]; x, y)$ as in (3.15) with two local centers and scales. Clearly if the centers are constrained to small sets Ω_α, the number of these ends $N_{(x,y)}$ is equal to the number

$$Q_4([\Sigma_1], [\Sigma_2], [\Sigma_3], [\Sigma_4])$$

of the introduction (mod 2).

We shall show in §IV that the local center maps are of maximal rank.

Thus, by Proposition (3.18) we can take the preimages by the local center maps of the surfaces in X to represent the cohomology classes over $W_{(x,y)}$—we can assume that the $V_{\Sigma_i} \cap W_{(x,y)}$ are the connections whose local centers lie exactly on Σ_i. Then $N_{(x,y)}$ consists of connections whose two local centers are at x, y. In §§IV, V, VI we will prove that $N_{(x,y)}$ may be described as follows (see Corollary 5.6 and §V(i)).

(3.21) Proposition. *The surfaces Σ_i and the metric on X can be chosen so that each of the ends $N_{(x,y)}$ has the form $\mathbf{R}^+ \times L_{(x,y)}$, where the "link" $L_{(x,y)}$ is a circle. Moreover $u_1(L_{(x,y)}) = 1$ (cf. Definition 2.25).*

Given this proposition we straightaway deduce (β) of the introduction, hence Theorem B, by the pairing

$$0 = u_1(\partial \hat{N}) = Q_4(\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4) \mod 2,$$

where N is the obvious truncation of N defined using the description of (3.21).

Theorem C. ($b^+ = 2$).

This follows exactly the same pattern. Represent homology classes in X by surfaces $\Sigma_1, \cdots, \Sigma_6$ and argue with

$$N = M_3 \cap V_{\Sigma_1} \cdots V_{\Sigma_6}$$

a 3-dimensional manifold. The same transversality arguments show that we do not need to consider the spaces M_1, M_2 when describing the ends of N. In just the same way we reduce to a collection of open sets $N_{(xyz)} \subset N$ associated to points of intersection $x \in \Sigma_i \cap \Sigma_j$, $y \in \Sigma_k \cap \Sigma_l$, $z \in \Sigma_m \cap \Sigma_n$ and the number of these is $Q_6(\Sigma_1, \cdots, \Sigma_6) \mod 2$.

We shall prove:

(3.22) Proposition. *The surfaces Σ_i and the metric can be chosen so that every set N_{xyz} has the form $\mathbf{R}^+ \times L_{(xyz)}$, where the link $L_{(xyz)}$ is a 2-torus. Moreover $u_2(L_{(xyz)}) = 1$.*

Then we deduce Theorem C, via γ, by the pairing

$$0 = u_2(\partial \hat{N}) = Q_6(\Sigma_1, \cdots, \Sigma_6).$$

III(v). Excision and the index of families. As a first step towards the proofs of Propositions (3.21), (3.22) stated in the previous section we will now calculate the mod 2 cohomology classes u_i defined in Definitions (2.24), (2.25) on a space of parameters which will contain the "links" of the moduli spaces.

Suppose that, in the set-up of §III(i), a 4-manifold X is a connected sum $X = X_0 \# X_1$ and that $U_i \subset X_i$ are the complements of balls with $X = U_0 \cup U_1$. If A_0, A_1 are connections on bundles P_0, P_1 over X_0, X_1 which are flat on the complements of the U_i and $U_0 \cap U_1$, we can form the sum of the connections

$A_0\#_\rho A_1$, determined by a bundle gluing map ρ. Let the base point of X be in X_0 and form the family $A_{(\rho)}$ of based connections over X parametrized by $(\pm\rho) \in SU(2)/\pm 1$—a copy of $\mathbb{RP}^3$.

If X_0 and X_1 are spin manifolds, each of the individual connections A_i over X_i defines a Dirac operator with a numerical index:

$$m_i = \operatorname{ind}(D_{A_i}) \in \mathbb{Z}.$$

(3.23) Lemma. *The index of the family $D_{(\rho)}$ of operators over X parametrized by $\mathbb{RP}^3$ is $m_0 + m_1 \cdot \eta$, where $\eta \in KO(\mathbb{RP}^3)$ is the Hopf line bundle.*

As a special case of this we can take the trivial connected sum of manifolds $X \cong X\#S^4$ to obtain the Taubes construction gluing in the bundle with $c_2 = 1$ over S^4. More generally we can consider a family parametrized by l gluing maps—defining a point in $(\mathbb{RP}^3)^l$—associated to points $x_1, \cdots, x_l$ in X. These attach l copies of the basic (approximate) "instanton" to some fixed connection A_0 over X. We have:

(3.24) Lemma. *The index of such a multi-instanton family is $\eta_1 + \eta_2 + \cdots + \eta_l + m_0$, where $\eta_i \in KO((\mathbb{RP}^3)^l)$ is the Hopf line bundle corresponding to the ith factor.*

To motivate the proofs of these lemmas consider the case in Lemma (3.24) when the index m_0 is 0 and we form the multi-instanton family with very small scale sizes, so that the curvature is highly concentrated near the l points x_i. Then one can show that $\operatorname{Ker} D^*_{A_{(\rho)}}$ is 1-dimensional (and the cokernel is 0). The norm of the harmonic spinors making up $\operatorname{Ker} D^*_{A_{(\rho)}}$ is concentrated near the points x_i and to each of these points we can associate a 1-dimensional subspace of $\operatorname{Ker} D^*_{A_{(\rho)}}$—consisting of sections which are small near x_j $(j \neq i)$. Then Lemma (3.24) asserts that as ρ varies each of these 1-dimensional spaces forms the nontrivial bundle over the corresponding copy of $\mathbb{RP}^3$. Direct arguments on these lines have been made by Atiyah and Jones [8, V] and by Taubes [26], and in the companion article [14] such methods will be used to study the orientation class of Yang-Mills modili spaces. But for the two lemmas here we can work quite formally using the Atiyah-Singer "excision axiom" for indices of families. This means that direct analysis is avoided; that being built into the Atiyah-Singer axiom and their proof using the machinery of pseudo-differential operators.

Proof of Lemma (3.23). According to [9, I, IV, V] the assignment of the index to a family y of (real) elliptic operators factors through a linear map

$$\operatorname{ind}: KR(Y \times TZ) \to KO(Y)$$

using the symbol of the family. The index map obeys an excision property. If U is an open subset of X which is identified with a subset of another manifold X^*, then the family index of any element in $KR(Y \times TU)$ can be computed equally via the composites:

$$\begin{tikzcd} & KR(Y \times TX) \arrow[dr, "\mathrm{ind}"] & \\ KR(Y \times TU) \arrow[ur] \arrow[dr] & & KO(Y) \\ & KR(Y \times TX') \arrow[ur, "\mathrm{ind}"'] & \end{tikzcd}$$

(the KR-theory here has compact supports so pushes forward under open inclusions).

Consider first the family $D'_{(\rho)}$ of operators defined over the *disconnected* manifold $X_0 \amalg X_1$ in the same fashion as the $D_{(\rho)}$ family. So connections are fixed on bundles P_0, P_1; the bundle P_0 has a fixed base point and the parameter ρ is a choice of identification $(P_0)_{x_0} \cong (P_1)_{x_1}$. Hence restricting to X_1, this family of *based* connections is precisely one of the fibers of $\tilde{\mathscr{B}}_{X_1} \to \mathscr{B}_{X_1}$. The index of the Dirac family has a constant contribution m_0 from X_0 and a twisted contribution $m_1 \cdot \eta$ from X_1. This twisting comes about precisely because the element -1 in the center of the gauge group acts as (-1) on the bundle E hence on the bundle valued harmonic spinors, just as in Proposition (2.20).

Using the excision property we can next see that for a fixed ρ_0 the numerical index

$$\mathrm{ind}\, D_{(\rho_0)} \in \mathbb{Z}$$

is $m_0 + m_1$. This is done by first excising the connections to reduce to the case of trivial bundles and ordinary Dirac operators, then comparing with the model case of the 4-sphere $S^4 = S^4 \# S^4$, whose Dirac operator has index 0.

Finally consider the difference of the symbols σ_ρ, σ_{ρ_0} of $D_{(\rho)}$, $D_{(\rho_0)}$ (σ_{ρ_0} viewed as a constant family over $\mathbb{RP}^3$). Using the identification of the bundles over $U_0 \subset X_0$ we can lift $[\sigma_\rho - \sigma_{\rho_0}]$ back to $KR(\mathbb{RP}^3 \times TU_0)$ and obtain the same element there as we do from $[\sigma'_\rho - \sigma'_{\rho_0}]$. Hence

$$\begin{aligned} \mathrm{ind}\, D_{(\rho)} &= \mathrm{ind}\big(D_{(\rho)} - D_{(\rho_0)}\big) + \mathrm{ind}\, D_{(\rho_0)} \\ &= \mathrm{ind}\big(D'_{(\rho)} - D'_{(\rho_0)}\big) + \mathrm{ind}\, D'_{(\rho_0)} \\ &= \mathrm{ind}\, D'_{(\rho)} = m_0 + m_1\eta, \end{aligned}$$

as required.

The multi-parameter version (3.24) is proved in exactly the same way (or by repeated application of (3.23)) when we recall that the basic (anti) instanton on S^4 has $c_2 = 1$, hence spinor index 1.

When the connection A_0 is the trivial flat connection θ with centralizer subgroup $SU(2) \subset \mathcal{G}$ we obtain in this way a multi-instanton family of *based* connections, flat away from $x_1, \cdots, x_l$, parametrized by

$$\underbrace{\mathbb{RP}^3 \times \cdots \times \mathbb{RP}^3}_{l} \hookrightarrow \tilde{\mathscr{B}}_X^*.$$

Inside this family the fibration $\tilde{\mathscr{B}}^* \to \mathscr{B}^*$, forgetting the base point, is represented by the left action of $SO(3)$ on $\mathbb{RP}^3 \times \cdots \times \mathbb{RP}^3$. So the quotient may be represented by (say) that first $l-1$ $\mathbb{RP}^3$ factors. Denote by t_i $(i = 1, \cdots, l-1)$ the mod 2 class corresponding to the cohomology generator in the ith factor.

(3.25) Lemma. *In this representation the classes defined in* (2.24), (2.25) *pull back so that*

(i) $l = 2p$ *even*, $u_1 = t_1 + t_2 + \cdots + t_{2p-1}$.

(ii) $l = 2p+1$ *odd, the term* $t_1 t_2 \cdots t_{2p}$ *appears in* u_l *with coefficient* 1.

Proof. (i) Over the transversal to the $SO(3)$ action defined by fixing the last coordinate:

$$\begin{aligned} u_1 &= w_1(\operatorname{ind} D) = w_1(\eta_1 + \eta_2 + \cdots + \eta_{2p-1} + m_0 \cdot 1) \quad (\text{by } (3.24)) \\ &= t_1 + t_2 + \cdots + t_{2p-1}. \end{aligned}$$

(ii) Over the transversal:

$$\begin{aligned} u_i &= w_i((\operatorname{ind} D) \otimes (\det \operatorname{ind} D)) \\ &= w_i\big((\eta_1 + \eta_2 + \cdots + \eta_{2p} + m_0 + 1)\eta_1 \cdots \eta_{2p}\big). \end{aligned}$$

Now m_0 is even since it is the index of 2 copies of the basic Dirac operator on X, hence the total Stiefel-Whitney class is

$$\begin{aligned} &W\big((\eta_1 + \cdots + \eta_{2p} + m_0 + 1)(\eta_1 \cdots \eta_{2p})\big) \\ &\qquad = (1 + \Sigma)(1 + \Sigma + t_1)(1 + \Sigma + t_2)\big(1 + \Sigma + t_{2p}\big), \end{aligned}$$

where $\Sigma = \Sigma_1^{2p} t_i$. The term in $t_1 t_2 \cdots t_{2p}$ of this is the same as the corresponding term in

$$(\Sigma + t_1)(\Sigma + t_2) \cdots \big(\Sigma + t_{2p}\big)$$

which has a coefficient equal (mod 2) to det A, where

$$A = \begin{pmatrix} 0 & 1 & 1 & \cdots \\ 1 & 0 & 1 & \cdots \\ 1 & 1 & 0 & \cdots \end{pmatrix}.$$

But $(1 + A)^2 = 0 \bmod 2$, so $A^2 = 1$ and $\det A = 1 \bmod 2$.

IV

IV(i). Deformation of instantons. In this section we develop techniques to handle the concentrated "particle-like" connections, introduced onto general 4-manifolds by Taubes, which arise in a natural way when we define the compactifications of §III(iii). The goal is to produce a combination of the moduli description in the simplest case of [12], [17] with more general construction considered in the second paper of Taubes [25]. This will give a way to describe the structure of the sets W of §III(iii) covering the ends of Yang-Mills moduli spaces; in at least enough generality for the proofs of Propositions (3.21) and (3.22).

The strategy adopted in [12] was designed to throw the main burden of analysis onto Taubes' Implicit function theorem and the Atiyah-Singer index theorem. The first of these gave a way to construct solutions and to deform a solution in a family; the second could be used to show that this family was maximal and so gave a local description of the moduli space. The same strategy could be extended to the more general problems that we face here, but the method becomes excessively cumbersome. Thus we adopt here a different approach based upon an "alternating method" for solving P.D.E.'s. J. Roe has pointed out that this is quite like the method used by Schwarz to construct harmonic functions on Riemann surfaces [1]. The method has the advantage of minimizing real analytical problems at the expense of some complexity of organization. We alternate between solving equations over a 4-manifold X and equations over the model S^4, exploiting conformal invariance to pass back and forth. This has the further advantage that the discussion applies equally well to connections over general connected sums, and for greater clarity we will carry out the main part of the work in this setting, to be explored further in [15].

Of course it is not at all surprising that the results we obtain are just what one would be led to expect from those of Taubes, and there are probably many different ways to arranging the material. In particular there is work of Itoh [20] going in this direction.

We begin with a review of the standard deformation theory of an ASD connection A, modelled on the Kuranishi theory of deformations of complex structures (for more details see [21], [7], [12], [17]). Associated to A is the

Atiyah-Hitchin-Singer deformation complex

$$\Omega^0(\mathfrak{g}_P) \xrightarrow{d_A} \Omega^1(\mathfrak{g}_P) \xrightarrow{d_A^+} \Omega^2_+(\mathfrak{g}_P), \tag{4.1}$$

which is an elliptic complex whose finite-dimensional cohomology H^1_A, H^2_A may be identified with the harmonic subspaces

$$\begin{aligned} H^1_A &\cong \operatorname{Ker}\big((d_A^* \oplus d_A^+): \Omega^1 \to \Omega^0 \oplus \Omega^2_+\big), \\ H^2_A &\cong \operatorname{Ker}\big(d_A^*: \Omega^2_+ \to \Omega^1\big). \end{aligned} \tag{4.2}$$

The covariant constant sections H^0_A of $\mathfrak{g}_P$ make up the Lie algebra of the centralizer Γ_A.

For any small $\tilde{a}$ the nonlinear gauge fixing equation

$$d_A^* a = d_A^*(\exp(u)(A + \tilde{a}) - A) = 0 \tag{4.3}$$

has a small solution u. Thus the slice $\operatorname{Ker} d_A^* \subset \Omega^1(\mathfrak{g}_P)$ cuts every orbit of the gauge group near $[A] \in (\mathscr{A}/\mathscr{G})$. It is a genuine local transversal if $H^0_A = 0$; otherwise one must divide further by the action of Γ_A. The ASD equations for $A + a$ on the slice are elliptic.

Let V_A be any lifting of H^2_A to $\Omega^2_+(\mathfrak{g}_P)$. Using the inverse function theorem it is possible to define smooth maps on a neighborhood of 0 in H^1_A:

$$\begin{aligned} p \in H^1_A &\to \tilde{p} \in \operatorname{Ker} d_A^*, \\ p \in H^1_A &\to \phi(p) \in V_A, \end{aligned} \tag{4.4}$$

which are Γ_A-equivariant and solve the equation

$$F_+(A + \tilde{p}) = d_A^+\tilde{p} + [\tilde{p} \wedge \tilde{p}]_+ = \phi(p). \tag{4.5}$$

Conversely any small solution a of

$$d_A^* a = 0, \qquad F_+(A + a) \in V_A$$

corresponds to a point p of $H^1_{A_0}$. Thus a local model for the moduli space of gauge equivalence classes of ASD connections, near $[A_0]$, is given by dividing the zeros of the finite-dimension map ϕ by the action of Γ_{A_0}.

The map $p \to \tilde{p}$ is close to the identity:

$$|p - \tilde{p}| \leqslant \operatorname{const}|p|^2, \qquad \left|\frac{\partial \tilde{p}}{\partial p} - 1\right| \leqslant \operatorname{const}|p|. \tag{4.6}$$

Dual to the parametrization by $p \to \tilde{p}$ it is possible to find a local coordinate π mapping from a neighborhood of the origin in $\Omega^1(\mathfrak{g}_P)$ to $H^1_{A_0}$ such that $p \mapsto \pi \circ \tilde{p}$ defines a local diffeomorphism. One way of doing this is to take the L^2-inner product with a suitable subspace U_A of the sections of $\Omega^1(\mathfrak{g}_P)$, not meeting the perpendicular complement of H^1_A. By the unique continuation of these harmonic forms [2] U_A, V_A can be chosen to have supports in any given open set.

Finally, the "leading" term in this local Kuranishi model can be described explicitly. If $\omega \in H^2_A$ is any solution of the adjoint equation $d^*_A\omega = 0$, then from (4.5)

$$\begin{aligned}\langle \omega, \phi(p)\rangle &= \langle \omega, d^+_A p + [\tilde{p} \wedge \tilde{p}]_+\rangle = \langle \omega, [\tilde{p} \wedge \tilde{p}]_+\rangle \\ &= \langle \omega, [p \wedge p]_+\rangle + O|p|^3,\end{aligned}$$

using (4.6). For small p the map $p \to \langle \omega, \phi(p)\rangle$ is approximated in C^1 by the quadratic term $q_\omega(p) = \langle \omega, [p \wedge p]_+\rangle$ and it follows that if zero is a regular value of the total map

$$q \colon H^1_{A_0}\setminus\{0\} \to \left(H^2_{A_0}\right)^*,$$

then the moduli space has an isolated singularity at $[A_0]$ with a neighborhood modelled on the zeros of q (having dividing by Γ_{A_0}).

We will show that there is a similar theory for describing the deformations of an idealized ASD connection associated to one of the other strata in the compactification $\overline{M}_k$ of §III(iii), at least away from the diagonals in the symmetric products. If W is a neighborhood of the point $([A]; x_1, \cdots, x_l)$ as in (3.15), with all the x_α distinct and l local center and scale maps, then

$$\bar{\lambda} = \sup_{1 \leqslant \alpha \leqslant l} \lambda_\alpha$$

is a measure of the size of W. Since we are only interested in a neighborhood system we can take $\bar{\lambda}$ small—all our results will apply under this assumption.

IV(ii). We now pass rapidly over the arguments needed to place an ASD connection A representing a point in W, with l local concentrations of curvature, in certain (nonunique) standard gauges. Then we can set up the problem of describing the moduli in a form which applies to general connected sums.

The same estimate of curvature decay used in ([12, Theorem 16], [17, §9]) implies that, for any given $\delta > 0$, there are small r-neighborhoods of the centers x_i in which the curvature is bounded by

$$|F_A(y)| \leqslant \text{const}\, \lambda_i^{-2}\left(\frac{\lambda_i}{d(y, x_i)}\right)^{4-\delta} + c$$

for $d(y, x_i) > \lambda_i^{3/4}$ (r, C independent of λ_i).

This means that for a $K > 0$ it is possible to construct an "exponential gauge"—just as in [17, Proposition 9.38]—over the intersection of such a neighborhood with $X \setminus B(x_i, K\sqrt{\lambda_i})$, representing the connection by matrices satisfying

$$
\begin{aligned}
|A(y)| &\leqslant \operatorname{const}\left(\lambda_i^{((1-\delta)/2)} + \int_{K\sqrt{\lambda_i}}^{d(y,x_i)} \left[\left(\frac{\lambda_i^{2-\delta}}{r^{4-\delta}}\right) + C\right] dr\right) \\
&\leqslant \operatorname{const}\left(\lambda_i^{((1-\delta)/2)} + Cr\right).
\end{aligned} \tag{4.7}
$$

So if $\delta < 1$ and r, $\bar{\lambda}$ are small, these connection matrices may be supposed uniformly small. Over the complement of fixed $2r$-neighborhoods of the x_i, A can be put in a gauge such that it becomes as close as we please to a connection A_0 representing a point in $U \subset M_{k-l}$. Patching the two gauges together as in [17, Proposition 9.33] gives:

(4.8) Proposition. *For any $K_1, \eta_1 > 0$, if $\bar{\lambda}$ is sufficiently small it is possible to choose representatives of the gauge equivalence classes $[A]$, $[A_0]$ over $X \setminus \bigcup_1^l B(x_i, K_1\sqrt{\lambda_i})$ such that: $A = A_0 + \tilde{a}$, $\|\tilde{a}\|_{L^\infty(X \setminus \bigcup B(x_i, K_1\sqrt{\lambda_i}))} < \eta_1$.*

The "gauge fixing" equation (4.3) is in divergence form. The solutions may be regarded as the zeros of a smooth map of the Banach spaces

$$
\begin{aligned}
&L_1^{2p} \times L^{2p} \to \left(\operatorname{Im} d_{A_0}^*\right) \subset L_{-1}^{2p}, \\
&(u, \tilde{a}) \to d_{A_0}^*(\exp(u)(A_0 + \tilde{a}) - A_0)
\end{aligned} \tag{4.9}
$$

for any fixed $p > 2$ (cf. [28, Theorem 2.5]). The derivative of this map at $(0,0)$ is

$$
(u, \tilde{a}) \to d_{A_0}^* \tilde{a} - \Delta_{A_0} u \tag{4.10}
$$

and the implicit function theorem assures the existence of a small solution u for any small $\tilde{a}$.

The difference element $\tilde{a}$ of Proposition (4.9) is initially defined over the complement of the balls $B(x_i, k_1\sqrt{\lambda_i})$, but in the space L^{2p} it may equally well be extended by zero over all of X. Choosing η_1 small enough for the implicit function theorem to operator over X, then restricting back to the complement of the $B(x_i, K_1\sqrt{\lambda_i})$, we get a connection $A_0 + a$, gauge equivalent to A with $d_{A_0}^* a = 0$ and $\|a\|_{L^{2p}} \leqslant \operatorname{const} \eta_1$.

The choice of the particular function space L^{2p} which we use throughout is not very important. Any fixed $p > 2$ would do but it is more convenient to take p large and we will suppose $p > 6$. The difference element a is the solution of an elliptic system, it is smooth over $X \setminus \bigcup B(x_i, K_1\sqrt{\lambda_i})$ and once the L^{2p} norm is small all other norms are essentially equivalent. Note that the "gauge fixing" above will *not* in fact fix a unique representation for the connection over $X \setminus \bigcup B(x_i, K_1\sqrt{\lambda_i})$; setting up the equation over X has the effect of imposing certain nonlocal boundary conditions.

For each center of concentration x_i the local scale λ_i can be used to define an "approximately conformal" map from a small ball $B(x_i, K_2\sqrt{\lambda_i})$ to the copy of the standard round 4-sphere $S_{x_i} = (TX)_{x_i} \cup \{\infty\}$, just as in [12, §§III.3, III.4]. The image of S_{x_i} is the complement of a ball centered at ∞ of radius $O(\sqrt{\lambda_i}/K_2)$. Over compact subsets of $S_{x_i} \setminus \{\infty\}$ the rescaled connections so obtained can be made arbitrarily close to the basic (anti) instanton connection $I^{(i)}$ on the negative spin bundle over S_{x_i}. Arguing over these 4-spheres exactly as over the original 4-manifold X above, shows that the rescaled connections can be placed in standard gauges. To summarize:

(4.11) Proposition. *Let $p > 2$ and let $K_1, K_2, \eta_2 > 0$ be fixed. Any gauge equivalence class $[A] \in W$ with $\bar{\lambda}$ sufficiently small may be expressed in the form of data*

$$\left(A_0 + a, I^{(i)} + a'_{(i)}, g_{(i)}\right),$$

where A_0 represents a connection in U and

(i) *a is defined over $X \setminus \bigcup B(x_i, K_1\sqrt{\eta_i})$ and satisfies $d^*_{A_0}a = 0$, $\|a\|_{L^{2p}(X)} \leqslant \eta_2$.*

(ii) *Each $a'_{(i)}$ is defined over the complement of a ball $B(\infty, K_2\sqrt{\lambda_i}) \subset S_{x_i}$ and satisfies*

$$d^*_{I^{(i)}}a'_{(i)} = 0, \ \left\|a'_{(i)}\right\|_{L^{2p}(S_{x_i})} < \eta_2.$$

(iii) *The transition function $g_{(i)}$ is a map from the annular region of common definition of the connections to the copy $\mathrm{Hom}((P_{(k-l)})_{x_i}, (W_+)_{X,x_i})$ of $SU(2)$. (Here we use the standard radial gauge fixing to spread frames for the bundles over small open sets, so in local trivializations:*

$$I + a'_{(i)} = -dg_{(i)}g_{(i)}^{-1} + g_{(i)}(A_0 + a)g_{(i)}^{-1}.)$$

In Proposition (4.11) the 4-spheres carrying the rescaled connections and the original 4-manifold X appear symmetrically. We will now make a long digression to analyze ASD connections in similar standard gauges over general connected sums. We start by imposing simplifying but restrictive hypotheses but it will be easy, in §IV(vi), to modify these so that the analysis applies to the connections of (4.11).

Suppose that X_0, X_1 are oriented, compact 4-manifolds with Riemannian metrics which are *flat* in fixed small balls and that we use an identification of these balls to define a "conformal connected sum" $X_0 \# X_1$. Thus if ξ is a local oriented Euclidean coordinate centered on x_0 in X_0 and η on x_1 in X_1 the identification map on an annular region is by the conformal equivalence

$$\eta = \lambda\bar{\xi}/|\xi|^2. \tag{4.12}$$

Here $\xi \to \bar{\xi}$ is any reflection and λ is a parameter which will eventually be made small. More invariantly the data needed to define such a sum is a pair (λ, σ), where $\lambda > 0$ and σ is an orientation *reversing* isometry $(TX_0)_{x_0} \xrightarrow{\sigma} (TX_1)_{x_1}$, analogous to the ρ of §III(i). Since the identification map is conformal the connected sum can be given a metric compatible with the conformal structures of the X_i.

Define a pair of thin "shells" R_1, R_{-1} in the manifold X_0, as in Diagram (4.13). The radii defining the walls are $k^{-1}N\sqrt{\lambda}$, $N\sqrt{\lambda}$, $N^{-1}\sqrt{\lambda}$, $kN^{-1}\sqrt{\lambda}$. Here the factor k defining the shell thickness is fixed at any convenient value—say $k = 0.9$—while the factor N defining the separation of the shells has to be chosen reasonably large for the proof to work; as we shall see any value of N

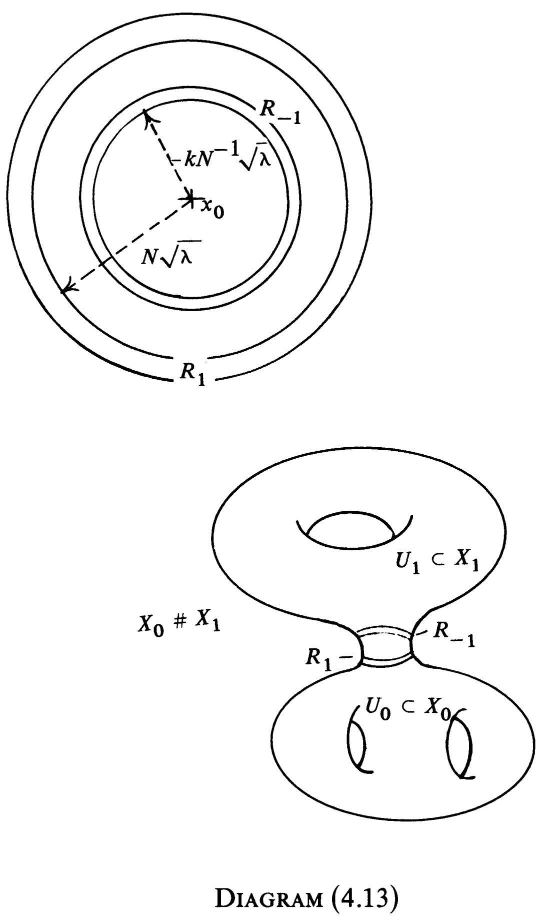

DIAGRAM (4.13)

with

$$K_N \equiv \left(\frac{k^{-4}-1}{8(1-k)}\right)\frac{N}{(N-N^{-1})^3} < 1 \tag{4.14}$$

will do.

The configuration (4.13) is preserved by the identifying map (4.12) so the shells R_1, R_{-1} in X_0 become identified with corresponding shells in X_1 *whose sizes are reserved*. For simplicity we use the same names R_1, R_{-1} for the shells in each manifold; R_1 the *outer* shell in X_0 and the *inner* shell in X_1. Let $U_0 \subset X_0$ be the complement of the ball $|\xi| \leqslant k\sqrt{\lambda}/N$ and $\hat{U}_0 = U_0 \setminus R_{-1}$ the complement of $|\xi| \leqslant \sqrt{\lambda}/N$. Define U_1, $\hat{U}_1 \subset X_1$ symmetrically, so that $X_0 \# X_1$ is covered by the pair of open sets $\hat{U}_0$, $\hat{U}_1$, overlapping in an annular region bordered by the thin shells.

Suppose that A_0, A_1 are ASD connections on $SU(2)$ bundles P_0, P_1 over X_0, X_1 satisfying the condition

(4.15) Hypothesis. $H^1_{A_i} = H^1_{A_i} = H^2_{A_i} = 0$.

For any $\eta_3 > 0$ consider the set of connections over $X_0 \# X_1$ representable in the standard form analogous to (4.12)

$$A \cong (A_0 + a,\ A_1 + a',\ g),$$

$$\begin{cases} a \text{ defined over } \hat{U}_0;\ d^*_{A_0}a = 0,\ \|a\|_{L^{2p}(X_0)} < \eta_3, \\ a' \text{ defined over } \hat{U}_1;\ d^*_{A_1}a' = 0,\ \|a'\|_{L^{2p}(X_1)} < \eta_3, \\ g\colon \hat{U}_0 \cap \hat{U}_1 \to \operatorname{Hom}((P_0)_{x_0}, (P_1)_{x_1}). \end{cases} \tag{4.16}$$

(So $A_1 + a' = -dg\,g^{-1} + g(A_0 + a)g^{-1}$ in local trivializations.)

(4.17) Theorem. *If A_0, A_1 satisfy Hypothesis* (4.15) *and if η_3 and λ are sufficiently small, the gauge equivalence classes of ASD connections over $X_0 \# X_1$ which can be represented in standard form* (4.16) *are smoothly parametrized by a copy of* $\operatorname{Hom}((P_0)_{x_0}, (P_1)_{x_1})/(\pm 1) \cong SO(3)$.

The next three sections make up the proof of (4.17). The proof will, of course, give more: this family of ASD connections is close to one of the families discussed in §III(i) parametrized by the identification map on the overlap.

IV(iii). Shifting the support of an error term. Fix $p > 6$ as in §IV(ii) and regard the map on sections of bundles over X_0,

$$\begin{aligned} &\Omega^2(\mathfrak{g}_{P_0}) \times \Omega^1(\mathfrak{g}_{P_0}) \to (\Omega^0 \oplus \Omega^2_+)(\mathfrak{g}_{P_0}), \\ &\qquad (b, a) \to (d^* \oplus d^+)_{A_0} b + [b, a]_+ + [b \wedge b]_+, \end{aligned} \tag{4.18}$$

as a smooth map of the Banach spaces $L_1^p \times L^{2p} \to L^p$. The partial derivative in the b-factor at $(0,0)$ is defined by the elliptic operator $(d^* \oplus d^+)_{A_0}$ which is supposed invertible by (4.15). So if a is small in L^{2p} and $\sigma \in \Omega^2_+(\mathfrak{g}_{P_0})$ is small in L^p, there is a small solution b to

$$d^*_{A_0} b = 0, \qquad d^+_{A_0} b + [b,a]_+ + [b \wedge b]_+ = -\sigma. \tag{4.19}$$

We will apply this when a is initially defined (and smooth) over $U_0 \subset X_0$ and such that the self-dual curvature $\sigma = F_+(A_0 + a)$ is supported in the outer shell $R_1 \subset X_0$. Put $\delta = \|F_+(A_0 + a)\|_{L^\infty(X_0)}$, so

$$\|F_+(A_0 + a)\|_{L^p(X_0)} \leqslant \text{const}\, \delta \cdot \lambda^{2/p}.$$

For any $h_1 > 1$ we may find a cut-off function ψ_{-1} with $d\psi_{-1}$ supported in the inner shell R_{-1}, $\operatorname{supp}\psi_{-1} \subset U_0$, $\psi_{-1} = 1$ on $\hat{U}_0$, such that

$$|d\psi_{-1}| \leqslant h_1\left(\frac{N}{(1-k)\cdot\sqrt{\lambda}}\right). \tag{4.20}$$

If δ is not large and λ is small, we extend a by 0 to X_0 and solve (4.19) for b. Then let τ be the self-dual curvature of the smooth connection $A_0 + a + \psi_{-1} \cdot b$ over X_0:

$$\begin{aligned} \tau &= F_+(A_0 + a + \psi_{-1}\cdot b) \\ &= F_+(A_0 + a) + d_{A_0}(\psi_{-1} b) + [a, \psi_{-1} b]_+ + \psi_{-1}^2 (b \wedge b)_+. \end{aligned}$$

This is supported in the shell R_{-1}, where

$$\begin{aligned} \tau &= d(\psi_{-1} \wedge b)_+ + \psi_{-1}\big(d_{A_0} + b + [a,b]_+ + \psi_{-1}(b \wedge b)_+\big) \\ &= (d\psi_{-1} \wedge b)_+ + \psi_{-1}(1 - \psi_{-1})(b \wedge b)_+. \end{aligned}$$

So

$$\begin{aligned} \|\tau\|_{L^\infty(X_0)} &\leqslant \frac{1}{\sqrt{2}}\left(\|d\psi_{-1}\|_{L^\infty}\big\|b|_{R_{-1}}\big\|_{L^\infty} + \big\|b|_{R_{-1}}\big\|^2_{L^\infty}\right) \\ &\leqslant \frac{h_1 N}{\sqrt{2}(1-k)\sqrt{\lambda}}\big\|b|_{R_{-1}}\big\|_{L^\infty} + \frac{1}{\sqrt{2}}\big\|b|_{R_{-1}}\big\|^2_{L^\infty}. \end{aligned} \tag{4.21}$$

Restricting to U_0, the effect of the procedure is to replace a connection with self-dual curvature σ supported in R_1 by another having self-dual curvature τ supported in R_{-1}.

(4.22) Lemma. *For any $h_2 > 1$ and $K_3 > 0$ we can find $\eta_4 > 0$ such that if $\delta < K_3$, $\|a\|_{L^{2p}} < \eta_4$, and λ is sufficiently small, then*:

(i) $\|\tau\|_{L^\infty(R_{-1})} \leqslant (h_2 K_N)\cdot N^4 \cdot \delta$.

(ii) $\|b\|_{L^{2p}(X_0)} \leqslant c_1 \cdot \lambda^{(p+2/2p)} \cdot \delta$ (*where c_1 depends on A_0, k, N, η_4*).

Proof. The more delicate estimate (i) will be derived from the corresponding linear elliptic equation.

Suppose $\tilde{b}$ is the solution to

$$d^*_{A_0}\tilde{b} = 0, \qquad d^+_{A_0}\tilde{b} = -\sigma.$$

Then $\tilde{b}$ is expressed in terms of σ and the kernel $L(x, y)$ of the operator $(d^* \oplus d^-)^{-1}_{A_0}$ over the compact manifold X_0:

$$\tilde{b}(x) = \int_{X_0} L(x, y)\sigma(y)\, d\mu_y.$$

So if $x \in R_{-1}$,

$$|\tilde{b}(x)| \leqslant \max_{y \in R_1} |L(x, y)| \mathrm{vol}(R_1) \cdot \|\sigma\|_{L^\infty}, \tag{4.23}$$

where $|L(x, y)|$ denotes the "operator" norm of the map $L(x, y)$: $(\Lambda^2_+)_y \to (\Lambda^1)_x$. The sup norm of σ is δ and

$$\mathrm{vol}(R_1) = \omega_3 \cdot \left(\frac{k^{-4} - 1}{4}\right) \cdot N^4 \cdot \lambda^2,$$

where ω_3 is the volume of the unit 3-sphere. On the other hand by making λ small the singular kernel $L(x, y)$ can be made arbitrarily close to the corresponding kernel L_0 over Euclidean space: $|L(x, y)|/|L_0(x, y)| \leqslant \sqrt{h_2}$, say, for $x \in R_{-1}$, $y \in R_1$.

Identifying the kernel L_0 is a matter of linear algebra. If G is the flat-space Greens function

$$G(x, y) = \frac{1}{2\omega_3} \cdot \frac{1}{|x - y|^2}, \quad x, y \in \mathbb{R}^4,$$

then $G \cdot \mathbb{1} = (\frac{1}{2})(d^+ d^*)^{-1}$ on Ω^2_+. (Here $\mathbb{1}$ denotes the flat space parallel transport of 2-forms and the factor $\frac{1}{2}$ appears because in Euclidean coordinates

$$d^+ d^* = -\tfrac{1}{2} \sum (\partial/\partial x_i)^2$$

on Ω^2_+.) Hence

$$\begin{aligned} L_0(0, y)(\Phi) &= \frac{1}{\omega_3} d^*_y\left(\mathbb{1}\phi \cdot \frac{1}{|y|^2}\right) \quad (\text{for } \Phi \in \Lambda^2_+) \\ &= \frac{1}{\omega_3} * \left(\frac{dr \wedge \mathbb{1}\Phi}{r^3}\right). \end{aligned}$$

so $|L_0(x, y)| = 1/\sqrt{2}\,\omega_3|x - y|^3$. Now $d(R_1, R_{-1}) = (N - N^{-1}) \cdot \sqrt{\lambda}$, hence

$$|\tilde{b}(x)| \leqslant \frac{\sqrt{h_2}}{\sqrt{2}}\left(\frac{k^{-4} - 1}{4}\right)\frac{N^4}{(N - N^{-1})^3}\sqrt{\lambda} \cdot \delta,$$

so, from (4.21),

$$\|\tau\|_{L^\infty} \leqslant h_1\sqrt{h_2}\left(\frac{k^{-4}-1}{8(1-k)}\right)\frac{N^5}{(N-N^{-1})^3}\cdot\delta$$
$$= h_1\sqrt{h_2}K_N\cdot N^4\cdot\delta.$$

To incorporate the nonlinearity in (4.19) put $b = \tilde{b} + \beta$ so that

$$(d^* + d^+)_{A_0}\beta = -([a,b]_+ + (b\wedge b)_+).$$

We only need a rough estimate of β: for example, we have

$$\begin{aligned}\|\beta\|_{L^\infty(X_0)} &\leqslant \text{const}\left\|(d^*\oplus d^+)_{A_0}\beta\right\|_{L^6}\\ &\leqslant \text{const}\|[a,b]_+ + (b\wedge b)_+\|_{L^6}\\ &\leqslant \text{const}\left(\|a\|_{L^{12}}\|b\|_{L^{12}} + \|b\|_{L^{12}}^2\right)\\ &\leqslant \text{const}\left(\eta_4\|b\|_{L_1^3} + \|b\|_{L_1^3}^2\right)\end{aligned}$$

by the Sobolev embedding $L_1^3 \to L^{12}$. Then using (4.19) we can certainly get

$$\|b\|_{L_1^3} \leqslant \text{const}\|\sigma\|_{L^3} \leqslant \text{const}\,\delta\cdot\lambda^{2/3}$$

for small λ, η_4. Then

$$\|\beta\|_{L^\infty} \leqslant \text{const}\,\delta\cdot\lambda^{2/3}.$$

Choose h_1 close to 1 and then absorb the nonlinear term $\|\beta\|$ into the leading term (since $\lambda^{2/3} \ll \sqrt{\lambda}$) to get (i). Similarly (ii) comes from:

$$\|b\|_{L^{2p}} \leqslant \text{const}\,\|b\|_{L_1^{4p/p+2}} \leqslant \text{const}\,\|\sigma\|^{4p/p+2} \leqslant \text{const}\lambda^{(p+2/2p)}\delta.$$

(4.24) Remarks. (i) If, in (4.19), the sections σ, a depend smoothly on a parameter t, the solution b will also, and

$$d_{A_0}^+\left(\frac{\partial b}{\partial t}\right) + \left[\frac{\partial a}{\partial t}, b\right]_+ + \left[a, \frac{\partial b}{\partial t}\right]_+ + \left(b\wedge\frac{\partial b}{\partial t} + \frac{\partial b}{\partial t}\wedge b\right)_+ = -\frac{\partial\sigma}{\partial t}.$$

Arguing with this differentiated expression just as before gives, for small enough λ, η_4,

$$\left\|\frac{\partial\tau}{\partial t}\right\|_{L^\infty} \leqslant \left((h_2K_N)N^4\cdot\left\|\frac{\partial\sigma}{\partial t}\right\|_{L^\infty} + c_2\lambda^{p+2/2p}\left\|\frac{\partial a}{\partial t}\right\|_{L^{2p}}\right),$$
$$\left\|\frac{\partial b}{\partial t}\right\|_{L^{2p}} \leqslant c_3\lambda^{p+2/2p}\left(\left\|\frac{\partial\sigma}{\partial t}\right\|_{L^\infty} + \left\|\frac{\partial a}{\partial t}\right\|_{L^{2p}}\right).$$

(ii) If M is a fixed number $< N$, then by arguing exactly as in the proof of Lemma (4.22) we get a bound on the restriction of b to the ball $B_M = B(x_0, M\sqrt{\lambda}) \subset X_0$:

$$\left\|b|_{B_M}\right\|_{L^\infty} \leqslant c(M)\cdot\sqrt{\lambda}.$$

IV(iv). Constructing solutions by the alternating method. If the parameter λ, defining the conformal structure of the connected sum, is sufficiently small, then Lemma (4.22) can be used repeatedly to construct a family of ASD connections $A^{(\infty)}(\rho)$ over $X_0 \# X_1$ parametrized by $\mathrm{Hom}((P_0)_{x_0}, (P_1)_{x_1})$.

We use an iterative cycle to improve an initial "approximate" solution to the equations (compare [24], [17, §7]). Suppose that at the nth stage we have defined a connection $A^{(n)}(\rho)$ for every $\rho \in \mathrm{Hom}((P_0)_{x_0}, (P_1)_{x_1})$, represented in the notation of (4.16) as

$$A^{(n)} = \left(A_0 + a_n, A_1 + a_n', \rho\right),$$

where we suppose a_n, a_n' defined over U_0, U_1 respectively, but *not* necessarily satisfying the gauge fixing equations $d_{A_0}^* a = 0$, $d_{A_1}^* a' = 0$ nor the ASD equations. ρ is the "constant" transition function in the sense of (4.16)(iii). Assume also that the following inductive hypotheses hold for fixed K_3 and h_2 such that $h_2 K_N \leqslant 1$.

(4.25) (i) The self-dual curvature of $A^{(n)}$, $\sigma_n = F_+(A^{(n)}$, is supported in $R_{(-1)^n}$ with uniform norm:

$$\delta_n = \|\sigma_n\|_{L^\infty(X_0)} \quad (n \text{ even})$$
$$= \|\sigma_n\|_{L^\infty(X_1)} \quad (n \text{ odd})$$

and $\delta_n \leqslant K_3$.

(ii) $\|a_n\|_{L^{2p}(X_0)}$ and $\|a_n'\|_{L^{2p}(X_1)}$ are less than η_4, where η_4 is the constant of Lemma (4.22), which we suppose also to give the symmetrical result over X_1.

If n is even and λ is small, pass to the $(n+1)$th stage by using Lemma (4.22) over X_0 to replace a_n by

$$a_{n+1} = a_n + \psi_{-1} b,$$

a modification supported in U_0. Define a_{n+1}' by modifying a_n' over $U_0 \cap U_1$ to preserve compatibility under the fixed transition function defined by ρ. If n is odd do the same with the roles of the manifolds X_0, X_1 reversed.

The inductive hypothesis (4.25) means that we can use (4.22) to estimate the change in norm of the "error term" σ_n. If n is even, then

$$\|\sigma_{n+1}\|_{L^\infty(X_0)} \leqslant (h_2 K_N) N^4 \cdot \delta_n,$$

$$\delta_{n+1} = \|\sigma_{n+1}\|_{L^\infty(X_1)} \leqslant N^{-4} \|\sigma_{n+1}\|_{L^\infty(X_0)},$$

since the conformal factor of the identifying map (4.12) exceeds N^2 on $R_{-1} = \mathrm{supp}(\sigma_{n+1})$, and σ_{n+1} is a 2-form. So

$$\delta_{n+1} \leqslant (h_2 K_N) \delta_n. \tag{4.26}$$

The calculation is the same for n odd. Thus if λ is sufficiently small and the inductive hypothesis (4.25) holds at all stages up to the nth, we get

$$\delta_n \leqslant (h_2 K_N)^n \delta_0 \tag{4.27}$$

—an improvement since $h_2 K_N < 1$.

The iteration can be started by defining a connection

$$A^{(0)}(\rho) = \left(A_0 + a_0,\ A_1 + a_0',\ \rho\right)$$

"cutting off" A_0, A_1 by a function ψ_1, symmetrical with ψ_{-1}, in the exponential gauges which, together with ρ, identify the bundles P_0, P_1 over $U_0 \cap U_1$:

$$\begin{aligned} a_0 &= (\psi_1 - 1) \cdot A_0 + (1 - \psi_1) \cdot \rho^{-1} A_1 \rho, \\ a_0' &= \psi_1 \cdot \rho A_0 \rho^{-1} - \psi_1 \cdot A_1. \end{aligned}$$

The connection $A^{(0)}$ has self-dual curvature supported in R_1 and, since the connection matrices of A_0, A_1 are $O(\sqrt{\lambda})$ on the overlap,

$$\begin{aligned} &\|a_0\|_{L^\infty(X_0)},\ \|a_0'\|_{L^\infty(X_0')} \leqslant c_4(A_0, A_1)\sqrt{\lambda}, \\ &\|F_+(A^{(0)})\|_{L^\infty(X_0)} \leqslant c_5(A_0, A_1). \end{aligned} \tag{4.28}$$

Take $K_3 = c_5(A_0, A_1)$. Once λ is small the inductive hypotheses are satisfied by $A^0(\rho)$ and the geometric decay (4.26) means that at no subsequent stage will they be violated through the failure of (4.25)(i).

Similarly, if (4.25) holds at stage n, then by Lemma (4.22)(ii)

$$\|a_{n+1} - a_n\|_{L^{2p}(X_0)} \leqslant c_1 \lambda^{(p+2/2p)} \delta_n. \tag{4.29}$$

Since the identifying map (4.12) has a bounded distortion factor on $U_0 \cap U_1$, independent of λ, the L^{2p} norms of X_0, X_1 compare uniformly on the overlapping region and we get a corresponding bound

$$\|a_{n+1}' - a_n'\|_{L^{2p}(X_1)} \leqslant c_6 \lambda^{(p+2/2p)} \delta_n. \tag{4.29$'$}$$

Thus, given K_3, if we take λ so small that

$$\begin{aligned} &c_4 \lambda^{p+2/2p} + (c_1 + c_6)\lambda^{p+2/2p}\left(\sum \delta_n\right) \\ &\qquad \leqslant c_4\sqrt{\lambda} + (c_1 + c_6)\lambda^{p+2/2p} \sum_{n=0}^{\infty} (h_2 K_N)^n \cdot K_3 \end{aligned}$$

is less than η_4, then using (4.28) the iteration continues indefinitely and it follows from (4.29) and (4.29)$'$ that a_n, a_n' converge to define an ASD connection

$$A^{(\infty)}(\rho) = \left(A_0 + a_\infty,\ A_1 + a_\infty',\ \rho\right).$$

Moreover (4.29) and (4.29)′ give

$$\|a_\infty\|_{L^{2p'}}\|a'_\infty\|_{L^{2p}} \leqslant c_7 \cdot \lambda^{(p+2/2p)} \quad \text{say.} \tag{4.30}$$

This effectively completes the constructive part of Theorem (4.17), and the family of ASD solutions here is completely analogous to those found, in the case when $X_1 = S^4$, by Taubes. It remains to prove that the solutions are distinct, can be put into standard form, and that any other ASD solutions in standard form are gauge equivalent to one of this family.

(4.31) Lemma. *Under the hypothesis (4.15) two of the ASD connections $A^{(\infty)}(\rho)$, $A^{(\infty)}(\bar\rho)$ are gauge equivalent if and only if $\bar\rho = \pm\rho$.*

Proof. On the one hand, it is clear that $A^\infty(\rho)$, $A^\infty(-\rho)$ are gauge equivalent, indeed we could work throughout with the associated $SO(3)$ bundles where the center ± 1 disappears.

Conversely if we apply the bound on derivatives with respect to parameters (4.24)(i) (with t parametrizing a path in $\mathrm{Hom}((P_0)_{x_0}, (P_1)_{x_1})$) at each stage of the iteration, then we deduce, just as above, that

$$\begin{aligned}\|a_\infty(\rho) - a_\infty(\bar\rho)\|_{L^{2p}(X_0)} &+ \|a'(\rho) - a'_\infty(\bar\rho)\|_{L^{2p}(X_1)} \\ &\leqslant \mathrm{const}\,\lambda^{(p+2/2p)}|p - \bar p|.\end{aligned} \tag{4.32}$$

On any compact 4-manifold, possibly with boundary, the Sobolev embedding $L_1^{2p} \to C^0$ is a compact map. It follows from this that if A is any connection over the "manifold with a hole" $X_0 \setminus B(X_0, kN^{-1}\sqrt{\lambda})$, there is a constant c_λ such that, for any section u of the Lie algebra bundle

$$\|u\|_{C^0} \leqslant c_\lambda\big(\|d_A u\|_{L^{2p}} + |\pi_x(u)|\big),$$

where $\pi_x(u)$ is the projection of the value of u in the fiber over some fixed point x to the vector space $(H_A^0)_x$ of covariant constant sections. It is a simple exercise to show that if the connection A extends over all of X_0 the constants c_λ can be made independent of λ (cf. the proof of Lemma (4.49) below). Similarly, by arguing as in [12], we get the corresponding statement for gauge transformations. If A is a connection over X, there is a constant $c(A)$ such that any bundle automorphism g defined over the complement of $B(x_0, kN^{-1}\sqrt{\lambda})$ factors as $g = \gamma \cdot \tilde g$ with $\gamma \in \Gamma_A$ and

$$\|\tilde g - 1\|_{C^0} \leqslant c(A)\|d_A g\|_{L^{2p}}.$$

Suppose there is a gauge transformation between $A^{(\infty)}(\rho)$ and $A^{(\infty)}(\bar\rho)$ represented by g_0 over U_0 and g_1 over U_1, compatible on the overlap. Thus

$$d_{A_0} g_0 = g_0 a_\infty(\rho) - a_\infty(\bar\rho) g_0^{-1} \quad \text{over } U_0,$$

$$d_{A_1} g_1 = g_1 a'_\infty(\rho) - a'_\infty S(\bar\rho) g_1^{-1} \quad \text{over } U_1.$$

Composing with $\{\pm 1\} = \Gamma_{A_0} = \Gamma_{A_1}$ if necessary we may without loss of generality suppose, using (4.30) and (4.32),

$$\begin{aligned}
&\|g_0 - 1\|_{C^0(U_0)} + \|g_1 - 1\|_{C^0(U_1)} \\
&\quad \leqslant \big(c(A_0) + c(A_1)\big)\big(\|d_{A_0}g_0\|_{L^{2p}} + \|d_{A_1}g_1\|_{L^{2p}}\big) \\
&\quad \leqslant \text{const}\big(\|a_\infty(\rho) - a_\infty(\rho)\|_{L^{2p}} + \|a'_\infty(\rho) + a'_\infty(\bar\rho)\|_{L^{2p}}\big) \\
&\quad \leqslant \text{const}\,\lambda^{(p+2/2p)}|\rho - \bar\rho|.
\end{aligned}$$

But compatibility on the overlap means that $\bar\rho = g_1\rho g_0^{-1}$ over $U_0 \cap U_1$ so

$$|\rho - \bar\rho| \leqslant \text{const}\,\lambda^{(p+2/2p)}|\rho - \bar\rho|$$

and, for small λ, this implies $\rho = \bar\rho$.

The construction in §IV(v) below will be slightly simplified if a small modification is made to the connections before putting them in the standard form. For any fixed M with $1 < M < kN$ we can prove, as in Remark (4.24)(ii), that in the annulus defined by the radii $kM^{-1}\sqrt{\lambda}$, $k^{-1}M\sqrt{\lambda}$ the uniform norms of $a_\infty(\rho)$, $a'_\infty(\rho)$ are each $O(\sqrt{\lambda})$. Thus $a_\infty(\rho)$ can be cut off, to $a^*_\infty(\rho)$ say, over X_0 so that $F_+(A_0 + a^*_\infty(\rho))$ is bounded by K_4, independent of λ, and supported in a shell defined by radii $kM^{-1}\sqrt{\lambda}$, $M^{-1}\sqrt{\lambda}$. We do the same over X_1, so that the connections $A_0 + a^*_\infty(\rho)$, $A_1 + a'^*_\infty(\rho)$ are defined over X_0, X_1 and the transition function between them is defined over an annulus slightly smaller than $U_0 \cap U_1$. However, the parameter N could always be taken arbitrarily large (cf. (4.14)) so, abusing notation, we will suppose by rechoosing N from now on that $A_0 + a^*_\infty$, $A_1 + a'^*_\infty$ have self-dual curvature supported in shells $\tilde R_1$, $\tilde R_{-1}$, where, in X_0, $\tilde R_{-1}$ has inner and outer radii $k^2\sqrt{\lambda}/N$, $k\sqrt{\lambda}/N$ and $\tilde R_1$ has radii $k^{-1}N\sqrt{\lambda}$, $k^{-2}N\sqrt{\lambda}$. In the construction scheme of §IV(v) below the old shells R_1, R_{-1} will carry the leading "error term" and the new shells $\tilde R_1$, $\tilde R_{-1}$ will carry a lower order term.

The bound (4.30) implies that

$$\|a^*_\infty\|_{L^{2p}(X_0)'}\|a'^*_\infty\|_{L^{2p}(X_1)} \leqslant c_7\lambda^{(p+2/2p)},$$

so using the implicit function theorem in (4.9) we find gauge transformations

$$\begin{aligned}
&\exp(u_0) \quad \text{of } P_0 \text{ over } X_0,\\
&\exp(u_1) \quad \text{of } P_1 \text{ over } X_1,
\end{aligned}$$

putting the connections into standard gauges

$$\begin{aligned}
A_0 + a(\rho) &= e^{u_0}\big(A_0 + a^*_\infty\big), \qquad d^*_{A_0}a(\rho) = 0,\\
A_1 + a'(\rho) &= e^{u_1}\big(A_1 + a'^*_\infty\big), \qquad d^*_{A_1}a'(\rho) = 0,
\end{aligned} \tag{4.33}$$

with

$$\|a(\rho)\|_{L^{2p}},\ \|a'(\rho)\|_{L^{2p}},\ \|u_0\|_{L_1^{2p}},\ \|u_1\|_{L_1^{2p}} \leqslant c_8 \lambda^{(p+2/2p)} \quad \text{say.} \tag{4.34}$$

So we have constructed a family of equivalence classes of ASD connections $(A_0 + a(\rho), A_1 + a'(\rho), g(\rho))$ in the standard form (4.16) parametrized by

$$(\pm\rho) \in \operatorname{Hom}\big((P_0)_{x_0}, (P_1)_{x_1}\big)/\pm 1.$$

(It is clear from the construction that this family is homotopic, and indeed close in L^{2p} norm, to the family of §III(i) formed by first distorting A_0, A_1 to be flat near x_0, x_1.)

Finally, we will show that the transition function

$$g(\rho) = e^{-u_1} \circ \rho \circ e^{u_0} \quad \text{over } \hat{U}_0 \cap \hat{U}_1$$

is very close to the original gluing parameter ρ. Indeed for any connection $(A_0 + a, A_1 + a', g)$ in the standard form (4.16), g is interpreted as a map

$$\hat{U}_0 \cap \hat{U}_1 \to \operatorname{Hom}\big((P_0)_{x_0}, (P_1)_{x_1}\big) \cong SU(2).$$

The transition relation gives

$$|dg| \leqslant |A_0| + |A_1| + |a| + |a'|.$$

(Here A_0, A_1 denote the "connection matrices" in radial gauges spread out from x_0, x_1.) Thus, using the Hölder inequality,

$$\begin{aligned} \|dg\|_{L^p(\hat{U}_0 \cap \hat{U}_1)} &\leqslant \operatorname{const}\Big(\sqrt{\lambda}\,\big[\operatorname{Vol}(\hat{U} \cap \hat{U}_1)\big]^{1/p} \\ &\quad + \big(\|a\|_{L^{2p}} + \|a'\|_{L^{2p}}\big) \cdot \operatorname{Vol}(\hat{U}_0 \cap \hat{U}_1)^{1/2p}\Big) \\ &\leqslant \operatorname{const} n_3 \cdot \lambda^{1/p}. \end{aligned} \tag{4.35}$$

Then the Sobolev embedding $L_1^p \to C^0$ implies that the variation of g is small:

$$|\operatorname{Var}(g)| \leqslant \operatorname{const} \eta_3 \cdot \lambda^{(p-2/2p)} \tag{4.36}$$

(compare [12], [17]). (The extra powers of λ can be found by rescaling $\hat{U}_0 \cap \hat{U}_1$ to a uniform size; cf. the proof of Lemma (4.43) below.)

Fix a rule for assigning an "average value" or center of mass $\hat{g} \in \operatorname{Hom}((P_0)_{x_0}, (P_1)_{x_1})$ to the map g. For example this could be defined by projecting the C. of M. in the copy of $\mathbb{R}^4$ containing $\operatorname{Hom}((P_0)_{x_0}, (P_1)_{x_1})$ onto the unit 3-sphere. The only property we need is that if $\hat{g} = \hat{h}$ for two maps g, h, then $\|\hat{g} - \hat{h}\|_{L^\infty}$ can be estimated by the derivatives of g, h.

The average value $\hat{g}(\rho)$ for our family of ASD connections (4.33) is close to ρ. The bounds (4.34) and (4.35) give

$$|\hat{g}(\rho) - \rho| \leqslant \operatorname{const} \lambda^{(p+2/2p)}. \tag{4.37}$$

Similarly, if the derivative $(\partial\hat{g}(\rho)/\partial\rho)$ is interpreted by using the action of $\operatorname{Aut}(P_0)_{x_0}$ to trivialize $T(\operatorname{Hom}((P_0)_{x_0}, (P_1)_{x_1}))$, then

$$\begin{aligned}\left|\frac{\partial\hat{g}(\rho)}{\partial\rho} - 1\right| &\leqslant \text{const}\left(\left\|\frac{\partial u_0}{\partial\rho}\right\|_{C^0} + \left\|\frac{\partial u_1}{\partial\rho}\right\|_{C^0} + \|u_1\|_{C^0}\right)\\ &\leqslant \text{const}\left(\left\|\frac{\partial a_\infty}{\partial\rho}\right\|_{L^{2p}} + \left\|\frac{\partial a'_\infty}{\partial\rho}\right\|_{L^{2p}} + \left\|a'_\infty\right\|_{L^{2p}}\right)\\ &\leqslant \text{const}\,\lambda^{(p+2/2p)},\end{aligned} \tag{4.38}$$

using Remark (4.24)(i) as in the proof of Lemma (4.31). It follows that $\rho \mapsto \hat{g}(\rho)$ defines a diffeomorphism of $\operatorname{Hom}((P_0)_{x_0}, (P_1)_{x_1})$.

IV(v). Constructing gauge transformations. The proof of Theorem (4.17) is finished in this section by constructing a gauge transformation between any ASD solution in the standard form (4.16) and one of the family produced in §IV(iv). This is done by an alternating construction in which the leading estimate is precisely complimentary to Lemma (4.22). The iterative cycle is now a little more complicated because we have to consider the gauge fixing equations as well as the ASD equations themselves. This reflects the fact that we are really working with the middle term of the deformation complex (4.1).

The iterative cycle will be based on the following procedure for transferring gauges between the manifolds, supposing always that λ is sufficiently small.

(4.39) The transfer procedure ("From X_1 to X_0''"). Let $B = (A_0 + b, A_1 + b', h)$ be any connection in the standard form. Recall that this means that b and b' are defined over $\hat{U}_0$ and $\hat{U}_1$ respectively, whereas in the family $(A_0 + a(\rho), A_1 + a'(\rho), g(\rho))$ constructed in §IV(iv), $a(\rho)$, $a'(\rho)$ define ASD connections over the larger sets U_0, U_1. Moreover we have extended $a(\rho)$, $a'(\rho)$ over X_0, X_1 to define connections whose ASD curvature is supported in the shells $\tilde{R}_{-1}$, $\tilde{R}_1$ and uniformly bounded by K_4.

Since the map $\rho \mapsto \hat{g}(\rho)$ is a diffeomorphism, there is a ρ such that $\hat{g}(\rho) = \hat{h}$. The composition $s = g(\rho)^{-1} \circ h$ is an automorphism of $P_0|_{\hat{U}_0 \cap \hat{U}_1}$. Use this to define a new gauge equivalent connection $s(A_0 + b)$ over $\hat{U}_0 \cap \hat{U}_1$. The difference of two connections transforms tensorially so

$$s(A_0 + b) = A_0 + a + g^{-1}(b' - a')g. \tag{4.40}$$

Use the RHS of this formula extend $s(A_0 + b)$ over $R_{-1} \subset X_0$ (g, b', a' are defined there).

Now make two modifications, by cut-offs, to this procedure to get a new connection $A_0 + b^*$ over all of X_0; gauge equivalent to $A_0 + b$ over $\hat{U}_0$. First, the equality of $\hat{g}(\rho)$, $\hat{h}$ together with (4.36) imply that s is everywhere close to 1_{P_0}. Put $\bar{s} = \exp(\Phi \log s)$, where Φ is a cut-off function equal to 0 on the complement of $\hat{U}$ and to 1 on the complement of $\hat{U}$. We can arrange that $|d\Phi| < \text{const}/\sqrt{\lambda}$. Extend $\bar{s}$ by 1 over $\hat{U}$; the connection $\bar{s}(A_0 + b)$ is equal to $s(A_0 + b)$ near the inner ring $R_{-1} \subset X_0$. Second, cut off the tensorial difference term in (4.40) to define a connection

$$A_0 + a + \psi_{-1} \cdot \left(g^{-1}(b' - a')g\right)$$

over R_{-1}. The two constructions piece together to give a globally defined connection $A_0 + \bar{b}$ over X_0. Finally if (as will be the case when η_3 is small; cf. the proof of Lemma (4.43) below) $\|\bar{b}\|_{L^{2p}(X_0)}$ is small enough to apply the implicit function theorem to (4.9), make a gauge transformation $\exp(u^*)$ over X_0 so that $\exp(u^*)(A_0 + \bar{b}) = A_0 + b^*$ and $d_{A_0}^* b^* = 0$.

In this way we get a new representation for the same gauge equivalence class $[B]$ over $X_0 \# X_1$ in standard form:

$$\left(A_0 + b^*|_{\hat{U}_0}, A_1 + b', he^{-u^*}\right).$$

Moreover $A_0 + b^*$ is defined over all of X_0 with self-dual curvature supported in $\tilde{R}_{-1} \cup R_1$, and $\|F^+(A_0 + b^*)|_{\tilde{R}_{-1}}\| \leqslant K_4$ since $A_0 + b^*$ is gauge equivalent to $A_0 + a(\rho)$ over $\tilde{R}_{-1}$. Applying the transfer procedure twice, once from X_1 to X_0 then from X_0 to X_1, we can without loss suppose that our original representatives b, b' were extended in this way over X_0, X_1.

More generally suppose that at the nth stage of the iteration we have found representations

$$\left(A_0 + b_n, A_1 + b_n', h_n\right) \tag{4.41}$$

in standard form for $[B]$, and that ρ_n is defined by $\hat{h}_n = \hat{g}_{\rho_n}$. For brevity write a_n, a_n' for $a(\rho_n)$, $a'(\rho_n)$. We assume that, in the same fashion as above, b_n, b_n' are defined over X_0, X_1 with $d_{A_0}^* b_n$, $d_{A_1}^* b_n' = 0$ and that $F_+(A_0 + b_n)$, $F_+(A_1 + b_n')$ are supported in $R_{-1} \cup \tilde{R}_{-1}$, $R_1 \cup \tilde{R}_1$ respectively. Put

$$\begin{aligned} \chi_n &= F_+(A_0 + b_n) - F_+(A_0 + a_n), \\ \varepsilon_n &= \|\chi_n|_{R_{-1}}\|_{L^\infty(X_0)}, \\ \zeta_n &= \|\chi_n|_{\tilde{R}_{-1}}\|_{L^\infty(X_0)}, \end{aligned} \tag{4.42}$$

and define χ_n', ε_n', ζ_n' symmetrically over X_1. Note that $\chi_n|_{R_{-1}} = F_+(A_0 + b_n)$ since the $A_0 + a(\rho)$ are ASD over U_0.

Start the iteration with b, b'. Then, if n is even, pass from the nth to the $(n+1)$th stage by applying the transfer procedure (4.39) from X_1 to X_0, and if n is odd pass to the next stage by the corresponding transfer from X_0 to X_1. The next lemma gives bounds on the error terms ε, ζ for the new representations; for simplicity suppose n is even.

(4.43) Lemma. *For any $h_3 > 1$, $K_5 > 0$ we can find $\eta_5 > 0$, $c_9 > 0$ such that if $\|b_n\|_{L^{2p}(X_0)}$, $\|b_n'\|_{L^{2p}(X_1)} < \eta_5$, $\varepsilon_n + \varepsilon_n' + \zeta_n + \zeta_n' < K_5$, and λ is sufficiently small, then the following bounds hold.*

(i) $\quad \varepsilon_{n+1} \leqslant (h_3 K_N)\varepsilon_n + c_9 \cdot \zeta_n'$.

(ii) (α) $\quad \zeta_{n+1} \leqslant c_9 \cdot \lambda^{(p+2/2p)}(\varepsilon_n + \varepsilon_n' + \zeta_n + \zeta_n')$.

(β) $\quad \zeta_{n+1}' \leqslant \zeta_n' + c_9 \cdot \lambda^{(p+2/2p)}(\varepsilon_n + \varepsilon_n' + \zeta_n + \zeta_n')$.

(iii) (α) $\quad \|b_n - a_n\|_{L^{2p}(X_0)} \leqslant c_9\lambda^{(p+2/2p)}(\varepsilon_n + \zeta_n)$.

(β) $\quad \|b_n' - a_n'\|_{L^{2p}(X_0)} \leqslant c_9\lambda^{(p+2/2p)}(\varepsilon_n' + \zeta_n')$.

(γ) $\quad \|b_{n+1} - a_{n+1}\|_{L^{2p}(X_0)} \leqslant c_9\lambda^{(p+2/2p)}(\varepsilon_{n+1} + \zeta_{n+1})$.

(δ) $\quad \|b_{n+1}' - a_{n+1}'\|_{L^{2p}(X_1)} \leqslant c_9\lambda^{(p+2/2p)}(\varepsilon_{n+1}' + \zeta_{n+1}')$.

Proof. Begin with (iii)(α), (β) which refer only to the connections at the nth stage. They follow exactly as in Lemma (4.22)(ii) from the equations

$$d_{A_0}^+(b_n - a_n) = \chi_n + (a_n \wedge a_n - b_n \wedge b_n)_+, \qquad d_{A_0}^*(b_n - a_n) = 0$$

(and symmetrically over X_1) once η_5 is small enough to get an estimate of the quadratic term.

Now pass to (i). ε_{n+1} is the norm

$$\begin{aligned}&\left\|F_+(A_0 + b_{n+1}) \mid R_{-1}\right\|_{L^\infty(X_0)}\\ &\quad = \left\|F_+\left(A_0 + \psi_{-1}\left(g_n^{-1}(b_n' - a_n')\right) \mid R_{-1}\right)\right\|_{L^\infty(X_0)}\\ &\quad \leqslant k^{-4}N^4\left\|F_+\left(A_1 + \psi_{-1}(b_n' - a_n')\right) \mid R_{-1}\right\|_{L^\infty(X_1)},\end{aligned}$$

where we have used the definition of b_{n+1} in the transfer procedure to pass back to X_1 by the transition function $g_n = g_{\rho_n}$. The factor $k^{-4}N^4$ enters from the comparison of the norms of X_0, X_1 restricted to R_{-1}.

On the manifolds X_1 the difference $b_n' - a_n'$ can be estimated, using the kernel L' for $(d^* \oplus d^+)_{A_1}^{-1}$, in terms of χ_n'. This is just the same as in Lemma (4.22) but with error term χ_n' supported now in an "inner" ring $R_1 \cup \tilde{R}_1 \subset X_1$. The calculation of the leading linear term is now

$$\left\|L'(\chi_n') \mid R_{-1}\right\| \leqslant \max_{X \in R_{-1}} |L'(x, y)| \times \left(\operatorname{vol}_{X_1}(R_1) \cdot \varepsilon_n' + \operatorname{vol}_X(\tilde{R}_1) \cdot \zeta_n'\right).$$

Only the coefficient of ε_n' is really important; this is

$$\max|L'|\operatorname{vol}_{X_1}(R_1)\varepsilon_n' \leqslant \sqrt{2}\sqrt{h_3}\left(\frac{1 - k^4}{4}\right)\frac{N^{-4}\sqrt{\lambda}}{(N - N^{-1})^3}\varepsilon_n'$$

which gives a contribution from ε'_n to

$$k^{-4}N^4\left\| F_+\left(A_1 + \psi_{-1}\left(b'_n - a'_n\right)\right) | _{R_{-1}}\right\|_{L^\infty(X_1)}$$

of

$$h_3 \frac{k^{-4} - 1}{4(k^{-1} - 1)} \cdot \frac{N}{(N - N^{-1})^3} \varepsilon'_n \leqslant h_3 K_N \varepsilon'_n$$

up to positive powers of λ. Similarly the contribution from the $\chi_n|_{\tilde{R}_1}$ is $\leqslant$ const ζ'_n.

Next consider the inequalities (ii)(α), (β), for the "secondary" ζ, ζ' terms. Let us adopt the notation of (4.39), so that $\mathscr{A}_c + b_{n+1}$ over X_0 (which is $A_0 + b^*$ in (4.39)) is found by first forming a connection $A_0 + \bar{b}$ and then making the gauge transformation $\exp(u^*)$. The ρ-variable then has to be adjusted from ρ_n to ρ_{n+1} to preserve the condition $\hat{h}_{n+1} = \hat{g}_{\rho_{n+1}}$ $(\equiv \hat{g}_{n+1})$.

The variations of g_n, h_n over $\hat{U}_0 \cap \hat{U}_1$ are small by (4.36). We will show that, together with the equality of $\hat{g}_n$, $\hat{h}_n$, this implies a bound:

(4.44) $$\left\| g_n^{-1}h_n - 1\right\|_{C^0} \leqslant \text{const}\, \lambda^{(p-2/2p)}\left(\left\| a_n - b_n\right\|_{L^{2p}(X_0)} + \left\| a'_n - b'_n\right\|_{L^{2p}(X_1)}\right).$$

This is completely straightforward but we will include the argument here since the corresponding step in the proof of Lemma (4.31) was not given in detail.

First rescale the small annulus $\hat{U}_0 \cap \hat{U}_1$ by a factor $1/\sqrt{\lambda}$ to a standard Riemannian model Π. On Π the Sobolev embedding $L_1^{2p} \to C^0$ implies the existence of a constant such that for functions (or vector-value functions) w:

$$\| W \|_{C^0} \leqslant \text{const}\left(\left(\int_\Pi \| w \|^{2p}\right)^{1/2p} + \left| \int_\Pi w \right|\right).$$

Now suppose that g, h are two maps $\Pi \to \text{Hom}((P_0)_{x_0}, (P_1)_{x_1}) \cong SU(2) \cong S^3$, with small variations and $\hat{g} = \hat{h} = 1$. Then $g = \exp(v)$, $h = \exp(w)$, say, and since exp defines a diffeomorphism with small metric distortion factor on the small ball containing $\text{Im}(v)$ and $\text{Im}(w)$, norms of g, h, $g - h$ are equivalent with those of v, w, $v - w$. Similarly

$$\hat{g}^{-1}\hat{h} = \exp\left[\int_\Pi w - v\right] + O\left(\| g^{-1}h - 1\|_{C^0}^2\right).$$

Thus over Π, if $\hat{g} = \hat{h}$,

$$\| g^{-1}h - 1\|_{C^0} \leqslant \text{const}\left(\int_\Pi |dgg^{-1} - dhh^{-1}|^{2p}\right).$$

Transforming back to $\hat{U}_0 \cap \hat{U}_1$ with $g = g_n$, $h = h_n$ gives

$$\| g_n h_n - 1\|_{C^0} \leqslant \text{const}\, \lambda^{(p-2/2p)}\left(\int_{\hat{U}_0 \cap \hat{U}_1} |dg_n g_n^{-1} - dh_n h_n^{-1}|^{2p}\right)^{1/2p}.$$

But the derivatives of g and h can be expressed in terms of the different connections over X_0, X_1 using the transition relationship

$$\left.\begin{array}{l} g_n(A_0+a_n)=A_1+a'_n \\ h_n(A_0+b_n)=A_1+b'_n \end{array}\right\} \Rightarrow \left(dg_ng_n^{-1}-dh_nh_n^{-1}\right)$$
$$=g_n(A_0+a_n)g_n^{-1}-\left(A_1+a'_n\right)$$
$$-h_n(A_0+b_n)h_n^{-1}+\left(A_1+b'_n\right)$$

so

$$\left|dg_ng_n^{-1}-dh_nh_n^{-1}\right|\leqslant\left|a'_n-b'_n\right|+\left|g_n(A_0+a_n)g_n^{-1}-h_n(A_0+b_n)h_n^{-1}\right|$$
$$\leqslant\left|a'_n-b'_n\right|+\left|a_n-b_n\right|+2\big(\left|A_0\right|+\left|a_n\right|\big)\left|g_n^{-1}h_n-1\right|.$$

Since we know that the L^{2p} norms of the (connection matrix) A_0 and of a_n are small (cf. (4.28), (4.34)), we can rearrange the terms to get (4.44).

Now the connection $A_0+\bar{b}$ was defined by cutting off the transform of A_0+b_n by $g_n^{-1}h_n$. Since the gradient of the cut off Φ is $O(1/\sqrt{\lambda})$ we have

$$\|\bar{b}-b_n\|_{L^{2p}(X_0)}\leqslant\operatorname{const}\Bigg(\|b'_n-a'_n\|_{L^{2p}(X_1)}+\|b_n-a_n\|_{L^{2p}(X_0)}$$
$$+\frac{1}{\sqrt{\lambda}}\|g_n^{-1}h_n-1\|_{C^0}\times\operatorname{Vol}(\hat{U}_0\cap\hat{U}_1)^{1/2p}\Bigg)$$
$$\leqslant\operatorname{const}\lambda^{(p+2/2p)}\left(\varepsilon_n+\zeta_n+\varepsilon'_n+\zeta'_n\right)$$

by (4.44) and part (ii)(α), (β). So

$$\|d^*_{A_0}b\|_{L^{2p}_{-1}(X_0)}\leqslant\operatorname{const}\lambda^{(p+2/2p)}\left(\varepsilon_n+\zeta_n+\varepsilon'_n+\zeta'_n\right)$$

and if η_5 is small enough that we can apply the implicit function theorem to fix the gauge we get a gauge transformation $\exp(u^*)$ with

$$\|u^*\|_{L^{2p}_1(X_0)}\leqslant\operatorname{const}\lambda^{(p+2/2p)}\left(\varepsilon_n+\zeta_n+\varepsilon'_n+\zeta'_n\right). \tag{4.45}$$

Our new triple $(A_0+b_{n+1}, A_1+b'_{n+1}, h_{n+1})$ is

$$\left(\exp(u^*)\left(A_0+\bar{b}\right),\ A_1+b'_n,\ h_n^{-1}\bar{s}\exp(-u^*)\right),$$

where $\bar{s}$ is $g_n^{-1}h_n$ cut off to 1_{P_0} as in (4.39). It follows from the bounds (4.44), (4.45) on $(s-1)$ (hence on $\bar{s}-1$) and on u^* that $\hat{h}_{n+1}-\hat{h}_n$ is bounded by a constant times $\lambda^{(p+2/2p)}(\varepsilon_n+\varepsilon'_n+\zeta_n+\zeta'_n)$. By (4.38) we get a similar bound on $\rho_{n+1}-\rho_n$:

$$\left|\rho_{n+1}-\rho_n\right|\leqslant\operatorname{const}\lambda^{(p+2/2p)}\left(\varepsilon_n+\varepsilon'_n+\zeta_n+\zeta'_n\right). \tag{4.46}$$

Now

$$\begin{aligned}\zeta_{n+1} &= \left\|\left(F_+(A_0 + a_{n+1}) - F_+(A_0 + b_{n+1})\right)\right\|_{L^\infty(X_0)} \\ &\leqslant \left\|F_+(A_0 + 1_{n+a}) - F_+(A_0 + a_{n+1})\right\|_{L^\infty(X_0)} \\ &\quad + \left\|\left(F_+(A_0 + a_n) - F_+(A_0 + b_{n+1})\right)\mid_{\tilde{R}_{-1}}\right\|_{L^\infty(X_0)}.\end{aligned}$$

By construction $F_+(A_0 + a_n) = F_+(A_0 + \bar{b})\mid_{\tilde{R}_{-1}}$ and $A_0 + \bar{b}$, $A_0 + b_{n+1}$ are gauge equivalent by u^*. On the other hand the first term is bounded by a constant times $|\rho_{n+1} - \rho_n|$. So we get (ii)(α):

$$\begin{aligned}\zeta_{n+1} &\leqslant \text{const}|\rho_{n+1} - \rho_n| + K_4 \cdot \|u^*\|_{L^\infty} \\ &\leqslant \text{const}\, \lambda^{(p+2/2p)}\left(\varepsilon_n + \varepsilon'_n + \zeta_n + \zeta'_n\right).\end{aligned}$$

The bound on ζ_n in (ii)(β) is similar and (iii)(γ), (δ) follow from (α), (β) replacing n by $n + 1$.

We may use this lemma, just as in §IV(iv), to show that if $h_3 K_N < 1$ and λ is small, the iteration continues indefinitely and the error terms tend to zero geometrically. Notice that only one of the pairs of error terms (ε_n, ζ_n), $(\varepsilon'_n, \zeta'_n)$ are improved at each step; but over two steps we get, for small λ,

$$\left(\varepsilon + \varepsilon' + \zeta + \zeta'\right)_{n+2} < r\left(\varepsilon + \varepsilon' + \zeta + \zeta'\right)_n$$

with $r < 1$. Finally (4.46) implies that the ρ_n's converge to a ρ_∞ and it follows immediately from the bounds of Lemma (4.43)(iii) that $A(\rho_\infty)$ is gauge equivalent to B over $X_0 \# X_1$. This completes the proof of Theorem (4.17).

IV(vi). Generalizations. The first of the conditions assumed for Theorem (4.17) which we will relax is that the metrics on X_0, X_1 be flat in the identifying region. If m_0, m_1 are any Riemannian metrics on X_0, X_1, then a connected sum, as an explicit manifold, can be defined in the same fashion as before using geodesic coordinate systems centered on x_0, x_1. The data required ($\lambda > 0$, σ: $(TX_0)_{x_0} \to (TX_1)_{x_1}$) is the same and we can define regions U_0, U_1 etc. just as before. The simplest notion of approximation of metrics to use is based upon "quasi-conformality." We will say that a metric on $X_0 \# X_1$ is conformally ε-close to m_0, m_1 if there are functions f_0, f_1 on U_0, U_1 such that

$$\left\|\left(m_i - f_i \cdot m\right)\mid_{U_i}\right\|_{L^\infty(X_i, m_i)} < \varepsilon. \tag{4.47}$$

It is easy to see that metrics m exist which are conformally $C \cdot \lambda$—close to the m_i, where C is a constant depending on the curvatures of the Riemannian metrics m_0, m_1. For convenience we can suppose that the metrics $f_i m|_{U_i}$ are extended over the compact manifolds X_i and that a uniform bound like (4.47) is preserved.

Consider the ASD connections over $X_0 \# X_1$, defined with respect to such a metric m, which can be put into the standard representation (4.16), where the gauge fixing equations $d^*_{A_0} a = 0$, $d^*_{A_1} a' = 0$ are formed by using the original metrics m_i. The discussion of §IV(ii) is modified so that in place of (4.19) we put the equations

$$d^*_{A_0, m_0} b = 0, \qquad d^+_{A_0, f_0 m} b + ([a, b] + b \wedge b)_{+, f_0 m} = -\sigma$$

having the indicated dependence on the metrics. The relevant linear operator is $d^*_{A_0, m_0} \oplus d^+_{A_0, f_0 m}$ which can be written as

$$\left(d^*_{A_0} \oplus d^+_{A_0}\right)_{m_0} + \mu \cdot d^-_{A_0, m_0},$$

where μ is a tensor whose uniform norm is of the order ε. Thus the operator norm of the perturbation $\mu \cdot d^-_{A_0, m_0}\colon L^p_1 \to L^p$ is $O(\varepsilon)$. If $(d^*_{A_0} \oplus d^+_{A_0})_{m_0}$ is an isomorphism, then so is the perturbed operator, once ε is small, and we can solve the nonlinear equation just as before. Suppose that $\varepsilon \leqslant K_6 \cdot \sqrt{\lambda}$ for fixed $K_6 > 0$.

In comparing the solution b to the nonlinear equation with the approximation $\tilde{b}$ given by the kernel of $(d^* \oplus d^+)^{-1}_{A_0, m_0}$ there is a new contribution coming from the change in metric. We have

$$(d^* \oplus d^+)_{A_0, m_0} \beta = -([a, b]_+ + (b \wedge b)_+) - \mu \cdot d^-_{A_0, m_0} b$$

from which we deduce

$$\begin{aligned} \|\beta\|_{L^\infty} &\leqslant \operatorname{const} \lambda^{2/3} \cdot \delta + \left\| (d^* \oplus d^+)^{-1}_{A_0} \left(\mu \cdot d^-_{A_0, m_0} b\right) \right\|_{L^\infty} \\ &\leqslant \operatorname{const} \lambda^{2/3} \cdot \delta + K_6 \cdot \sqrt{\lambda} \cdot \left\| d^-_{A_0, m_0} b \right\|_{L^8} \\ &\leqslant \operatorname{const}\left(\lambda^{2/3} \cdot \delta + K_6 \cdot \sqrt{\lambda} \cdot \lambda^{1/3} \cdot \delta\right) \\ &\leqslant \operatorname{const} \lambda^{2/3} \cdot \delta \end{aligned}$$

so the statement of Lemma (4.22)(i) is still correct. The alternating construction of §IV(iv) gives the same geometric decay. To start the construction make initial modifications to the connections A_0, A_1 to get nearby ASD connections with respect to the metrics $f_i \cdot m$. The modifications needed have L^{2p}_1 norm of the order $\varepsilon \cdot \|F_{A_i}\|_{L^{2p}}$, hence of order $\sqrt{\lambda}$. This means that if we now define $A^{(0)}$ by cutting off these modified connections, the bounds (4.28) are preserved. So the construction of connections in §IV(iv) goes through as before and the construction of gauge changes in §IV(v) needs essentially no modification. In sum we have:

(4.48) Proposition. *Given metrics m_0, m_1 on X_0, X_1, a constant $K_6 > 0$, and m_i-ASD connections A_i satisfying hypothesis* (4.15), *then the statement of Theorem* (4.17) *holds for ASD connections defined with respect to any metric m which is conformally $K_6 \cdot \sqrt{\lambda}$ close to the m_i.*

Next we relax hypothesis (4.15)—that the deformation complexes of the connections A_i be acylic. This involves combining the construction described above with the descriptions, summarized in §IV(i), of the moduli of ASD connections over the individual manifolds.

Choose liftings V_{A_0}, V_{A_1} of $H^2_{A_0}$, $H^2_{A_1}$ to forms supported away from the points X_0, X_1 and let $p_0 \mapsto \tilde{p}_0$, $p_1 \mapsto \tilde{p}_1$ be the local deformations of (4.4), parametrized by $H^1_{A_0}$, $H^1_{A_1}$. Let π_0, π_1 be maps from $\Omega^1(\mathfrak{g}_{P_0})$, $\Omega^1(\mathfrak{g}_{P_1})$ to $H^1_{A_0}$, $H^1_{A_1}$ transverse to the deformations. These should depend only on the restrictions to regions away from the points x_i—for example induced by L^2 projection with subspaces U_{A_0}, U_{A_1} supported away from the identification region.

We wish to describe the equivalence classes of connections over $X_0 \# X_1$ representable in the standard form. Generalizing §IV(i) we solve, in place of (4.19), equations of the form

$$(4.49) \qquad d^*_{A_0} b = 0, \qquad d^+_{A_0} b + [b, a]_+ + [b \wedge b]_+ + \phi = \sigma,$$

where $\phi \in V_{A_0}$. The extra variable means that the linearized equation has a unique solution so the inverse function theorem applies again. Lemma (4.22) is unaltered since the kernel representation has the same form. In addition we get that $|\phi|$ is $O(\delta)$. The alternating construction of §IV(iii) goes through to construct a family of solutions to the corresponding "infinite dimensional part" of the ASD equations (compare [25, §III]). But now rather than having a single parameter ρ, the iteration is started with a family $A^0(\rho, p_0, p_1)$ defined by cutting off $A_0 + \tilde{p}_0$, $A_1 + \tilde{p}_1$. For small p_0, p_1 we get a limiting connection

$$A^\infty(\rho, p_0, p_1) = \left(A_0 + \tilde{p}_0 + a_\infty,\ a_1 + \tilde{p}_1 + a'_\infty,\ \rho\right)$$

such that

$$(4.50) \qquad F_+(A^\infty(\rho, p_0, p_1)) = \phi_0(\rho, p_0, p_1) + \phi_1(\rho, p_0, p_1),$$

where $\phi_i(\rho, p_0, p_1) \in V_{A_i}$. The norms of a_∞, a'_∞ satisfy the same estimates (4.30), similarly for the derivatives with respect to the parameters ρ, p_0, p_1.

The holonomy centralizers Γ_{A_0}, Γ_{A_1} can be made to act as symmetries of this family $A^{(\infty)}(\rho, p_0, p_1)$ and in place of Lemma (4.31) we put

(4.51) Lemma. *Once λ, $|p_i|$, $|\bar{p}_i|$ are sufficiently small a pair of connections $A^{(\infty)}(\rho, p_0, p_1)$, $A^\infty(\bar{\rho}, \bar{p}_0, \bar{p}_1)$ are gauge equivalent if and only if the parameters differ by the action of $\Gamma_{A_0} \times \Gamma_{A_1}$.*

Proof. On the compact manifold X_0 we can find a constant c such that for any element $u \in \Omega^0(\mathfrak{g}_{P_0})$ and decomposition $d_{A_0} u = \alpha + h$ with $h \in \operatorname{Ker}(d^*_{A_0} \oplus d^+_{A_0}) \cong H^1_{A_0}$, the inequality $||h| + |u||_{C^0} < c(\|\alpha\|_{L^{2p}} + |\pi_x(u)|)$ holds (cf. Lemma (4.31)). Now suppose that u and α are only defined over $U_0 \subset X_0$.

By rescaling the shell R_{-1} to a standard size (as in Lemma (4.43)) we get a bound on the variation of u relative to the exponential gauge:

$$\operatorname{Var}\left(u|_{R_{-1}}\right) \leqslant \text{const}\, \lambda^{1/2-1/p}\left(\|\alpha\|_{L^{2p}} + \|h\|\right).$$

So if $\hat{u}$ is the mean value of $u|_{R_{-1}}$ and we cut off u to get $u' = u + \psi_{-1}(u - \hat{u})$, then

$$d_{A_0}u' - \psi_{-1}d_{A_0}u = (1 - \psi_{-1})[A_0, u] + (d\psi_{-1})(u - \hat{u}).$$

So $d_{A_0}u' = h + \alpha'$, where

$$\alpha' = \psi_{-1}\alpha + (\psi_{-1} - 1)(h - [A_0, u]) + d\psi_{-1}(u - \hat{u}).$$

So, since $\operatorname{Vol}(R_{-1}) \leqslant \text{const.}\, \lambda^2$,

$$\|\alpha'\|_{L^{2p}} \leqslant \|\alpha\|_{L^{2p}} + \text{const.}\left(\|h\| \cdot \lambda^{1/p} + \|u\|_{C^0}\lambda^{1/2+1/p} + \|\alpha\|_{L^{2p}}\right),$$

and if λ is small we deduce that for a constant c', independent of λ,

$$\|h\| + \|u\|_{C^0} \leqslant c'\left(\|\alpha\|_{L^{2p}} + |\pi_x(u)|\right),$$

since the uniform norms of u, u' differ by at most $\operatorname{Var}(u|_{R_{-1}})$.

This implies the corresponding nonlinear result: If $g_0 = \exp(u_0)$ is a gauge transformation between connections over U_0:

$$A_0 + p_0 + a,\ A_0 + \bar{p}_0 + \bar{a},$$

and if $|p_0|$, $|\bar{p}_0|$, $\|a\|_{L^{2p}}$, $\|\bar{a}\|_{L^{2p}}$ are small, we can write $g_0 = \gamma_0 \cdot \tilde{g}_0$ with $\gamma_0 \in \Gamma_{A_0}$ and

$$|p_0 - \bar{p}_0| + \|\tilde{g}_0 - 1\| \leqslant \text{const.}\|a - \bar{a}\|_{L^{2p}}.$$

Then, just as in Lemma (4.31), if there is a gauge equivalence between

$$A^{(\infty)}(\rho, p_0 \cdot p_1),\ A^{(\infty)}(\bar{\rho}, \bar{p}_0, \bar{p}_1)$$

we may, without loss, suppose it represented by g_0, g_1 with

$$\begin{aligned} |p_0 - \bar{p}_0| &+ |p_1 - \bar{p}_1| + \|g_0 - 1\| + \|g_1 - 1\| \\ &\leqslant \text{const}\left(\|a_\infty - \bar{a}_\infty\|_{L^{2p}} + \|a'_\infty - \bar{a}'_\infty\|_{L^{2p}}\right) \\ &\leqslant \text{const}\, \lambda^{p+2/2p}\left(|\rho - \bar{\rho}| + |p_0 - \bar{p}_0| + |p_1 - \bar{p}_1|\right) \end{aligned}$$

and since $\rho = g_1 \rho g_0^{-1}$ when λ is small the only solution is $\rho = \bar{\rho}$, $p_0 = \bar{p}_0$, $p_1 = \bar{p}_1$.

Next find small gauge transformations putting this family of connections into standard gauges just as in (4.33), (4.34). Everything is now parametrized by ρ, p_0, p_1 where $p_i \in H^1_{A_i}$ are small. In place of (4.37), (4.38) we prove now that the map

$$(\rho, p_0, p_1) \xrightarrow{\Pi} \left(\hat{g}(\rho, p_0, p_1),\ \pi_0 a(\rho, p_0, p_1),\ \pi_1 a'(\rho, p_0, p_1)\right)$$

defines a diffeomorphism close to the identify from one neighborhood of

$$\operatorname{Hom}\big((P_0)_{x_0},(P_1)_{x_1}\big)\times\{0\}\times\{0\}\subset\operatorname{Hom}\big((P_0)_{x_0},(P_1)_{x_1}\big)\times H^1_{A_0}\times H^1_{A_1}$$

to another. Then §IV(v), which was the hardest part of the proof of Theorem (4.17), needs almost no change. If B is any solution of the ASD equations representable in standard form, we define a sequence of representations (4.41) inductively and define $(\rho_n,(p_0)_n,(p_1)_n)$ so that

$$\Pi\big(\rho_n,(p_0)_n,(p_1)_n\big)=\big(h_n,\pi_0(b_n),\pi_1(b'_n)\big)$$

—which will always be possible if $\|b_n\|_{L^{2p}}$, $\|b'_n\|_{L^{2p}}$, and λ are small. The error terms ε_n, ζ_n, ε'_n, ζ'_n are defined in just the same way as before.

The statement of Lemma (4.43) needs no change. In the proof we have to consider the equations

$$\begin{aligned}&d^+_{A_0}(b_n-a_n)=\chi_n+(a_n\wedge a_n-b_n\wedge b_n)-e^{u_0}\phi e^{-u_0},\\&d^*_{A_0}(b_n-a_n)=0.\end{aligned}\tag{4.52}$$

Here $\phi_0\in V_{A_0}$ (with $|\phi_0|=O(\eta_5)$) is the extra term of the kind that arose in (4.49) and e^{u_0} is the gauge transformations used as in (4.33) to fix the gauge over X_0. The norm $\|u_0\|_{C^0}$ is $o(\lambda)$ by (4.34). We also know that $\pi_0(b_n)=\pi_0(a_n)$.

It is clear that, once η_5 is small and so e^{u_0} uniformly close to 1, the finite-dimensional subspace $e^{u_0}V_{A_0}e^{-u_0}$ is transverse to $\operatorname{Im}d^+_{A_0}$. If we define linear maps

$$\alpha\colon V_{A_0}\to V_{A_0},\qquad t\colon V_{A_0}\to\operatorname{Ker}d^*_{A_0}\cap\big(\operatorname{Ker}d^+_{A_0}\big)^\perp\subset\Omega^1(\mathfrak{g}_{P_0})$$

by the equation $v=d^+_{A_0}(tv)+e^{u_0}(\alpha v)e^{-u_0}$, then

$$\|tv\|_{L^{2p}_1}\leqslant\operatorname{const}\|u_0\|_{C^0}\cdot\|v\|,\qquad\|\alpha v-v\|\leqslant\operatorname{const}\|u_0\|_{C^0}\cdot\|v\|.$$

(All norms on V_{A_0} are, of course, equivalent.) For fixed u_0 there is a kernel $L^{(u_0)}$ such that $c=L^{(u_0)}\cdot\chi$ solves the linearized version of (4.51):

$$d^+_{A_0}c-\chi\in u^{u_0}\vee_{A_0}e^{-u},\quad d^*_{A_0}c=0,\quad(d\pi_0)c=0.$$

And since $L^{(u_0)}\chi=L^{(0)}\chi+t(\chi-d^+_{A_0}L^{(0)}\chi)$ the kernels $L^{(u_0)}$, $L^{(0)}$ differ by a smooth kernel whose uniform norm is bounded by a fixed multiple of $\|u_0\|_{C^0}$, hence is $o(\lambda)$. This means that we get the same bounds on b_n-a_n as before and the construction of gauge transformations by the alternating method can proceed.

Finally this discussion can be combined with that for varying metrics and proves the following generalization of Theorem (4.17).

(4.53) Theorem. *Let m_0, m_1 be Riemannian metrics on X_0, X_1 and A_0, A_1 be ASD connections defined with respect to those metrics. Suppose $K_6 > 0$; then if η_3 and λ are sufficiently small and if m is a metric on $X_0 \# X_1$ which is conformally $K_6 \cdot \sqrt{\lambda}$-close to the m_i, there are*

(i) *A $\Gamma_{A_0} \times \Gamma_{A_1}$ invariant neighborhood N of*

$$\mathrm{Hom}\big((P_0)_{x_0}, (P_1)_{x_1}\big) \times \{0\} \times \{0\} \subset \mathrm{Hom}\big((P_0)_{x_0}, (P_1)_{x_1}\big) \times H^1_{A_0} \times H^1_{A_1}$$

and

(ii) *A $\Gamma_{A_0} \times \Gamma_{A_1}$ equivariant map*

$$\Phi = (\phi_0, \phi_1) \colon N \to H^2_{A_0} \times H^2_{A_1}$$

such that the gauge equivalence classes of m-ASD connections over $X_0 \# X_1$, representable in standard form (4.16), *are parametrized by*

$$\Phi^{-1}(0)/\Gamma_{A_0} \times \Gamma_{A_1}.$$

A leading term in the local defining equation $\Phi = 0$ of Theorem (4.53) can be identified explicitly. In addition to the quadratic part in the parameter p_0, p_1 there is a new contribution, involving the bundle clutching parameter ρ, of precisely the same kind as that considered in [25, §5].

The family $A^{(\infty)}(\rho, p_0, p_1)$ is expressed in the form

$$\big(A_0 + \tilde{p}_0 + a_\infty,\ A_1 + \tilde{p}_1 + a'_\infty, \rho\big)$$

and $F_+(A^{(\infty)}) = \phi_0 + \phi_1$, where $\phi_i(\rho, p_0, p_1) \in V_{A_i}$. We will use the following bounds on the small deformations a_∞, a'_∞ (all can be proved readily by cutting off and using estimates for the elliptic equations over the compact manifolds, just as we have done many times above):

$$\begin{aligned} &\|a_\infty\|_{L^2}, \|a'_\infty\|_{L^2} \leqslant \text{const}\,\lambda^{3/2}, \\ &\|a_\infty\|_{L^1}, \|a'_\infty\|_{L^1} \leqslant \text{const}\,\lambda^2, \\ &|a_\infty(\xi)| \leqslant \text{const}\,\lambda^2/|\xi|^2 \quad \text{if } \sqrt{\lambda} \leqslant |\xi|, \\ &|a'_\infty(\eta)| \leqslant \text{const}\,\lambda^2/|\eta|^3 \quad \text{if } \sqrt{\lambda} \leqslant |\eta|. \end{aligned} \tag{4.54}$$

For simplicity suppose that the Riemannian metrics m_0, m_1 are flat in the identifying region, as in §IV(ii). Let ω be an element of $(\mathrm{Ker}\, d^*_{A_0} \subset \Omega^2_+(\mathfrak{g}_{P_0})) \cong H^2_{A_0}$ and S be the 3-sphere in X_0 with center x_0 and radius $\sqrt{\lambda}$. The value of the map ϕ_0 can be detected by the pairing with all such harmonic forms and

$$\langle \phi_0, \omega \rangle = -\int_{X_0} \mathrm{Tr}(\phi_0 \wedge \omega) = -\int_{X'_0} \mathrm{Tr}\big(F_+(A_0 + \tilde{p}_0 + a_\infty) \wedge \omega\big)$$

by (4.50). Here X_0' is the subset of X_0 obtained by removing the small ball interior to S, and we have used the fact that ϕ_0 is supported away from x_0. Now, since $\tilde{p}_0 = p_0 + O(|p_0|^2)$,

$$\langle \phi_0, \omega \rangle = -\int_{X_0'} \operatorname{Tr}\Big(\big(d^+_{A_0}(\tilde{p}_0 + a_\infty) + p_0 \wedge p_0\big) \wedge \omega\Big)$$
$$+ O\big(|p_0|^3\big) + O\big(\lambda^2 |p_0|\big) + O(\lambda^3).$$

The quadratic term in p_0 is the same, up to $O(|p_0|^2 \cdot \lambda^2)$, as that in the description of the moduli around $[A_0]$ itself. A new term comes from the fact that X_0' has boundary S, so, integrating by parts,

$$\int_{X_0'} \operatorname{Tr}\big(d^+_{A_0}(\tilde{p}_0 + a_\infty) \wedge \omega\big) = \int_S \operatorname{Tr}\big((\tilde{p}_0 + a_\infty) \wedge \omega\big).$$

The 3-sphere S lies within the annulus $U_0 \cap U_1$ where, relative to the local trivializations,

$$A_1 + \tilde{p}_1 + a'_\infty = \rho(A_0 + \tilde{p}_0 + a_\infty)\rho^{-1}.$$

Now $\int_S \operatorname{Tr}(A_0 \wedge \omega)$ is $O(\lambda^3)$ since the connection matrix A_0 satisfies

$$d^+ A_0 = O|A_0|^2 = O(\lambda) \quad \text{in } X_0 \setminus X_0'$$

and, in the local trivialization,

$$d\omega = O|A_0| = O(\sqrt{\lambda})$$

so we may throw the integral over S onto an integral over S onto an integral over $X_0 \setminus X_0'$ with uniformly small integrand. Thus

$$\int_S \operatorname{Tr}\big((\tilde{p}_0 + a_\infty) \wedge \omega\big) = \int_S \operatorname{Tr}\big((A_1 + \tilde{p}_1 + a'_\infty) \wedge \rho^{-1}\omega\rho\big) + O(\lambda^3).$$

To evaluate the last integral we regard S as a small 3-sphere in X_1, so A_1, $\tilde{p}_1$ are Lie algebra valued 1-forms defined over the interior $X_1 \setminus X_1'$ of S. All the 1-forms are $O(\sqrt{\lambda})$, the 3-volume of S is $O(\lambda^{3/2})$ so to find the integral to order λ^3 we need to know $\omega|_S$ to order λ. In the local coordinate system on X_0 (cf. (4.12)) we can write

$$\omega = \omega(x_0) + M(\xi) + O(\lambda),$$

where $\omega(x_0) \in \Lambda^2_+(T^*X_0)_{x_0} \otimes (\mathfrak{g}_{P_0})_{x_0}$ and $M\colon (TX_0)_{x_0} \to \Lambda^2_+(T^*X_0)_{x_0} \otimes (\mathfrak{g}_{P_0})_{x_0}$ is linear. We define a pair of 2-forms ω', ω' on a neighborhood of x_1 in X_1 both equal (to order λ) to $\rho^{-1}\omega\rho$ when restricted to S.

First, let $\omega'(\eta) = \rho^{-1}[\omega(x_0)^\sigma + M^\sigma\eta]\rho$. Here $\rho^{-1}\omega(x_0)^\sigma\rho$ is the *anti*-self-dual form on X_1 obtained from $\omega(x_0)$ by the orientation reversing map σ: $(TX_0)_{x_0} \to (TX_1)_{x_0}$ and the bundle identification ρ. Similarly M^σ is defined by conjugating M with σ. The form ω' is closed and since a point of S represented by ξ in one coordinate system is represented by $\sigma(\xi)$ in the other,

$$\omega'|_S = \rho^{-1}\big(\omega(x_0) + M(\xi)\big)\rho|_S.$$

Thus

$$\begin{aligned}\int_S &\operatorname{Tr}\big((A_1 + \tilde{p}_1) \wedge \rho^{-1}\omega\rho\big)\\ &= \int_S \operatorname{Tr}\big((A_1 + \tilde{p}_1) \wedge \omega'\big) + O(\lambda^3)\\ &= \int_{X\setminus X'} \operatorname{Tr}\big(d^-(A_1 + \tilde{p}_1) \wedge \omega'\big) \quad (\text{since } d\omega' = 0)\\ &= \int_{X\setminus X'} \operatorname{Tr}\big(F_-(A_1 + \tilde{p}_1) \wedge \omega'\big) + O(\lambda^3) + O\big(\lambda^2|p_1|\big).\end{aligned}$$

But

$$F_-(A_1 + \tilde{p}_1)(\eta) = F_-(A_1 + \tilde{p}_1)(x_1) + (\text{linear in } \eta) + O(\lambda),$$

and in the integrals over the 4-ball the contributions from linear terms in η vanish by symmetry. So

$$\begin{aligned}\int_S &\operatorname{Tr}\big((A_1 + \tilde{p}_1) \wedge \rho^{-1}\omega\rho\big)\\ &= (4\omega_3)\operatorname{Tr}\big(F_-(A_1 + \tilde{p}_1)(x_1) \cdot \rho^{-1}\omega^{\sigma}p\big) \cdot \lambda^2 + O(\lambda^3) + O\big(\lambda^2|p_1|\big).\end{aligned}$$

Second, let τ be the orientation *preserving* identification map $\tau(\eta) = \lambda\sigma^{-1}(\eta)/|\eta|^2$, and let ω'' be the closed self-dual form

$$\omega''(\eta) = \rho^{-1}\tau^*\big(\omega(x_0) + M(\xi)\big)\rho,$$

defined over the fixed region $0 < |\eta| \leqslant 1$, say. Thus $|\omega''(\eta)| \leqslant \text{const}\,\lambda^2/|\eta|^4$ and, integrating again,

$$\int_S \operatorname{Tr}\big(a'_\infty \wedge \rho^{-1}\omega\rho\big) = \int_{|\eta|=1} \operatorname{Tr}\big(a'_\infty \wedge \omega''\big)^+ \int_{\sqrt{\lambda} \leqslant |\eta| \leqslant 1} \operatorname{Tr}\big(a' \wedge d_{A'}\omega''\big).$$

The first-term is $O(\lambda^4)$ and since $d_{A'}(\omega'') \leqslant |A'| \cdot |\omega''| \leqslant \text{const}\,\lambda^2/|\eta|^3$, the second is bounded by a multiple of

$$\int_{\sqrt{\lambda}}^1 \eta^4/r\,dr = O(\lambda^3).$$

In sum,

$$\begin{aligned}\langle \phi_0(\rho, p_0, p_1), \omega\rangle &= q_\omega(\rho, p_0, p_1)\\ &\quad + O\big(\lambda^3 + |p_0|^3 + (|p_0| + |p_1|)\lambda^2\big),\end{aligned} \tag{4.55}$$

where

$$\begin{aligned}q_\omega(\rho, p_0, p_1) &= \langle p_0 \cap p_0, \omega\rangle\\ &\quad - 4(\omega_3) \cdot \lambda^2 \operatorname{Tr}\big(F(A_1 + \tilde{p}_1)(x_1)\rho^{-1}\omega^{\sigma}\rho\big).\end{aligned} \tag{4.56}$$

Rather than using this directly on connected sums we now end this long digression and return to the concentrated instanton connections which motivated the description in this section.

V. Algebraic models for the ends of moduli spaces

V(i). The discussion of §IV(iii)–(vi) extends in the obvious way to describe connections over connected sums $X_0\#(X_1\# \cdots \#X_l)$, where $X_1,\cdots, X_l$ are attached to X_0 by identifications around distinct points in X_0. In particular we can take X_0 to be the Riemannian 4-manifold X of §III and $X_1,\cdots, X_l$ to be 4-spheres. The connection A_0 is some ASD connection, when Chern class $k - l$, over X_0 and on each copy of S^4 we take the standard instanton with $c_2 = 1$. Then we are back in the setting of Proposition (4.11). It is easy to see that the metric on the connected sum which is defined to be conformally equivalent to the fixed metric on X is conformally (const $\bar{\lambda}$)-close to the standard metrics on the round 4-spheres, S_{x_i}. Here $\bar{\lambda} = \max_{1\leqslant i\leqslant l}\lambda_i$ is the basic measure of "maximal concentration." If $\bar{\lambda}$ is small, then the construction of §IV(iv) goes through to produce a "small" family of solutions to the ASD equations modulo the obstruction space $H^2_{A_0}$. The vector space H^2_I is zero so there are no obstructions from the 4-spheres.

Take the transversal maps $\pi_1,\cdots,\pi_l$, giving local coordinates on the moduli spaces over the S_{x_i}'s to be those defined by the local centers and scales on X. Transferring Theorem (4.53) to this framework we find some fixed $r > 0$ and a Γ_{A_0}-equivalent family of connections $A^\infty(n)$ parametrized by a manifold

$$
\begin{aligned}
N_r = &\left(\prod_{i=1}^{l} \operatorname{Hom}\big(E_{x_i},(W_+)_{X,x_i}\big)\right)\times\big(B_r(0)\subset H^1_A\big)\\
&\times\left(\prod_{i=1}^{l}\big(B_{\lambda_i r}(x_i)\subset X_0\big)\times\big((1-r)\lambda_i,(1+r)\lambda_i\big)\right).
\end{aligned}
\tag{5.1}
$$

Here x_i, λ_i are fixed, as in §IV. The local centers and scales of the concentrated connections move because of the parameters $p_1,\cdots,p_l$ in the S_{x_i} moduli spaces but the motion is restricted to be of size $r\lambda_i$ by the scaling in the construction.

We define a projection Π as in §IV(vi). Any ASD connection B represented in standard form with $\Pi(B)\in N_{r/2}$, say, is gauge equivalent to one of the $A^{(\infty)}(n)$. Since $n \mapsto \Pi A^{(\infty)}(n)$ defines a diffeomorphism close to the identity, there is in $N_{r/2}$ a copy of

$$
\left(\prod_{i=1}^{l}\operatorname{Hom}\big(E_{x_i}(W_{+,X})_{x_i}\big)\right)\times\big(B_{r/2}(0)\subset H^1_A\big)
$$

parametrizing connections with centers and scales equal to the x_i, λ_i. The ASD connections in this family correspond to the solutions of a $\Gamma_{A_0} \times (\pm 1) \times \cdots \times (\pm 1)$ equivariant equation: ($\phi = 0 \in H^2_{A_0}$). By Proposition (4.11) any concentrated ASD connection in W^-_λ with these centers and scales can be put in standard form and so is gauge equivalent to one in the family constructed.

Now we can let the centers x_i move over disjoint open sets $\Omega_i \subset X$ and the λ_i move in an interval $(0, \varepsilon)$ to get a "large family" parametrized by a manifold N. It is convenient to exploit the fact that the bundle identification spaces $\operatorname{Hom}(E_{x_i}, (W_{+,X})_{x_i})$ are copies of the 3-sphere so that the identification parameters and scales can be combined into a number of copies $G_{x_i} \setminus \{0\}$ of $\mathbb{R}^4 \setminus \{0\}$. If we adjoin an extra point in each factor corresponding to $\lambda_i = 0$ we get a parameter space

$$(5.2) \quad N \subset \overline{N} = \Omega_1 \times \cdots \times \Omega_l \times L_1 \times L_2 \times \cdots \times L_l \times \left(B_{r/2}(0) \subset H^1_{A_0}\right)$$

with symmetry group $\Gamma_{A_0} \times \{\pm 1\} \times \{\pm 1\} \times \cdots \times \{\pm 1\}$. The extra points in $\overline{N}$ give exactly the parameters needed to describe concentrated connections in the other strata, i.e., points of $\overline{W} \cap M_{k,j}$ for $j = l+1, \cdots, k$. Moreover the natural extension of the map Φ to $\overline{N}$, obtained by ignoring points whose associated scale λ_i is zero, is continuous as one sees by arguing in the manner of (4.55) (cf. Lemma (5.4) below).

Let ω be an element of $H^2_{A_0}$. The explicit "quadratic part," (4.56), of Φ simplifies, since over S^4 the curvature of the basic instanton is minus the identity map. Let $\operatorname{Tr}(\rho_i \omega(x_i))$ denote the natural pairing between

$$\omega(x_i) \in \Lambda^2_+(x_i) \otimes \mathfrak{g}_{P_0}(x_i)$$

and

$$s^2(\rho)\colon \mathfrak{g}_P(x_i) \to \Lambda^2_+(x_i).$$

Then put

$$(5.3) \qquad q_\omega(n) = 8\pi^2 \sum_{i=1}^{l} \lambda_i^2 \operatorname{Tr}\left(\rho_i \omega(x_i)\right) + \langle p_0 \wedge p_0, \omega \rangle.$$

We know that part of the end of M_k is parametrized by the zero set of Φ in N. Similarly, if some of the points x_i are removed we have other local models for the ends of the moduli spaces M_j ($j < k$). A priori these are unrelated to Φ but it is not hard to check that in fact the extension of Φ to $\overline{N}$ behaves nicely with respect to the stratification. A neighborhood of $([A_0], x_1, \cdots, x_l)$ in the topological space $\overline{M}_k$ is modelled on the zero set of Φ in $\overline{N}$, divided by $\Gamma_{A_0} \times (\pm 1) \times \cdots \times (\pm 1)$. Like Φ the functions $q_\omega(n)$ extend to $\overline{N}$ continuously; indeed for fixed centers x_i ($i = 1, \cdots, l$) the map q_ω is quadratic in the vector space parameters G_{x_i} and $H^1_{A_0}$, and the variation with the centers is smooth.

(5.4) Lemma. *The function* $\langle \Phi, \omega \rangle$ *on* $\bar{N}$ *is* C^1 *and if we put* $\langle \Phi, \omega \rangle = q_\omega + h$, *then*

(i) $|h|$ *is* $O(\bar{\lambda}^3 + |p_0|^3 + |p_0| \cdot \bar{\lambda}^2)$,

(ii) $|\partial h/\partial p_0|$ *is* $O(\bar{\lambda}^3 + |p_0|^2 + |p_0| \cdot \bar{\lambda}^2)$,

(iii) $|\partial h/\partial p_i|$, $|\partial h/\partial x_i|$, $\lambda_i|\partial h/\partial \lambda_i|$ *are* $O(\lambda_i^3 + |p_0|\lambda_i^2)$.

Proof. The details of this are left to the reader. First, to establish (i) the account at the end of §IV(vi) must be modified to take account of the small difference in metrics on the overlapping region. Second, the function Φ here differs slightly from that in §IV(vi) because of the changes $p_1, \cdots, p_l$ in the "small family" which were made to get connections with prescribed centers and scales but these are $O(\bar{\lambda}^2)$ which can be absorbed into the stated error term.

Third, get the derivative bounds in parts (ii), (iii) by arguing just as in §IV(vi) with the differentiated versions of (4.54) obtained as in Remark (4.24)(i). This is simplest for the derivatives $\partial h/\partial p_0$, $\partial h/\partial p_i$. To bound $\partial h/\partial x_i$ and $\partial h/\partial \lambda_i$ it is best to change our point of view—since translations of space do not define differentiable actions on L^p functions. Rather than explicitly changing the centers and scales, use a family of diffeomorphisms to keep the configuration R_1, R_{-1} fixed and vary, in compensation, the metric on the manifold and the connection A_0. Then differentiate through the construction scheme again. The bounds of part (iii) essentially imply that Φ is differentiable on $\bar{N}$, since the q_ω are.

To summarize, we have the following general description:

(5.5) Theorem. *A neighborhood of the point* $([A_0], x_1, \cdots, x_l)$ *in* $\overline{M}_k$ *is modelled on the quotient by* $\Gamma_{A_0} \times (\pm 1) \times (\pm 1) \times \cdots \times (\pm 1)$ *of the zero set of a map*

$$\Phi\colon \bar{N} \to V_{A_0} \cong H^2_{A_0},$$

stratified by the coordinate hyperplanes. Moreover for $\omega \in H^2_{A_0}$, $\langle \Phi, \omega \rangle$ *is approximated by* q_ω *in the sense of Lemma* (5.4).

Using this and standard arguments it is possible to read off a description of the manifolds themselves in many cases, given sufficient transversality in the algebraic model. We will deduce Propositions (3.21), (3.22) in this way using the next corollary.

(5.6) Corollary. *Suppose* $H^1_{A_0} = 0$ *and the points* $x_1, \cdots, x_l$ *are such that the zero set* $\tilde{L}$ *of the associated projective equations*

$$(q_\omega = 0)_{\omega \in H^2_{A_0}} \subset \mathbb{P}(G_{x_1} \times \cdots \times G_{x_l})$$

is cut out transversally. Then if $\Omega_1, \cdots, \Omega_l$ *are small, a neighborhood of* $([A_0], x_1, \cdots, x_l)$ *in* $\overline{M}_k$ *has the form of the quotient by* $\Gamma_{A_0} \times \{\pm 1\} \times \{\pm 1\} \times \cdots \times \{\pm 1\}$ *of*

$$(\text{cone over } L) \times \Omega_1 \times \cdots \times \Omega_l,$$

where the Ω_i *factors represent the local centers in* X.

This is an immediate consequence of (5.4) and (5.5). If the centers are fixed and the maps are restricted to small spheres ($\Sigma \lambda_i^2 =$ constant), then Φ is close in C^1 to the map obtained from the q_ω's under the linear isomorphism $V_{A_0} = H^2_{A_0}$ defined by L^2 projection. The latter vanishes transversally so the zero sets are isotopic in the ambient sphere by a small projection.

V(ii). If the connection A_0 is the flat product connection θ, then the model equations (5.3) take a particularly simple form. The harmonic forms in $\Omega^2_+(\mathfrak{g}_{P_0}) = \Omega^2_+ \otimes SU(2)$ are tensor products of the ordinary self-dual harmonic forms with constant sections of $SU(2)$ and the identification parameters ρ_i and scales λ_i define maps

$$R_i = \lambda_i^2 s^2(\rho_i)^{-1}\colon \left(\Lambda^2_{+,X}\right)_{x_i} \to SU(2).$$

The system of equations $q_\omega = 0$ (5.3) is equivalent to the equation

$$\sum_{i=1}^{l} R_i(\omega(x_i)) = 0, \qquad \omega \in \mathscr{H}^2_+(X) \subset \Omega^2_+(X). \tag{5.7}$$

Suppose that there are no solutions of this equation with any of the R_i vanishing (that is $\lambda_i = 0$ except for the obvious case when all are zero. Then the only strata $M_j \times S^{l-j}(X)$ meeting a small neighborhood of $([\theta], x, \cdots, x_l)$ in $\overline{M}_k$ are M_l and $[\theta] \times S^l(X)$. Choosing bases, if we normalize the last variable R_l to be 1 we get simultaneously *a* transversal to the $\Gamma_\theta \cong SU(2)$ symmetry and to the scale invariance $\lambda_i \to t\lambda_i$. Provided the zero set L of these normalized algebraic equations is cut out transversally, then by Corollary (5.6) the link of the symmetric product close to $(x_1, \cdots, x_l)$ in the moduli space will be modelled on L. It is now very straightforward to find what this link should be in the two cases $l = 2$, $l = 3$ of Propositions (3.21), (3.22). In each case we have to assume some general position properties of the harmonic forms to get a good link and in the next section we check that these can always be supposed to hold. Then the proofs of Theorems B and C will be complete.

Case 1. $b_2^+(X) = 1$, $l = 2$. The self-dual harmonic forms are spanned by a single element ω and the points x_1, x_2 are in good position if ω does not vanish at either. Under this assumption the link L is a circle, for the normalized equations (5.7) take the shape

$$R_1(\omega(x_1)) = -\omega(x_2).$$

The solution set corresponds to a translate in $SO(3)$ of a 1-parameter subgroup $\gamma = \mathrm{Stab}(\omega(x_1))$ which represents a generator of $\pi_1(SO(3)) \cong \mathbb{Z}/2$. Hence if x_1, x_2 are in good position we deduce Proposition (3.21) from Lemma (3.25)(i) since $u_1(L) = t_1(\gamma) = 1$.

Case 2. $b_2^+(X) = 2$, $l = 3$. If ω_1, ω_2 span $\mathscr{H}_+^2(X)$, then the points x_1, x_2, x_3 are in good position relative to the harmonic forms so long as no two of the pairs of vectors $(\omega_1(x_i), \omega_2(x_i))_{i=1,2,3}$ are homothetic. In this case there are no solutions with vanishing scales. We show that this condition also implies that the link is a torus, homologous to a product $\gamma_1 \times \gamma_2$ of generators γ_i in the $\mathbb{R}^+ \times SO(3)$ factors. Proposition (3.22) follows from this since

$$\begin{aligned} u_2(L) &= t_1 t_2(\gamma_1 \times \gamma_2) \quad \text{(by Lemma (3.25)(ii))} \\ &= t_1(\gamma_1) t_2(\gamma_2) = 1. \end{aligned}$$

Write the equations (5.7), relative to bases, as

$$M_1 v_1 + M_2 v_2 + v_3 = 0, \qquad M_1 w_1 + M_2 w_2 + w_3 = 0 \tag{5.8}$$

(so $v_i, w_i \in \mathbb{R}^3$ and $M_1, M_2 \in \mathbb{R}^+ \times SO(3)$). Consider first the special choice of vectors

$$\begin{aligned} v_1 &= e_1, \quad v_2 = 0, \quad v_3 = -e_1, \\ w_1 &= 0, \quad w_2 = e_2, \quad w_3 = -e_2. \end{aligned} \tag{5.9}$$

Then the equations (5.8) become $M_1 e_1 = e_1$, $M_2 e_2 = e_2$, and define a torus $\gamma_1 \times \gamma_2$. The set T of vectors $(v_1, v_2, v_3; w_1, w_2, w_3)$ such that two pairs (v_i, w_i) are homothetic has codimension 2, hence it does not separate the space $(\mathbb{R}^3)^6$ of these "coefficients" in the algebraic equations for M_1, M_2. So if the equations are everywhere regular—for "good" coefficients not in T—the general result will follow by continuity.

This last calculation is very like that in [25, Lemma 8.5]. Fix v_1, v_2, w_1, w_2 in $\mathbb{R}^3$ and consider the map

$$F\colon (SO(3) \times \mathbb{R}^+)^2 \to \mathbb{R}^3 \times \mathbb{R}^3$$

$$F(M_1, M_2) = \begin{pmatrix} M_1 v_1 + M_2 v_2 \\ M_1 w_1 + M_2 w_2 \end{pmatrix}.$$

We must show that if (v_1, w_1), (v_2, w_2) are not homothetic, then F is everywhere of maximal rank. Without loss consider the point M_1, $M_2 = 1$ (using the symmetry of the problem). Here the differential of F is represented by

$$dF\colon \mathbb{R} \times \mathbb{R}^3 \times \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3 \times \mathbb{R}^3,$$

$$\begin{aligned} dF(h_1, y_1, h_2, y_2) = (&h_1 v_1 + y_1 \times v_1 + h_2 v_2 + y_2 \times v_2, \\ &h_1 w_1 + y_1 \times w_1 + h_2 w_2 + y_2 \times w_2) \end{aligned}$$

with adjoint

$$(dF)^*(p,q) = \big(v_1 \cdot p + w_1 \cdot q,\, v^1 \times p + w_1 \times q,\\ v_2 \cdot p + w_2 \cdot q,\, v_2 \times p + w_2 \times q\big).$$

It is an easy exercise to show that for any pair of vectors p, q in $\mathbb{R}^3$, not both zero, all the solutions of the simultaneous equations

$$v \cdot p + w \cdot q = 0,\ v \times p + w \times q = 0$$

are mutually homothetic. Thus since (v_1, w_1), (v_2, w_2) are not homothetic, dF is surjective and our algebraic equations are of maximal rank.

VI. Variation of harmonic forms

To finish the proofs of Theorems B and C we check that the moduli spaces satisfy two genericity properties. For generic Riemannian metrics on a manifold whose intersection form is negative definite Freed and Uhlenbeck [17, §3] prove that the moduli spaces are smooth manifolds except for the presence of quotient singularities, inherited from the ambient space $\mathscr{B}_X$, which are associated to reductions of the bundle. We have seen in §III(ii) how these reductions fit into our cohomology point of view. There is a simple reason why the corresponding singularities are absent for typical Riemannian metrics on manifolds with indefinite forms. This can be understood in terms of the "periods" of the harmonic representatives of the two-dimensional cohomology, and one should expect these reductions to have some topological significance in *families* of metrics of dimension b_{+2} (cf. [13]).

Similarly, in §V(ii) the "links" in the moduli spaces were described under the assumption that the pointwise values of these harmonic forms were in good position. We check that this will be true for typical metrics and points. Discussions of both these properties exist already: Freed and Uhlenbeck's transversality theorem proves the absence of reductions and there is material on the variation of the pointwise values of the harmonic forms in [27].

Note that these general position properties are not especially significant for our topological deductions since the "explicit" models (e.g., Theorem (5.54)) can always be perturbed by hand without affecting the homology discussion. But it is tidier not to have do this.

Consider a Riemannian 4-manifold X with $b_2^+(X) = p$ and so a p-dimensional subspace $\mathscr{H}_+^2 \subset H^2(X; \mathbb{R})$ consisting of the cohomology classes whose harmonic representative, relative to the metric, is self-dual. If L is a line bundle over X the unique Yang-Mills connection on L has curvature the harmonic representative of $c_1(L)$. The line bundle admits an ASD connection if and only if $c_1(L)$ lies in the subspace $\mathscr{H}^2_-$; the annihilator of $\mathscr{H}^2_+$ under the fixed cup product form on H^2. If $\mathscr{H}^2_-$ contains no points in the integer lattice $H^2(X;\mathbb{Z})/\text{Torsion} \subset H^2(X;\mathbb{R})$ (except 0), then no ASD connection on an $SU(2)$ bundle with $c_2 > 0$ can reduce to an S^1 bundle. More precisely there are no self-dual reductions of a bundle with $c_2 = h$ so long as $\mathscr{H}^2_-$ contains no lattice points e with $e^2 = -h$.

The harmonic subspace $\mathscr{H}^2_+$ defines a "period map" P from the space of conformal structures on X to the open subset G of the Grassmannian $\mathrm{Gr}_p(H^2)$ consisting of p-planes on which the cup product form is positive definite. Each lattice point e ($e^2 < 0$) defines a codimension p submanifold $W_e \subset G$ made up of the p-planes annihilating e, and the union of the W_e with e^2 fixed is locally finite in G. We will show that the period map is transverse to the submanifolds W_e. It then follows in a routine way that if the form is indefinite any metric can be perturbed slightly to avoid singularities from reductions and similarly that if, say, $b_2^+(X) = 1$, then any path of metrics can be perturbed so that reductions occur at a discrete set of points (cf. [13]).

To verify this transversality we compute the derivative of the period map P. A conformal structure on X is specified by its self-dual subbundle

$$\Lambda^2_+ \subset \Lambda^2_X, \qquad \alpha \wedge \alpha > 0 \quad \forall \alpha \in \Lambda^2_+.$$

Relative to a fixed base metric the space of conformal structures is identified—via the graph of the map—with bundle maps

$$\mu\colon \Lambda^2_+ \to \Lambda^2_-, \qquad |\mu(\alpha)| < |\alpha|. \tag{6.1}$$

So the tangent space to the conformal structures at the given point is the vector space $\mathrm{Hom}(\Lambda^2_+, \Lambda^2_-)$ (cf. [11], [27]).

Similarly, the tangent space to $G \subset \mathrm{Gr}_p(H_2)$ at the corresponding point is the space $\mathrm{Hom}(\mathscr{H}^2_+, \mathscr{H}^2_-)$ and if the point lies in W_e, so $e \in \mathscr{H}^2_-$, then the normal bundle of W_e is identified with $\mathscr{H}^2_+$. The quotient map $TG \to NW_e$ is $M \to M^*(e)$.

(6.2) Lemma. *The derivative of the period map at this point is*

$$dP\colon \mathrm{Hom}(\Lambda^2_+, \Lambda^2_-) \to \mathrm{Hom}(\mathscr{H}^2_+, \mathscr{H}^2_-), \qquad dP(\mu)(\omega) = \pi_-\mu(\omega),$$

where $\pi_-\colon \Omega^2_- \to \mathscr{H}^2_-$ *is* L^2 *projection.*

The transversality of P and W_e follows directly from this and the fact that a harmonic form which vanishes on an open set is identically zero [2]. The composite of dP with the quotient map to the normal bundle is

$$\mu \to \pi_+(\mu^* e)$$

and an element $\omega \in \mathscr{H}_+^2$ is in the cokernel if and only if

$$\langle \mu^* e, \omega \rangle = \langle \mu, e \otimes \omega \rangle = 0 \quad \forall \mu$$

which implies $\omega = 0$.

The calculation for Lemma (6.2) can be combined with that for the pointwise variation of harmonic forms. Let S: $\Omega_-^2 \to \Omega_+^2$ be the operator, defined with respect to a given metric

$$S(\alpha) = G_+ \left(d_+^* d \right) \alpha \tag{6.3}$$

where d_+^*: $\Omega^3 \to \Omega_+^2$ and G_+ is the Greens operator for $d_+^* d$ on Ω_+^2. S is a conformally invariant singular integral operator. As the conformal structure varies the self-dual harmonic forms vary, both globally as a subspace in H^2 and pointwise as sections of Λ_-^2. Using L^2 orthogonal projection to identify the $\mathscr{H}_+^2$ for nearby metrics with that for a fixed base metric, and similarly fiberwise projection to identify the Λ_+^2 subspaces, the derivative of the SD harmonic forms pointwise can be interpreted as a map

$$\operatorname{Hom}\left(\Lambda_+^2, \Lambda_-^2 \right) \to \operatorname{Hom}\left(\mathscr{H}_+^2, \Lambda_+^2 \right).$$

(6.4) Lemma. *The derivative of the projections of the pointwise values of the SD harmonic forms is*

$$\mu \to (\omega \to -S \circ \mu(\omega)).$$

To establish (6.2) and (6.4) let $t \cdot \mu$ define a 1-parameter family of conformal structures and ω be a self-dual harmonic form with respect to the base point $t = 0$. We find a lifted family ω_t of closed 2-forms such that $\omega_0 = \omega$, $\dot{\omega}_t|_{t=0} \perp \mathscr{H}_+^2$ and $*_t \omega_t = \omega_t$. Thus, differentiating and evaluating at $t = 0$:

$$*_0 \dot{\omega} + \dot{*}_0 \omega_0 = \dot{\omega}_0.$$

But $\dot{*}_0 = \mu$ and $\omega_0 = \omega$ so

$$\dot{\omega}_0 = \mu\omega - \phi$$

with $\phi \in \Omega_+^2$, $\phi \perp \mathscr{H}_+^2$. The variation in the periods $dP(\mu)(\omega)$ is $\pi(\dot{\omega}_0) = \pi(\mu\omega)$ while the variation in the projection of the pointwise values is $-\phi$. Since the ω_t are closed;

$$d\phi = d(\mu\omega)$$

and since $\phi \perp \mathscr{H}_+^2$ it follows that

$$\phi = S(\mu\omega).$$

Consider first the case $b_2^+(X) = 1$. It is obvious that the self-dual harmonic form ω is nonzero on some dense open set, so there is no problem in choosing a pair of points in X to get the good link of §V(ii). We will see that moreover, for generic metrics, ω vanishes transversally on a 1-dimensional submanifold in X. This will be the case if the "universal" section Ω of the bundle $\pi_1^*(\Lambda^2_+)$ over the product

$$X \times \{\text{conformal structures}\}$$

vanishes transverally. Here we are using L^2 projection to identify harmonic spaces $\mathscr{H}_+^2$ for nearby metrics and pointwise projection to identify the Λ^2_+ bundles, as in Lemma (6.4). Then the derivative of Ω in the μ direction at a point x is, by (6.4),

$$(\nabla\Omega)_x \colon \Omega^2_+ \otimes \Omega^2_S - \to (\Lambda^2_+)_x$$
$$\mu \to -(S \circ \mu(\omega))(x)$$

which is surjective. If $\alpha \in (\Lambda^2_+)_x$ were perpendicular to the image and δ_α denotes the corresponding distributional section of Λ^2_+, then every section μ of $\Lambda^2_+ \otimes \Lambda^2_-$ would be perpendicular to $\omega \otimes S^*(\delta_\alpha)$, which is absurd.

Similarly, if $b_2^+(X) = 2$, fix a base ω_1, ω_2 for $\mathscr{H}_+^2$ and consider the section Ω of the bundled

$$\pi_1^*(\Lambda^2_+ \oplus \Lambda^2_+) \oplus \pi_2^*(\Lambda^2_+ \oplus \Lambda^2_+)$$

over $X \times X \times \{\text{conformal structures}\}$. In each fiber of this $(\mathbb{R}^3)^4$ bundle there is an algebraic subvariety of codimension 2 (with singularities in codimension 8) representing homothetic pairs of vectors. If we know that for all (x, y) in $X \times X$ the map

$$(\nabla\Omega)_{x,y} \colon \Omega^2_+ \otimes \Omega^2_- \to (\Lambda^2_+)^2_x \oplus (\Lambda^2_+)^2_y$$

is surjective, it will follow that for generic metrics the set of (x, y) which do not lead to a good link form a codimension 2 set in $X \times X$ (having isolated singularities). In particular, we can always choose the surfaces Σ_i in Proposition (3.22) to get good links over all configurations of intersection points.

Elements $\alpha_1, \alpha_2 \in (\Lambda^2_+)_x$, $\beta_1, \beta_2 \in (\Lambda^2_+)_y$ are perpendicular to $(\nabla\Omega)_{x,y}\mu$ if

$$\left\langle \mu, \omega_1 \otimes \left[S^*(\delta_{\alpha_1} + \delta_{\alpha_2})\right] + \omega_2 \otimes \left[S^*(\delta_{\beta_1} + \delta_{\beta_2})\right]\right\rangle = 0.$$

The reader will be able to show that this cannot happen for all μ, using the unique continuation theorem.

Acknowledgment. The bulk of this work was carried out while the author was a member of the Institute for Advanced Studies, Princeton, whose generous hospitality is gratefully acknowledged. The author also thanks particularly M. Atiyah, R. Bott, M. Freedman, D. Sullivan, C. Taubes, and L. Taylor for encouraging discussions and suggestions.

Appendix

The motivation for this work was not so much to prove negative results on nonexistence (or nonsmoothability) of manifolds as to understand what features suddenly change as the dimension b_2^+ grows. Here we explain why the scheme of proof used for Theorems A, B, C breaks down when $b_2^+ \geqslant 3$.

Extrapolating from these proofs one could consider a bundle with $c_2 = k$ over a simply connected manifold with $b_2 = k - 1$. We represent $2k$ homology classes by surface the Σ_i and cut the Yang-Mills moduli space M_k by the associated submanifolds V_{Σ_i} representing $\mu(\Sigma_i)$ to get a manifold N of dimension $k + 1$. Suppose the metric on X can be chosen so that there are $Q_{2k}(\Sigma_1, \cdots, \Sigma_{2k})$ (mod 2) ends of N corresponding to configurations of intersection points whose "links" all carry the same nonzero homology class in the space of connections. Then *if there were no other ends of* N we deduce that the intersection form is a sum of $k - 1$ copies of the hyperbolic form.

This last point is the crucial one and it is here that the picture changes as b_2^+ grows. The ends of N can be covered by open sets W associated to all the intermediate strata:

$$M_{k-1} \times X,\ M_{k-2} \times s^2(X), \cdots, s^k(X).$$

We can avoid any stratum $M_{k-l} \times s^l(X)$ so long as M_{k-l} does not simultaneously intersect any $2k - 2l$ of the V_Σ's, which can be achieved by general position if

$$\dim M_{k-l} = 5k - 8l < 2(2k - 2l), \quad \text{i.e. if } k < 4l.$$

This is the case if $b_2^+ = 0, 1, 2$ (so $k < 4$); but for larger values of k there may be additional ends of N associated to the intermediate strata and the homology contribution of these is a new unknown.

Consider, for example, the manifold $X = K\#(s^2 \times s^2)$ with intersection form $-2E_8 + 4\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. A choice of Riemannian metric on X defines harmonic forms and moduli spaces of ASD connections. By using a theorem of Freedman and Taylor [18] one can show that there exists a metric on X and ten surfaces Σ_i such that $Q_{10}(\Sigma_1, \cdots, \Sigma_{10}) \neq 0$ but the links of the moduli space M_5 over the configurations of intersection points are each homologous to a product $\gamma_1 \times \gamma_2 \times \gamma_3 \times \gamma_4$. Here the γ_1 represent 1-parameter subgroups in the $SO(3)$ parameters, as in §V(ii), and we know by Lemma (3.25)(ii) that this product carries nonzero homology since $\langle u_4, \gamma_1 \times \cdots \times \gamma_4 \rangle = 1$.

We deduce that for such a metric and representatives Σ_i of the homology classes the moduli space M_4 must simultaneously intersect 8 of the V_{Σ_i}'s, say $\sigma_{ij} = M_4 \cap V_{\Sigma_1} \cap \cdots \cap \hat{V}_{\Sigma_i} \cap \cdots \cap \hat{V}_{\Sigma_j} \cap \cdots \cap V_{\Sigma_{10}}$. If the intersections are transverse the σ_{ij} are 1-dimensional submanifolds and our description of §V(i) shows that the end of N associated to σ_{ij} is a fibration over σ_{ij} with fiber $SO(3) \times \mathbb{R}^+$. One can deduce from the equation $\langle u_4, \partial N \rangle = 0$ that $[\Sigma \sigma_{ij}]$ is nonzero in $H^1(\mathscr{B}_x)$ for these surfaces and this metric.

In certain situations combinations of topological invariants from the harmonic forms and moduli spaces give differentiable invariants of the underlying 4-manifold [13]. It seems that the natural generalizations of our Theorems A, B, C give constraints on these invariants rather than on the homology of the manifold.

References

[1] L. Ahlfors & L. Sario, *Riemann surfaces*, Princeton University Press, Princeton, NJ, 1960.
[2] N. Aronszajn, *A unique continuation theorem for solutions of elliptic partial differential equations of the second order*, J. Math. Pures Appl. **36** (1957) 235–249.
[3] M. F. Atiyah, *Bott periodicity and the index of elliptic operators*, Quart. J. Math. (2) **19** (1968) 113–40.
[4] ______, *Global theory of elliptic operators*, Proc. Conf. Functional Analysis, Tokyo, 1969, 21–31.
[5] M. F. Atiyah & R. Bott, *The Yang-Mills equations over Riemann surfaces*, Philos. Trans. Roy. Soc. London Ser. A **308** (1982) 523–615.
[6] M. F. Atiyah, R. Bott & A. Shapiro, *Clifford modules*, Topology **3** (1964) 3–38.
[7] M. F. Atiyah, N. J. Hitchin & I. M. Singer, *Self duality in four dimensional Riemannian geometry*, Proc. Roy. Soc. London Ser. A **362** (1978) 425–461.
[8] M. F. Atiyah & J. D. S. Jones, *Topological aspects of Yang-Mills theory*, Comm. Math. Phys. **61** (1978) 97–118.
[9] M. F. Atiyah & I. M. Singer, *The index of elliptic operators*. I, III, IV, V, Ann. of Math. **87** (1968) 484–530, 546–604; **93** (1971) 119–149.
[10] ______, *Dirac operators coupled to vector potentials*, Proc. Nat. Acad. Sci. U.S.A. **81** (1984) 2597.
[11] A. Besse, *Géométrie Riemannienne en dimension* 4, CEDIC Fernand Nathan, Paris 1981.
[12] S. K. Donaldson, *An application of gauge theory to four dimensional topology*, J. Differential Geometry **18** (1983) 279–315.
[13] ______, *La topologie differentielle des surfaces complexes*, C. R. Acad. Sci. Paris Sér. I Math., **301** (1985) 317–320.
[14] ______, *The orientation of Yang-Mills moduli spaces and* 4-*manifold topology*, to appear.
[15] ______, *Polynomial invariants for smooth* 4-*manifolds*, in preparation.
[16] R. Fintushel & R. J. Stern, $SO(3)$-*connections and the topology of* 4-*manifolds*, J. Differential Geometry **20** (1984) 523–539.
[17] D. S. Freed & K. K. Uhlenbeck, *Instantons and* 4-*manifolds*, Math. Sci. Res. Inst. Publ., Springer, New York, 1984.
[18] M. H. Freedman & L. Taylor, Λ-*splitting* 4-*manifolds*, Topology **16** (1977) 181–184.
[19] F. Hirzebruch & H. Hopf, *Felder von Flächenelement in* 4-*dimensionalen Mannigfaltikkeiten*, Math. Ann. **136** (1956) 156–172.
[20] M. Itoh, *Self-dual Yang-Mills equations and Taubes' Theorem*, Tsukaba J. Math. **8** (1984) 1–29.
[21] M. Kuranishi, *A new proof for the existence of locally complete families of complex structures*, Proc. Complex Analysis (Minneapolis), Springer, Berlin, 1964, 142–56.

[22] S. Lang, *Differential manifolds*, Addison-Wesley, Reading, MA, 1972.

[23] J. P. Serre, *A course in arithmetic*, Springer, Berlin, 1973.

[24] C. H. Taubes, *Self dual connections on non self-dual manifolds*, J. Differential Geometry **17** (1982) 139–170.

[25] ______, *Self dual connections on manifolds with indefinite intersection matrix*, J. Differential Geometry **19** (1984) 517–60.

[26] ______, *Monopoles and maps from S^2 to S^2: the topology of the configuration space*, Comm. Math. Phys. **95** (1984) 345–391.

[27] ______, preprint, 1983 (Preliminary version of [25]).

[28] K. K. Uhlenbeck, *Removable singularities in Yang-Mills fields*, Comm. Math. Phys. **83** (1982) 11–30.

UNIVERSITY OF OXFORD, ENGLAND