CONTENTS

1. FOUR-DIMENSIONAL MANIFOLDS

1.1. **Yang-Mills Equations and BPST Instantons.** In this section we will introduce the *Yang-Mills equations* and the *anti-self-dual equations* on $\mathbb{R}^4$ and then describe the simplest example of an *instanton*. This is the so-called *BPST instanton*, or *pseudoparticle*, first introduced in [BPST]. We will do this in some detail in the hope of motivating the deep work of Donaldson [Don1], [Don2] on the application of gauge theory to the topology of smooth 4-manifolds. We will turn to the work of Donaldson in Sections 1.2 and 1.4.

1.1.1. *The Quaternionic Hopf Bundle.* Generally we will identify $\mathbb{R}^4$ as a real vector space and as a C^∞-manifold with the division algebra $\mathbb{H}$ of quaternions. Sp(1) will denote the Lie group of unit quaternions (those $g \in \mathbb{H}$ with $\|g\| = 1$). As a submanifold of $\mathbb{H}$, Sp(1) is diffeomorphic to the 3-sphere $\mathbb{S}^3$. It is also isomorphic to the Lie group SU(2) of 2×2 complex matrices U that are unitary ($U^{-1} = \overline{U}^T$) and satisfy $\det U = 1$. Indeed, every such U can be written uniquely in the form $U = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}$, where $a = a^1 + a^2\mathbf{i}$ and $b = b^1 + b^2\mathbf{i}$ are complex numbers satisfying $|a|^2 + |b|^2 = 1$. Then the map

$$\begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} \mapsto a + b\mathbf{j} = a^1 + a^2\mathbf{i} + (b^1 + b^2\mathbf{i})\mathbf{j} = a^1 + a^2\mathbf{i} + b^1\mathbf{j} + b^2\mathbf{k}$$

is a Lie group isomorphism. We will allow ourselves the luxury of adopting whichever view of this Lie group is most convenient in any given situation. We can therefore view its Lie algebra in one of two ways. It is isomorphic to the Lie algebra $\operatorname{Im}\mathbb{H}$ of pure imaginary quaternions with the commutator bracket $[x, y] = xy - yx = 2\operatorname{Im}(xy)$ determined by quaternion multiplication and also to the Lie algebra $\mathfrak{su}(2)$ of 2×2 complex matrices A that are skew-Hermitian ($\overline{A}^T = -A$) and tracefree ($\operatorname{tr} A = 0$) with the matrix commutator bracket $[A, B] = AB - BA$. A convenient basis for $\mathfrak{su}(2)$ consists of $T_j = -\frac{1}{2}\mathbf{i}\,\sigma_j$, $j = 1, 2, 3$, where $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_2 = \begin{pmatrix} 0 & -\mathbf{i} \\ \mathbf{i} & 0 \end{pmatrix}$ and $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ are the Pauli spin matrices. $\{T_1, T_2, T_3\}$ is an orthonormal basis relative to the inner product $\langle A, B \rangle = -2\operatorname{tr}(AB)$ on $\mathfrak{su}(2)$.

Now we begin our investigation by recalling the construction of the *quaternionic Hopf bundle*

$$\mathrm{Sp}(1) \hookrightarrow \mathbb{S}^7 \xrightarrow{\pi} \mathbb{HP}^1.$$

We identify the 7-sphere $\mathbb{S}^7$ with the subset of $\mathbb{H}^2$ consisting of all pairs $p = (q^1, q^2)$ of quaternions satisfying $\|q^1\|^2 + \|q^2\|^2 = 1$. Then we can define a smooth right action $\sigma : \mathbb{S}^7 \times \mathrm{Sp}(1) \to \mathbb{S}^7$ of Sp(1) on $\mathbb{S}^7$ by $\sigma(p, g) = p \cdot g = (q^1, q^2) \cdot g = (q^1 g, q^2 g)$. This action is free and the orbits $p \cdot \mathrm{Sp}(1)$ are submanifolds of $\mathbb{S}^7$ diffeomorphic to $\mathbb{S}^3$. Define an equivalence relation on $\mathbb{S}^7$ that identifies two points if and only if they are on the same orbit and denote the equivalence class of $(q^1, q^2) \in \mathbb{S}^7$ by $[q^1, q^2]$. The orbit space $\mathbb{S}^7/\mathrm{Sp}(1)$ consisting of all of these equivalence classes is, by definition, the *quaternionic projective line* $\mathbb{HP}^1$. We let $\pi : \mathbb{S}^7 \to \mathbb{HP}^1$ be the natural projection $\pi(q^1, q^2) = [q^1, q^2]$ and provide $\mathbb{HP}^1$ with the quotient topology determined by π. $\mathbb{HP}^1$ is provided with a differentiable structure determined by the atlas consisting of two charts (U_k, φ_k), $k = 1, 2$, defined as follows.

$$\begin{aligned} U_k &= \{\, [q^1, q^2] \in \mathbb{HP}^1 : q^k \neq 0 \,\}, \quad k = 1, 2 \\ \varphi_k &: U_k \to \mathbb{H}, \quad k = 1, 2 \\ \varphi_1([q^1, q^2]) &= q^2 (q^1)^{-1} \\ \varphi_2([q^1, q^2]) &= q^1 (q^2)^{-1} \end{aligned}$$

Notice that each U_k covers all but one point of $\mathbb{HP}^1$, but φ_k maps U_k *onto* $\mathbb{H}$. Relative to this differentiable structure the map $\pi : \mathbb{S}^7 \to \mathbb{HP}^1$ is smooth. Moreover, each of the maps

$$\Psi_k : \pi^{-1}(U_k) \to U_k \times \mathrm{Sp}(1), \quad k = 1, 2,$$
$$\Psi_k(p) = (\pi(p), \psi_k(p)), \quad k = 1, 2,$$

where

$$\psi_k(p) = \psi_k(q^1, q^2) = q^k / \|q^k\|, \quad k = 1, 2,$$

is a local trivialization. Specifically, each is a diffeomorphism that is equivariant with respect to the given right action of Sp(1) on $\mathbb{S}^7$ and the natural right action of Sp(1) on $U_k \times \mathrm{Sp}(1)$, that is,

$$\Psi_k(p \cdot g) = (\pi(p), \psi_k(p)g), \quad k = 1, 2.$$

Consequently, $\mathbb{S}^7$ is a principal Sp(1)-bundle over $\mathbb{HP}^1$ (see Chapter I, Section 5, page 50, of [KN1]) which we denote symbolically by $\mathrm{Sp}(1) \hookrightarrow \mathbb{S}^7 \xrightarrow{\pi} \mathbb{HP}^1$. It is called the *quaternionic Hopf bundle*.

Remark 1.1. $\mathbb{HP}^1$ is diffeomorphic to the 4-sphere $\mathbb{S}^4$. One can see this in the following way. Let (U_S, φ_S) and (U_N, φ_N) be the charts on $\mathbb{S}^4$ corresponding to stereographic projection from the north and south poles of $\mathbb{S}^4$, respectively. If $\overline{\varphi}_1$ denotes the map $\overline{\varphi}_1([q^1, q^2]) = \overline{\varphi_1([q^1, q^2])}$, then $\varphi_S^{-1} \circ \varphi_2$ and $\varphi_N^{-1} \circ \overline{\varphi}_1$ are diffeomorphisms of $\mathbb{HP}^1$ minus a point to $\mathbb{S}^4$ minus a point. On the intersection of their domains they agree so they determine a global diffeomorphism of $\mathbb{HP}^1$ onto $\mathbb{S}^4$. Composing with $\pi : \mathbb{S}^7 \to \mathbb{HP}^1$ gives a principal Sp(1)-bundle

$$\mathrm{Sp}(1) \hookrightarrow \mathbb{S}^7 \xrightarrow{\pi_1} \mathbb{S}^4$$

over $\mathbb{S}^4$ which is also often called the quaternionic Hopf bundle. Some care should be exercised, however, since reversing the roles of φ_1 and φ_2 gives another identification of $\mathbb{HP}^1$ with $\mathbb{S}^4$, but the corresponding Sp(1)-bundle over $\mathbb{S}^4$ is *not* equivalent to the one we just described. This is most readily shown by computing their Chern numbers which turn out to be 1 in the first case and -1 in the second.

1.1.2. *Yang-Mills Connections on the Hopf Bundle.* Now fix a point $p = (q^1, q^2) \in \mathbb{S}^7 \subseteq \mathbb{H}^2$. The orbit of our Sp(1)-action containing p (that is, the fiber of $\pi : \mathbb{S}^7 \to \mathbb{HP}^1$ above $\pi(p) = [q^1, q^2]$) is a submanifold of $\mathbb{S}^7$ diffeomorphic to $\mathbb{S}^3$. The subset of the tangent space $T_p(\mathbb{S}^7)$ to $\mathbb{S}^7$ at p consisting of tangent vectors to this fiber is called the *vertical space* at p and is denoted $\mathrm{Vert}_p(\mathbb{S}^7)$. It is a 3-dimensional linear subspace of $T_p(\mathbb{S}^7)$ which, in turn, can be identified with a linear subspace of $T_p(\mathbb{H}^2) \cong T_p(\mathbb{R}^8) \cong \mathbb{R}^8$. Thus, relative to the usual Euclidean inner product on $\mathbb{R}^8$, $\mathrm{Vert}_p(\mathbb{S}^7)$ has an orthogonal complement and we will call the intersection of this orthogonal complement with $T_p(\mathbb{S}^7)$ the *(natural) horizontal space* at p and denote it $\mathrm{Hor}_p(\mathbb{S}^7)$. At each $p \in \mathbb{S}^7$ we therefore have a natural orthogonal decomposition

$$T_p(\mathbb{S}^7) \cong \mathrm{Vert}_p(\mathbb{S}^7) \oplus \mathrm{Hor}_p(\mathbb{S}^7).$$

If one fixes $g \in \mathrm{Sp}(1)$ and explicitly computes the derivative at p of the diffeomorphism $\sigma_g : \mathbb{S}^7 \to \mathbb{S}^7$ given by $\sigma_g(p) = p \cdot g$ one finds that

$$(\sigma_g)_{*p}(\mathrm{Hor}_p(\mathbb{S}^7)) = \mathrm{Hor}_{p \cdot g}(\mathbb{S}^7).$$

One also checks that the distribution $p \mapsto \mathrm{Hor}_p(\mathbb{S}^7)$ is C^∞ in the sense that, if X is any smooth vector field on $\mathbb{S}^7$ and if one decomposes each $X_p = X(p)$ into vertical and horizontal parts

$$X_p = \mathrm{Vert}(X_p) + \mathrm{Hor}(X_p) \quad \forall p \in \mathbb{S}^7,$$

then the vector fields $\mathrm{Vert}(X)$ and $\mathrm{Hor}(X)$ defined by $\mathrm{Vert}(X)_p = \mathrm{Vert}(X_p)$ and $\mathrm{Hor}(X)_p = \mathrm{Hor}(X_p)$ are also smooth. Consequently, the distribution $p \mapsto \mathrm{Hor}_p$ of 4-dimensional spaces is a connection on the quaternionic Hopf bundle (see Chapter II, Section 1, page 63, of [KN1]). We call this the *natural connection* on $\mathrm{Sp}(1) \hookrightarrow \mathbb{S}^7 \xrightarrow{\pi} \mathbb{HP}^1$.

Any connection on a principal G-bundle arises from a connection 1-form, that is, a $\mathfrak{g}$-valued 1-form on the bundle space whose kernel at each point is the horizontal space at that point (see Proposition 1.1, Chapter II, Section 1, page 64, of [KN1]). Identifying the Lie algebra of Sp(1) with $\mathrm{Im}\,\mathbb{H}$ and defining an $\mathrm{Im}\,\mathbb{H}$-valued 1-form $\tilde{\omega}$ on $\mathbb{H}^2$ by

$$\tilde{\omega} = \mathrm{Im}\,(\,\bar{q}^1 dq^1 + \bar{q}^2 dq^2\,)$$

one can show that the connection 1-form ω for the natural connection on the quaternionic Hopf bundle is the restriction of $\tilde{\omega}$ to $\mathbb{S}^7$, that is,

$$\omega = \iota^* \tilde{\omega},$$

where $\iota : \mathbb{S}^7 \hookrightarrow \mathbb{H}^2$ is the inclusion map (see Section 6.1, pages 334-335, of [Nab4]).

In the physics literature it is more common to describe connections locally on the base manifold by pulling back the connection 1-form by a section of the bundle corresponding to some trivialization (these pullbacks are called *gauge potentials*). For the trivializations $(U_k, \Psi_k), k = 1, 2$, of the Hopf bundle each U_k covers all but one point of $\mathbb{HP}^1$ so that the connection is uniquely determined by either one of the corresponding pullbacks. For example, the inverse of $\Psi_1 : \pi^{-1}(U_1) \to U_1 \times \mathrm{Sp}(1)$ is given by $\Psi_1^{-1}([q^1, q^2], g) = (\|q^1\|\, g, q^2(q^1/\|q^1\|)^{-1} g) \in \mathbb{S}^7 \subseteq \mathbb{H}^2$ so the associated section $s_1 : U_1 \to \pi^{-1}(U_1)$ is

$$s_1([q^1, q^2]) = \Psi_1^{-1}([q^1, q^2], 1) = (\,\|q^1\|, q^2(q^1/\|q^1\|)^{-1}\,).$$

Since U_1 is also the domain of the standard chart (U_1, φ_1) on $\mathbb{HP}^1$ we can write $s_1^*\omega$ in terms of these coordinates on $\mathbb{HP}^1$. More precisely, we pull $s_1^*\omega$ back to $\mathbb{H}$ by φ_1^{-1} to obtain an $\mathrm{Im}\,\mathbb{H}$-valued 1-form $(s_1 \circ \varphi_1^{-1})^*\omega$ on $\mathbb{H}$. For reasons that will become clear shortly we will denote this 1-form $\mathcal{A}_{1,0}$. Computing the pullback explicitly gives

$$\mathcal{A}_{1,0} = \mathrm{Im}\left(\frac{\bar{q}}{1 + \|q\|^2}\, dq \right)$$

on $\mathbb{H}$ (see Chapter 5, Section 9, pages 297-301, of [Nab4]). Regarded simply as an $\mathrm{Im}\,\mathbb{H}$-valued 1-form on $\mathbb{H} \cong \mathbb{R}^4$, $\mathcal{A}_{1,0}$ first appeared in a quite different context in the physics literature [BPST] where it was initially referred to as a *pseudoparticle*. We will have more to say about this shortly.

Thus far we know only one connection on the Hopf bundle and we would now like to produce some more. For this we recall that an *automorphism* of the principal bundle $\mathrm{Sp}(1) \hookrightarrow \mathbb{S}^7 \xrightarrow{\pi} \mathbb{HP}^1$ is a diffeomorphism $f : \mathbb{S}^7 \to \mathbb{S}^7$ of $\mathbb{S}^7$ onto itself that respects the action of Sp(1) on $\mathbb{S}^7$ ($f(p \cdot g) = f(p) \cdot g$). Each such automorphism induces a diffeomorphism $f_{\mathbb{HP}^1} : \mathbb{HP}^1 \to \mathbb{HP}^1$ of $\mathbb{HP}^1$ onto itself by $\pi \circ f = f_{\mathbb{HP}^1} \circ \pi$. If $f_{\mathbb{HP}^1}$ happens to be the identity map, then f is called a *(global) gauge transformation* of $\mathrm{Sp}(1) \hookrightarrow \mathbb{S}^7 \xrightarrow{\pi} \mathbb{HP}^1$. Now, if f is any automorphism and ω is any connection 1-form, then the pullback $f^*\omega$ is also a connection

1-form. If f is a gauge transformation, the the connections ω and $f^*\omega$ are said to be *gauge equivalent*. One can show (see Chapter 6, Section 1, pages 336-341, of [Nab4]) that, by judiciously choosing automorphisms of the Hopf bundle by which to pull back the natural connection ω, one can produce new connections $\omega_{\lambda,n}$ for $(\lambda, n) \in (0,\infty) \times \mathbb{H}$ that are uniquely determined by the $\operatorname{Im}\mathbb{H}$-valued 1-forms

$$\mathcal{A}_{\lambda,n} = \operatorname{Im}\left(\frac{\bar{q} - \bar{n}}{\lambda^2 + \|q-n\|^2}\, dq \right)$$

on $\mathbb{H}$. For reasons that we will discuss shortly, $\mathcal{A}_{\lambda,n}$ is called the *BPST instanton* with *center n* and *spread* λ. Although all of the $\omega_{\lambda,n}$ differ from the natural connection $\omega = \omega_{1,0}$ by an automorphism, distinct pairs (λ, n) give rise to connections that are *not* gauge equivalent so the $\omega_{\lambda,n}$ all represent distinct gauge equivalence classes of connections on the Hopf bundle (see Chapter 6, Section 3, page 357, of [Nab4]). We will see that a great deal more is true.

Any connection ω on any principal bundle has a curvature Ω that can be calculated from the Cartan Structure Equation $\Omega = d\omega + \frac{1}{2}[\omega,\omega]$ and is uniquely determined by a family of pullbacks $\mathcal{F} = s^*\Omega$, called *gauge field strengths*, by sections corresponding to some trivializing cover (see Chapter II, Theorem 5.2, page 77, of [KN1]. For the connection $\omega_{\lambda,n}$ on the Hopf bundle the curvature $\Omega_{\lambda,n}$ is uniquely determined by the single gauge field strength $\mathcal{F}_{\lambda,n} = d\mathcal{A}_{\lambda,n} + \frac{1}{2}[\mathcal{A}_{\lambda,n}, \mathcal{A}_{\lambda,n}]$. A rather tedious, but routine calculation (see Chapter 5, Section 11, pages 324-328, of [Nab4]) gives

$$\begin{aligned}\mathcal{F}_{\lambda,n} &= \frac{\lambda^2}{(\lambda^2 + \|q-n\|^2)^2}\, d\bar{q} \wedge dq \\ &= \frac{2\lambda^2}{(\lambda^2 + \|q-n\|^2)^2}\Big[(dx^1 \wedge dx^2 - dx^3 \wedge dx^4)\,\mathbf{i} \\ &\qquad + (dx^1 \wedge dx^3 + dx^2 \wedge dx^4)\,\mathbf{j} + (dx^1 \wedge dx^4 - dx^2 \wedge dx^3)\,\mathbf{k} \Big],\end{aligned}$$

where x^1, x^2, x^3 and x^4 are standard coordinates on $\mathbb{R}^4$.

The $\operatorname{Im}\mathbb{H}$-valued 2-forms $\mathcal{F}_{\lambda,n}$ on $\mathbb{H} \cong \mathbb{R}^4$ have a number of properties that are crucial to our story. We will denote by ${}^*\nu$ the Hodge dual of the p-form ν, $0 \le p \le 4$, on $\mathbb{R}^4$ corresponding to the usual orientation and inner product on $\mathbb{R}^4$. Extending * to $\operatorname{Im}\mathbb{H}$-valued forms componentwise one finds that each $\mathcal{F}_{\lambda,n}$ is *anti-self-dual* in the sense that

$${}^*\mathcal{F}_{\lambda,n} = -\mathcal{F}_{\lambda,n}.$$

The Hodge star operator * on $\mathbb{R}^4$ also defines a pointwise inner product $\langle \mu, \nu \rangle$ on the real valued p-forms on $\mathbb{R}^4$ ($\mu \wedge^* \nu = \langle \mu, \nu \rangle \mathrm{vol}_{\mathbb{R}^4}$). Combined with an inner product on the Lie algebra this gives a pointwise inner product on Lie algebra-valued 1-forms. The most useful way to describe this is to identify $\operatorname{Im}\mathbb{H}$ with $\mathfrak{su}(2)$ and use the inner product $\langle A, B \rangle = -2\,\mathrm{tr}(AB)$ on $\mathfrak{su}(2)$. Then the squared norm $\|\mathcal{F}\|^2$ of a gauge field strength is given by

$$-\mathrm{tr}\,(\mathcal{F} \wedge^* \mathcal{F}) = \frac{1}{2}\,\|\mathcal{F}\|^2\,\mathrm{vol}_{\mathbb{R}^4},$$

where the wedge product is computed as a matrix product with the entries multiplied by the ordinary wedge product. For $\mathcal{F}_{\lambda,n}$ this gives

$$\|\mathcal{F}_{\lambda,n}\|^2 = \frac{96\lambda^4}{(\lambda^2 + \|q-n\|^2)^4}.$$

Notice that $\|\mathcal{F}_{\lambda,n}\|^2$ has a maximum value of $96/\lambda^4$ at $q = n$ and that, for a fixed n, its variation with λ is such that the *total field strength*

$$\frac{1}{2}\int_{\mathbb{R}^4} \|\mathcal{F}_{\lambda,n}\|^2 \operatorname{vol}_{\mathbb{R}^4} = \int_{\mathbb{R}^4} \frac{48\lambda^4}{(\lambda^2 + \|q-n\|^2)^4}\, d^4q = 8\pi^2$$

remains constant at $8\pi^2$. Thus, the BPST gauge potential $\mathcal{A}_{\lambda,n}$ on $\mathbb{R}^4$ has field strength that is "centered" at n in $\mathbb{R}^4$ with a "spread" that is determined by λ (hence the terminology introduced earlier) and a total field strength that is independent of n and λ (see Figure 1). One can think of this in the following way. Fix the center n. Then, as a function of λ, the maximum value $96/\lambda^4$ of $\|\mathcal{F}_{\lambda,n}\|^2$ approaches infinity as $\lambda \to 0$ in such a way as to maintain a constant value for the total field strength. Thus, as $\lambda \to 0$, the field strength concentrates more and more at n. We will see that the reason for this as well as its consequences are deeper than one might expect.

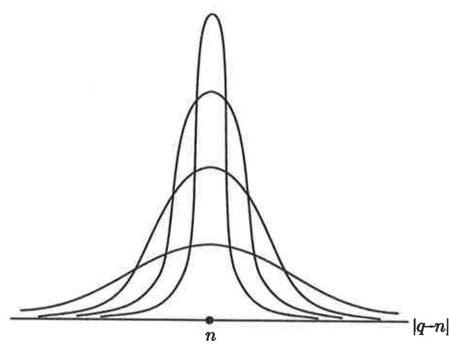

FIGURE 1. BPST Field Strengths $\|\mathcal{F}_{\lambda,n}\|^2$

Now we would like to temporarily suppress from our minds where the potentials $\mathcal{A}_{\lambda,n}$ came from, that is, the Hopf bundle, and regard them simply as Lie algebra-valued 1-forms on $\mathbb{R}^4$. Any such Lie algebra-valued 1-form $\mathcal{A}$ can be thought of as a gauge potential for a connection on the trivial Sp(1)-bundle over $\mathbb{R}^4$ and so has a gauge field strength $\mathcal{F} = d\mathcal{A} + \frac{1}{2}[\mathcal{A}, \mathcal{A}]$. We define the *Yang-Mills action* $\mathcal{YM}(\mathcal{A})$ of $\mathcal{A}$ by

$$\mathcal{YM}(\mathcal{A}) = \int_{\mathbb{R}^4} -\operatorname{tr}(\mathcal{F} \wedge^* \mathcal{F}) = \frac{1}{2}\int_{\mathbb{R}^4} \|\mathcal{F}\|^2 \operatorname{vol}_{\mathbb{R}^4}. \tag{1}$$

The motivation for the terminology comes from the attempts of Yang and Mills in [YM] to construct a non-Abelian generalization of classical electromagnetic theory to describe the isotopic spin of a nucleon. The integral in (1) might well be infinite, but if it is not we will say that $\mathcal{A}$ has *finite action* and then think of $\mathcal{YM}(\mathcal{A})$ as the total field strength of the gauge potential $\mathcal{A}$.

The Euler-Lagrange equations for the Yang-Mills action $\mathcal{YM}(\mathcal{A})$ under variations of $\mathcal{A}$ are

$$d^{\mathcal{A}*}\mathcal{F} = d^*\mathcal{F} + [\mathcal{A}, {}^*\mathcal{F}] = 0, \tag{2}$$

where $d^{\mathcal{A}*}\mathcal{F}$ is the covariant exterior derivative of ${}^*\mathcal{F}$ associated with $\mathcal{A}$. These are the *Yang-Mills Equations* on $\mathbb{R}^4$. Quite independently of $\mathcal{YM}$, any gauge field strength $\mathcal{F}$, being the pullback of a curvature, satisfies a purely geometrical condition called the *Bianchi Identity*

$$d^{\mathcal{A}}\mathcal{F} = d\mathcal{F} + [\mathcal{A}, \mathcal{F}] = 0. \tag{3}$$

(see Chapter II, Theorem 5.4, page 78, of [KN1]. Now notice that, if the gauge field strength $\mathcal{F}$ of $\mathcal{A}$ happens to be anti-self-dual (${}^*\mathcal{F} = -\mathcal{F}$), then the Bianchi identity implies that the Yang-Mills equations are automatically satisfied. A finite action gauge potential $\mathcal{A}$ on $\mathbb{R}^4$ with anti-self-dual gauge field strength $\mathcal{F}$ is called an *instanton* on $\mathbb{R}^4$. In particular, an instanton on $\mathbb{R}^4$ is a solution to the Yang-Mills equations and so is a critical point of the Yang-Mills action. In fact, however, instantons are absolute minima for the Yang-Mills action (see Chapter 6, Section 3, page 361, of [Nab4]).

We have described a family of instantons $\mathcal{A}_{\lambda,n}$ parametrized by $(\lambda, n) \in (0, \infty) \times \mathbb{R}^4$ and have lately been thinking of them simply as $\mathrm{Im}\,\mathbb{H}$-valued 1-forms on $\mathbb{R}^4$ and forgetting where they came from. In order to unearth the motivation for Donaldson's work we will now want to identify $\mathbb{HP}^1$ with $\mathbb{S}^4$ in the manner described in Remark 1.1 and view the $\mathcal{A}_{\lambda,n}$ as gauge potentials corresponding to connections $\omega_{\lambda,n}$ on

$$\mathrm{Sp}(1) \hookrightarrow \mathbb{S}^7 \xrightarrow{\pi_1} \mathbb{S}^4.$$

Stated otherwise, the $\mathcal{A}_{\lambda,n}$ can be viewed as connection 1-forms on the *trivial* Sp(1)-bundle over $\mathbb{R}^4$

$$\mathrm{Sp}(1) \hookrightarrow \mathbb{R}^4 \times \mathrm{Sp}(1) \to \mathbb{R}^4$$

that "come from" connection 1-forms $\omega_{\lambda,n}$ on the *nontrivial* Hopf bundle over $\mathbb{S}^4$. Turning matters about, one might say that these connections on the trivial bundle over $\mathbb{R}^4$ "extend to $\mathbb{S}^4$" in the sense that $\mathbb{S}^4 = \mathbb{R}^4 \cup \{\infty\}$ is the 1-point compactification of $\mathbb{R}^4$ and, due to their asymptotic behavior as $\|x\| \to \infty$ in $\mathbb{R}^4$ (they have finite action), the connections extend to the point at infinity. Notice, however, that the extension process involves not only the connections, but the bundle on which they are defined as well. A consequence of the famous *Removable Singularities Theorem* of Karen Uhlenbeck [Uhl] asserts that this interpretation is not as fanciful as it might seem. Specifically, Uhlenbeck proved the following. Let $\mathcal{A}$ be an $\mathrm{Im}\,\mathbb{H}$-valued 1-form on $\mathbb{R}^4$ that satisfies the Yang-Mills equations (2) and has finite Yang-Mills action $\mathcal{YM}(\mathcal{A}) < \infty$. Then there exists a unique (up to equivalence) Sp(1)-bundle $\mathrm{Sp}(1) \hookrightarrow P \xrightarrow{\pi} \mathbb{S}^4$ over $\mathbb{S}^4$, a connection 1-form ω on P and a section $s : \mathbb{S}^4 - \{N\} \to \pi^{-1}(\mathbb{S}^4 - \{N\})$ such that $\mathcal{A} = (s \circ \varphi_S^{-1})^*\omega$, where $\varphi_S : \mathbb{S}^4 - \{N\} \to \mathbb{R}^4$ is stereographic projection from the north pole N.

Thus, finite action Yang-Mills potentials on $\mathbb{R}^4$ extend to connections on *some* principal Sp(1)-bundle over $\mathbb{S}^4$. But a principal Sp(1)-bundle $\mathrm{Sp}(1) \hookrightarrow P \xrightarrow{\pi} \mathbb{S}^4$ over $\mathbb{S}^4$ is uniquely determined up to equivalence by its second Chern number $c_2(P)[\mathbb{S}^4]$ (see Chapter 6, Section 4, pages 328-334, of [Nab5]) which can be calculated from

$$c_2(P)[\mathbb{S}^4] = \frac{1}{8\pi^2} \int_{\mathbb{R}^4} \mathrm{tr}\,(\mathcal{F} \wedge \mathcal{F}),$$

where $\mathcal{F} = d\mathcal{A} + \frac{1}{2}[\mathcal{A}, \mathcal{A}]$. Now suppose that $\mathcal{A}$ is an instanton so that ${}^*\mathcal{F} = -\mathcal{F}$. Then $-\mathrm{tr}\,(\mathcal{F} \wedge {}^*\mathcal{F}) = \mathrm{tr}\,(\mathcal{F} \wedge \mathcal{F})$ so

$$c_2(P)[\mathbb{S}^4] = \frac{1}{8\pi^2}\, \mathcal{YM}(\mathcal{A}).$$

Thus, the Yang-Mills action of an instanton $\mathcal{A}$ on $\mathbb{R}^4$ directly encodes the topology of the bundle over $\mathbb{S}^4$ to which $\mathcal{A}$ extends. But $\mathcal{YM}(\mathcal{A})$ is determined by the asymptotic behavior of the field strength $\mathcal{F}$ on $\mathbb{R}^4$ so it is this physical characteristic of the field that is represented by the Chern number. Physicists call $-c_2(P)[\mathbb{S}^4]$ the *instanton number* or *topological charge* of $\mathcal{A}$. In particular, each $\mathcal{A}_{\lambda,n}$ has instanton number -1. The "reason" that all of the BPST instantons $\mathcal{A}_{\lambda,n}$ have the same total field strength $\mathcal{YM}(\mathcal{A}_{\lambda,n})$ is now clear; they all extend to (that is, come from) the same Sp(1)-bundle over $\mathbb{S}^4$, that is, the Hopf bundle. Notice also that, because the Chern number is always an integer (see Chapter 6, Section 4, pages 328-334, of [Nab5]), the topological charge of an instanton on $\mathbb{R}^4$ is "quantized". In particular, its total field strength, being an integral multiple of $8\pi^2$, cannot be altered by a continuous variation of the field and so is "conserved", but for purely topological reasons, unlike the more common Noether conserved quantities.

1.1.3. *The Moduli Space of Instantons on the Hopf Bundle.* The motivation for Donaldson's approach to the topology of smooth 4-manifolds comes not from the instantons themselves, but rather from their gauge equivalence classes. If $\mathcal{A}$ is an instanton on $\mathbb{R}^4$ with topological charge -1 we will denote by $[\mathcal{A}]$ its gauge equivalence class and by $\mathcal{M}$ the collection of all such gauge equivalence classes. The BPST instantons $\mathcal{A}_{\lambda,n}$ each determine a point $[\mathcal{A}_{\lambda,n}]$ in $\mathcal{M}$ and, moreover, $[\mathcal{A}_{\lambda,n}] = [\mathcal{A}_{\lambda',n'}]$ if and only if $(\lambda, n) = (\lambda', n')$. Much deeper is the fact, proved by Atiyah, Hitchin and Singer in [AHS], that *every* instanton on $\mathbb{R}^4$ with topological charge -1 is gauge equivalent to some BPST instanton $\mathcal{A}_{\lambda,n}$ so that the $[\mathcal{A}_{\lambda,n}]$ exhaust all of $\mathcal{M}$.

$$\mathcal{M} = \left\{ [\mathcal{A}_{\lambda,n}] : (\lambda, n) \in (0,\infty) \times \mathbb{R}^4 \right\}$$

$\mathcal{M}$ is called the *moduli space of instantons on* $\mathbb{R}^4$ *with topological charge* -1. One can identify this with the moduli space of connections with anti-self-dual curvature on the Sp(1)-bundle over $\mathbb{S}^4$ with Chern number 1, that is, on the quaternionic Hopf bundle $\mathrm{Sp}(1) \hookrightarrow \mathbb{S}^7 \xrightarrow{\pi_1} \mathbb{S}^4$. The map $(\lambda, n) \in (0,\infty) \times \mathbb{R}^4 \mapsto [\mathcal{A}_{\lambda,n}] \in \mathcal{M}$ is a bijection. Points that are "nearby" in the usual topology of $(0,\infty) \times \mathbb{R}^4 \subseteq \mathbb{R}^5$ give rise to connections whose gauge potentials are "close" in the sense that their field strengths are centered at nearby points and have approximately the same scale so it would seem appropriate to identify $\mathcal{M}$ topologically with the open subspace $(0,\infty) \times \mathbb{R}^4$ of $\mathbb{R}^5$. We will see in Section 1.3 that any two differentiable structures on $\mathbb{R}^5$ (which is homeomorphic to $(0,\infty) \times \mathbb{R}^4$) are necessarily diffeomorphic so there is no ambiguity about the appropriate differentiable structure for $\mathcal{M}$. We therefore identify $\mathcal{M}$ as a smooth manifold with $(0,\infty) \times \mathbb{R}^4$. A still more instructive picture is obtained by noting that there is an orientation preserving, conformal diffeomorphism of $(0,\infty) \times \mathbb{R}^4$ onto the open 5-dimensional ball B^5 in $\mathbb{R}^5$ and that, with this, one can supply $\mathcal{M}$ with "spherical coordinates" (see Chapter 6, Section 5, pages 372-375, of [Nab4]). The picture of $\mathcal{M}$ that emerges from this is as follows. $\mathcal{M}$ is viewed as the open 5-ball B^5 with the gauge equivalence class $[\mathcal{A}_{1,0}]$ of the natural connection at the center. Proceeding radially outward from $[\mathcal{A}_{1,0}]$ toward a point on $\partial B^5 = \mathbb{S}^4$ one encounters potentials all of which have the same center n, but become more and more concentrated, that is, for which the spread $\lambda \to 0$. This boundary point in $\mathbb{S}^4$ can then, at least intuitively, be identified with a connection/gauge potential concentrated entirely at n, that is, one whose curvature/field strength is a sort of

Dirac delta. A particularly pleasing aspect of this picture is that the base manifold $\mathbb{S}^4$ of the bundle to which the connections $\mathcal{A}_{\lambda,n}$ extend emerges as the boundary of the moduli space in some compactification of $\mathcal{M}$ ($\mathcal{M} \cong B^5 \hookrightarrow \overline{B}^5 = B^5 \cup \mathbb{S}^4$) and that a point in it can be viewed as a sort of singular connection. The moral is that the topologies of the underlying 4-manifold $\mathbb{S}^4$ and the moduli space $\mathcal{M}$ are inextricably linked. In the next section we will begin to see how Donaldson exploited this idea.

1.2. **Donaldson's Theorem on Intersection Forms.** We will now provide a very broad sketch of the profound generalization, due to Donaldson [Don1], of the scenario suggested by the example in the preceding section. The idea is to study the topology of a smooth 4-manifold X by identifying X with the boundary in a compactification of a certain moduli space of connections on some principal bundle over X.

1.2.1. *Topological Background.* Throughout this section X will denote a compact, connected, simply connected, oriented, smooth 4-manifold.

Remark 1.2. A simply connected 4-manifold is always orientable, but much of what we will have to say is very sensitive to the specific orientation chosen so we will tend to emphasize this by specifying that our manifolds are *oriented*.

Typical examples include the 4-sphere $\mathbb{S}^4$, the product $\mathbb{S}^2 \times \mathbb{S}^2$ of two 2-spheres, the complex projective plane $\mathbb{CP}^2$ with its natural orientation as a complex manifold, the same manifold $\overline{\mathbb{CP}}^2$ with the opposite orientation, the connected sum $\mathbb{CP}^2 \# \overline{\mathbb{CP}}^2$ of $\mathbb{CP}^2$ and $\overline{\mathbb{CP}}^2$ and the *Kummer surface* $K3$ which can be thought of as the submanifold (complex algebraic surface) in $\mathbb{CP}^3$ whose homogeneous coordinates x_1, x_2, x_3, x_4 satisfy $x_1^4 + x_2^4 + x_3^4 + x_4^4 = 0$.

We begin by enumerating a few classical results on the topology of such manifolds. We will be particularly interested in the homology of X. Since X is assumed connected, $H_0(X;\mathbb{Z}) \cong \mathbb{Z}$ (see Theorem 3.2.3, page 171, of [Nab4]). Being simply connected, $\pi_1(X)$ is trivial so the Hurewicz Theorem (see Corollary 12.2, page 48, of [Green]) implies that $H_1(X;\mathbb{Z}) \cong 0$ as well and this implies that $H_2(X;\mathbb{Z})$ is torsion free (see Proposition E.1, page 217, of [FU1]). Poincaré Duality (see Theorem 26.6, page 164, of [Green]) then implies that $H_4(X;\mathbb{Z}) \cong \mathbb{Z}$ and $H_3(X;\mathbb{Z}) \cong 0$. Thus, manifolds of the sort we are considering can differ homologically only in $H_2(X;\mathbb{Z})$ which is a finitely generated, free Abelian group ($\mathbb{Z}$-module) and therefore completely determined by the second Betti number $b_2(X)$, that is, the rank of $H_2(X;\mathbb{Z})$. Since there is no torsion in $H_2(X;\mathbb{Z})$,

$$H^2(X;\mathbb{Z}) \cong \mathrm{Hom}\,(H_2(X;\mathbb{Z}),\mathbb{Z}) \cong H_2(X;\mathbb{Z})$$

(see Corollary 23.14, page 134, of [Green]).

Another very special property of this type of manifold is that every homology class α in $H_2(X;\mathbb{Z})$ can be represented by a smoothly embedded, compact, oriented surface (2-manifold) Σ in X. This means that there exists a smooth embedding $\iota : \Sigma \hookrightarrow X$ of Σ in X such that $\iota_*([\Sigma]) = \alpha$, where $[\Sigma]$ is the fundamental class of Σ in $H_2(\Sigma;\mathbb{Z})$ and $\iota_* : H_2(\Sigma;\mathbb{Z}) \to H_2(X;\mathbb{Z})$ is the map induced by ι. (see Proposition E.9, page 224, of [FU1] or Theorem 1.1, page 20, of [Kirby]). A given $\alpha \in H_2(X;\mathbb{Z})$ can be represented by a variety of surfaces of different genus. The elements of $H_2(X;\mathbb{Z})$ can also be represented by complex line bundles $L \to X$ over X because these line bundles are determined up to equivalence by their 1^{st} Chern class $c_1(L) \in H^2(X;\mathbb{Z})$. We point out as well that, if X_1 and X_2 are two manifolds of the type under consideration here, the Mayer-Vietoris sequence (see Part II, Section 17, page 74, of [Green]) implies that $H_2(\,\cdot\,;\mathbb{Z})$ is additive on connected sums, that is,

$$H_2(X_1\#X_2;\mathbb{Z}) \cong H_2(X_1;\mathbb{Z}) \oplus H_2(X_2;\mathbb{Z}).$$

The *intersection form* on X is a certain integer-valued, symmetric, bilinear form $Q_X : H_2(X;\mathbb{Z}) \times H_2(X;\mathbb{Z}) \to \mathbb{Z}$ on $H_2(X;\mathbb{Z})$ (see Chapters 3 and 4 of [Scor]). This can be defined in a variety of ways, but we will opt for the following. If $H_2(X;\mathbb{Z}) \cong 0$, then Q_X takes the value 0 at the only point in $H_2(X;\mathbb{Z}) \times H_2(X;\mathbb{Z})$; this is generally called the *empty form* and denoted $\emptyset$. This is the case, for example, when $X = \mathbb{S}^4$. Otherwise, suppose α_1 and α_2 are two homology classes in $H_2(X;\mathbb{Z})$. Select smoothly embedded, compact, oriented surfaces Σ_1 and Σ_2 in X representing them. By the Transversality Theorem (see Chapter 3, Section 2, of [Hirsch]), one can assume that these surfaces intersect transversally, that is, the tangent spaces to Σ_1 and Σ_2 span the tangent space to X at each point in the intersection $\Sigma_1 \cap \Sigma_2$, so that $\Sigma_1 \cap \Sigma_2$ is a finite set of isolated points in X. A point $p \in \Sigma_1 \cap \Sigma_2$ is assigned the value 1 if an oriented basis for $T_p(\Sigma_1)$ together with an oriented basis for $T_p(\Sigma_2)$ gives an oriented basis for $T_p(X)$; otherwise it is assigned the value -1. Then $Q_X(\alpha_1, \alpha_2)$ is the sum of these values over all of the intersection points.

Remark 1.3. Identifying the second homology $H_2(X;\mathbb{Z})$ with the second cohomology $H^2(X;\mathbb{Z})$ by Poincaré duality, the intersection form $Q_X : H^2(X;\mathbb{Z}) \times H^2(X;\mathbb{Z}) \to \mathbb{Z}$ is given by pairing the cup product with the fundamental class of X

$$Q_X(\alpha_1, \alpha_2) = \langle\, \alpha_1 \smile \alpha_2, [X]\,\rangle$$

and this serves to define the intersection form if X is merely a topological rather than a smooth 4-manifold.

If $H_2(X;\mathbb{Z}) \neq 0$, then Q_X is unimodular in the sense that its matrix relative to any basis for $H_2(X;\mathbb{Z})$ has determinant ± 1 (see Chapter 3, Section 2, page 116, of [Scor]). Q_X is said to be *even* if $Q_X(\alpha, \alpha)$ is even for all $\alpha \in H_2(X;\mathbb{Z})$ and *odd* otherwise. Q_X is *positive definite* (respectively, *negative definite*) if $Q_X(\alpha, \alpha) > 0$ (respectively, $Q_X(\alpha, \alpha) < 0$) for every nonzero $\alpha \in H_2(X;\mathbb{Z})$. Q_X is *definite* if it is either positive or negative definite and *indefinite* otherwise. If $H_2(X;\mathbb{Z}) \neq 0$ then the maximal rank of a $\mathbb{Z}$-submodule of $H_2(X;\mathbb{Z})$ on which Q_X is positive (respectively, negative) definite is denoted $b_2^+(X)$ (respectively, $b_2^-(X)$). If $H_2(X;\mathbb{Z}) = 0$, then both of these are taken to be zero. Then the second Betti number of X is $b_2(X) = b_2^+(X) + b_2^-(X)$. The *signature* of X (or of Q_X) is $\sigma(X) = \sigma(Q_X) = b_2^+(X) - b_2^-(X)$. The signature is additive for both disjoint unions and connected sums. Two intersection forms Q_{X_1} and Q_{X_2} are said to be *equivalent* if and only if there exist bases for $H_2(X_1;\mathbb{Z})$ and $H_2(X_2;\mathbb{Z})$ relative to which Q_{X_1} and Q_{X_2} have the same matrix. The intersection form is additive on connected sums, that is,

$$Q_{X_1\#X_2} = Q_{X_1} \oplus Q_{X_2}$$

(see Chapter 3, Section 2, page 118, of [Scor]).

Example 1.1. In each of the following examples an unspecified basis is chosen for the $\mathbb{Z}$-module $H_2(X;\mathbb{Z})$ and Q_X is given as a matrix relative to this basis (see Chapter 1, Section 1, pages 3-5, of [DK]).

X	$H_2(X;\mathbb{Z})$	Q_X	$b_2^+(X)$	$\sigma(X)$
$\mathbb{S}^4$	0	$\emptyset$	0	0
$\mathbb{CP}^2$	$\mathbb{Z}$	(1)	1	1
$\overline{\mathbb{CP}}^2$	$\mathbb{Z}$	(-1)	0	-1
$\mathbb{S}^2 \times \mathbb{S}^2$	$\mathbb{Z} \oplus \mathbb{Z}$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	1	0
$\mathbb{CP}^2 \# \overline{\mathbb{CP}}^2$	$\mathbb{Z} \oplus \mathbb{Z}$	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	1	0
$K3$	$22\mathbb{Z}$	$2(-E_8) \oplus 3\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	3	-16

where E_8 is the even, positive definite, unimodular matrix given by

$$E_8 = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 2 \end{pmatrix}$$

The following classical result of Rokhlin [Rokh] will be useful in the next section when we discuss the existence of exotic differentiable structures.

Theorem 1.1. *(Rokhlin) Let X be a compact, connected, simply connected, oriented, smooth 4-manifold with even intersection form. Then the signature $\sigma(X)$ is divisible by 16.*

Remark 1.4. Rokhlin's Theorem is generally stated for spin 4-manifolds which always have even intersection form. For simply connected 4-manifolds the converse is also true, that is, an even intersection form implies the existence of a spin structure. For a proof of Rokhlin's Theorem 1.1 see Chapter X, Section 1, pages 64-65, of [Kirby].

It has been known for some time that the intersection form is a basic invariant of compact 4-manifolds. In 1949, J.H.C. Whitehead [White] proved the following.

Theorem 1.2. *(Whitehead) Two compact, connected, simply connected, topological 4-manifolds X_1 and X_2 have the same homotopy type if and only if their intersection forms are equivalent.*

1.2.2. *Donaldson's 1983 Theorem.* In 1982, the situation took a dramatic turn when Freedman [Freed] proved that every unimodular, symmetric, $\mathbb{Z}$-valued bilinear form on a finitely generated, free Abelian group is the intersection form of at least one (and at most two) compact, connected, simply connected, oriented *topological* 4-manifolds, thereby giving a complete classification of such manifolds and, in particular, proving the 4-dimensional Poincaré conjecture (see Section 1.3.1). This is, in particular, true of the vast, impenetrable maze of definite forms. If you do not believe that this is a vast, impenetrable maze, consider the fact that when the rank is 40 there are at least 10^{51} equivalence classes of positive definite forms (see Chapter II, Section 6, page 28, of [MH]). In 1983, however, Donaldson [Don1] showed that the differential topologist need not venture into this maze because there is only *one* positive/negative definite unimodular, symmetric, $\mathbb{Z}$-valued bilinear form that can arise as the intersection form of a compact, connected, simply connected, oriented *smooth* 4-manifold. In particular, there are enormous quantities of topological 4-manifolds that can admit no smooth structure at all.

Theorem 1.3. *(Donaldson's Theorem on Intersection Forms) Let X be a compact, connected, simply connected, oriented, smooth 4-manifold with positive (respectively, negative) definite intersection form Q_X. Then there is a basis for $H_2(X;\mathbb{Z})$ relative to which the matrix of Q_X is the identity matrix (respectively, minus the identity matrix).*

Donaldson's Theorem is remarkable, but even more remarkable is its proof which is a byproduct of the analysis of an instanton moduli space quite like the BPST instanton moduli space discussed in Section 1.1.3. The proof is an extremely delicate blend of topology, geometry, and analysis (see [Mor2], [FU1], [Law], or Chapter 8 of [DK]). We will offer only a minimalist sketch of the ideas that lie behind the proof in the negative definite case. Thus, we assume that our compact, connected, simply connected, oriented, smooth 4-manifold X has

$$b_2^+(X) = 0,$$

but $b_2(X) \neq 0$. Denote by

$$\mathrm{Sp}(1) \hookrightarrow P \xrightarrow{\pi} X$$

the principal Sp(1)-bundle over X with second Chern number 1. Note that when $X = \mathbb{S}^4$ this is just the quaternionic Hopf bundle. Now choose a Riemannian metric g for X. With g and the given orientation of X one can define a Hodge star operator * on the differential forms defined on X. Since $\dim X = 4$ the Hodge dual of a 2-form is a 2-form.

Now let ω be a connection on $\mathrm{Sp}(1) \hookrightarrow P \xrightarrow{\pi} X$. The curvature Ω of ω is a Lie algebra-valued 1-form on P, not on X. However, pulling Ω back by the local sections corresponding to some trivializing cover of X gives a family of Lie algebra-valued 2-forms on X that are related by the adjoint action of Sp(1) on its Lie algebra. As a result these locally defined 2-forms on X with values in the Lie algebra piece together into a globally defined 2-form $\mathcal{F}_\omega$ on X with values, not in the Lie algebra, but in the adjoint bundle $\mathrm{ad}\,P$ (the vector bundle associated to $\mathrm{Sp}(1) \hookrightarrow P \xrightarrow{\pi} X$ by the adjoint representation). This 2-form $\mathcal{F}_\omega$ is also often referred to as the curvature of ω. The Hodge star extends to such ad P-valued forms so ${}^*\mathcal{F}_\omega$ is another 2-form on X with values in the adjoint bundle. We will say that the connection ω is *g-anti-self-dual* if

$${}^*\mathcal{F}_\omega = -\mathcal{F}_\omega.$$

Now, it is not at all clear that any such 2-forms exist on $\mathrm{Sp}(1) \hookrightarrow P \xrightarrow{\pi} X$. It is, in fact, a deep theorem of Taubes [Taub1] that, because $b_2^+(X) = 0$, there do exist g-anti-self-dual connections on the (Chern number 1) bundle $\mathrm{Sp}(1) \hookrightarrow P \xrightarrow{\pi} X$ for *any* choice of g. However, if $X = \mathbb{S}^2 \times \mathbb{S}^2$, where $b_2^+(X) = 1$, with its standard metric and orientation there are no such connections on the Sp(1)-bundle with Chern number 1. The same is true of $\mathbb{CP}^2$, but, oddly enough, not of $\overline{\mathbb{CP}}^2$ (Donaldson theory is highly orientation dependent). A g-anti-self-dual connection on $\mathrm{Sp}(1) \hookrightarrow P \xrightarrow{\pi} X$ is called an *instanton* on $\mathrm{Sp}(1) \hookrightarrow P \xrightarrow{\pi} X$.

A *(global) gauge transformation* of our bundle is a diffeomorphism $f : P \to P$ of P onto itself that respects the right action of Sp(1) on P ($f(p \cdot g) = f(p) \cdot g$ for all $p \in P$ and all $g \in \mathrm{Sp}(1)$) and covers the identity map on X ($\pi \circ f = \pi$). These form a group under composition, called the *gauge group* of $\mathrm{Sp}(1) \hookrightarrow P \xrightarrow{\pi} X$ and denoted $\mathcal{G}(P)$. The gauge group acts on connections by pullback ($\omega \cdot f = f^*\omega$) and two connections ω and ω' are said to be *gauge equivalent* if there exists an $f \in \mathcal{G}(P)$ such that $\omega' = f^*\omega$. The collection of all gauge equivalence classes of g-anti-self-dual connections on $\mathrm{Sp}(1) \hookrightarrow P \xrightarrow{\pi} X$ is denoted

$$\mathcal{M} = \mathcal{M}(P, g)$$

and called the *moduli space* of g-anti-self-dual connections on $\mathrm{Sp}(1) \hookrightarrow P \xrightarrow{\pi} X$.

Remark 1.5. We are about to describe what purports to be a picture of the moduli space of g-anti-self-dual connections on $\mathrm{Sp}(1) \hookrightarrow P \xrightarrow{\pi} X$. We must preface this by saying that the amount of technical work that must be done in order to produce this picture is prodigious and cannot really be carried out in the smooth context we have just described. In order to obtain analytically tractable spaces in which to work one must introduce various Sobolev completions of the space of connections and the gauge group. There are standard procedures for doing this, but for our very modest sketch we will be content to suppress these completions and simply recommend a few sources for the material we omit . The constructions are outlined in Section 1, pages 33-49, of [Nab1]; more thorough discussions are available on pages 85-107 of [Mor2] and in Chapter II, Section 7, of [Law].

We have *chosen* the Riemannian metric g and it turns out that if we choose badly the moduli space $\mathcal{M}(P, g)$ might well have no decent mathematical structure at all. However, Donaldson has shown that, for a "generic" choice of g, $\mathcal{M}(P, g)$ has quite a nice structure that we will now describe (for a sketch of how this structure comes about see Section B.2 of [Nab4] and for all of the details see [Mor2], [FU1], [Law], or [DK]). We will simply describe the structure. Thus, we let g denote some generic metric and denote the moduli space $\mathcal{M}(P, g)$ simply as $\mathcal{M}$. Then, in a word, $\mathcal{M}$ looks like Figure 2. More precisely, Donaldson proves all of the following.

(1) If m is one-half the number of $\alpha \in H_2(X; \mathbb{Z})$ with $Q_X(\alpha, \alpha) = -1$, then there exist points $\{p_1, \ldots, p_m\}$ in $\mathcal{M}$ such that $\mathcal{M} - \{p_1, \ldots, p_m\}$ is a smooth, orientable, 5-dimensional manifold.

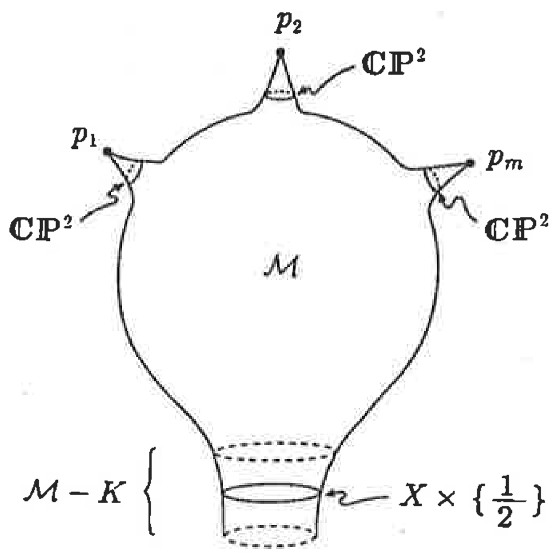

FIGURE 2. Moduli Space

(2) Each $p_i, i = 1, \ldots, m$, has a neighborhood in $\mathcal{M}$ that is homeomorphic to a cone over $\mathbb{CP}^2$ with vertex at p_i and these neighborhoods are smooth away from p_i (the cone over $\mathbb{CP}^2$ is obtained from the cylinder $\mathbb{CP}^2 \times [0, 1]$ by identifying $\mathbb{CP}^2 \times \{1\}$ to a point, which is then called the vertex of the cone).

(3) There is a compact subspace K of $\mathcal{M}$ such that $\mathcal{M} - K$ is an open submanifold of $\mathcal{M} - \{p_1, \ldots, p_m\}$ diffeomorphic to the cylinder $X \times (0, 1)$.

Now construct from $\mathcal{M}$ another space $\mathcal{M}_0$ by cutting off the open top half of each cone and the open bottom half of the cylinder (see Figure 3). Then $\mathcal{M}_0$ is a compact manifold with boundary that inherits an orientation from $\mathcal{M}$. The boundary is the disjoint union of a copy of X and m copies of $\mathbb{CP}^2$ or $\overline{\mathbb{CP}}^2$. Let us suppose that there are p copies of $\mathbb{CP}^2$ and q copies of $\overline{\mathbb{CP}}^2$. Thus $\mathcal{M}_0$ is a cobordism between X and the disjoint union

$$\mathbb{CP}^2 \sqcup \overset{p}{\cdots} \sqcup \mathbb{CP}^2 \sqcup \overline{\mathbb{CP}}^2 \sqcup \overset{q}{\cdots} \sqcup \overline{\mathbb{CP}}^2 = p\,\mathbb{CP}^2 \sqcup q\,\overline{\mathbb{CP}}^2$$

(see Chapter 1, Section 1, page 28, of [Scor]). As it happens, the signature of the intersection form is a cobordism invariant (see Chapter 3, Section 2, page 123 of [Scor]). Since the signature is additive on disjoint unions and $b_2^+(X) = 0$, $b_2^+(X) - b_2^-(X) = p - q$ gives $b_2(X) = b_2^-(X) = q - p$ and therefore

$$b_2(X) \leq q + p = m.$$

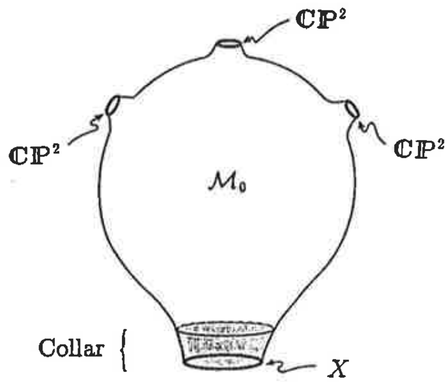

FIGURE 3. Cobordism

Next we will show that $b_2(X) \geq m$ as well. Select some $\alpha_1 \in H_2(X;\mathbb{Z})$ with $Q_X(\alpha_1, \alpha_1) = -1$. Note that there must be at least one such since $b_2^+(X) = 0$, but $b_2(X) \neq 0$. Then there is a Q_X-orthogonal decomposition

$$H_2(X;\mathbb{Z}) = \mathbb{Z}\alpha_1 \oplus G_1.$$

Suppose first that there is no $\alpha_2 \in H_2(X;\mathbb{Z})$ with $Q_X(\alpha_2, \alpha_2) = -1$ and $\alpha_2 \neq \pm\alpha_1$. Then $G_1 = \emptyset$ so $H_2(X;\mathbb{Z}) = \mathbb{Z}\alpha_1$ and $\{\alpha_1\}$ is a basis for $H_2(X;\mathbb{Z})$ relative to which the matrix of Q_X is (-1). In this case we are done. Now suppose there is such an α_2. The Schwartz Inequality gives

$$(Q_X(\alpha_1, \alpha_2))^2 < Q_X(\alpha_1, \alpha_1)Q_X(\alpha_2, \alpha_2) = 1.$$

But $Q_X(\alpha_1, \alpha_2)$ is an integer so $Q_X(\alpha_1, \alpha_2) = 0$ and therefore $\alpha_2 \in G_1$. Now repeat the argument inside G_1 and continue inductively until you run out of $\alpha \in H_2(X;\mathbb{Z})$ with $Q_X(\alpha, \alpha) = -1$. There are $2m$ such α and they occur in pairs $(\alpha, -\alpha)$ so the result is a Q_X-orthogonal decomposition

$$H_2(X;\mathbb{Z}) = \mathbb{Z}\alpha_1 \oplus \cdots \oplus \mathbb{Z}\alpha_m \oplus G,$$

where G is either empty or the Q_X-orthogonal complement of $\mathbb{Z}\alpha_1 \oplus \cdots \oplus \mathbb{Z}\alpha_m$. In particular, $m \leq b_2(X)$ so

$$b_2(X) = m = p + q.$$

But $H_2(X;\mathbb{Z})$ is a finitely-generated, free Abelian group so G must be empty and

$$H_2(X;\mathbb{Z}) = \mathbb{Z}\alpha_1 \oplus \cdots \oplus \mathbb{Z}\alpha_m.$$

Since $Q_X(\alpha_i, \alpha_i) = -1$ for every $i = 1, \ldots, m$, the matrix of Q_X relative to the basis $\{\alpha_1, \ldots, \alpha_m\}$ is minus the identity and this is Donaldson's Theorem.

1.3. **Exotic Differentiable Structures.** A differentiable structure (or smooth structure) on a (Hausdorff, second countable) topological space X is a maximal atlas of C^∞-related charts. It is a simple matter to define different differentiable structures on the same space X. For example, if $X = \mathbb{R}$, then the standard differentiable structure is the unique maximal atlas determined by the single chart $(\mathbb{R}, id_{\mathbb{R}})$. The chart $(\mathbb{R}, \varphi)$, where $\varphi : \mathbb{R} \to \mathbb{R}$ is given by $\varphi(x) = x^3$ is not in this standard differentiable structure because $\varphi^{-1}(x) = x^{1/3}$ is not differentiable with respect to the standard structure at $x = 0$. Consequently, the chart $(\mathbb{R}, \varphi)$ on $\mathbb{R}$ determines another differentiable structure that is not the same as the standard one. Nevertheless, the differentiable manifolds determined by $(\mathbb{R}, id_{\mathbb{R}})$ and $(\mathbb{R}, \varphi)$, call them $\mathbb{R}$ and $\mathbb{R}'$, respectively, are not *too* different. Specifically, the map $\varphi : \mathbb{R}' \to \mathbb{R}$ is a diffeomorphism. More generally, one can show that any two differentiable structures on the topological space $\mathbb{R}$ are necessarily diffeomorphic. Still more generally, one has the following result (see the Appendix in [Miln3]).

Theorem 1.4. *Any smooth, connected, 1-dimensional manifold (without boundary) is diffeomorphic either to the circle $\mathbb{S}^1$ if it is compact, or to $\mathbb{R}$ if it is not compact.*

Up to diffeomorphism, $\mathbb{R}$ has a unique differentiable structure. Rado [Rado] proved, in 1925, that the same is true of $\mathbb{R}^2$ and, in 1952, Moise [Moi] proved the same thing for $\mathbb{R}^3$. Eventually, the combined work of a number of people showed that, in fact, the same is true of $\mathbb{R}^n$ for $n = 1, 2, 3, 5, 6, 7, \ldots$. It will not have escaped your attention that $n = 4$ is conspicuously absent from this list. Indeed, our objective in this section is to very briefly sketch how Donaldson's Theorem 1.3 on intersection forms, when combined with deep results of Freedman [Freed] on topological 4-manifolds, implies the existence of smooth 4-manifolds that are homeomorphic to $\mathbb{R}^4$, but not diffeomorphic to $\mathbb{R}^4$ with its standard differentiable structure. These are called *exotic* or *fake* $\mathbb{R}^4$s.

Remark 1.6. We will be concerned only with the case of $\mathbb{R}^4$, but we point out that, in 1956, Milnor [Miln1] proved the existence of smooth 7-manifolds that are homeomorphic, but not diffeomorphic to the 7-sphere $\mathbb{S}^7$. In fact, there are 28 such exotic differentiable structures on $\mathbb{S}^7$.

1.3.1. *Freedman's Theorem.* We will need to borrow a number of Freedman's results in order to outline the construction of an exotic $\mathbb{R}^4$ so we might as well begin with the major result of [Freed] (see also [FQ]). It is a complete classification of all compact, connected, simply connected topological 4-manifolds.

Theorem 1.5. *(Freedman) Let Q be any $\mathbb{Z}$-valued, symmetric, unimodular bilinear form on a finitely generated, free Abelian group. Then there exists a compact, connected, simply connected topological 4-manifold whose intersection form is equivalent to Q.*

(1) *If Q is even, there is exactly one such manifold X_Q up to homeomorphism.*
(2) *If Q is odd, there are exactly two such manifolds up to homeomorphism, at least one of which does not admit a differentiable structure.*

Example 1.2. The form $Q = E_8$ is even so Freedman's Theorem 1.5 guarantees the existence of a unique topological 4-manifold X_{E_8} whose intersection form is E_8. The signature of E_8 is 8 so Rokhlin's Theorem 1.1 implies that X_{E_8} cannot admit a differentiable structure. On the other hand, the form $E_8 \oplus E_8$ is even and has signature 16 so Rokhlin's Theorem 1.1 does not prohibit Freedman's topological 4-manifold $X_{E_8 \oplus E_8}$ from admitting a differentiable structure. Nevertheless, since $E_8 \oplus E_8$ is positive definite, but not equivalent to the identity form, Donaldson's Theorem 1.3 implies that $X_{E_8 \oplus E_8}$ cannot admit a smooth structure.

The combination of Freedman's Theorem 1.5 and Donaldson's Theorem 1.3 yields a huge number of compact, connected, simply connected, topological 4-manifolds that can admit no smooth structure at all (one for every definite form Q other than $\pm id$). One should contrast this with the following result of Frank Quinn (see Chapter 8, Section 2, of [FQ]) which implies, in particular, that removing a single point from *any* of these examples results in a 4-manifold that *does* admit a smooth structure.

Theorem 1.6. *(Quinn) Any connected, noncompact 4-manifold admits a differentiable structure.*

Freedman's Theorem is quite remarkable and it has many equally remarkable consequences of which we will record just a few. Suppose first that X_1 and X_2 are two compact, connected, simply connected, *smooth* 4-manifolds with the same intersection form Q, that is, Q_{X_1} and Q_{X_2} are both equivalent to Q. Whitehead's Theorem 1.2 implies that X_1 and X_2 have the same homotopy type, but now we see that much more is true. Indeed, Freedman's Theorem 1.5 implies that, up to homeomorphism, there is exactly one topological 4-manifold with this intersection form. Consequently, we have the following.

Theorem 1.7. *Let X_1 and X_2 be two compact, connected, simply connected, smooth 4-manifolds with equivalent intersection forms. Then X_1 and X_2 are homeomorphic.*

Combining this with the Whitehead Theorem 1.2 and applying it to the 4-sphere $\mathbb{S}^4$ one obtains the *4-Dimensional Poincaré Conjecture*.

Corollary 1.8. *If a topological 4-manifold X is homotopy equivalent to the 4-sphere $\mathbb{S}^4$, then it is homeomorphic to $\mathbb{S}^4$.*

Example 1.3. According to Donaldson's Theorem, if X is a compact, connected, simply connected, smooth 4-manifold and the intersection form Q_X is definite, then Q_X is either the identity or minus the identity, that is, either $(1) \oplus \overset{b(X)}{\cdots} \oplus (1)$ or $(-1) \oplus \overset{b(X)}{\cdots} \oplus (-1)$. We know that the first of these is the intersection form of $\mathbb{CP}^2 \# \overset{b(X)}{\cdots} \# \mathbb{CP}^2$ and the second is the intersection form of $\overline{\mathbb{CP}}^2 \# \overset{b(X)}{\cdots} \# \overline{\mathbb{CP}}^2$ so Freedman's Theorem implies that X is homeomorphic to the first of these of Q_X is positive definite and to the second if Q_X is negative definite. If Q_X is indefinite and odd, then it is equivalent to $(1) \oplus \overset{b^+(X)}{\cdots} \oplus (1) \oplus (-1) \oplus \overset{b^-(X)}{\cdots} \oplus (-1)$ (see Chapter II, Section 4, of [MH]) and so X is homeomorphic to $\mathbb{CP}^2 \# \overset{b^+(X)}{\cdots} \# \mathbb{CP}^2 \# \overline{\mathbb{CP}}^2 \# \overset{b^-(X)}{\cdots} \# \overline{\mathbb{CP}}^2$.

If Q_X is indefinite and even, then it is equivalent to $\pm(E_8 \oplus \overset{2m}{\cdots} \oplus E_8) \oplus n \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ for integers $m \geq 0$ and $n > 0$ (see Chapter II, Section 5, of [MH]). For example, $-(E_8 \oplus E_8) \oplus 3 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is represented by the $K3$-surface. Finally we point out that *homeomorphism* can generally not be strengthened to *diffeomorphism* here. Indeed, Donaldson [Don2] has shown that $\mathbb{CP}^2 \# 9\, \overline{\mathbb{CP}}^2$ itself admits nonstandard differentiable structures so that a smooth 4-manifold with intersection form $(1) \oplus 9(-1)$ need not be diffeomorphic to $\mathbb{CP}^2 \# 9\, \overline{\mathbb{CP}}^2$.

1.3.2. *An Exotic* $\mathbb{R}^4$. Now we will sketch one construction of an exotic $\mathbb{R}^4$, that is, a 4-dimensional, smooth manifold that is homeomorphic, but not diffeomorphic to $\mathbb{R}^4$ with its usual differentiable structure. There are other constructions (see Chapter 9, Sections 3 and 4, of [GS]) and many other examples (uncountably many according to a theorem of Taubes [Taub2]). The constructions are all quite involved and we must be content with a very superficial sketch of just one of them relying on a great deal of heavy artillery.

Once we have a 4-manifold that is homeomorphic to $\mathbb{R}^4$ we will need some criterion to guarantee that it is *not* diffeomorphic to $\mathbb{R}^4$ with its standard smooth structure. This much at least is easy. In standard $\mathbb{R}^4$ there are arbitrarily large smoothly embedded 3-spheres so *every compact set is contained inside some smoothly embedded 3-sphere*.

Lemma 1.9. *Let R be a smooth 4-manifold that is homeomorphic to $\mathbb{R}^4$. If R is diffeomorphic to $\mathbb{R}^4$ with its standard smooth structure, then every compact set in R is contained inside some smoothly embedded 3-sphere in R. Consequently, if there exists a compact set C in R that is not contained inside any smoothly embedded 3-sphere in R, then R is an exotic $\mathbb{R}^4$, that is, R is homeomorphic, but not diffeomorphic to $\mathbb{R}^4$.*

We begin the construction with $X = \mathbb{CP}^2 \# 9\, \overline{\mathbb{CP}}^2$. This is a 4-manifold and, as we pointed out above, Donaldson [Don2] has shown that X itself admits non-standard differentiable structures. Nevertheless, we will consider only the standard smooth structure on X coming from $\mathbb{CP}^2$ and $\overline{\mathbb{CP}}^2$. The intersection form of X can be written as $Q_X = (1) \oplus 9\,(-1) = (1) \oplus 8\,(-1) \oplus (-1)$ relative to some basis $\{e_0, e_1, \ldots, e_9\}$ for $H_2(X;\mathbb{Z})$, where $Q_X(e_0, e_0) = 1, Q_X(e_j, e_j) = -1$ for $j = 1, \ldots, 9$, and $Q_X(e_j, e_k) = 0$ for $j \neq k$. Now, consider the element

$$\alpha = 3e_0 + e_1 + \cdots + e_8 \in \operatorname{Span}\{e_0, \ldots, e_8\} \subseteq H_2(X;\mathbb{Z}).$$

Then

$$Q_X(\alpha, \alpha) = 1$$

and, for any $\beta = x^0 e_0 + x^1 e_1 + \cdots + x^8 e_8 + x^9 e_9 \in H_2(X;\mathbb{Z})$,

$$Q_X(\alpha, \beta) = 3x^0 - x^1 - \cdots - x^8.$$

Consequently, β is in the Q_X-orthogonal complement of α if and only if

$$3x^0 = x^1 + \cdots + x^8 \quad \text{and} \quad x^9 \in \mathbb{Z} \quad \text{is arbitrary.}$$

Write $\langle\alpha\rangle^\perp$ for the Q_X-orthogonal complement of α in $H_2(X;\mathbb{Z})$. The eight elements

$$\{ e_2 - e_1,\ e_2 - e_3,\ e_4 - e_3,\ e_4 - e_5,\ e_6 - e_5,\ e_6 - e_7,\ e_8 - e_7,\ e_0 + e_6 + e_7 + e_8 \}$$

of Span $\{e_0, \ldots, e_8\}$ are easily seen to form a basis for $\langle\alpha\rangle^\perp \cap \mathrm{Span}\{e_0, \ldots, e_8\}$. Computing Q_X of its various pairs of elements one finds that, relative to this basis, the matrix of the restricted intersection form Q_X is just $-E_8$. The restriction of Q_X to all of $\langle\alpha\rangle^\perp$ is therefore $-E_8 \oplus (-1)$. Consequently, Q_X can be written as

$$Q_X = -E_8 \oplus (-1) \oplus (1).$$

According to Freedman's Theorem 1.5 this corresponds to a *topological* splitting

$$X = N \# \mathbb{CP}^2$$

of X into the connected sum of a topological 4-manifold N with intersection form $Q_N = -E_8 \oplus (-1)$ and $\mathbb{CP}^2$. Notice, however, that this cannot be a *smooth* splitting since Q_N is negative definite, but not equivalent to minus the identity so that, by Donaldson's Theorem 1.3, N cannot not admit a smooth structure at all.

The form $-E_8 \oplus (-1)$ is defined on $\langle\alpha\rangle^\perp = H_2(N;\mathbb{Z})$ so α generates the homology of the $\mathbb{CP}^2$-summand in $X = N \# \mathbb{CP}^2$. We have already pointed out that every homology class in $H_2(X;\mathbb{Z})$ can be represented by a smoothly embedded, compact, oriented surface in X. However, the genus of such a representing surface is not unique and determining the minimal such genus is generally difficult. In the case of our $\alpha \in H_2(X;\mathbb{Z})$ we would like to argue that this minimal genus is greater than zero, that is, that *α cannot be represented by a smoothly embedded 2-sphere in X.* We will provide only a brief sketch of the argument. Let us suppose to the contrary that α is represented by some smooth embedding of a 2-sphere Σ_s in X. Identify Σ_s with a smooth submanifold of X diffeomorphic to $\mathbb{S}^2$. Then Σ_s has a tubular neighborhood in X and any two such are isotopic (see Theorems 5.2 and 5.3 of [Hirsch]). More specifically, we construct a neighborhood of Σ_s in X as follows. Fix some Riemannian metric g on X. Then for $\epsilon > 0$ sufficiently small, we define

$$\mathcal{N}_g(\epsilon) = \{ x \in X : dist_g(x, \Sigma_s) < \epsilon \}$$

and

$$\overline{\mathcal{N}}_g(\epsilon) = \{ x \in X : dist_g(x, \Sigma_s) \leq \epsilon \}.$$

Define a projection map

$$\pi_{\mathcal{N}} : \overline{\mathcal{N}}_g(\epsilon) \to \Sigma_s$$

by letting $\pi_{\mathcal{N}}(x)$ be the point in Σ_s closest to x for every $x \in \overline{\mathcal{N}}_g(\epsilon)$. Then $\mathcal{N}_g(\epsilon)$ is an open neighborhood of Σ_s in X and $(\overline{\mathcal{N}}_g(\epsilon), \pi_{\mathcal{N}})$ is a disc bundle over Σ_s. Each such disc bundle over Σ_s is characterized by an integer, generally called the *Euler number*, which can be computed in a variety of ways, one of which is the self-intersection $Q_X(\alpha, \alpha)$ of the homology class α represented by Σ_s (see Chapter 4, Section 1, page 103, and Exercise 1.4.11 (b), page 28, of [GS]). In our case $Q_X(\alpha, \alpha) = 1$ so the Euler number of $(\overline{\mathcal{N}}_g(\epsilon), \pi_{\mathcal{N}})$ is 1. Now we would like to compare this with an analogous situation in $\mathbb{CP}^2$. The homology $H_2(\mathbb{CP}^2;\mathbb{Z})$ of $\mathbb{CP}^2$ is isomorphic to $\mathbb{Z}$ and is generated by the fundamental class $[\mathbb{CP}^1]$ of any copy of $\mathbb{CP}^1 \cong \mathbb{S}^2$ in $\mathbb{CP}^2$. The self-intersection $Q_{\mathbb{CP}^2}([\mathbb{CP}^1], [\mathbb{CP}^1])$ is 1 so the disc bundle over $\mathbb{CP}^1$ constructed as above for Σ_s also has Euler number 1. It follows, in particular, that the neighborhood $\mathcal{N}_g(\epsilon)$ of Σ_s in X is diffeomorphic to a smooth tubular neighborhood of $\mathbb{CP}^1$ in $\mathbb{CP}^2$ which has boundary diffeomorphic to $\mathbb{S}^3$ (because the boundary of a disc bundle is a circle bundle and Euler number 1 implies that this must be the Hopf bundle) and complement diffeomorphic to a standard 4-dimensional closed ball. From this it follows that we can cut out Σ_s and its tubular neighborhood from X and smoothly glue in that 4-ball instead. Since the homology of this 4-ball is trivial, this gives a new compact, connected, simply connected 4-manifold in which the homology in X that was generated by α has been killed. As a result, this new smooth 4-manifold has intersection form $-E_8 \oplus (-1)$ and we have already seen that this contradicts Donaldson's Theorem 1.3.

Thus, α cannot be represented by a smoothly embedded 2-sphere in X. On the other hand, a deep result of Freedman on Casson handles (Theorem 1.1 of [Freed]) implies that α can be represented by a *topologically* embedded 2-sphere Σ_t in X and that Σ_t is has an open neighborhood U in X that is *homeomorphic* to a disc bundle over $\mathbb{S}^2$ with Euler number 1. As above, U is then homeomorphic to an open subset of $\mathbb{CP}^2$. Using a topological embedding of U in $\mathbb{CP}^2$ we can transfer the topological sphere Σ_t to $\mathbb{CP}^2$. We will use the same symbols U and Σ_t for these subsets of $\mathbb{CP}^2$. Freedman showed that the complement $\mathbb{CP}^2 - \Sigma_t$ of Σ_t in $\mathbb{CP}^2$ is homeomorphic to an open 4-dimensional ball and therefore homeomorphic to $\mathbb{R}^4$. This open ball $\mathbb{CP}^2 - \Sigma_t$ inherits a smooth structure from $\mathbb{CP}^2$ and we would like to argue that, with this smooth structure, $\mathbb{CP}^2 - \Sigma_t$ cannot be diffeomorphic to $\mathbb{R}^4$ with its standard smooth structure, that is, that the submanifold $\mathbb{CP}^2 - \Sigma_t$ of $\mathbb{CP}^4$ is an exotic $\mathbb{R}^4$. The idea is to apply Lemma 1.9 to $\mathbb{CP}^2 - \Sigma_t$.

We will argue that the compact subset $\mathbb{CP}^2 - U$ of $\mathbb{CP}^2 - \Sigma_t$ cannot be enclosed by any smoothly embedded 3-sphere in $\mathbb{CP}^2 - \Sigma_t$. Suppose to the contrary that $\mathbb{CP}^2 - U$ is contained inside the smoothly embedded 3-sphere S in $\mathbb{CP}^2 - \Sigma_t$. Then S is contained in U. Now return to $X = \mathbb{CP}^2 \# 9\, \overline{\mathbb{CP}}^2$. The 3-sphere S also embeds smoothly in X. Thus, Σ_t is enclosed in X by a smoothly embedded 3-sphere. Now we can proceed exactly as in the case of Σ_s by cutting the 3-sphere out of X and smoothly gluing in a 4-dimensional ball, thereby killing the homology in X generated by α. The result is again a compact, connected, simply connected 4-manifold with intersection form $-E_8 \oplus (-1)$ and this violates Donaldson's Theorem 1.3. This contradiction implies that $\mathbb{CP}^2 - \Sigma_t$ is not diffeomorphic to $\mathbb{R}^4$ and is therefore an exotic $\mathbb{R}^4$.

1.4. **Donaldson Invariants.** In Section 1.2 we witnessed the emergence of the Yang-Mills equations and gauge-theoretic techniques in differential topology. Donaldson [Don1] proved his Theorem 1.3 on the intersection form of a compact, connected, simply connected, oriented, smooth 4-manifold X with $b_2^+(X) = 0$ (or $b_2^-(X) = 0$) by studying the moduli space of gauge equivalence classes of anti-self-dual connections on the Sp(1)-bundle over X with Chern number 1. Here we will briefly describe the program carried out by Donaldson [Don3] to refine these techniques in order to produce a family of very sensitive invariants for smooth 4-manifolds. The technical work required to do this is prodigious and, in a sense, the invariants themselves were superseded with the appearance of Seiberg-Witten theory (Section 1.6) so we will not belabor the technical issues. A more detailed outline and an account of many of the details is to be found in [Mor2] with still more details in Chapter III of [FM]; for the full story, one should consult [DK].

1.4.1. *Moduli Spaces of Connections.* Throughout this section X will denote a compact, connected, simply connected, oriented, smooth 4-manifold. When the need arises we will also specify assumptions concerning $b_2^+(X)$. We will write

$$\mathrm{Sp}(1) \hookrightarrow P_k \stackrel{\pi_k}{\to} X$$

for the principal Sp(1)-bundle over X with Chern number k.

Remark 1.7. The Lie group Sp(1) is isomorphic to SU(2) which, in turn, is the universal double cover of the rotation group SO(3). The entire program we intend to describe can be carried out equally well with SO(3)-bundles over X. Locally these two approaches are entirely equivalent, but there can be important global differences. The reason is that, whereas SU(2)-bundles over X are characterized by the second Chern class c_2, SO(3)-bundles over X are characterized by the first Pontrjagin class p_1 and the second Stiefel-Whitney class w_2. The flexibility of being able to choose w_2 can be quite useful (see Section II (vi) of [Don3], Sections 9.1.1, 9.1.2 and 9.1.3 of [DK] and also [FS1]). We will consider only the Sp(1) case here, but [DK] treats both in detail.

$\mathcal{A}(P_k)$ will denote the set of all smooth connection 1-forms on P_k and $\mathcal{G}(P_k)$ is the group of all difeomorphisms f of P_k onto itself that respect the right action of Sp(1) on P_k ($f(p \cdot g) = f(p) \cdot g$ for all $p \in P_k$ and all $g \in \mathrm{Sp}(1)$) and cover the identity map on X ($\pi_k \circ f = \pi_k$). $\mathcal{G}(P_k)$ is called the *gauge group*, its elements are called *(global) gauge transformations*, and it acts on $\mathcal{A}(P_k)$ on the right by pullback, that is, for every $\omega \in \mathcal{A}(P_k)$, $\omega \cdot f = f^*\omega \in \mathcal{A}(P_k)$. Two connections $\omega, \omega' \in \mathcal{A}(P_k)$ are said to be *gauge equivalent* if there exists an $f \in \mathcal{G}(P_k)$ for which $\omega' = f^*\omega$. The set of all gauge equivalence classes $[\omega]$ for $\omega \in \mathcal{A}(P_k)$ is called the *moduli space of connections* on $\mathrm{Sp}(1) \hookrightarrow P_k \xrightarrow{\pi_k} X$ and is written

$$\mathcal{B}(P_k) = \mathcal{A}(P_k)/\mathcal{G}(P_K).$$

Remark 1.8. As we mentioned in Remark 1.5 such moduli spaces generally have no reasonable mathematical structure due to the assumed smoothness of the connections and gauge transformations. To obtain objects that one can study analytically it is necessary to introduce various Sobolev completions. As we did earlier we will suppress these completions here and simply send those interested in seeing how it is done to the references provided in Remark 1.5. We should point out, however, that our next two Propositions are the keys to defining Sobolev completions of $\mathcal{A}(P_k)$ and $\mathcal{G}(P_k)$.

For each integer $i \geq 0$ we will denote by $\Omega^i(P_k, \mathfrak{sp}(1))$ the real vector space of all i-forms on P_k with values in the Lie algebra $\mathfrak{sp}(1)$ of Sp(1). $\Omega^i_{\mathrm{ad}}(P_k, \mathfrak{sp}(1))$ will denote the linear subspace of $\Omega^i(P_k, \mathfrak{sp}(1))$ consisting of those elements φ of $\Omega^i(P_k, \mathfrak{sp}(1))$ that are *tensorial of type ad*, that is, that satisfy the following two conditions (see Chapter II, Section 5, page 75, of [KN1]).

(1) φ vanishes if one of its arguments is vertical (that is, tangent to a fiber of P_k).

(2) If $g \in \mathrm{Sp}(1)$ and $\sigma_g : P_k \to P_k$ is the diffeomorphism given by $\sigma_g(p) = p \cdot g$, then $\sigma_g^*\varphi = g^{-1}\varphi g$.

Next let $\Omega^i(X, \mathrm{ad}\, P_k)$ denote the vector space of all i-forms on P_k with values in the adjoint bundle ad P_k (the vector bundle associated to P_k by the adjoint (conjugation) action of Sp(1) on its Lie algebra $\mathfrak{sp}(1)$). This is easily seen to be isomorphic to $\Omega^i_{\mathrm{ad}}(X, \mathfrak{sp}(1))$. If ω and ω_0 are two connection 1-forms on P_k, then $\omega - \omega_0$ is tensorial of type ad so we have the following.

Proposition 1.10. *$\mathcal{A}(P_k)$ is an affine space modeled on the vector space $\Omega^1_{ad}(X, \mathfrak{sp}(1)) \cong \Omega^1(X, ad\, P_k)$. In particular, if ω_0 is any fixed element of $\mathcal{A}(P_k)$, then*

$$\mathcal{A}(P_k) = \{\, \omega_0 + \varphi : \varphi \in \Omega^1_{ad}(P_k, \mathfrak{sp}(1)\,\}.$$

Similarly, one can consider the nonlinear adjoint bundle $\mathrm{Ad}\,P_k$. This is the fiber bundle associated to P_k by the adjoint (conjugation) action of Sp(1) on itself. Let $\Omega^0(X, \mathrm{Ad}\,P_k)$ be the set of smooth sections of $\mathrm{Ad}\,P_k$. This is a group under pointwise multiplication in the fibers and is easily seen to be isomorphic to the group $\Omega^0_{\mathrm{Ad}}(P_k, \mathrm{Sp}(1))$ of smooth maps $\psi : P_k \to \mathrm{Sp}(1)$ that satisfy $\psi(p \cdot g) = g^{-1}\psi(p)g$ for all $p \in P_k$ and all $g \in \mathrm{Sp}(1)$. One can then prove the following (see Lemma 4.1.2 of [Mor2]).

Proposition 1.11. $\mathcal{G}(P_k) \cong \Omega^0_{Ad}(P_k, Sp(1)) \cong \Omega^0(X, Ad\,P_k)$

If ω is in $\mathcal{A}(P_k)$, then the *stabilizer* $\mathrm{Stab}(\omega)$ of ω is the subgroup of $\mathcal{G}(P_k)$ that leaves ω fixed.

$$\mathrm{Stab}(\omega) = \{ f \in \mathcal{G}(P_k) : \omega \cdot f = \omega \}$$

Any such stabilizer contains the subgroup $\mathbb{Z}_2$ of $\mathcal{G}(P_k)$ generated by the constant sections through ± 1 in Sp(1). If $\mathrm{Stab}(\omega) = \mathbb{Z}_2$, then ω is said to be *irreducible*; otherwise, ω is *reducible*. The reason for the terminology is that, when ω is reducible, $\mathrm{Stab}(\omega)/\mathbb{Z}_2 \cong U(1)$ and ω is induced from a connection on a $U(1)$-bundle (see Corollary 4.3.5 of [Mor2]). We will see that reducible and irreducible connections play very different roles in the analysis of the moduli spaces. We will write $\hat{\mathcal{A}}(P_k)$ for the open subset of $\mathcal{A}(P_k)$ consisting of the irreducible connections and will introduce, in addition to $\mathcal{B}(P_k) = \mathcal{A}(P_k)/\mathcal{G}(P_k)$ the *moduli space of irreducible connections*

$$\hat{\mathcal{B}}(P_k) = \hat{\mathcal{A}}(P_k)/\mathcal{G}(P_k).$$

The objects of real interest in Donaldson theory are certain subspaces of $\mathcal{B}(P_k)$ and $\hat{\mathcal{B}}(P_k)$. One describes them by first choosing a Riemannian metric g on X. Together with the orientation of X this gives a Hodge star operator * on forms defined on X. We will say that a connection ω on P_k is *g-anti-self dual* (g-ASD) if the curvature form $F_\omega \in \Omega^2(X, \mathrm{ad}\,P_k)$ satisfies $^*F_\omega = -F_\omega$. Note, however, that such connections can exist only if the Chern number is non-negative. Indeed,

$$k = \frac{1}{8\pi^2}\int_X \mathrm{tr}\,(F_\omega \wedge F_\omega) = \frac{1}{8\pi^2}\int_X (\,|F^-_\omega|^2 - |F^+_\omega|^2\,)\,\mathrm{vol}, \tag{4}$$

where $F^\pm_\omega = \frac{1}{2}(F_\omega \pm {}^*F_\omega)$ are the self-dual and anti-self-dual parts of F_ω. Thus, $F^+_\omega = 0$ if and only if $k \geq 0$. We will therefore restrict our attention henceforth to bundles P_k with

$$k \geq 0.$$

For any such k we let $\mathcal{A}_{ASD}(P_k, g)$ denote the subset of $\mathcal{A}(P_k)$ consisting of all g-ASD connections and $\hat{\mathcal{A}}_{ASD}(P_k, g)$ the subset of $\mathcal{A}_{ASD}(P_k, g)$ consisting of the irreducible elements. The corresponding moduli spaces of gauge equivalence classes are

$$\mathcal{M}(P_k, g) = \mathcal{A}_{ASD}(P_k, g)/\mathcal{G}(P_k)$$

and

$$\hat{\mathcal{M}}(P_k, g) = \hat{\mathcal{A}}_{ASD}(P_k, g)/\mathcal{G}(P_k).$$

For a given X, g, and $k \geq 0$, $\mathcal{A}_{ASD}(P_k, g)$ (and therefore $\mathcal{M}(P_k, g)$) might well be empty. This is the case, for example, when $k = 1$ if X is either $\mathbb{CP}^2$ or $\mathbb{S}^2 \times \mathbb{S}^2$ and g is the standard metric (Fubini-Study in the case of $\mathbb{CP}^2$), but not if $k = 2$ (see Examples 4.1.3 and 4.1.5 of [DK]). In Section 1.1 we have described $\mathcal{M}(P_k, g)$

when $X = \mathbb{S}^4$, $k = 1$, and g is the standard metric on $\mathbb{S}^4$. One can show that, in this case, every connection in $\mathcal{A}_{ASD}(P_1, g)$ is irreducible so $\mathcal{M}(P_1, g) = \hat{\mathcal{M}}(P_1, g)$.

For any X and any g the $k = 0$ bundle is trivial and it follows from (4) that any g-ASD connection is necessarily flat. Conversely, a flat connection is certainly g-ASD. Since flat connections exist on any trivial bundle, the moduli space $\mathcal{M}(P_0, g)$ is nonempty. However, since we have assumed X is simply connected, any two flat connections on P_0 are gauge equivalent so $\mathcal{M}(P_0, g)$ consists of a single point (see Chapter II, Section 9, of [KN1]). For this reason we will henceforth restrict our attention to

$$k > 0.$$

An example of particular interest is $\overline{\mathbb{CP}}^2$ with its standard metric and $k = 1$ (see Example 4.1.2 of [DK]). In this case, one can write out explicit formulas for representatives of each gauge equivalence class of g-ASD connections very much as we did for $\mathbb{S}^4$ in Section 1.1.2. The moduli space $\hat{\mathcal{M}}(P_1, g)$ turns out to be an open cone on $\mathbb{CP}^2$. The base of the closed cone is a copy of $\overline{\mathbb{CP}}^2$ each point of which corresponds to a sequence of irreducible g-ASD connections becoming ever more concentrated at a point. The vertex of the closed cone corresponds to a reducible g-ASD connection. Reducible connections always give rise to such a cone singularity in the moduli space. By adding the base and the vertex one obtains a natural compactification of $\hat{\mathcal{M}}(P_1, g)$. One should compare this example with both the discussion of the 4-sphere in Section 1.1.3 and the scenario described in Section 1.2.2 for Donaldson's Theorem 1.3 (see Figure 2).

Sorting out the mathematical structure of the moduli spaces $\mathcal{M}(P_k, g)$ and $\hat{\mathcal{M}}(P_k, g)$ when $k > 0$ requires an enormous amount of very delicate technical work. The path one must follow is outlined in [Nab1] (pages 44-49), while a more thorough discussion can be found in [Mor2] (pages 95-107) and the whole story is in Chapters 4 and 5 of [DK]. We must be content with a simple statement of the results we need. For these we require one more preliminary result.

We will denote by $\mathcal{R}(X)$ the space of all Riemannian metrics on X. This is a space of sections of a fiber bundle over X and can be given the structure of a pathwise connected Banach manifold (see Lemma 5.6.1 of [Mor2]). With this structure one can prove the following (see Theorem 5.3.1 of [Mor2]).

Theorem 1.12. *Let X be a compact, connected, simply connected, oriented, smooth 4-manifold with $b_2^+(X) > 0$ and let $k > 0$ be an integer. Then there is a dense G_δ-set $\mathcal{R}_G(X)$ in $\mathcal{R}(X)$, the elements of which are called generic Riemannian metrics on X, such that for every generic $g \in \mathcal{R}_G(X)$, $\mathcal{M}(P_k, g) = \hat{\mathcal{M}}(P_k, g)$ is a smooth manifold of dimension*

$$8k - 3(1 + b_2^+(X)).$$

Very roughly, here is how one might go about showing that $\hat{\mathcal{M}}(P_k, g)$ is a smooth manifold for a generic choice of g. Define

$$\hat{\mathcal{M}}(P_k, \mathcal{R}(X)) = \{\, ([\omega], g) \in \hat{\mathcal{B}}(P_k) \times \mathcal{R}(X) : \omega \text{ is } g\text{-ASD}\,\}.$$

This is an infinite-dimensional smooth submanifold of $\hat{\mathcal{B}}(P_k) \times \mathcal{R}(X)$. One shows that the projection map

$$\hat{\mathcal{M}}(P_k, \mathcal{R}(X)) \to \mathcal{R}(X)$$

is smooth with Fredholm derivative at each point. Then the Banach space version of Sard's Theorem (see Theorem 3.6.12 of [AMR]) implies that the set of regular values of the projection map is a dense G_δ-set in $\mathcal{R}(X)$. But then, for any g in this set, the inverse image of g

$$\left(\hat{\mathcal{B}}(P_k) \times \{g\}\right) \cap \hat{\mathcal{M}}(P_k, \mathcal{R}(X)) = \hat{\mathcal{M}}(P_k, g)$$

under the projection map is a smooth submanifold of $\hat{\mathcal{M}}(P_k, \mathcal{R}(X))$. Calculating the dimension of this manifold is much more subtle. One associates with each g-ASD ω a certain elliptic complex, computes its index from the Atiyah-Singer Index Theorem, and shows that, under the assumptions we have made, this is just the dimension of $\hat{\mathcal{M}}(P_k, g)$ (see pages 44-47 of [Nab1] for a sketch and Section 5.2, pages 96-98, of [Mor2] for more details).

A few remarks are in order. If $8k - 3(1 + b_2^+(X))$ is negative, then the moduli space is generically empty. The restriction on $b_2^+(X)$ arises because the subset of $\mathcal{R}(X)$ consisting of those g for which reducible g-ASD connections on P_k exist is a countable union of smooth submanifolds of codimension $b_2^+(X)$. Thus, if $b_2^+(X) = 0$, reducibles are generically unavoidable and one expects cone singularities in the moduli space $\mathcal{M}(P_k, g)$. As we saw in Section 1.2.2, this is a good, not a bad thing since the singularities lead to Donaldson's Theorem 1.3.

Crudely put, the idea behind defining the Donaldson invariants is to regard the moduli space $\hat{\mathcal{M}}(P_k, g)$ as a cycle in $\hat{\mathcal{B}}(P_k)$ over which to integrate certain differential forms on $\hat{\mathcal{B}}(P_k, g)$. Somewhat less crudely, the invariants are obtained by pairing the cohomology of $\hat{\mathcal{B}}(P_k, g)$ with the fundamental class $[\hat{\mathcal{M}}(P_k, g)]$ of $\hat{\mathcal{M}}(P_k, g)$. In order to carry out such a program a great many questions need to be answered: What are these cohomology classes of $\hat{\mathcal{B}}(P_k, g)$? Does $\hat{\mathcal{M}}(P_k, g)$ have a fundamental class? At very least, $\hat{\mathcal{M}}(P_k, g)$ must be orientable and, if the result is to be a differential topological invariant, the integrals must be independent of the choice of generic metric g. The following result of Donaldson addresses the first of these issues.

Theorem 1.13. *Let X be a compact, connected, simply connected, oriented, smooth 4-manifold with $b_2^+(X) > 0$. Let $g \in \mathcal{R}_G(X)$ be a generic Riemannian metric on X and $k > 0$ an integer. Then the smooth $(8k - 3(1 + b_2^+(X))$-dimensional manifold $\hat{\mathcal{M}}(P_k, g)$ is orientable. An orientation for $\hat{\mathcal{M}}(P_k, g)$ is determined by the given orientation of X and a choice of orientation for the vector space $H_+^2(X; \mathbb{R})$ of real-valued, self-dual 2-forms on X.*

The proof amounts to constructing an explicit model of the determinant line bundle (top exterior power) of the tangent bundle of $\hat{\mathcal{M}}(P_k, g)$ from a family of differential operators on X and exhibiting a nonzero section (see Theorem 5.5.1 of [Mor2]).

Regarding the metric independence one proceeds in the following way. Let g_0 and g_1 be two elements of $\mathcal{R}_G(X)$. Since $\mathcal{R}(X)$ is a pathwise connected Banach manifold we can join g_0 and g_1 with a smooth path $\{g_t : 0 \leq t \leq 1\}$ in $\mathcal{R}(X)$. Donaldson has shown that, if $b_2^+(X) > 1$, then an arbitrarily small perturbation of the path $\{g_t : 0 \leq t \leq 1\}$ will encounter no reducible connections, that is, for a generic path from g_0 to g_1, $\mathcal{M}(P_k, g_t) = \hat{\mathcal{M}}(P_k, g_t)$ for each $t \in [0, 1]$. We can then define the corresponding *parametrized moduli space* $\hat{\mathcal{M}}(P_k, \{g_t\})$ to be the subspace of $\hat{\mathcal{B}}(P_k) \times [0, 1]$ given by

$$\hat{\mathcal{M}}(P_k, \{g_t\}) = \left\{ ([\omega], t) \in \hat{\mathcal{B}}(P_k) \times [0, 1] : [\omega] \in \hat{\mathcal{M}}(P_k, g_t) \right\}$$

The proof of the following is another application of the infinite-dimensional version of Sard's Theorem (see Theorems 5.6.2 and 5.6.3 of [Mor2]).

Theorem 1.14. *Let X be a compact, connected, simply connected, oriented smooth 4-manifold with $b_2^+(X) > 1$ and let $k > 0$ be an integer. Let $g_0, g_1 \in \mathcal{R}_G(X)$ be generic Riemannian metrics on X. Then, for a generic path $\{g_t : 0 \leq t \leq 1\}$ joining g_0 and g_1 in $\mathcal{R}(X)$, the parametrized moduli space $\hat{\mathcal{M}}(P_k, \{g_t\})$ is a smooth, orientable manifold with boundary. A choice of orientation $\mathcal{O}$ for $H^2_+(X; \mathbb{R})$ determines an orientation for $\hat{\mathcal{M}}(P_k, \{g_t\})$. The oriented boundary of $\hat{\mathcal{M}}(P_k, \{g_t\})$ is the disjoint union of $\hat{\mathcal{M}}(P_k, g_1)$ with the orientation induced by $\mathcal{O}$ and $\hat{\mathcal{M}}(P_k, g_0)$ with the orientation opposite to the one induced by $\mathcal{O}$.*

A rough, intuitive interpretation of this result is that, if $b_2^+(X) > 1$, then a generic variation of g varies $\hat{\mathcal{M}}(P_k, g)$ within a single homology class of $\hat{\mathcal{B}}(P_k)$ so pairing either $\hat{\mathcal{M}}(P_k, g_0)$ or $\hat{\mathcal{M}}(P_k, g_1)$ with a cohomology class of $\hat{\mathcal{B}}(P_k)$ will give the same result. Since $g_0, g_1 \in \mathcal{R}_G(X)$ were arbitrary it will follow that the Donaldon invariants (once they are defined in the manner suggested above) are independent of the choice of generic Riemannian metric on X. For this reason we will assume henceforth that

$$b_2^+(X) > 1.$$

1.4.2. *The Uhlenbeck Compactification.* There is much that remains to be done in order to carry out the program we have described for constructing the Donaldson invariants. Since the idea is to pair certain cohomology classes of $\hat{\mathcal{B}}(P_k)$ with a homology class determined by the $\hat{\mathcal{M}}(P_k, g)$ for generic g one must isolate

(1) appropriate cohomology classes of $\hat{\mathcal{B}}(P_k)$, and
(2) an appropriate homology class of $\hat{\mathcal{B}}(P_k)$ determined by $\hat{\mathcal{M}}(P_k, g)$.

An initial step toward the first of these issues is, in Donaldson's view, "an exercise in algebraic topology" (see page 9 of [Don4]) in which he constructs maps

$$\mu : H_2(X; \mathbb{Z}) \to H^2(\hat{\mathcal{B}}(P_k); \mathbb{Z}) \tag{5}$$

and

$$\mu : H_0(X; \mathbb{Z}) \to H^4(\hat{\mathcal{B}}(P_k); \mathbb{Z}) \tag{6}$$

and, by restriction, maps

$$\mu : H_2(X; \mathbb{Z}) \to H^2(\hat{\mathcal{M}}(P_k, g); \mathbb{Z}) \tag{7}$$

and

$$\mu : H_0(X; \mathbb{Z}) \to H^4(\hat{\mathcal{M}}(P_k, g); \mathbb{Z}) \tag{8}$$

for which it is customary to use the same symbol. Any of these is likely to be referred to as the *Donaldson μ-map* (see pages 59-61 of [Nab1] for an informal sketch of the construction, Section 7.2 of [Mor2] or Sections 3.1.5 - 3.1.10 of [FM] for more details, and Sections 5.1 and 5.2 of [DK] for the whole story). For any $\alpha \in H_2(X; \mathbb{Z})$ we then have a 2-dimensional cohomology class $\mu(\alpha) \in H^2(\hat{\mathcal{M}}(P_k, g); \mathbb{Z})$ and, roughly speaking, the idea is to form a sufficiently large product $\mu(\alpha) \smile \cdots \smile \mu(\alpha)$ that can be paired with $\hat{\mathcal{M}}(P_k, g)$. In order for this to succeed, of course, the dimension $8k - 3(1 + b_2^+(X))$ of the moduli space must be even since every $\mu(\alpha)$ has degree two. This is the case if and only if $b_2^+(X)$ is odd so, from this point on, we will assume

$$b_2^+(X) \;\text{ is odd and }\; b_2^+(X) \geq 3.$$

Under these assumptions it will be convenient to write

$$2d_k = 8k - 3(1 + b_2^+(X)) \tag{9}$$

so that the cohomology classes we would like to pair with $\mathcal{M}(P_k, g)$ are $\mu(\alpha) \smile \overset{d_k}{\cdots} \smile \mu(\alpha)$ for $\alpha \in H_2(X; \mathbb{Z})$. For reasons that we will discuss next the μ-map is not quite up to the task we have in mind for it and we will need to make some adjustments in the program we have been describing.

Dealing with the second issue is not by any means an "exercise". If we think, intuitively, of a second homology class as a surface Σ and a second cohomology class as a 2-form, then we would ideally like to "integrate $\mu(\Sigma) \wedge \overset{d_k}{\cdots} \wedge \mu(\Sigma)$ over $\hat{\mathcal{M}}(P_k, g)$". We have seen that we can assume $\hat{\mathcal{M}}(P_k, g)$ is a smooth, finite-dimensional manifold. If $\hat{\mathcal{M}}(P_k, g)$ were compact we could integrate over it. More precisely, if $\hat{\mathcal{M}}(P_k, g)$ were compact it would carry a fundamental class $[\hat{\mathcal{M}}(P_k, g)]$ which we could pair with any $\mu(\alpha) \smile \overset{d_k}{\cdots} \smile \mu(\alpha)$ for $\alpha \in H_2(X; \mathbb{Z})$. . However, as we have seen, the manifolds $\hat{\mathcal{M}}(P_k, g)$ are generally not compact so such a fundamental class need not exist. To rectify the situation Donaldson employed deep analytical results of Uhlenbeck to introduce what has come to be known as the *Uhlenbeck compactification* $\overline{\mathcal{M}}(P_k, g)$ of $\mathcal{M}(P_k, g)$. We will have a very brief look at how this is done.

Since we have assumed that $b_2^+(X) \geq 3$ and g is generic, $\mathcal{M}(P_k, g) = \hat{\mathcal{M}}(P_k, g)$. To construct a compactification one must understand how a sequence in $\hat{\mathcal{M}}(P_k, g)$ can fail to have a convergent subsequence in $\hat{\mathcal{M}}(P_k, g)$ so we consider a sequence $\{\omega_j\}$ of irreducible g-ASD connections on P_k that gives rise to a sequence $\{[\omega_j]\}$ in $\hat{\mathcal{M}}(P_k, g)$ without a convergent subsequence.

Remark 1.9. If our hypotheses had not excluded the presence of reducible g-ASD connections, then the sequence $\{\omega_j\}$ of irreducible g-ASD connections could converge to a reducible g-ASD connection ω. Then ω would correspond to a point $[\omega]$ in $\mathcal{M}(P_k, g)$ at which there is a cone singularity of the sort we saw in Donaldson's Theorem 1.3 (see Figure 2). In our present context this cannot occur.

The Compactness Theorem of Karen Uhlenbeck (see Theorem 6.1.1 of [Mor2]) implies that, after passing to a subsequence, the sequence $\{\mathcal{F}_{\omega_j}\}$ of curvatures must have pointwise norms $\| \mathcal{F}_{\omega_j}(x) \|$ which, like the BPST instantons of Section 1.1.2, become increasingly concentrated at a point (or finite set of points) in X. Intuitively, these connections "converge to a connection with curvature supported on a finite set of isolated points in X" which, because of its singular nature, does not correspond to a point in the moduli space $\mathcal{M}(P_k, g)$. The compactification adds these points and this essentially amounts to adding a copy of X. For example, it will add the $\mathbb{S}^4$ boundary of the moduli space of BPST instantons in Section 1.1.3 and the copy of X at the bottom of Figure 2.

To describe the Uhlenbeck compactification $\overline{\mathcal{M}}(P_k, g)$ of $\mathcal{M}(P_k, g)$ more precisely we recall that, for any topological space X and any integer $t \geq 1$, the t^{th} *symmetric product* of X is the topological quotient $s^t(X)$ of

$X \times \overset{t}{\cdots} \times X$ by the action of the symmetric group on t letters that permutes the coordinates. It is therefore the space of unordered and possibly non-distinct t-tuples of points in X. Intuitively, one should think of these as the points at which the connections can concentrate. In particular, $s^1(X)$ is just X. Now consider the disjoint union

$$\mathcal{M}(P_k, g) \quad \sqcup \quad \mathcal{M}(P_{k-1}, g) \times s^1(X) \quad \sqcup \quad \mathcal{M}(P_{k-2}, g) \times s^2(X) \quad \sqcup \cdots \sqcup \quad \mathcal{M}(P_0, g) \times s^k(X). \tag{10}$$

The elements of $\mathcal{M}(P_{k-t}, g) \times s^t(X)$ for any $t = 1, 2, \ldots, k$ are called *ideal g-ASD connections* on X. For any point of the form $(\,[\omega],\ \{x_1, \ldots, x_t\}\,)$ in $\mathcal{M}(P_{k-t}, g) \times s^t(X)$, ω is called a *background connection*. Uhlenbeck's Compactness Theorem (Theorem 6.1.1 of [Mor2]) motivates the appropriate way of defining sequential convergence in the disjoint union (10). For any point $(\,[\omega],\ \{x_1, \ldots, x_t\}\,)$ in $\mathcal{M}(P_{k-t}, g) \times s^t(X)$, for some $t = 1, 2, \ldots, k$, we introduce an associated measure on X called the *curvature density* and defined by

$$\|\,\mathcal{F}_\omega(x)\,\|^2\, d\,\mathrm{vol}_g + \sum_{l=1}^{t} 8\pi^2 \delta_{x_l},$$

where vol_g is the volume form on X determined by the metric g and δ_{x_l} is the point measure at x_l of mass 1. We will say that a sequence $(\,[\omega_n], \{x_1^n, \ldots, x_{t(n)}^n\}\,)$, $n = 1, 2, \ldots$, of points in the disjoint union (10) converges to $(\,[\omega], \{x_1, \ldots, x_t\}\,) \in \mathcal{M}(P_{k-t}, g) \times s^t(X)$ if the following two conditions are satisfied.

(1) The curvature densities for the $(\,[\omega_n], \{x_1^n, \ldots, x_{t(n)}^n\}\,)$ converge weakly to the curvature density for $(\,[\omega], \{x_1, \ldots, x_t\}\,)$ as $n \to \infty$, that is, for any continuous real-valued function φ on X,

$$\int_X \|\,\mathcal{F}_{\omega_n}(x)\,\|^2 \varphi(x)\, d\,\mathrm{vol}_g(x) + \sum_{l=1}^{t(n)} 8\pi^2 \varphi(x_l^n) \to \int_X \|\,\mathcal{F}_\omega(x)\,\|^2 \varphi(x)\, d\,\mathrm{vol}_g(x) + \sum_{l=1}^{t} 8\pi^2 \varphi(x_l).$$

(2) For every compact set $K \subseteq X - \{x_1, \ldots, x_t\}$ the restrictions of the connections ω_n to K converge, up to gauge equivalence, to the restriction of ω to K.

Entirely analogous definitions cover the cases in which some (or all) of the terms of the sequence are in $\mathcal{M}(P_k, g)$ (drop the corresponding sums to the left of the arrow) or the limit is in $\mathcal{M}(P_k, g)$ (drop the sum to the right of the arrow). These notions of convergence supply the disjoint union (10) with a topology and one can show that this topology is Hausdorff, second countable and metrizable. Moreover, $\mathcal{M}(P_k, g)$ and each of the "strata" $\mathcal{M}(P_{k-t}, g) \times s^t(X)$ inherits its usual topology as a subspace. The *Uhlenbeck compactification* $\overline{\mathcal{M}}(P_k, g)$ of $\mathcal{M}(P_k, g)$ is defined to be the closure of $\mathcal{M}(P_k, g)$ in this topology on the disjoint union (10). A detailed proof of Theorem 1.15 is available in Chapter 4, Section 4, pages 158-170, of [DK].

Theorem 1.15. *Let X be a compact, connected, simply connected, oriented, smooth 4-manifold with $b_2^+(X) > 1$, g a generic Riemannian metric on X, and $Sp(1) \hookrightarrow P_k \xrightarrow{\pi_k} X$ the principal Sp(1)-bundle over X with Chern number k. Then $\overline{\mathcal{M}}(P_k, g)$ is compact and contains $\mathcal{M}(P_k, g)$ as a dense, open subset.*

Remark 1.10. We should point out that our references [Mor2] and [FM] take the Uhlenbeck compactification to be the entire disjoint union (10) with the topology we have described. This is also compact, but $\mathcal{M}(P_k, g)$ is not dense unless k is sufficiently large so it may not really qualify as a "compactification" of $\mathcal{M}(P_k, g)$ in the usual sense.

The Uhlenbeck compactification $\overline{\mathcal{M}}(P_k, g)$ is therefore a separable, compact, metrizable topological space. It generally does not admit a manifold structure, however, so we are not *a priori* guaranteed the existence of a fundamental class. However, being separable and metrizable, $\overline{\mathcal{M}}(P_k, g)$ has a *topological dimension* that can be described in a number of equivalent ways (see Chapter 9, Section 2, pages 356-357, of [DK] where it is defined as the *covering dimension* of the topological space $\overline{\mathcal{M}}(P_k, g)$). Similarly, each "stratum" $\mathcal{M}(P_k, g) \times s^t(X)$, $t \geq 1$, has a topological dimension and one can deduce from general results in algebraic topology that if each $\mathcal{M}(P_k, g) \times s^t(X)$, $t \geq 1$, has codimension at least two in $\overline{\mathcal{M}}(P_k, g)$, then $\overline{\mathcal{M}}(P_k, g)$ carries a fundamental class

$$[\overline{\mathcal{M}}(P_k, g)] \in H_{2d_k}(\overline{\mathcal{M}}(P_k, g); \mathbb{Z}).$$

Writing out all of the relevant dimensions one finds that this codimension condition is satisfied only if the Chern number k is in the so-called *stable range* of X, that is,

$$k \geq \frac{1}{4}\left(5 + 3b_2^+(X) \right). \tag{11}$$

This is equivalent to

$$d_k \geq \frac{1}{2}\left(7 + 3b_2^+(X) \right) \tag{12}$$

so one can equally well say that the codimension condition is satisfied if d_k is in the stable range for X given by (12).

Even for k in the stable range there is one more obstacle to overcome before one can define the Donaldson invariants. The μ-map $\mu : H_2(X; \mathbb{Z}) \to H^2(\hat{\mathcal{M}}(P_k, g); \mathbb{Z})$ carries each $\alpha \in H_2(X; \mathbb{Z})$ to a cohomology class $\mu(\alpha)$ of $\hat{\mathcal{M}}(P_k, g)$, not $\overline{\mathcal{M}}(P_k, g)$, so one cannot pair $\mu(\alpha) \smile \overset{d_k}{\cdots} \smile \mu(\alpha)$ with $[\overline{\mathcal{M}}(P_k, g)]$. There are a number of ways to deal with this (see Section 9.2 of [DK]), but we will simply note that one can construct a map

$$\overline{\mu} : H_2(X; \mathbb{Z}) \to H^2(\overline{\mathcal{M}}(P_k, g); \mathbb{Z})$$

which, when followed by the restriction map,

$$H^2(\overline{\mathcal{M}}(P_k, g); \mathbb{Z}) \to H^2(\hat{\mathcal{M}}(P_k, g); \mathbb{Z})$$

is equal to μ. The construction of this extension of the μ-map proceeds one stratum $\mathcal{M}(P_{k-t}, g) \times s^t(X)$ at a time and requires a detailed understanding of the Taubes [Taub1] gluing procedure (see Section 7.4, page 127, of [Mor2] for a sketch).

1.4.3. *The Invariants.* With this one can finally begin introducing the *Donaldson invariants*. Lest they be lost in the turmoil of the preceding pages we will begin by laying down all of our hypotheses. Let X denote a compact, connected, simply connected, oriented, smooth 4-manifold with $b_2^+(X)$ odd and at least 3. Let $g \in \mathcal{R}_G(X)$ be a generic Riemannian metric for X and fix an orientation of $H_+^2(X; \mathbb{R})$. Let $k \geq \frac{1}{4}\left(5 + 3b_2^+(X) \right)$ be an integer in the stable range of X and write the dimension of the smooth moduli space $\mathcal{M}(P_k, g) = \hat{\mathcal{M}}(P_k, g)$ as $2d_k = 8k - 3(1 + b_2^+(X))$. With all of these assumptions in place we will define a *Donaldson invariant* $\gamma_{d_k}(X)$ of X. This can be thought of either as a d_k-multilinear, $\mathbb{Z}$-valued map on $H_2(X;\mathbb{Z})$ or, equivalently, as a $\mathbb{Z}$-valued, homogeneous polynomial of degree d_k on $H_2(X;\mathbb{Z})$ and it is customary to use the same symbol $\gamma_{d_k}(X)$ for both. As a multilinear map

$$\gamma_{d_k}(X) : H_2(X;\mathbb{Z}) \times \overset{d_k}{\cdots} \times H_2(X;\mathbb{Z}) \to \mathbb{Z}$$

it is given by

$$\gamma_{d_k}(X)(\alpha_1,\ldots,\alpha_{d_k}) = \langle\, \overline{\mu}(\alpha_1) \smile \cdots \smile \overline{\mu}(\alpha_{d_k}),\ [\overline{\mathcal{M}}(P_k,g)]\,\rangle.$$

The corresponding homogeneous polynomial

$$\gamma_{d_k}(X) : H_2(X;\mathbb{Z}) \to \mathbb{Z}$$

is defined by

$$\gamma_{d_k}(X)(\alpha) = \langle\, \overline{\mu}(\alpha) \smile \overset{d_k}{\cdots} \smile \overline{\mu}(\alpha),\ [\overline{\mathcal{M}}(P_k,g)]\,\rangle.$$

This is not quite the end, however. It will soon be important to have the definition of $\gamma_{d_k}(X)$ extended to operate on $H_0(X;\mathbb{Z})$ as well as $H_2(X;\mathbb{Z})$. One begins by returning to our, as yet unused, μ-map (8).

$$\mu : H_0(X;\mathbb{Z}) \to H^4(\hat{\mathcal{M}}(P_k,g);\mathbb{Z})$$

Unfortunately, unlike $\mu : H_2(X;\mathbb{Z}) \to H^2(\hat{\mathcal{M}}(P_k,g);\mathbb{Z})$, this map does not admit an extension to the Uhlenbeck compactification. More precisely, if we denote by x the generator of $H_0(X;\mathbb{Z}) \cong \mathbb{Z}$, the class $\mu(x) \in H^4(\hat{\mathcal{M}}(P_k,g);\mathbb{Z})$ does not extend to a cohomology class on $\overline{\mathcal{M}}(P_k,g)$. However, $\mu(x)$ does extend to a class $\overline{\mu}(x)$ on the complement of the set of points in $\overline{\mathcal{M}}(P_k;g)$ with trivial background connection and this is enough, under certain restrictions, to extend our definition of the Donaldson invariants to include the 4-dimensional classes. Basically, the restriction is that there must be "enough" 2-dimensional classes as well. Specifically, if a and b are non-negative integers with $d_k = 4k - \frac{3}{2}(1+b_2^+(X)) = a+2b$ and $a \geq \frac{1}{4}(5+3b_2^+(X))$, then, for $\alpha_1,\ldots,\alpha_a \in H_2(X;\mathbb{Z})$, the class

$$\overline{\mu}(\alpha_1) \smile \cdots \smile \overline{\mu}(\alpha_a) \smile \overline{\mu}(x)^b = \overline{\mu}(\alpha_1) \smile \cdots \smile \overline{\mu}(\alpha_a) \smile \overline{\mu}(x) \smile \overset{b}{\cdots} \smile \overline{\mu}(x)$$

can be paired with the fundamental class $[\overline{\mathcal{M}}(P_k,g)]$ and we can define

$$\gamma_{d_k}(X) : H_2(X;\mathbb{Z}) \times \overset{a}{\cdots} \times H_2(X;\mathbb{Z}) \times H_0(X;\mathbb{Z}) \times \overset{b}{\cdots} \times H_0(X;\mathbb{Z}) \to \mathbb{Z}$$

by

$$\gamma_{d_k}(X)\,(\alpha_1,\ldots,\alpha_a,n_1x,\ldots,n_bx) = n_1\cdots n_b\,\langle\, \overline{\mu}(\alpha_1) \smile \cdots \smile \overline{\mu}(\alpha_a) \smile \overline{\mu}(x)^b,\ [\overline{\mathcal{M}}(P_k,g)]\,\rangle.$$

For any non-negative integers a and b with $d_k = 4k - \frac{3}{2}(1 + b_2^+(X)) = a + 2b$ and $a \geq \frac{1}{4}(5 + 3b_2^+(X))$ the elements of

$$H_2(X;\mathbb{Z}) \times \overset{a}{\cdots} \times H_2(X;\mathbb{Z}) \times H_0(X;\mathbb{Z}) \times \overset{b}{\cdots} \times H_0(X;\mathbb{Z})$$

are called *k-stable* for X and $\gamma_{d_k}(X)$ is defined on any k-stable element.

At this point we have defined a Donaldson invariant $\gamma_d(X)$ for any $d \equiv -\frac{3}{2}(1+b_2^+(X))\, mod\, 4$ and $d = a+2b$ with a and b non-negative integers and $a \geq \frac{1}{4}(5 + 3b_2^+(X))$. These are called the *stable range Donaldson invariants*. Before proceeding with the definitions we will take a moment to list a few properties of these $\gamma_d(X)$ that remain true for all of those we still have to define (see Chapter III of [FM] and Chapter 9 of [DK]).

(1) $\gamma_d(X)$ does not depend on the choice of the generic metric g.
(2) The $\gamma_d(X)$ are invariant under orientation preserving diffeomorphisms of X that also preserve the chosen orientation for $H^2_+(X;\mathbb{R})$.
(3) Reversing the orientation of $H^2_+(X;\mathbb{R})$ reverses the sign of $\gamma_d(X)$.
(4) Reversing the orientation of X can have a much more dramatic effect on the Donaldson invariants. In many examples, the invariants all vanish for one of the two possible orientations for X.

Now we will briefly address the problem of extending the definition of $\gamma_d(X)$ to include still more values of d. Doing so depends on what are called *blowup formulas*. By definition, the *blowup* of a 4-manifold X is the 4-manifold $X\#\overline{\mathbb{CP}}^2$ obtained by forming the connected sum of X and $\overline{\mathbb{CP}}^2$. For any non-negative integer n, the *n-fold blowup* of X is

$$X\#\,\overline{\mathbb{CP}}^2\#\overset{n}{\cdots}\#\,\overline{\mathbb{CP}}^2 = X\#\,n\overline{\mathbb{CP}}^2.$$

Note that, since $b_2^+(\overline{\mathbb{CP}}^2) = 0$, $b_2^+(X) = b_2^+(X\#\,n\overline{\mathbb{CP}}^2)$ so the stable ranges of X and $X\#\,n\overline{\mathbb{CP}}^2$ are the same. Moreover, since $Q_{\overline{\mathbb{CP}}^2}$ is negative definite, there are no self-dual harmonic 2-forms on $\overline{\mathbb{CP}}^2$, so there is a natural identification of $H^2_+(X;\mathbb{R})$ and $H^2_+(X\#\,n\overline{\mathbb{CP}}^2;\mathbb{R})$. In particular, orienting $H^2_+(X;\mathbb{R})$ orients $H^2_+(X\#\,n\overline{\mathbb{CP}}^2;\mathbb{R})$ as well. The simplest of the blowup formulas ($n = 1$) can be stated in the following way. If d_k is in the stable range of X, then $d_{k+1} = d_k + 4$ is in the stable range of $X\#\overline{\mathbb{CP}}^2$. The blowup formula asserts that, if $e \in H_2(X\#\overline{\mathbb{CP}}^2;\mathbb{Z})$ is a generator for $H_2(\overline{\mathbb{CP}}^2;\mathbb{Z}) \subseteq H_2(X\#\overline{\mathbb{CP}}^2;\mathbb{Z}) \cong H_2(X;\mathbb{Z}) \oplus H_2(\overline{\mathbb{CP}}^2;\mathbb{Z})$, then

$$\gamma_{d_{k+1}}(X\#\overline{\mathbb{CP}}^2)(\alpha_1,\ldots,\alpha_{d_k},e,e,e,e) = -2\gamma_{d_k}(X)(\alpha_1,\ldots,\alpha_{d_k}) \tag{13}$$

(see Chapter III, Theorem 8.1, of [FM]). The point is that it is entirely possible for d_{k+1} to be in the stable range of $X\#\overline{\mathbb{CP}}^2$ even if d_k is not in the stable range of X and in this case we can *define* $\gamma_{d_k}(X)$ by (13), that is,

$$\gamma_{d_k}(X)(\alpha_1,\ldots,\alpha_{d_k}) = -\frac{1}{2}\gamma_{d_{k+1}}(X\#\overline{\mathbb{CP}}^2)(\alpha_1,\ldots,\alpha_{d_k},e,e,e,e). \tag{14}$$

Notice, however, that with this definition $\gamma_{d_k}(X)$ is no longer $\mathbb{Z}$-valued.

The general result is as follows (see Chapter III, Theorem 8.3, of [FM]). Let

$$s = (\alpha_1,\ldots,\alpha_a,n_1x,\ldots,n_bx)$$

be k-stable for X. Then $\gamma_{d_k}(X)(s)$ is defined. Consider, for some $n \geq 1$, the n-fold blowup $X\#\,n\overline{\mathbb{CP}}^2$ of X. For each $i = 1,\ldots,n$, let $e_i \in H_2(X\#\,n\overline{\mathbb{CP}}^2;\mathbb{Z})$ be the generator of the i^{th} $\overline{\mathbb{CP}}^2$-summand in $X\#\,n\overline{\mathbb{CP}}^2$. Since $d_{k+n} = d_k + 4n = (a + 4n) + 2b$,

$$(s,e_1,e_1,e_1,e_1,\ldots,e_n,e_n,e_n,e_n)$$

is $(k+n)$-stable for $X\#\,n\overline{\mathbb{CP}}^2$ so

$$\gamma_{d_{k+n}}(X\#\,n\overline{\mathbb{CP}}^2)(s,e_1,e_1,e_1,e_1,\ldots,e_n,e_n,e_n,e_n)$$

is defined. The blowup formula of Donaldson asserts that

$$\gamma_{d_k}(X)(s) = \left(-\frac{1}{2}\right)^n \gamma_{d_{k+n}}(X\#\,n\overline{\mathbb{CP}}^2)(s,e_1,e_1,e_1,e_1,\ldots,e_n,e_n,e_n,e_n). \tag{15}$$

Once again, the point is that, even if s is not k-stable for X because $a < \frac{1}{4}(5+3b_2^+(X))$ one can certainly choose n sufficiently large that $a+4n \geq \frac{1}{4}(5+3b_2^+(X)) = \frac{1}{4}(5+3b_2^+(X\# n\overline{\mathbb{CP}}^2))$ and so $(s, e_1, e_1, e_1, e_1, \ldots, e_n, e_n, e_n, e_n)$ is $(k+n)$-stable for $X\# n\overline{\mathbb{CP}}^2$. Checking that the value of $\gamma_{d_{k+n}}(X\# n\overline{\mathbb{CP}}^2)(s, e_1, e_1, e_1, e_1, \ldots, e_n, e_n, e_n, e_n)$ does not depend on the choice of such a sufficiently large n one can then *define* $\gamma_{d_k}(X)(s)$ by the right-hand side of (15). Notice that, with this definition, the Donaldson invariants $\gamma_{d_k}(X)$ take their values in $\mathbb{Z}[\frac{1}{2}]$.

We have now defined $\gamma_d(X)$ for any d satisfying $d \equiv -\frac{3}{2}(1 + b_2^+(X))\, mod\, 4$. There is one final trick for extending the range of d-values for which we can define $\gamma_d(X)$. Suppose d satisfies

$$d \equiv -\frac{3}{2}(1 + b_2^+(X))\, mod\, 2, \tag{16}$$

but

$$d \not\equiv -\frac{3}{2}(1 + b_2^+(X))\, mod\, 4. \tag{17}$$

Then

$$d = 2k - \frac{3}{2}(1 + b_2^+(X))$$

for some *odd* integer k. Then $k + 1$ is even and we can write

$$d + 2 = 4\left(\frac{k+1}{2}\right) - \frac{3}{2}(1 + b_2^+(X)) = d_{(k+1)/2} \equiv -\frac{3}{2}(1 + b_2^+(X))\, mod\, 4.$$

Consequently, γ_{d+2} is defined. Since $d_{(k+1)/2} = a+2b$, where $a = d$ and $b = 1$, it is defined on $(\alpha_1, \ldots, \alpha_d, x)$, where $\alpha_1, \ldots, \alpha_d \in H_2(X; \mathbb{Z})$ and x is the generator of $H_0(X; \mathbb{Z})$. When d satisfies (16) and (17) we define

$$\gamma_d(X)(\alpha_1, \ldots, \alpha_d) = \frac{1}{2}\gamma_{d+2}(X)(\alpha_1, \ldots, \alpha_d, x).$$

Remark 1.11. We would like to make an observation that will be of use to us in the next section. The dimension of the moduli space $\mathcal{M}(P_k, g)$ is given by $\dim \mathcal{M}(P_k, g) = 8k - 3(1 + b_2^+(X))$. By choosing $b_2^+(X)$ and k appropriately one can arrange that $\dim \mathcal{M}(P_k, g) = 0$ so that $\mathcal{M}(P_k, g)$ is a set of isolated points. This cannot occur for k in the stable range since $k \geq \frac{1}{4}(5 + 3b_2^+(X))$ implies $\dim \mathcal{M}(P_k, g) \geq 7 + 3b_2^+(X) > 0$. On the other hand, if we take X to be a manifold with $b_2^+(X) = 7$ and then consider the SU(2)-bundle over X with Chern number $k = 3$ we do, indeed, have $\dim \mathcal{M}(P_3, g) = 0$. Such 4-manifolds are easy enough to find since $b_2^+(p\,\mathbb{CP}^2 \# q\,\overline{\mathbb{CP}}^2) = p - q$ so, in particular, $b_2^+(7\,\mathbb{CP}^2) = 7$. In this case it turns out that all of the Donaldson invariants of $7\,\mathbb{CP}^2$ are identically zero (see Proposition 9.3.1 of [DK]). More interesting examples of 0-dimensional moduli spaces (for SO(3)-bundles) can be found in Section 9.1 of [DK].

This completes our list of the Donaldson invariants. There are infinitely many such invariants and, as one might expect, computing any one of them for a given X is generally a Herculean task (see [O'G], for example). As it happens, however, many of the most significant applications of Donaldson theory arise from general results on the vanishing or non-vanishing of the invariants. These results are themselves deep and technically very demanding, but have the advantage of being applicable to a wide range of problems. We will illustrate this with just one example. The first result we need is called the *Connected Sum Theorem* (see Theorem 8.1.1 of [Mor2] and Theorem 9.3.4 of [DK] for sketches and Theorem 4.9 of [Don3] for a complete proof).

Theorem 1.16. *(Connected Sum Theorem) Let X be a compact, connected, simply connected, oriented, smooth 4-manifold with $b_2^+(X)$ odd and suppose that X is orientation preserving diffeomorphic to $X_1 \# X_2$ where $b_2^+(X_1) > 0$ and $b_2^+(X_2) > 0$. Then each $\gamma_d(X)$ is identically zero.*

Next we recall that a *complex surface* is a complex manifold of complex dimension two and therefore real dimension four. These all have natural orientations as complex manifolds. A *complex algebraic surface* is a complex surface that is defined by homogeneous polynomial equations in some complex projective space $\mathbb{CP}^n$. The $K3$ surface is an example, but there are a great many others (see [Beau]). The following is a *Non-Vanishing Theorem* for complex algebraic surfaces (see Theorem 8.1.2 of [Mor2] for a sketch, Chapter 10 of [DK] for a more detailed discussion and [Don3] for a complete proof).

Theorem 1.17. *(Non-Vanishing Theorem) Let X be a compact, connected, simply connected, complex algebraic surface with $b_2^+(X)$ odd and greater than 1. Then, for all sufficiently large d, $\gamma_d(X)$ is not identically zero.*

As an application we will consider the $K3$ surface. We know that $b_2^+(K3) = 3$ and that the intersection form of $K3$ is $2(-E_8) \oplus 3\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Note that $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is the intersection form of $\mathbb{S}^2 \times \mathbb{S}^2$ and $b_2^+(\mathbb{S}^2 \times \mathbb{S}^2) = 1$. Consequently, the intersection form of $3(\mathbb{S}^2 \times \mathbb{S}^2)$ is $3\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $b_2^+(\mathbb{S}^2 \times \mathbb{S}^2) = 3$. According to Freedman's Theorem 1.5, $K3$ is homeomorphic to $2X_{-E_8} \# 3(\mathbb{S}^2 \times \mathbb{S}^2)$. Topologically, $K3$ decomposes into a connected sum, one summand of which is $3(\mathbb{S}^2 \times \mathbb{S}^2)$. We will show that $K3$ *cannot* be written as a smooth connected sum, one summand of which is $3(\mathbb{S}^2 \times \mathbb{S}^2)$, unless the other summand is has trivial second homology (which $2X_{-E_8}$ clearly does not). In fact, we will deduce from Theorems 1.3, 1.16 and 1.17 that this, and much more, is true not only for $K3$, but for any compact, connected, simply connected, complex algebraic surface with even intersection form and $b_2^+(X)$ odd and greater than 1. Results such as these were quite inaccessible before the advent of Donaldson theory.

Corollary 1.18. *Let X be a compact, connected, simply connected, complex algebraic surface with even intersection form and $b_2^+(X)$ odd and greater than 1. Suppose that X is orientation preserving diffeomorphic to $X_1 \# X_2$, where $b_2^+(X_1) > 0$ and $b_2^+(X_2) > 0$. Then either X_1 or X_2 has trivial second homology.*

Proof. Since X is compact, connected, and simply connected, so are X_1 and X_2. By Theorem 1.17, the Donaldson invariants $\gamma_d(X)$ are not identically zero for sufficiently large d. But then, by Theorem 1.16, either $b_2^+(X_1) = 0$ or $b_2^+(X_2) = 0$. Without loss of generality we can assume that $b_2^+(X_1) = 0$. Thus, either $b_2(X) = 0$ or X_1 has negative definite intersection form. Suppose $b_2(X) \neq 0$. Then, by Donaldson's Theorem 1.3, the intersection form of X_1 is equivalent to the diagonal form with all diagonal entries -1. But, since X has even intersection form, so does X_1 and there can be no $\alpha \in H_2(X_1; \mathbb{Z})$ with $Q_{X_1}(\alpha, \alpha) = -1 \equiv 1 \, mod \, 2$ and this is a contradiction. Thus, $b_2(X) = 0$ and $H_2(X; \mathbb{Z})$ is trivial. □

Between 1983 and 1994 the differential topology of 4-manifolds was dominated by the ideas of Donaldson and, after 1990, this amounted to the calculation of Donaldson invariants. Much progress was made, but the computations were enormously complicated, there were infinitely many invariants to be computed, and there were no known relations among the invariants to lighten the load. This all changed in 1994, but in two quite different ways. The first event was a remarkable breakthrough in the computation of Donaldson invariants by Kronheimer and Mrowka [KM1]. Their procedure was to consolidate all of the $\gamma_d(X)$ into a single formal power series, called the *Donaldson series* $\mathcal{D}_X(\alpha)$, on the second homology. This is to be thought of as the generating function for the Donaldson invariants. Next they introduced a condition called *KM-simple type*, satisfied by all known examples of the type of 4-manifold we are considering here, which guaranteed that the invariants themselves could be retrieved from a knowledge of the Donaldson series. All of this would be rather uninteresting, of course, if it were not possible to obtain information about $\mathcal{D}_X(\alpha)$ independently of actually calculating the Donaldson invariants. The *Structure Theorem* of Kronheimer and Mrowka (proved in [KM1]) asserts that, in fact, the entire Donaldson series is completely determined by a finite set of data. The following statement of the Structure Theorem is taken from [FS2] which contains a simplification of the proof in [KM1].

Theorem 1.19. *(Structure Theorem) Let X be a compact, connected, simply connected, oriented, smooth 4-manifold with $b_2^+(X) > 1$ and odd. Suppose that X is of simple type and some Donaldson invariant of X is not identically zero. Then there exist finitely many cohomology classes $K_1, \ldots, K_s \in H_2(X;\mathbb{Z})$, called* KM-basic classes*, and nonzero rational numbers $a_1, \ldots, a_s$, called* KM-coefficients*, such that*

$$\mathcal{D}_X(\alpha) = e^{Q_X(\alpha,\alpha)/2} \sum_{i=1}^{s} a_i e^{K_i(\alpha)}.$$

With the appearance of this result in early 1994 Donaldson theory seemed to have turned a corner. Long and complex calculations of an apparently infinite set of independent invariants were suddenly replaced by the (certainly nontrivial, but at least finite) problem of determining the basic classes and coefficients. Initially, this resulted in very considerable progress (see [Stern] for the state of the art as of July, 1994). As fate would have it, however, the fall of 1994 witnessed another event which effectively rendered this triumph of Kronheimer and Mrowka moot. Edward Witten, at the end of a lecture at M.I.T. on $N = 2$ supersymmetric Yang-Mills theory, made a conjecture which, within weeks, brought about the demise of Donaldson theory and initiated an entirely new approach to the study of smooth 4-manifolds. The story of this event and the ensuing frenzy is told by one who was there in [Taub3]. In Section 1.5 we will take just one, ever so modest, step toward understanding how one might go about finding the Donaldson invariants in quantum field theory and then, in Section 1.6, we will have a look at the revolution to which this led.

1.5. **The Mathai-Quillen Formalism and Witten's TQFT Partition Function.** In 1988, Edward Witten [Witt2], prompted by Atiyah, constructed a quantum field theory in which the Donaldson invariants appeared as expectation values of certain observables and thus ushered in the study of what has come to be known as *topological quantum field theory (TQFT)*. This construction of Witten's was a remarkable achievement and provided the most direct sort of link between topology and physics. Topologists were (and still are) eager to understand and exploit the insights that gave rise to such a radically new view of smooth 4-manifold invariants. These insights, however, sprang from the deepest regions of theoretical physics and were largely inaccessible to mathematicians. In this section we will not presume to make Witten's ideas palatable to those trained in modern mathematics, but will only provide a brief hint as to how the simplest of the Donaldson invariants might fit into the path integral framework of quantum field theory. That this is possible at all is, in large measure, the result of a beautiful paper of Atiyah and Jeffrey [AJ] that we will discuss in due course.

1.5.1. *The Euler Number.* We begin with a few general remarks. We consider an oriented, real vector bundle $\xi = (\pi_E : E \to X)$ of even rank (fiber dimension) $n = 2k$ over a compact, oriented, smooth manifold X of dimension n. Then E is an oriented manifold of dimension $2n$. Examples include the tangent and cotangent bundles of an oriented, even dimensional manifold. We will denote by V the typical fiber of ξ and will assume that V comes equipped with a positive definite inner product coming from a fiber metric on ξ. The exterior algebra of V will be denoted

$$\wedge V = \oplus_p \wedge^p V = (\oplus_p \wedge^{2p} V) \oplus (\oplus_p \wedge^{2p+1} V) = V_{\mathbf{0}} \oplus V_{\mathbf{1}}$$

where $\mathbb{Z}_2 = \{\mathbf{0}, \mathbf{1}\}$. Select some oriented, orthonormal basis $\{\psi^1, \ldots, \psi^n\}$ for V. These basis elements are then odd generators for $\wedge V$ and we think of the elements of $\wedge V$ as polynomials with real coefficients in the anti-commuting variables $\{\psi^1, \ldots, \psi^n\}$. The corresponding volume form for V is defined by $\mathrm{vol} = \psi^1 \cdots \psi^n$ (we suppress the customary wedge $\wedge$ and write the product in $\wedge V$ by juxtaposition). Notice that one can define the exponential map on $\wedge V$ by the usual power series, noting that the series necessarily terminates for any element of $\wedge V$ due to the anti-commutativity of exterior multiplication.

Remark 1.12. Soon it will be useful to adopt some terminology that arose in supersymmetric physics so we will pause to formulate a few definitions. A *superalgebra* is a real or complex vector space A with a $\mathbb{Z}_2$-grading $A = A_{\mathbf{0}} \oplus A_{\mathbf{1}}$ and an associative, bilinear multiplication $A \times A \to A$ that satisfies $A_{\mathbf{i}} A_{\mathbf{j}} \subseteq A_{\mathbf{i}+\mathbf{j}}$ and which contains an element $1 \in A$ that satisfies $1a = a1 = a$ for all $a \in A$. The elements of $A_{\mathbf{0}}$ are said to be *even* and have *degree* $\mathbf{0} \in \mathbb{Z}_2$, while the elements of $A_{\mathbf{1}}$ are *odd* and have *degree* $\mathbf{1} \in \mathbb{Z}_2$. A is said to be *supercommutative* if, for all $a_1, a_2 \in A$, $a_1 a_2 = (-1)^{\mathbf{d}_1 \mathbf{d}_2} a_2 a_1$, where $\mathbf{d}_1 = \deg a_1$ and $\mathbf{d}_2 = \deg a_2$ (moving an odd thing past an odd thing costs a minus sign). In particular, $1 \in A_{\mathbf{0}}$. Thus, $\wedge V$, with exterior multiplication, is a supercommutative superalgebra. Other examples include the algebra $\Omega^*(V) = \wedge V^*$ of forms on V or, more generally, the algebra of forms $\Omega^*(X)$ on a smooth manifold X. Another class of examples that will play an important role soon is constructed in the following way. Let G be some compact, connected, matrix Lie group with Lie algebra $\mathfrak{g}$ (for example, $\mathrm{SO}(V)$ and $\mathfrak{so}(V)$). The vector space dual of $\mathfrak{g}$ will be written $\mathfrak{g}^*$ and $\mathrm{Sym}(\mathfrak{g}^*)$ will denote the symmetric algebra of $\mathfrak{g}^*$. We will identify $\mathrm{Sym}(\mathfrak{g}^*)$ with the algebra of polynomial functions with real coefficients on $\mathfrak{g}$. More precisely, we let $\{\zeta_1, \ldots, \zeta_n\}$ denote a basis for $\mathfrak{g}$ and $\{x^1, \ldots, x^n\}$ the corresponding dual basis for $\mathfrak{g}^*$. Each x^j is then a real-valued linear function on $\mathfrak{g}$ and $\mathrm{Sym}(\mathfrak{g}^*)$ is isomorphic to the polynomial algebra $\mathbb{R}[x^1, \ldots, x^n]$ generated by $x^1, \ldots, x^n$. Then

$$\mathbb{C}[\mathfrak{g}] = \mathrm{Sym}(\mathfrak{g}^*) \otimes \mathbb{C}$$

is identified with the algebra of polynomial functions on $\mathfrak{g}$ with complex coefficients. Each element $\mathcal{P}$ of $\mathbb{C}[\mathfrak{g}]$ has an algebraic degree $\deg \mathcal{P}$ as a polynomial, but, since $\mathbb{C}[\mathfrak{g}]$ is commutative , it will be more convenient to regard $\mathbb{C}[\mathfrak{g}]$ as a supercommutative superalgebra in which all of the elements are even by *doubling the degrees*. Thus,

$$\mathbb{C}[\mathfrak{g}] = \mathbb{C}[\mathfrak{g}]_{\mathbf{0}} \oplus \mathbb{C}[\mathfrak{g}]_{\mathbf{1}},$$

where $\mathbb{C}[\mathfrak{g}]_{\mathbf{0}} = \mathbb{C}[\mathfrak{g}]$ and $\mathbb{C}[\mathfrak{g}]_{\mathbf{1}}$ is the trivial subspace of $\mathbb{C}[\mathfrak{g}]$.

If A and B are two supercommutative superalgebras, then their *supertensor product*, denoted $A\hat{\otimes}B$ has as its underlying vector space the ordinary tensor product $A \otimes B$ with the $\mathbb{Z}_2$-grading

$$(A \otimes B)_{\mathbf{0}} = A_{\mathbf{0}} \otimes B_{\mathbf{0}} \oplus A_{\mathbf{1}} \otimes B_{\mathbf{1}}$$
$$(A \otimes B)_{\mathbf{1}} = A_{\mathbf{0}} \otimes B_{\mathbf{1}} \oplus A_{\mathbf{1}} \otimes B_{\mathbf{0}}$$

and with multiplication defined, not in the usual way ($(a_1 \otimes b_1)(a_2 \otimes b_2) = (a_1 b_1) \otimes (a_2 b_2)$) for homogeneous elements), but rather by

$$(a_1 \otimes b_1)(a_2 \otimes b_2) = (-1)^{\deg b_1 \deg a_2}(a_1 b_1) \otimes (a_2 b_2).$$

With this structure the maps $a \mapsto a\otimes 1$ and $b \mapsto 1\otimes b$ are algebra monomorphisms preserving the $\mathbb{Z}_2$-grading and such that

$$(a \otimes 1)(1 \otimes b) = a \otimes b, \quad \text{and} \quad (1 \otimes b)(a \otimes 1) = (-1)^{\deg b \deg a} a \otimes b.$$

Using these embeddings to identify A and B with subalgebras of $A\hat{\otimes}B$ and writing ab rather than $a \otimes b$ this gives

$$ba = (-1)^{\deg b \deg a} ab.$$

Moreover, if A and B are supercommutative, then so is $A\hat{\otimes}B$. An example with which we will deal shortly is

$$\mathbb{C}[\mathfrak{g}] \hat{\otimes} \Omega^*(X).$$

Notice that, since everything in $\mathbb{C}[\mathfrak{g}]$ is even, this coincides with the usual tensor product $\mathbb{C}[\mathfrak{g}] \otimes \Omega^*(X)$ except that it is $\mathbb{Z}_2$-graded by

$$(\,\mathbb{C}[\mathfrak{g}] \hat{\otimes} \Omega^*(X)\,)_{\mathbf{0}} = \mathbb{C}[\mathfrak{g}] \otimes \Omega(X)_{\mathbf{0}}$$
$$(\,\mathbb{C}[\mathfrak{g}] \hat{\otimes} \Omega^*(X)\,)_{\mathbf{1}} = \mathbb{C}[\mathfrak{g}] \otimes \Omega(X)_{\mathbf{1}}$$

The elements of $\mathbb{C}[\mathfrak{g}] \hat{\otimes} \Omega^*(X)$ are sums of *homogeneous terms* of the form $\mathcal{P} \otimes \varphi$, where $\mathcal{P} \in \mathbb{C}[\mathfrak{g}]$ and $\varphi \in \Omega^*(X)$ and should be thought of as $\Omega^*(X)$-valued polynomial functions on $\mathfrak{g}$. For example, $(\mathcal{P} \otimes \varphi)(\zeta) = \mathcal{P}(\zeta)\varphi$ for every $\zeta \in \mathfrak{g}$.

Now we return to the general development. The vector bundle ξ has associated with it an *Euler number*

$$\chi(\xi)$$

that can be defined in a variety of ways, of which we will begin with two.

(1) The Euler number $\chi(\xi)$ of ξ is the intersection number of any generic section of ξ with the zero-section of ξ. We should say what all of this means (see Theorems 21.9 and 21.11 of [MT]). Let $s_0 : X \to E$ be the zero section of ξ ($s_0(x) = 0_x$, where 0_x is the zero element of the fiber $\pi_E^{-1}(x)$). Let $s : X \to E$ be another section of ξ. Both s_0 and s are diffeomorphisms of X onto closed submanifolds $s_0(X)$ and $s(X)$ of E. Now suppose that these submanifolds intersect, that is, $s(x) = s_0(x) = 0_x$ for some $x \in X$. Then s is said to be *transverse* to s_0 at x if

$$(s_0)_{*x}(T_x(X)) \oplus s_{*x}(T_x(X)) = T_{s(x)}(E), \tag{18}$$

where $(s_0)_{*x}$ and s_{*x} are the derivatives of s_0 and s at x, respectively. By the Transversality Theorem (see Chapter 3, Section 2, of [Hirsch]), there is a dense set of sections s with the property that they are transverse to s_0 at *every* $x \in X$ for which $s(x) = 0_x$. These are called *generic sections*. For any generic section s the submanifolds $s_0(X)$ and $s(X)$ of E intersect transversally so $s_0(X) \cap s(X)$ is a submanifold. Since dim $s_0(X) =$ dim $s(X) = n$ and dim $E = 2n$, the intersection has dimension zero and is therefore a set of isolated points. Since $s_0(X) \cap s(X)$ is compact, it is a finite set of isolated points. Next notice that the diffeomorphisms s_0 and s and the given orientation of X provide $s_0(X)$ and $s(X)$ with orientations. Because of (18) we can then assign an index $\operatorname{Ind}(s; p)$ to every $p = s(x) \in s_0(X) \cap s(X)$ as follows. We take $\operatorname{Ind}(s; p) = 1$ if an oriented basis for $T_p(s_0(X)) = (s_0)_{*x}(T_x(X))$ together with an oriented basis for $T_p(s(X)) = s_{*x}(T_x(X))$ gives an oriented basis for $T_p(E)$; otherwise, $\operatorname{Ind}(s; p) = -1$. The sum of these indices over all $p \in s_0(X) \cap s(X)$ is the intersection number of s with s_0 when $s_0(X) \cap s(X) \neq \emptyset$ and the sum is independent of the choice of generic section s. In this case the Euler number of ξ is given by

$$\chi(\xi) = \sum_{p \in s_0(X) \cap s(X)} \operatorname{Ind}(s; p). \tag{19}$$

If $s_0(X) \cap s(X) = \emptyset$, then the intersection number of s with s_0 is taken to be 0, as is $\chi(\xi)$.

(2) The Euler number of ξ is

$$\chi(\xi) = \int_X e(\xi), \tag{20}$$

where $e(\xi) \in H^n(X; \mathbb{R})$ is the Euler class of ξ. Again, we should elaborate just a bit. The Euler class $e(\xi)$ can be defined in a number of ways. We will first describe how it can be viewed as a Chern-Weil characteristic class (see Chapter XII, Section 5, of [KN2]). The Chern-Weil construction of characteristic classes requires two ingredients. One needs a principal bundle $G \hookrightarrow P \stackrel{\pi_P}{\to} X$ and a polynomial $\mathcal{P}$ on the Lie algebra $\mathfrak{g}$ that is $\operatorname{ad} G$-invariant ($\mathcal{P}(gAg^{-1}) = \mathcal{P}(A)$ for all $A \in \mathfrak{g}$ and all $g \in G$). One chooses a connection ω on $G \hookrightarrow P \stackrel{\pi_P}{\to} X$, computes the curvature Ω and evaluates the polynomial on the curvature to obtain a differential form $\mathcal{P}(\Omega)$ on P. This form is then the pullback by π_P of a form $\overline{\mathcal{P}}(\Omega)$ on X which can be shown to be closed and whose cohomology class $[\overline{\mathcal{P}}(\Omega)]$ does not depend on the choice of ω. Making various choices for $\mathcal{P}$ gives rise to various characteristic classes for $G \hookrightarrow P \stackrel{\pi_P}{\to} X$. One obtains characteristic classes for vector bundles by regarding them as associated to some principal bundle $G \hookrightarrow P \stackrel{\pi_P}{\to} X$ by a representation of G.

Now, to define the Euler class of the oriented, real vector bundle $\xi = (\pi_E : E \to X)$ one chooses a fiber metric on ξ and considers the corresponding oriented, orthonormal frame bundle

$$SO(2k) \hookrightarrow F_{SO}(\xi) \stackrel{\pi_{SO}}{\to} X. \tag{21}$$

Then ξ is the vector bundle associated to the principal bundle (21) by the defining action of SO(2k) on $V \cong \mathbb{R}^{2k}$. For the ad SO(2k)-invariant polynomial one chooses the Pfaffian Pf : $\mathfrak{so}(2k) \to \mathbb{R}$ defined as follows. For each skew-symmetric matrix $A = (A_{ij})_{i,j=1,\ldots,2k} \in \mathfrak{so}(2k)$ we associate an element of $\wedge^2 V \subseteq \wedge V$ given by

$$\sum_{i<j} A_{ij}\psi^i\psi^j = \frac{1}{2}\psi^T A\psi,$$

where $\psi = \begin{pmatrix} \psi^1 \\ \vdots \\ \psi^{2k} \end{pmatrix}$ and "T" means transpose.. Then $(\frac{1}{2}\psi^T A\psi)^k$ is in $\wedge^{2k}V$, which is 1-dimensional, so it is a multiple of vol $= \psi^1 \cdots \psi^{2k}$. We define Pf$(A)$ by

$$\frac{1}{k!}\,(\frac{1}{2}\psi^T A\psi)^k = \mathrm{Pf}\,(A)\,\mathrm{vol}. \tag{22}$$

By expanding $(\frac{1}{2}\psi^T A\psi)^k$ one finds that

$$\mathrm{Pf}\,(A) = \frac{1}{2^k k!}\sum_{\sigma\in S_{2k}} \mathrm{sgn}(\sigma)\,A_{\sigma(1)\sigma(2)}\cdots A_{\sigma(2k-1)\sigma(2k)},$$

where S_{2k} is the group of permutations of $\{1, 2, \ldots, 2k-1, 2k\}$ and sgn(σ) is the sign of the permutation σ. Then Pf is, indeed, ad SO(2k)-invariant and satisfies $(\mathrm{Pf}\,(A))^2 = \det A$ (see Appendix B of [MT]). Now we choose a connection ω on the SO(2k)-bundle (21) with curvature Ω. Then the 2k-form on $F_{SO}(\xi)$ given by

$$\mathrm{Pf}\,(-\frac{1}{2\pi}\Omega) = \frac{(-1)^k}{2^k\pi^k}\mathrm{Pf}\,(\Omega) = \frac{(-1)^k}{2^{2k}\pi^k k!}\sum_{\sigma\in S_{2k}} \mathrm{sgn}(\sigma)\,\Omega_{\sigma(1)\sigma(2)}\cdots\Omega_{\sigma(2k-1)\sigma(2k)}$$

descends to 2k-form $\overline{\mathrm{Pf}}\,(-\frac{1}{2\pi}\Omega)$ on X whose cohomology class is the Euler class of ξ.

$$e(\xi) = [\,\overline{\mathrm{Pf}}\,(-\frac{1}{2\pi}\Omega)\,]. \tag{23}$$

Remark 1.13. It is not obvious, but is nevertheless true, that $e(\xi)$ does not depend on the initial choice of a fiber metric for ξ (see Theorem 5.1, page 318, of [KN2]). We should point out also that, if s is any local section of the principal bundle (21), then the form $\overline{\mathrm{Pf}}\,(-\frac{1}{2\pi}\Omega)$ can be described locally by

$$\frac{(-1)^k}{2^k\pi^k}\,\mathrm{Pf}\,(s^*\Omega) = (2\pi)^{-k}\mathrm{Pf}\,(-s^*\Omega). \tag{24}$$

Furthermore, under a change of local section the forms $s^*\Omega$ transform by conjugation and the Pfaffian is invariant under conjugation so (24) is also invariant and these local descriptions piece together into global description of $\overline{\mathrm{Pf}}(-\frac{1}{2\pi}\Omega)$ and so to a representative of the Euler class. It is customary to notationally suppress the "piecing together" aspect of this and refer to (24) as a representative of $e(\xi)$.

Example 1.4. Consider the 2-sphere $\mathbb{S}^2$ with its usual orientation and Riemannian metric. Let $\pi : T\mathbb{S}^2 \to \mathbb{S}^2$ be the tangent bundle of $\mathbb{S}^2$. The corresponding oriented, orthonormal frame bundle is $\mathrm{SO}(2) \hookrightarrow F_{\mathrm{SO}}(T\mathbb{S}^2) \overset{\pi_{\mathrm{SO}}}{\to} \mathbb{S}^2$. If θ and ϕ are the usual angular coordinates on $\mathbb{S}^2$, then $\{e_1, e_2\} = \{\frac{\partial}{\partial\phi}, \frac{1}{\sin\phi}\frac{\partial}{\partial\theta}\}$ is an oriented, orthonormal frame field on $\mathbb{S}^2$, that is, a section s of $F_{\mathrm{SO}}(T\mathbb{S}^2)$. The dual oriented, orthonormal field of 1-forms is $\{e^1, e^2\} = \{d\phi, \sin\phi\, d\theta\}$. Thus, the metric volume form is $e^1 \wedge e^2 = \sin\phi\, d\phi\, d\theta$. One computes $de^1 = 0 = 0\,(e^1 \wedge e^2)$ and $de^2 = \cos\phi\, d\phi\, d\theta = \cot\phi\,(e^1 \wedge e^2)$. Thus, the Levi-Civita connection ω on the frame bundle is determined by

$$s^*\omega = \begin{pmatrix} 0 & \cos\phi\, d\theta \\ -\cos\phi\, d\theta & 0 \end{pmatrix}.$$

Since SO(2) is Abelian, the curvature of ω is $\Omega = d\omega$ and so

$$s^*\Omega = \begin{pmatrix} 0 & -\sin\phi\, d\phi\, d\theta \\ \sin\phi\, d\phi\, d\theta & 0 \end{pmatrix}.$$

A representative of the Euler class is therefore

$$(2\pi)^{-1}\mathrm{Pf}\,(-s^*\Omega) = \frac{1}{2\pi}\sin\phi\, d\phi\, d\theta.$$

Consequently, the Euler number of the tangent bundle of $\mathbb{S}^2$ is

$$\chi(T\mathbb{S}^2) = \frac{1}{2\pi}\int_{\mathbb{S}^2} \sin\phi\, d\phi\, d\theta = 2.$$

Notice that this coincides with the Euler characteristic $\chi(\mathbb{S}^2)$ of $\mathbb{S}^2$.

The vector bundle $\xi = (\pi_E : E \to X)$ gives rise to another approach to the Euler class that will soon be important to us. We will say that a smooth p-form α on E has *compact vertical support* if, for every $x \in X$, the restriction $\alpha|_{\pi_E^{-1}(x)}$ of α to the fiber $\pi_E^{-1}(x)$ has compact support. The space of all such is denoted $\Omega^p_{CV}(E;\mathbb{R})$. Then $d : \Omega^p_{CV}(E;\mathbb{R}) \to \Omega^{p+1}_{CV}(E;\mathbb{R})$ and the resulting co-complex is denoted $\Omega^*_{CV}(E;\mathbb{R})$. The cohomology of this co-complex is denoted $H^*_{CV}(E;\mathbb{R})$ and called the *compact vertical cohomology* of E. According to the Thom Isomorphism Theorem (see Theorem 6.17 of [BT]) there is an isomorphism $\mathcal{T} : H^*(X;\mathbb{R}) \to H^{*+2k}_{CV}(E;\mathbb{R})$ and this carries the generator of $H^0(X;\mathbb{R})$ onto a cohomology class $U(E) \in H^{2k}_{CV}(E;\mathbb{R})$ whose integral over each fiber is 1 (see Proposition 6.18 of [BT]). This is called the *Thom class* and it has the property that, if s is any section of ξ, then

$$e(\xi) = s^*U(\xi) \tag{25}$$

(see Proposition 6.41 of [BT]).

Before proceeding any further with the technical issues we will try to see where this is going. Consider the special case in which X is a compact, oriented manifold and ξ is its tangent bundle $\pi_{TX} : TX \to X$. In this case the Euler number $\chi(\xi)$ of ξ is just the Euler characteristic $\chi(X)$ of X, usually defined as the alternating sum of the dimensions of the cohomology groups $H^p(X;\mathbb{R})$ (see Proposition 11.24 of [BT]). A section of ξ is then just a vector field on X and the statement that $\chi(X)$ is the intersection number of a generic vector field with the vector field that is identically zero is called the *Poincaré-Hopf Theorem*. On the other hand, the *Chern-Gauss-Bonnet Theorem* is the statement that $\chi(X)$ can be expressed as the integral over X of the

Euler class of the tangent bundle. The point is that $\chi(X)$ is a *topological invariant* and can be calculated either as *a sum of signed points* or as *the integral of a cohomology class*.

To see what this has to do with Donaldson invariants we would like to adopt a somewhat different perspective on the simplest of these invariants (see Remark 1.11). For this and for all of the subsequent discussions we will adopt a somewhat more economical notation, writing P for P_k, $\hat{\mathcal{M}}$ for $\hat{\mathcal{M}}(P_k, g)$, $\mathcal{G}$ for $\mathcal{G}(P_k)$, and so on. The gauge group $\mathcal{G}$ does not act freely on the space $\hat{\mathcal{A}}$ if irreducible connections since even irreducible connections have a $\mathbb{Z}_2$ stabilizer. However, $\hat{\mathcal{G}} = \mathcal{G}/\mathbb{Z}_2$ does act freely on $\hat{\mathcal{A}}$ so we have an infinite-dimensional principal bundle

$$\hat{\mathcal{G}} \hookrightarrow \hat{\mathcal{A}} \to \hat{\mathcal{A}}/\hat{\mathcal{G}} = \hat{\mathcal{B}}$$

over the Banach manifold $\hat{\mathcal{B}}$ (note that the orbits in $\hat{\mathcal{A}}$ of the $\hat{\mathcal{G}}$-action are the same as those of the $\mathcal{G}$-action so the quotient is still $\hat{\mathcal{B}}$). We will build a vector bundle associated to this principal bundle as follows. Consider the (infinite-dimensional) vector space $\Omega^2_+(X, \operatorname{ad} P)$ of self-dual 2-forms on X with values in the adjoint bundle ad P. We claim that there is a smooth left action of $\hat{\mathcal{G}}$ on $\Omega^2_+(X, \operatorname{ad} P)$. To see this we first think of $\mathcal{G}$ as the group of sections of the nonlinear adjoint bundle Ad P under pointwise multiplication. Since the elements of $\Omega^2_+(X, \operatorname{ad} P)$ take values in the $\mathfrak{sp}(1)$-fibers of ad P and the elements of $\mathcal{G}$ take values in the Sp(1)-fibers of Ad P, $\mathcal{G}$ acts on $\Omega^2_+(X, \operatorname{ad} P)$ by pointwise conjugation. Moreover, conjugation takes the same value at $\pm f \in \mathcal{G}$ so this $\mathcal{G}$-action on $\Omega^2_+(X, \operatorname{ad} P)$ descends to a $\hat{\mathcal{G}}$-action. Thus, we have an associated vector bundle

$$\hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega^2_+(X, \operatorname{ad} P), \tag{26}$$

the elements of which are equivalence classes $[\omega, \alpha] = [\omega \circ f, f^{-1} \circ \alpha]$ with $\omega \in \hat{\mathcal{A}}$, $\alpha \in \Omega^2_+(X, \operatorname{ad} P)$, and $f \in \hat{\mathcal{G}}$. Now recall that sections of associated vector bundles can be identified with equivariant maps from the principal bundle space into the typical fiber of the vector bundle. In our case we have an obvious map

$$F^+ : \hat{\mathcal{A}} \to \Omega^2_+(X, \operatorname{ad} P)$$

of $\hat{\mathcal{A}}$ into $\Omega^2_+(X, \operatorname{ad} P)$ that sends a connection to the self-dual part of its curvature.

$$F^+(\omega) = F^+_\omega = \frac{1}{2}(F_\omega + {}^*F_\omega)$$

Since the action of $\hat{\mathcal{G}}$ on $\hat{\mathcal{A}}$ is by conjugation and curvatures transform by conjugation under a gauge transformation, F^+ is equivariant so F^+ can be identified with a section

$$s_+ : \hat{\mathcal{B}} \to \hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega^2_+(X, \operatorname{ad} P),$$

given explicitly by

$$s_+([\omega]) = [\omega, F^+_\omega].$$

Now notice that the moduli space $\hat{\mathcal{M}}$ of irreducible anti-self-dual connections is precisely the *zero set* of the section s_+. Thus, if

$$s_0 : \hat{\mathcal{B}} \to \hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega^2_+(X, \operatorname{ad} P),$$

is the zero section, $\hat{\mathcal{M}}$ can be identified with

$$\hat{\mathcal{M}} = s_0(\hat{\mathcal{B}}) \cap s_+(\hat{\mathcal{B}}).$$

Now suppose that $\hat{\mathcal{M}}$ is 0-dimensional and compact, that is, a finite set of points each equipped with a sign (for concrete examples, see Section 9.1 of [DK]). Then the corresponding Donaldson invariant is the sum of these signs and one sees quite clearly the sense in which this invariant can be regarded as an infinite-dimensional analogue of the Poincaré-Hopf version of the Euler characteristic. Taking this analogy seriously would suggest the possibility of an integral representation of the 0-dimensional Donaldson invariant analogous to the Gauss-Bonnet-Chern Theorem. Notice, however, that such an "integral" would necessarily be over the infinite-dimensional moduli space $\hat{\mathcal{B}}$ and such integrals are notoriously difficult to define rigorously. But, as Nigel Hitchin [Hitch] has phrased it,

"This is such stuff as quantum field theory is made of."

Indeed, it was Edward Witten who first produced such an integral representation, not directly, but as the partition function for the quantum field theory introduced in [Witt2]. The integrals are purely formal and the arguments leading to the partition function and its identification with the Donaldson invariant are physical and require a deep understanding of the methods of (supersymmetric) quantum field theory. A still deeper understanding of physics is required to see how this quantum field-theoretic interpretation of the Donaldson invariants eventually pointed toward technically much simpler invariants that were destined to supplant Donaldson theory. We will not presume to describe the physics behind all of this, but will follow instead a path forged by Atiyah and Jeffrey [AJ]. The arguments are still formal in that they apply rigorous finite-dimensional theorems due to Mathai and Quillen [MQ] in an infinite-dimensional context, but they are at least geometrical. These finite-dimensional theorems are quite nontrivial and live most naturally in the context of equivariant cohomology, while the computation required to implement the beautiful idea in [AJ] is relatively straightforward, but lengthy. In both cases we merely sketch what is involved and refer to Section 5 of [Nab2] for a more detailed discussion; additional sources of background information include [BGV], [Blau], [BT], and [GuSt].

1.5.2. *The Universal Thom Form.* We will begin the discussion by introducing what may seem at first glance to be a rather strange notational device. We consider again the oriented, real vector bundle $\xi = (\pi_E : E \to X)$ of rank $n = 2k$ with typical fiber V. Motivated by the definition (22) of the Pfaffian we define the *Berezin* (or *fermionic*) *integral* of any $f \in \wedge V$ to be the real coefficient f_{vol} of $\text{vol} = \psi^1 \cdots \psi^{2k}$ in the polynomial f and write this as

$$\int f \, \mathcal{D}\psi = f_{\text{vol}}.$$

Here $\mathcal{D}\psi$ is just a formal reminder of the name we have given to the odd generators of $\wedge V$. With this (22) can be written

$$\int e^{\frac{1}{2}\psi^T A \psi} \, \mathcal{D}\psi = \text{Pf}\,(A) \tag{27}$$

for any $A \in \mathfrak{so}(2k)$. In particular, the representative (24) of the Euler class can now be written

$$(2\pi)^{-k} \int e^{\frac{1}{2}\psi^T(-s^*\Omega)\psi} \, \mathcal{D}\psi.$$

We will also need to extend this notion of Berezin integration in the following way. Let A be any supercommutative superalgebra and consider the supertensor product $A \,\hat{\otimes} \wedge V$ (see Remark 1.12). Regard the elements of $A \,\hat{\otimes} \wedge V$ as polynomials in the odd variables $\psi^1, \ldots, \psi^{2k}$ with coefficients in A and define the Berezin integral of $F \in A \,\hat{\otimes} \wedge V$ to be the coefficient in A of $\mathrm{vol} = \psi^1 \cdots \psi^{2k}$.

$$\int F \, \mathcal{D}\psi = F_{\mathrm{vol}}$$

Example 1.5. Introduce coordinate functions $u_1, \ldots, u_{2k}$ on V corresponding to the basis $\{\psi^1, \ldots, \psi^{2k}\}$. Thus, $\{u_1, \ldots, u_{2k}\}$ is the basis for V^* dual to $\{\psi^1, \ldots, \psi^{2k}\}$. Let $A = \Omega^*(V)$ be the algebra of complex-valued differential forms on V. Then each du_j, $j = 1, \ldots, 2k$, is in $\Omega^1(V)$ and $-idu_j\psi^j = i\psi^j du_j$ (sum over $j = 1, \ldots, 2k$) is in $\Omega^*(V) \,\hat{\otimes} \wedge V$. Writing du for the column vector $(du_1 \cdots du_{2k})^T$ we will show that

$$\int e^{i\psi^T du} \, \mathcal{D}\psi = \int e^{i\psi^j du_j} \, \mathcal{D}\psi = du_1 \cdots du_{2k},$$

where we again suppress the customary $\wedge$. To see this we compute as follows.

$$\int e^{i\psi^T du} \, \mathcal{D}\psi = \int e^{i(\psi^1 u_1 + \cdots + \psi^{2k} u_{2k})} \, \mathcal{D}\psi = \int e^{i\psi^1 u_1} \cdots e^{i\psi^{2k} u_{2k}} \, \mathcal{D}\psi$$

since the elements $\psi^j u_j$ are all even in $\Omega^*(V) \,\hat{\otimes} \wedge V$ and therefore commute. Next, using the (terminating) exponential series and discarding terms that do not contribute to the coefficient of $\mathrm{vol} = \psi^1 \cdots \psi^{2k}$, we obtain

$$\begin{aligned}\int e^{i\psi^T du} \, \mathcal{D}\psi &= \int (1 + i\psi^1 u_1) \cdots (1 + i\psi^{2k} u_{2k}) \, \mathcal{D}\psi = \int (i\psi^1 u_1) \cdots (i\psi^{2k} u_{2k}) \, \mathcal{D}\psi \\ &= i^{2k} \int (-1)^{\frac{1}{2}(2k)(2k+1)} du_1 \cdots du_{2k} \, \psi^1 \cdots \psi^{2k} \, \mathcal{D}\psi \\ &= du_1 \cdots du_{2k}.\end{aligned}$$

Notice also that, if we write $\|u\|^2 = u_1^2 + \cdots + u_{2k}^2 \in \Omega^0(V)$ and identify this with the even element $\|u\|^2 \otimes 1$ in $\Omega^*(V) \,\hat{\otimes} \wedge V$, then

$$(2\pi)^{-k} \int e^{-\frac{1}{2}\|u\|^2 + i\psi^T du} \, \mathcal{D}\psi = (2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} du_1 \cdots du_{2k} \tag{28}$$

and this is a form that integrates to 1 over V.

The Berezin integral $(2\pi)^{-k} \int e^{-\frac{1}{2}\|u\|^2 + i\psi^T du} \, \mathcal{D}\psi$ in (28) is often referred to as a *Gaussian representative* of the Thom class of V, thought of as a vector bundle over a point. It does not have compact support, but is *rapidly decreasing* and we will find that this is enough for our purposes. Our objective is construct similar Gaussian representatives of the Thom class for the vector bundle $\xi = (\pi_E : E \to X)$ and from them representatives of the Euler class (see (25)).

Remark 1.14. We will make no particular use of the fact, but we should point out that a Gaussian representative of the Thom class is easily converted into an actual representative of the Thom class, that is, into a form with compact support (see Section 10.3, page 156, of [GuSt]).

The first step is the construction of what is called a *universal Thom form*. Once again we let V denote the typical fiber of ξ. As in the previous example we choose an oriented, orthonormal (with respect to the fiber metric) basis $\{\psi^1, \ldots, \psi^{2k}\}$ for V with coordinate functions $\{u_1, \ldots, u_{2k}\}$. The basis identifies $SO(V)$ with $SO(2k)$ and therefore $\mathfrak{so}(V)$ with $\mathfrak{so}(2k)$. Choose a basis $\{\zeta_1, \ldots, \zeta_n\}$, $n = k(2k-1)$, for $\mathfrak{so}(2k)$ and let $\{x^1, \ldots, x^n\}$ be the dual basis for $\mathfrak{so}(2k)^*$. Now let

$$A = \mathbb{C}[\mathfrak{so}(2k)] \hat{\otimes} \Omega^*(V)$$

(see Remark 1.12). Every element of A is a sum of terms of the form $\mathcal{P} \otimes \varphi$, where $\mathcal{P}$ is a polynomial in $x^1, \ldots, x^n$ and φ is a form on V. We regard these elements as polynomial functions on $\mathfrak{so}(2k)$ with values in $\Omega^*(V)$. One should keep in mind that we have doubled the degrees of the polynomial parts so that $\mathbb{C}[\mathfrak{so}(2k)]$ is purely even. Now consider the supercommutative superalgebra

$$A \hat{\otimes} \wedge V = \mathbb{C}[\mathfrak{so}(2k)] \hat{\otimes} \Omega^*(V) \hat{\otimes} \wedge V.$$

The elements are polynomials in the odd variables $\psi^1, \ldots, \psi^{2k}$ with coefficients in A. Consequently, Berezin integration of an element of $\mathbb{C}[\mathfrak{so}(2k)] \hat{\otimes} \Omega^*(V) \hat{\otimes} \wedge V$ yields an element of $\mathbb{C}[\mathfrak{so}(2k)] \hat{\otimes} \Omega^*(V)$. Our universal Thom form will be obtained in this way.

Notice first that $\Omega^*(V) \hat{\otimes} \wedge V$ can be identified with the subalgebra of $\mathbb{C}[\mathfrak{so}(2k)] \hat{\otimes} \Omega^*(V) \hat{\otimes} \wedge V$ for which the polynomial part $\mathcal{P}$ is $1 \in \mathbb{C}[\mathfrak{so}(2k)]$. In particular, the exponent in (28) is in $\mathbb{C}[\mathfrak{so}(2k)] \hat{\otimes} \Omega^*(V) \hat{\otimes} \wedge V$.

$$-\frac{1}{2}\|u\|^2 + i\psi^T du \in \mathbb{C}[\mathfrak{so}(2k)] \hat{\otimes} \Omega^* V) \hat{\otimes} \wedge V \tag{29}$$

We intend to add just one more term and then exponentiate. For this extra term we observe the following. Each $\zeta \in \mathfrak{so}(2k)$ is represented on V by a linear transformation M_ζ whose matrix relative to the basis $\{\psi^1, \ldots, \psi^{2k}\}$ is skew-symmetric. Specifically,

$$M_\zeta(v) = \frac{d}{dt}\big[\,(\exp(-t\zeta)(v)\,\big]\Big|_{t=0}$$

for each $v \in V$, where $\exp : \mathfrak{so}(2k) \to SO(2k)$ is the exponential map on $\mathfrak{so}(2k)$. We will write (M_ζ) for the matrix of M_ζ relative to $\{\psi^1, \ldots, \psi^{2k}\}$. Since this is skew-symmetric,

$$\frac{1}{2}\psi^T(M_\zeta)\psi = -\frac{1}{2}\sum_{j=1}^{2k}\psi^j M_\zeta \psi^j = \frac{1}{2}\sum_{j=1}^{2k}(M_\zeta\psi^j)\psi^j. \tag{30}$$

Writing M_a for M_{ζ_a}, $a = 1, \ldots, n$, and $\zeta = x^a(\zeta)\zeta_a$ (summation convention over $a = 1, \ldots, n$), we have $M_\zeta = x^a(\zeta)M_a$ since $\zeta \mapsto M_\zeta$ is an isomorphism. Thus,

$$-\frac{1}{2}\sum_{j=1}^{2k}\psi^j M_\zeta \psi^j = -\frac{1}{2}x^a(\zeta)\sum_{j=1}^{2k}\psi^j M_a \psi^j \tag{31}$$

Now notice that $-\frac{1}{2}x^a \in \mathbb{C}[\mathfrak{so}(2k)]$ and $\sum_{j=1}^{2k} \psi^j M_a \psi^j \in \wedge V$. Thus,

$$-\frac{1}{2}x^a \otimes \Big(\sum_{j=1}^{2k} \psi^j M_a \psi^j \Big) \in \mathbb{C}[\mathfrak{so}(2k)] \,\hat{\otimes} \wedge V$$

or, better yet,

$$-\frac{1}{2}x^a \otimes 1 \otimes \Big(\sum_{j=1}^{2k} \psi^j M_a \psi^j \Big) \in \mathbb{C}[\mathfrak{so}(2k)] \,\hat{\otimes}\, \Omega^*(V) \,\hat{\otimes} \wedge V.$$

It is customary to write this even element of the algebra simply as

$$-\frac{1}{2} \sum_{j=1}^{2k} \psi^j x^a M_a \psi^j \in \mathbb{C}[\mathfrak{so}(2k)] \,\hat{\otimes}\, \Omega^*(V) \,\hat{\otimes} \wedge V. \tag{32}$$

Now one can exponentiate to obtain

$$e^{-\frac{1}{2}\sum_{j=1}^{2k} \psi^j x^a M_a \psi^j} \in \mathbb{C}[\mathfrak{so}(2k)] \,\hat{\otimes}\, \Omega^*(V) \,\hat{\otimes} \wedge V$$

and then perform a Berezin integration to get

$$\int e^{-\frac{1}{2}\sum_{j=1}^{2k} \psi^j x^a M_a \psi^j} \, \mathcal{D}\psi \in \mathbb{C}[\mathfrak{so}(2k)] \,\hat{\otimes}\, \Omega^*(V).$$

Since the $\Omega^*(V)$-part is 1, this is essentially a $\mathbb{C}$-valued polynomial on $\mathfrak{so}(2k)$. Indeed, by (27), (30) and (31), its value at any $\zeta \in \mathfrak{so}(2k)$ is given by

$$(2\pi)^{-k} \int e^{-\frac{1}{2}\sum_{j=1}^{2k} \psi^j x^a(\zeta) M_a \psi^j} \, \mathcal{D}\psi = \mathrm{Pf}\,(M_\zeta).$$

Now we combine (29) and (32) to obtain

$$-\frac{1}{2}\|u\|^2 + i\psi^T du - \frac{1}{2}\sum_{j=1}^{2k} \psi^j x^a M_a \psi^j \in \mathbb{C}[\mathfrak{so}(2k)] \,\hat{\otimes}\, \Omega^* V \,\hat{\otimes} \wedge V \tag{33}$$

Exponentiating, computing the Berezin integral and then multiplying by $(2\pi)^{-k}$ gives an element of $\mathbb{C}[\mathfrak{so}(2k)] \,\hat{\otimes}\, \Omega^*(V)$ denoted ν and called the *(Mathai-Quillen) universal Thom form* of V.

$$\begin{aligned} \nu &= (2\pi)^{-k} \int e^{-\frac{1}{2}\|u\|^2 + i\psi^T du - \frac{1}{2}\sum_{j=1}^{2k} \psi^j x^a M_a \psi^j} \, \mathcal{D}\psi \\ &= (2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \int \exp\Big(i\psi^T du - \frac{1}{2}\sum_{j=1}^{2k} \psi^j x^a M_a \psi^j \Big) \mathcal{D}\psi \end{aligned} \tag{34}$$

We will see the ν is basically a machine for producing representatives of the Thom class of any vector bundle ξ with typical fiber V.

Remark 1.15. It is an instructive exercise to compute ν explicitly when $V = \mathbb{R}^2$ with its usual orientation and inner product. Letting $\{\psi^1, \psi^2\}$ be the standard basis for $\mathbb{R}^2$ and taking $\{\zeta_1\} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ as the basis for $\mathfrak{so}(2)$, the result is the following non-homogeneous element of $\mathbb{C}[\mathfrak{so}(2)] \hat{\otimes} \Omega^*(\mathbb{R}^2)$.

$$\nu = (2\pi)^{-1}\, e^{-\frac{1}{2}(u_1^2+u_2^2)}\, du_1 du_2 + (2\pi)^{-1}\, x^1\, e^{-\frac{1}{2}(u_1^2+u_2^2)}. \tag{35}$$

Notice that each term has degree 2 in $\mathbb{C}[\mathfrak{so}(2)] \hat{\otimes} \Omega^*(\mathbb{R}^2)$ and the first term integrates to 1 over $\mathbb{R}^2$.

1.5.3. *Equivariant Cohomology.* The construction of the universal Thom form no doubt appears rather *ad hoc*. As it happens, ν lives quite naturally in the context of *equivariant cohomology* and we would like to say just a few words about this; the standard reference for all of this material is [GuSt]. Equivariant cohomology arose from attempts to understand the topology of the orbit space X/G of a topological space X on which some topological group G acts. We will be concerned only with the case in which X is a smooth manifold and G is a compact, connected, matrix Lie group acting smoothly on X on the left. For this action we will write

$$\sigma : G \times X \to X$$
$$\sigma(g, x) = g \cdot x = \sigma_g(x) = \sigma_x(g).$$

We denote by $\mathfrak{g}$ the Lie algebra of G, by $\mathbb{C}[\mathfrak{g}]$ the graded algebra of complex-valued polynomial functions on $\mathfrak{g}$, and by $\Omega^*(X)$ the graded algebra of complex-valued, smooth differential forms on X. We consider the supertensor product $\mathbb{C}[\mathfrak{g}] \hat{\otimes} \Omega^*(X)$, every element of which is a sum of terms of the form $\mathcal{P} \otimes \varphi$, where $\mathcal{P} \in \mathbb{C}[\mathfrak{g}]$ and $\varphi \in \Omega^*(X)$. These are best thought of as $\Omega^*(X)$-valued polynomials on $\mathfrak{g}$.

$$(\mathcal{P} \otimes \varphi)(\zeta) = \mathcal{P}(\zeta)\varphi$$

The degree of a homogeneous element $\alpha = \mathcal{P} \otimes \varphi$ is taken to be

$$\deg\, \alpha = \deg\,(\mathcal{P} \otimes \varphi) = 2 \deg\, \mathcal{P} + \deg\, \varphi,$$

where $\deg\, \mathcal{P}$ is the polynomial degree of $\mathcal{P}$ and $\deg\, \varphi$ is the cohomological degree of the form φ. Thus,

$$\mathbb{C}[\mathfrak{g}] \hat{\otimes} \Omega^*(X) = \bigoplus_{k=2i+j} \mathbb{C}^i[\mathfrak{g}] \otimes \Omega^j(X).$$

The action of G on X together with the adjoint action of G on $\mathfrak{g}$ give a natural action of G on $\mathbb{C}[\mathfrak{g}] \hat{\otimes} \Omega^*(X)$ defined as follows. If $\alpha = \mathcal{P} \otimes \varphi$ is a homogeneous element of $\mathbb{C}[\mathfrak{g}] \hat{\otimes} \Omega^*(X)$ and $g \in G$, then $g \cdot \alpha$ is the element of $\mathbb{C}[\mathfrak{g}] \hat{\otimes} \Omega^*(X)$ whose value at any $\zeta \in \mathfrak{g}$ is given by

$$(g \cdot \alpha)(\zeta) = (g \cdot (\mathcal{P} \otimes \varphi))(\zeta) = \mathcal{P}(g^{-1}\zeta g)\sigma^*_{g^{-1}}\varphi$$

An element α of $\mathbb{C}[\mathfrak{g}] \hat{\otimes} \Omega^*(X)$ is said to be *G-invariant* if $g \cdot \alpha = \alpha$ for every $g \in G$ or, equivalently,

$$\alpha(g\zeta g^{-1}) = \sigma^*_{g^{-1}}\alpha(\zeta)$$

for every $g \in G$ and every $\zeta \in \mathfrak{g}$. The graded algebra of all G-invariant elements of $\mathbb{C}[\mathfrak{g}] \hat{\otimes} \Omega^*(X)$ is denoted

$$\Omega^*_G(X) = [\, \mathbb{C}[\mathfrak{g}] \hat{\otimes} \Omega^*(X)\,]^G$$

and its elements are called *G-equivariant differential forms* on X. Notice that $\mathbb{C}[\mathfrak{g}]^G$ (those $\mathcal{P}$ in $\mathbb{C}[\mathfrak{g}]$ satisfying $\mathcal{P}(g\zeta g^{-1}) = \mathcal{P}(\zeta)$ for all $g \in G$ and all $\zeta \in \mathfrak{g}$) and $\Omega^*(X)^G$ (those $\varphi \in \Omega^*(X)$ satisfying $\sigma^*_{g^{-1}}\varphi = \varphi$ for all $g \in G$) can both be identified with subalgebras of $\Omega^*_G(X)$. Our grading of $\mathbb{C}[\mathfrak{g}]\,\hat{\otimes}\,\Omega^*(X)$ gives

$$\Omega^*_G(X) = \bigoplus_{k=0}^{\infty} \Omega^k_G(X) = \bigoplus_{k=2i+j} [\, \mathbb{C}^i[\mathfrak{g}]\,\hat{\otimes}\,\Omega^j(X)\,]^G.$$

Note that the universal Thom form (35) for $\mathbb{R}^2$ is clearly SO(2)-equivariant since the adjoint action of SO(2) on $\mathfrak{so}(2)$ is trivial (SO(2) is Abelian) and the exponential factor $e^{-\frac{1}{2}(u_1^2+u_2^2)}$ is rotationally invariant. It is not quite so obvious, but the analogous statement is true of every universal Thom form, that is,

$$\nu = (2\pi)^{-k} \int e^{-\frac{1}{2}\|u\|^2 + i\psi^T du - \frac{1}{2}\sum_{j=1}^{2k} \psi^j x^a M_a \psi^j}\, \mathcal{D}\psi \in [\, \mathbb{C}[\mathfrak{so}(V)]\,\hat{\otimes}\,\Omega^*(V)\,]^{\mathrm{SO}(V)}.$$

To build a cohomology from $\Omega^*_G(X)$ we define the *G-equivariant exterior derivative* d_G on $\Omega^*_G(X)$ as follows. For any $\alpha \in \Omega^*_G(X)$ the value of $d_G\alpha$ at $\zeta \in \mathfrak{g}$ is given by

$$(d_G\alpha)(\zeta) = d(\alpha(\zeta)) - \iota_{\zeta^\#}(\alpha(\zeta)),$$

where $\zeta^\#$ is the vector field on X defined at each $x \in X$ by

$$\zeta^\#(x) = \frac{d}{dt}\,(\exp(-t\zeta)\cdot x)\big|_{t=0}$$

and $\iota_{\zeta^\#}$ denotes interior multiplication (contraction) by $\zeta^\#$.

Example 1.6. We will show that $d_{\mathrm{SO}(2)}\nu = 0$, where ν is the universal Thom form of $\mathbb{R}^2$ in (35). Any $\zeta \in \mathfrak{so}(2)$ can be written

$$\zeta = \begin{pmatrix} 0 & \lambda \\ -\lambda & 0 \end{pmatrix} = \lambda\zeta_1.$$

Then $x^1(\zeta) = \lambda$ so

$$\nu(\zeta) = \frac{1}{2\pi}\, e^{-\frac{1}{2}(u_1^2+u_2^2)} du_1 du_2 + \frac{1}{2\pi}\,\lambda\, e^{-\frac{1}{2}(u_1^2+u_2^2)}.$$

Thus,

$$d(\nu(\zeta)) = 0 + (2\pi)^{-1}\lambda\, d(e^{-\frac{1}{2}(u_1^2+u_2^2)}) = (2\pi)^{-1} e^{-\frac{1}{2}(u_1^2+u_2^2)}\lambda(-u_1du_1 - u_2du_2).$$

Next we compute

$$\begin{aligned}\iota_{\zeta^\#}(\nu(\zeta)) &= (2\pi)^{-1} e^{-\frac{1}{2}(u_1^2+u_2^2)}\iota_{\zeta^\#}(du_1du_2) + 0 \\ &= (2\pi)^{-1} e^{-(u_1^2+u_2^2)}(\,\iota_{\zeta^\#}(du_1)du_2 - du_1\iota_{\zeta^\#}(du_2)\,).\end{aligned}$$

The proof will be complete if we show

$$\iota_{\zeta^\#}(du_1) = u_1(\zeta^\#) = -\lambda u_2 \quad \text{and} \quad \iota_{\zeta^\#}(du_2) = u_2(\zeta^\#) = \lambda u_1.$$

Let $v = v_1\psi^1 + v_2\psi^2 \in \mathbb{R}^2$ be arbitrary. Then

$$\begin{aligned}\zeta^{\#}(v) &= \frac{d}{dt}\left(exp(-t\zeta)\cdot v \right)\Big|_{t=0} \\ &= \frac{d}{dt}\left((1 - t\zeta + \frac{1}{2}t^2\zeta^2 - \cdots)\cdot v \right)\Big|_{t=0} \\ &= -\zeta v = -\lambda \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \\ &= \begin{pmatrix} -\lambda v_2 \\ \lambda v_1 \end{pmatrix} = \begin{pmatrix} -\lambda u_2(v) \\ \lambda u_1(v) \end{pmatrix}\end{aligned}$$

as required.

An element $\alpha \in \Omega^*_G(X)$ with $d_G\alpha = 0$ is said to be *G-equivariantly closed* and the previous example generalizes to show that this is true of any universal Thom form (see pages 86-87 of [GuSt]).

$$d_{\mathrm{SO}(V)}\nu = 0$$

One can show also that, for any $\alpha \in \Omega^*_G(X)$,

$$d_G \circ d_G(\alpha) = 0$$

so $d_G^2 = 0$ on $\Omega^*_G(X)$ and, due to the grading we have introduced on $\Omega^*_G(X)$,

$$d_G : \Omega^k_G(X) \to \Omega^{k+1}_G(X).$$

Consequently,

$$(\Omega^*_G(X), d_G)$$

is a co-chain complex. The cohomology of this co-chain complex is called the *Cartan model* of the *G-equivariant cohomology* of X and is denoted

$$H^*_G(X).$$

Every G-equivariantly closed form $\alpha \in \Omega^*_G(X)$, such as the universal Thom form ν, determines a G-equivariant cohomology class that is denoted $[\alpha]$. If $\alpha = d_G\beta$ for some $\beta \in \Omega^*_G(X)$, then α is said to be *G-equivariantly exact* and in this case the cohomology class $[\alpha]$ is trivial.

Remark 1.16. Let $\mathbb{S}^3 = \{(z^1, z^2) \in \mathbb{C}^2 : |z^1|^2 + |z^2|^2 = 1\}$ be the 3-sphere and $\mathbb{S}^1 = \{e^{i\theta} : \theta \in \mathbb{R}\}$ the Lie group of complex numbers of modulus one. Define a free left action of $\mathbb{S}^1$ on $\mathbb{S}^3$ by

$$e^{i\theta}\cdot(z^1, z^2) = (e^{i\theta}z^1, e^{i\theta}z^2).$$

The orbit space $\mathbb{S}^3/\mathbb{S}^1$ is, by definition, the complex projective line $\mathbb{CP}^1$, which is diffeomorphic to the 2-sphere $\mathbb{S}^2$ (see (1.2.8), page 53, of [Nab4]). The $\mathbb{S}^1$-equivariant cohomology $H^*_{\mathbb{S}^1}(\mathbb{S}^3)$ of $\mathbb{S}^3$ is computed on pages 90-93 of [Nab2] and is found to coincide with the de Rham cohomology of the orbit space $\mathbb{S}^2$. According to the following theorem of Cartan, this is no accident (see Theorem 2.5.1 of [GuSt]).

Theorem 1.20. *(Cartan) Let X be a smooth manifold and G a compact, connected Lie group. Suppose there is a smooth, free, left action of G on X. Then the orbit space X/G is a smooth manifold and the G-equivariant cohomology $H^*_G(X)$ of X is isomorphic to the de Rham cohomology $H^*_{deRham}(X/G;\mathbb{C})$ of X/G with complex coefficients.*

Finally we will need to introduce a notion of integration for equivariant differential forms and cohomology classes. For this we will now assume that X is compact and oriented and that the G-action preserves the orientation of X in the sense that each of the diffeomorphisms $\sigma_g : X \to X$ is orientation preserving. Let α be an element of $\Omega^*_G(X)$. For each $\zeta \in \mathfrak{g}$ we write $\alpha(\zeta)$ as

$$\alpha(\zeta) = \alpha(\zeta)_{[0]} + \alpha(\zeta)_{[1]} + \cdots + \alpha(\zeta)_{[n]},$$

where $\alpha(\zeta)_{[k]} \in \Omega^k(X)$ and $n = \dim\, X$. Now we define an element

$$\int_X \alpha \in \mathbb{C}[\mathfrak{g}]^G$$

by

$$\Big(\int_X \alpha\Big)(\zeta) = \int_X \alpha(\zeta) \overset{def}{=} \int_X \alpha(\zeta)_{[n]}.$$

Notice that $\int_X \alpha$ really is G-invariant since

$$\Big(\int_X \alpha\Big)(g\zeta g^{-1}) = \int_X \alpha(g\zeta g^{-1})_{[n]} = \int_X \sigma^*_{g^{-1}}(\alpha(\zeta)_{[n]}) = \int_X \alpha(\zeta)_{[n]} = \Big(\int_X \alpha\Big)(\zeta).$$

Thus,

$$\int_X : \Omega^*_G(X) \to \mathbb{C}[\mathfrak{g}]^G. \tag{36}$$

Example 1.7. Let ν be the universal Thom form for $\mathbb{R}^2$ given by (35). Then, for every $\zeta \in \mathfrak{so}(2)$,

$$\Big(\int_{\mathbb{R}^2} \nu\Big)(\zeta) = \int_{\mathbb{R}^2} \nu(\zeta)_{[2]} = \int_{\mathbb{R}^2} (2\pi)^{-1} e^{-\frac{1}{2}(u_1^2+u_2^2)} du_1 du_2 = 1$$

so

$$\int_{\mathbb{R}^2} \nu = 1 \quad \text{(the constant function in } \mathbb{C}[\mathfrak{so}(2)]^{\mathrm{SO}(2)}).$$

The result in the previous Example generalizes to any universal Thom form (see pages 87-88 of [GuSt]).

$$\nu = (2\pi)^{-k} \int e^{-\frac{1}{2}\|u\|^2 + i\psi^T du - \frac{1}{2}\sum_{j=1}^{2k} \psi^j x^a M_a \psi^j}\, \mathcal{D}\psi \Rightarrow \int_V \nu = 1$$

Finally, note that, if $\alpha = d_G\beta$ is G-equivariantly exact, then, for each $\zeta \in \mathfrak{g}$, $\alpha(\zeta)_{[n]} = d(\beta(\zeta)_{[n-1]})$ so, by Stokes' Theorem, $\int_X \alpha = 0 \in \mathbb{C}[\mathfrak{g}]^G$. Consequently, the integration map (36) descends to cohomology.

$$\int_X : H^*_G(X) \to \mathbb{C}[\mathfrak{g}]^G$$

1.5.4. *Gaussian Representatives of the Thom Class.* With this digression on equivariant cohomology behind us we will return to the general development. Recall that our interest in the universal Thom class $[\nu]$ resides in the fact, asserted earlier, that its representative ν can be used to manufacture Gaussian representatives of the Thom class of any oriented, real vector bundle with typical fiber V. We will now briefly sketch the procedure.

We begin by once again recalling our basic assumptions. X will denote a compact, oriented, smooth manifold of even dimension $n = 2k$ and we will consider an oriented, real vector bundle over X whose typical fiber V has dimension $n = 2k$ and that has been equipped with a positive definite fiber metric. We will now use the fact that any such vector bundle can be viewed as the vector bundle $\xi_\rho = (\pi_\rho : P \times_\rho V \to X)$ associated to some principal bundle

$$G \hookrightarrow P \stackrel{\pi_P}{\to} X$$

by a representation

$$\rho : G \to \mathrm{SO}(V) \quad (\rho(g)(v) = g \cdot v)$$

of G on V. We briefly recall the construction of ξ_ρ. The actions of G on P and on V combine to give an action of G on $P \times V$ defined by

$$(p, v) \cdot g = (p \cdot g, g^{-1} \cdot v).$$

The action is free so the orbit space $P \times_\rho V$ is a smooth manifold and the projection

$$P \times V \stackrel{Q}{\to} P \times_\rho V$$
$$Q(p, v) = [p, v]$$

defines a principal G-bundle. Let $\pi_1 : P \times V \to P$ be the projection onto the first factor and define π_ρ by the commutativity of the following diagram.

$$\begin{array}{ccc} P \times V & \xrightarrow{\pi_1} & P \\ {\scriptstyle Q}\downarrow & & \downarrow{\scriptstyle \pi_P} \\ P \times_\rho V & \xrightarrow{\pi_\rho} & X \end{array}$$

We are trying to produce representatives of the Thom class on $P \times_\rho V$. Notice that the right-hand side of the diagram is the arena in which the Chern-Weil procedure produces characteristic classes. As motivation for what is to come we will briefly recall how this goes. Choose a connection ω on P with curvature Ω and a $\mathcal{P} \in \mathbb{C}[\mathfrak{g}]^G$. Evaluate $\mathcal{P}(\Omega)$ to obtain a form on P. $\mathcal{P}(\Omega)$ is basic, that is, G-invariant and horizontal with respect to ω, so it descends to a form on X. This form on X is closed and its cohomology class does not depend on the choice of ω so it represents a characteristic class of $G \hookrightarrow P \stackrel{\pi_P}{\to} X$. The entire process can be encapsulated in the existence of a map, called the *Chern-Weil map*

$$CW_\omega : \mathbb{C}[\mathfrak{g}]^G \to \Omega^*(P)_{BASIC} \cong \Omega^*(X)$$
$$CW_\omega(\mathcal{P}) = \mathcal{P}(\Omega)$$

The idea is to carry out a similar procedure on the left-hand side $G \hookrightarrow P \times V \stackrel{Q}{\to} P \times_\rho V$ of the diagram.

Again choose a connection ω on P with curvature Ω. Then $\pi_1^*\omega$ is a connection on $P \times V$ with curvature $\pi_1^*\Omega$. Identify $T_{(p,v)}(P \times V)$ with $T_p(P) \oplus T_v(V)$. Since $\pi_1^*\omega$ and $\pi_1^*\Omega$ depend only on the $T_p(P)$-summands and there agree with ω and Ω, respectively, we will use the same symbols to denote the forms on $P \times V$. The ω-horizontal spaces in $P \times V$ are given by $\mathrm{Hor}_{(p,v)}(\omega) = \mathrm{Hor}_p(\omega) \oplus T_v(V)$. The decomposition

$$T_{(p,v)}(P \times V) = \left(\mathrm{Hor}_p(\omega) \oplus T_v(V)\right) \oplus \mathrm{Vert}_{(p,v)}(P \times V)$$

determines a projection $\pi_{\mathrm{Hor}}(\omega)$ of forms on $P \times V$ to ω-horizontal forms on $P \times V$ (evaluate on ω-horizontal parts of the tangent vectors). Now let $\alpha = \mathcal{P} \otimes \varphi$ be a homogeneous element of $\Omega_G^*(V) = [\mathbb{C}[\mathfrak{g}] \hat{\otimes} \Omega^*(V)]^G$. One can evaluate the polynomial part on the curvature Ω as in the ordinary Chern-Weil procedure to obtain $\mathcal{P}(\Omega) \otimes \varphi$ in $\Omega^*(P) \otimes \Omega^*(V)$. This one can identify with the form $\mathcal{P}(\Omega) \wedge \varphi$ on $P \times V$ which, because $\mathcal{P} \otimes \varphi$ is G-invariant, is in $\Omega^*(P \times V)^G$. It is generally not ω-horizontal, however, so to obtain a basic form on $P \times V$ one must compose with the horizontal projection $\pi_{Hor}(\omega)$. This gives a map

$$\alpha = \mathcal{P} \otimes \varphi \in \Omega_G^*(V) \mapsto \pi_{Hor}(\omega)\left[\mathcal{P}(\Omega) \wedge \varphi\right] \in \Omega^*(P \times V)_{\mathrm{BASIC}} \cong \Omega^*(P \times_\rho V)$$

and extending by linearity gives the *generalized Chern-Weil map*

$$CW_\omega : \Omega_G^*(V) \to \Omega^*(P \times V)_{\mathrm{BASIC}} \cong \Omega^*(P \times_\rho V).$$

One can show that CW_ω is a co-chain map

$$d \circ CW_\omega = CW_\omega \circ d_G$$

so it carries G-equivariantly closed forms on V to ordinary closed forms on $P \times V$ which then descend to closed forms on $P \times_\rho V$.

Notice that the representation $\rho : G \to \mathrm{SO}(V)$ induces a Lie algebra homomorphism $\rho_* : \mathfrak{g} \to \mathfrak{so}(V)$ (just the derivative of ρ at the identity in G). With this any $\mathcal{P} \otimes \varphi$ in $\Omega^*_{\mathrm{SO}(V)}(V)$ gives rise to an element $(\mathcal{P} \circ \rho_*) \otimes \varphi$ in $\Omega_G^*(V)$ and therefore, by linearity, so does every element of $\Omega^*_{\mathrm{SO}(V)}(V)$. In particular, the universal Thom form gives rise to

$$\nu_G = (2\pi)^{-k} \int e^{-\frac{1}{2}\|u\|^2 + i\psi^T du - \frac{1}{2}\sum_{j=1}^{2k} \psi^j [x^a \circ \rho_*] M_a \psi^j} \, \mathcal{D}\psi$$

in $\Omega_G^*(V)$. Applying the map CW_ω to ν_G gives a basic, closed form on $P \times V$ that is *the horizontal projection of*

$$(2\pi)^{-k} \int e^{-\frac{1}{2}\|u\|^2 + i\psi^T du - \frac{1}{2}\sum_{j=1}^{2k} \psi^j (x^a [\rho_* \Omega]) M_a \psi^j} \, \mathcal{D}\psi.$$

It will often be more convenient to work with a more compact form of this that we obtain as follows. We will write $(\rho_*\Omega)$ for the skew-symmetric matrix image of the curvature under ρ_*. In somewhat more detail, the values of the 2-form Ω are in $\mathfrak{g}$ and ρ_* carries each of these to $\mathfrak{so}(V)$. Each of these values corresponds, via the action of $\mathfrak{so}(V)$ on V, to a linear transformation $x^a[\rho_*\Omega]M_a$ on V. This we can then identify with a skew-symmetric matrix $(x^a[\rho_*\Omega]M_a)$. In this way we think of $(\rho_*\Omega)$ as a matrix of 2-forms on P with values in $\mathfrak{so}(V)$. Now use the fact that, for any skew-symmetric matrix S, $\psi^T S \psi = -\sum_{j=1}^{2k} \psi^j S \psi^j$. Moreover, we can implement the horizontal projection by insisting that the form be evaluated only on the horizontal parts of tangent vectors to $P \times V$. With this we define a basic, closed form $\tilde{U}$ on $P \times V$ by

$$\tilde{U} = (2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \int \exp\left(i\psi^T du + \frac{1}{2}\psi^T (\rho_*\Omega)\,\psi\right) \mathcal{D}\psi \quad \text{(evaluated on horizontal parts).} \tag{37}$$

Then $\tilde{U}$ descends to a closed form U on $P \times_\rho V$ ($\tilde{U} = Q^* U$). This is, in fact, a Gaussian representative of the Thom class of $P \times_\rho V$ (see Chapter 10 and, in particular, Section 10.4 of [GuSt])

Remark 1.17. One can write out an explicit, albeit rather cumbersome, formula for the ω-horizontal projection of any form φ. Writing $\omega = \omega^a \zeta_a$ and $\iota_a = \iota_{\zeta_a^\#}$ one has

$$\pi_{\text{Hor}}(\omega)(\varphi) = \varphi - \omega^1 \wedge \iota_1 \varphi - \omega^2 \wedge \iota_2 \varphi - \cdots - \omega^n \wedge \iota_n \varphi + \sum_{1 \le a_1 < \cdots < a_r \le n} (-1)^{r(r+1)/2} \omega^{a_1} \wedge \cdots \wedge \omega^{a_r} \wedge (\iota_{a_1} \circ \cdots \circ \iota_{a_r})(\varphi) \tag{38}$$

We will see an instance of this in the next example. However, it will generally be most useful for our purposes to view this as evaluating on horizontal parts. For example, if we recall the Cartan formula $\Omega = d\omega + \frac{1}{2}[\omega, \omega]$ and notice that the second term vanishes on ω-horizontal parts by definition, the right-hand side of (37), when evaluated on horizontal parts, gives the same result whether or not $\frac{1}{2}[\omega, \omega]$ is present. We can therefore simplify (37) and write

$$\tilde{U} = (2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \int \exp\left(i\psi^T du + \frac{1}{2}\psi^T (\rho_*(d\omega))\, \psi \right) \mathcal{D}\psi \quad \text{(evaluated on horizontal parts).} \tag{39}$$

We will have considerably more cosmetic surgery to perform on $\tilde{U}$ and U as we proceed.

Since our primary concern is with representatives of the Euler class, we will want to pull U back by a section of the vector bundle $\pi_\rho : P \times_\rho V \to X$. Now, any section of the associated bundle $P \times_\rho V$ can be written as

$$x \in X \xrightarrow{(s, S \circ s)} (s(x), S(s(x))) \in P \times V \xrightarrow{Q} [s(x), S(s(x))] \in P \times_\rho V,$$

where s is a section of $G \hookrightarrow P \xrightarrow{\pi_P} X$ and $S : P \to V$ is an equivariant map, that is, $S(p \cdot g) = \rho(g^{-1})(s(p))$ for all $p \in P$ and all $g \in G$ (see Section 6.8, page 384, of [Nab4]). Thus,

$$(Q \circ (s, S \circ s))^* U = (s, S \circ s)^* (Q^* U) = ((1, S) \circ s)^* \tilde{U} = s^* ((1, S)^* \tilde{U})$$

so to pull U back by a section of $P \times_\rho V$ we compute $(1, S)^* \tilde{U}$, which simply pulls the V-factors of $\tilde{U}$ back by the equivariant map S to get a form on P which we then pull back to X by a section s of the principal bundle $G \hookrightarrow P \xrightarrow{\pi_P} X$. We never need to explicitly compute U since $\tilde{U}$ contains all of the necessary information. The essential fact, particularly when we later mimic these ideas in the infinite-dimensional context of quantum field theory, is that S and s are entirely arbitrary and *one can obtain very different looking representatives of the Euler class by making different choices for S and s*. We will illustrate all of this with an example.

Example 1.8. We begin with the universal Thom form for $\mathbb{R}^2$ using the notation established in Remark 1.15, Example 1.6, and Example 1.7.

$$\nu = (2\pi)^{-1} e^{-\frac{1}{2}(u_1^2 + u_2^2)} (du_1 du_2 + x^1)$$

As the vector bundle with typical fiber $\mathbb{R}^2$ we take the tangent bundle $T\mathbb{S}^2$ of the 2-sphere $\mathbb{S}^2$, which we need to regard as a vector bundle associated to some principal bundle. We provide $\mathbb{S}^2$ with its usual orientation and Riemannian metric and consider the oriented, orthonormal frame bundle

$$\mathrm{SO}(2) \hookrightarrow F_{\mathrm{SO}}(\mathbb{S}^2) \xrightarrow{\pi_{\mathrm{SO}}} \mathbb{S}^2.$$

If $\rho : \mathrm{SO}(2) \to \mathrm{SO}(2)$ is the identity representation of SO(2) on SO(2) ($\rho = id_{\mathrm{SO}(2)}$), then

$$F_{\mathrm{SO}}(\mathbb{S}^2) \times_\rho \mathbb{R}^2 \cong T\mathbb{S}^2.$$

In this case, $\rho_* : \mathfrak{so}(2) \to \mathfrak{so}(2)$ is also the identity so $x^1 \circ \rho_* = x^1$. Now let $\omega = \omega^1 \zeta_1$ be a connection on the frame bundle with curvature $\Omega = \Omega^1 \zeta_1$ (soon we will be more specific and take ω to be the Levi-Civita connection, but for the moment we will leave it arbitrary). Then $x^1(\rho_* \Omega) = \Omega^1$ so $CW_\omega(\nu_{\mathrm{SO}})$ is the horizontal projection of

$$\varphi = (2\pi)^{-1} e^{-\frac{1}{2}(u_1^2+u_2^2)} \, (du_1 du_2 + \Omega^1)$$

on $F_{\mathrm{SO}}(\mathbb{S}^2) \times \mathbb{R}^2$ with respect to the connection ω on $F_{\mathrm{SO}}(\mathbb{S}^2) \times \mathbb{R}^2$ (keep in mind that this really means the pullback to $F_{\mathrm{SO}}(\mathbb{S}^2) \times \mathbb{R}^2$ of the chosen connection on $F_{\mathrm{SO}}(\mathbb{S}^2)$ by the projection $\pi_1 : F_{\mathrm{SO}}(\mathbb{S}^2) \times \mathbb{R}^2 \to F_{\mathrm{SO}}(\mathbb{S}^2)$ onto the first factor). We need to compute the ω-horizontal projection of φ. The explicit formula (38) in this special case is

$$\pi_{\mathrm{Hor}}(\omega)(\varphi) = \varphi - \omega^1 \wedge \iota_1 \varphi.$$

Here $\iota_1 = \iota_{\zeta_1^\#}$, where $\zeta_1^\#$ is the vector field on $F_{\mathrm{SO}}(\mathbb{S}^2) \times \mathbb{R}^2$ determined by

$$\begin{aligned}
\zeta_1^\#(p, v) &= \frac{d}{dt} \left(p \cdot \exp(t\zeta_1),\, \exp(-t\zeta_1) \cdot v \right)\Big|_{t=0} \\
&= \frac{d}{dt} \left(p \cdot \exp(t\zeta_1) \right)\Big|_{t=0} \oplus \frac{d}{dt} \left(\exp(-t\zeta_1) \cdot v \right)\Big|_{t=0} \\
&= \zeta_1^\#(p) \oplus \zeta_1^\#(v),
\end{aligned}$$

where we identify $T_{(p,v)}(F_{\mathrm{SO}}(\mathbb{S}^2) \times \mathbb{R}^2) = T_p(F_{\mathrm{SO}}(\mathbb{S}^2)) \oplus T_v(\mathbb{R}^2)$. As in Example 1.6 we have

$$\iota_{\zeta_1}(du_1 du_2) = -u_1 du_1 - u_2 du_2.$$

Moreover, $\iota_1 \Omega^1 = 0$ since $\zeta_1^\#$ is vertical and Ω is $d\omega$ evaluated on horizontal parts. Thus,

$$CW_\omega(\nu_{\mathrm{SO}}) = (2\pi)^{-1} e^{-\frac{1}{2}(u_1^2+u_2^2)} \left(\Omega^1 + du_1 du_2 + \omega^1 \wedge (u_1 du_1 + u_2 du_2) \right). \tag{40}$$

Next, we have seen that to get a representative of the Euler class of the tangent bundle of $\mathbb{S}^2$ we must choose a section s of the frame bundle and an equivariant map S of $F_{\mathrm{SO}}(\mathbb{S}^2)$ to $\mathbb{R}^2$. As a section of the frame bundle we choose the oriented, orthonormal frame field corresponding to the spherical coordinate chart, that is,

$$s(\phi, \theta) = \left(\phi, \theta, \partial_\phi, \frac{1}{\sin\phi}\, \partial_\theta \right).$$

We choose an equivariant map $F_{\mathrm{SO}}(\mathbb{S}^2) \xrightarrow{S} \mathbb{R}^2$ by beginning with a vector field on $\mathbb{S}^2$ (that is, a section of $T\mathbb{S}^2$). This can be chosen arbitrarily and we will take $V = \gamma \sin\theta\, \partial_\phi + \gamma \cos\theta \cot\phi\, \partial_\theta$, where γ is an arbitrary real parameter (when $\gamma = 1$ this is the infinitesimal generator for rotations about the x-axis). Relative to the frame field s the components of V are $\gamma\, sin\,\theta$ and $\gamma\, cos\,\theta\, cos\,\phi$. The map $S : F_{SO}(\mathbb{S}^2) \to \mathbb{R}^2$ is defined on the image of s by

$$S \circ s(\phi, \theta) = (sin\,\theta, cos\,\theta\, cos\,\phi)$$

and elsewhere by equivariance. We have seen that to pull back $CW_\omega(\nu_{SO})$ by the section $x \mapsto [s(x), S(s(x))]$ one proceeds in two steps. First pull back the $\mathbb{R}^2$-parts by S, that is, substitute $u_1 = \gamma\, sin\,\theta$ and $u_2 = \gamma\, cos\,\theta\, cos\,\phi$, to obtain a form on $F_{SO}(\mathbb{S}^2)$ and then pull this form back to $\mathbb{S}^2$ by s. To make matters quite concrete we will now choose a specific connection on the frame bundle. The most natural choice is the Levi-Civita connection which is given in terms of spherical coordinates by

$$\begin{pmatrix} 0 & -cos\,\phi\, d\theta \\ cos\,\phi\, d\theta & 0 \end{pmatrix} = (-cos\,\phi\, d\theta)\, \zeta_1$$

so that $s^*\omega^1 = -cos\,\phi\, d\theta$ and therefore $s^*\Omega^1 = sin\,\phi\, d\phi\, d\theta$. The result of all of these substitutions and a bit of trigonometry is the following representative of the Euler class of the tangent bundle of $\mathbb{S}^2$.

$$(2\pi)^{-1} e^{-\frac{1}{2}\gamma^2(sin^2\theta + cos^2\theta\, cos^2\phi)}\, sin\,\phi\, (1 + \gamma^2 cos^2\theta\, sin^2\phi)\, d\phi\, d\theta$$

In particular, since the Euler characteristic of $\mathbb{S}^2$ is 2, we obtain the following not altogether obvious integral formula.

$$\frac{1}{2\pi} \int_0^{2\pi} \int_0^{\pi} e^{-\frac{1}{2}\gamma^2(sin^2\theta + cos^2\theta\, cos^2\phi)}\, sin\,\phi\, (1 + \gamma^2 cos^2\theta\, sin^2\phi)\, d\phi\, d\theta = 2$$

1.5.5. *Witten's Partition Function.* Now recall that our interest in the Mathai-Quillen formalism stems from the fact that Atiyah and Jeffrey [AJ] have shown how it can be adapted and *formally* applied to the infinite-dimensional vector bundle $\hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega^2_+(X, \mathrm{ad}\, P)$ to yield an integral representation of the 0-dimensional Donaldson invariant and which coincides with the partition function of Witten's topological quantum field theory [Witt2]. We will now sketch the appropriate adaptation of the formula (39), still working in the finite-dimensional context, and will then move on to $\hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega^2_+(X, \mathrm{ad}\, P)$. There is quite a lot to be verified along the way, but we will be content to illustrate a few of the computations and refer to [AJ] and Section 5 of [Nab2] for more details.

We will begin augmenting the assumptions we have made thus far and making some specific choices of the various ingredients we require for the construction. We will assume that the principal bundle $G \hookrightarrow P \xrightarrow{\pi_P} X$ has bundle space P that is oriented and that the action of G on P preserves the orientation in the sense that each of the diffeomorphisms $\sigma_g : P \to P$ is orientation preserving. Next we recall that for any action of a compact Lie group G on a manifold P it is always possible to choose a Riemannian metric on P that is G-invariant, that is, for which the diffeomorphisms σ_g are all isometries (see Chapter IV, Example 1.4, page 155, of [KN1]). We fix, once and for all, a G-invariant Riemannian metric on P, denoted simply $\langle\, ,\, \rangle$. Similarly, we fix an ad-invariant inner product $(\, ,\,)$ on the Lie algebra $\mathfrak{g}$. This, in turn, determines by translation a bi-invariant Riemannian metric on G which can be normalized so that the volume of G with respect to the corresponding metric volume form is one.

At each $p \in P$ the Riemannian metric $\langle\, ,\, \rangle$ defines an orthogonal complement to the vertical subspace of $T_p(P)$ (the tangent space to the G-orbit through p) and, since G acts by isometries, these orthogonal complements are invariant under the action of G and so determine a connection ω on $G \hookrightarrow P \stackrel{\pi_P}{\to} X$ (see Chapter II, Section 1, page 63, of [KN1]). Henceforth we will use this connection exclusively. In particular, the pullback connection $\pi_1^*\omega$ on $P \times V$, which we continue to denote by the same symbol ω, has

$$\mathrm{Hor}_{(p,v)}(\omega) = T_p(p \cdot G)^\perp \oplus T_v(V).$$

at each point.

Now define, for each $p \in P$, a linear map

$$C_p : \mathfrak{g} \to \mathrm{Vert}_p(P) \subseteq T_p(P)$$

by

$$C_p(\zeta) = \zeta^\#(p) = \frac{d}{dt}\,(p \cdot \exp(t\zeta))\Big|_{t=0}.$$

This is an isomorphism onto $\mathrm{Vert}_p(P)$, but we wish to regard it as a map into $T_p(P)$. Relative to the invariant inner products on $\mathfrak{g}$ and $T_p(P)$, C_p has an adjoint $C_p^* : T_p(P) \to \mathfrak{g}$. The map

$$R_p = C_p^* \circ C_p : \mathfrak{g} \to \mathfrak{g}$$

is self-adjoint with trivial kernel so there is an inverse

$$R_p^{-1} : \mathfrak{g} \to \mathfrak{g}.$$

Moreover, as a map onto $\mathrm{Vert}_p(P)$, C_p has an inverse

$$C_p^{-1} : \mathrm{Vert}_p(P) \to \mathfrak{g}.$$

Note that, on $\mathrm{Vert}_p(P)$, C_p^{-1} agrees with ω_p, that is,

$$C_p^{-1}(w) = \omega_p(w), \quad w \in \mathrm{Vert}_p(P).$$

Indeed, if $w \in \mathrm{Vert}_p(P)$, then $w = \eta^\#(p) = C_p(\eta)$ for a unique $\eta \in \mathfrak{g}$ and, since ω is a connection, $\omega_p(\eta^\#(p)) = \eta$ for every $\eta \in \mathfrak{g}$. Thus, $\omega_p(w) = \eta = C_p^{-1}(w)$. Similarly, one checks that

$$C_p^*(v) = R_p(\omega_p(v)), \quad v \in T_p(P)$$

(see page 76 of [Nab2]) so C^* can be thought of as a $\mathfrak{g}$-valued 1-form on P. Selecting bases we can identify each R_p with an invertible matrix and ω with a matrix of real-valued 1-forms on P so this becomes a matrix equation $C^* = R\omega$ and we will write it as

$$\omega = R^{-1}C^*.$$

From this we compute

$$d\omega = dR^{-1} \wedge C^* + R^{-1}dC^*.$$

Noting that the first term vanishes on horizontal vectors we find that, in the expression (39), we can replace $d\omega$ by $R^{-1}dC^*$ to obtain

$$\tilde{U} = (2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \int \exp\Big(i\psi^T du + \frac{1}{2}\psi^T(\rho_*(R^{-1}dC^*))\,\psi\Big)\mathcal{D}\psi \quad \text{(evaluated on horizontal parts).} \qquad (41)$$

The next objective is to remove the explicit appearance of the inverse in (41) by using the Fourier inversion formula. We will follow the conventions of [RS2]. Thus, we let W be an oriented, real vector with inner product (,) and orthonormal coordinates $w = (w_1, \ldots, w_n)$. If $f(w) = f(w_1, \ldots, w_n)$ is in the Schwartz space $\mathcal{S}(W)$ of rapidly decreasing functions of $w_1, \ldots, w_n$, then the Fourier transform $\mathcal{F}[f] = \hat{f}$ of f is a function of another set of coordinates $y = (y_1, \ldots, y_n)$ in W defined by

$$\mathcal{F}[f](y) = \hat{f}(y) = (2\pi)^{-n/2} \int_W e^{-i\,(w,y)} f(w)\, dw.$$

The Fourier Inversion Formula then asserts that $\hat{f}$ is a Schwartz function of $y_1, \ldots, y_n$ and

$$f(w) = (2\pi)^{-n/2} \int_W e^{i\,(w,y)} \hat{f}(y)\, dy.$$

Combining these two gives

$$f(w) = (2\pi)^{-n} \int_W \int_W e^{i\,(w,y)} e^{-i\,(z,y)} f(z)\, dz\, dy.$$

If R is a self-adjoint matrix with positive determinant, then one can use this formula to compute $f(R^{-1}w)$. To get an integral that does not depend explicitly on R^{-1}, however, we also make the change of variable $y \mapsto Ry$. Then $(R^{-1}w, Ry) = (w, y)$ and $d(Ry) = \det R\, dy$ so

$$f(R^{-1}w) = (2\pi)^{-n} \int_W \int_W e^{i(w,y)} e^{-i(z,Ry)} f(z) \det R\, dz\, dy. \tag{42}$$

Now return to the expression (41) for $\tilde{U}$. We take W to be the Lie algebra $\mathfrak{g}$ with the chosen ad-invariant inner product (,). Letting $\zeta = (\zeta_1, \ldots, \zeta_n)$ denote orthonormal coordinates on $\mathfrak{g}$ we consider the function on $\mathfrak{g}$ defined by

$$f(\zeta) = (2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \int \exp\Big(i\psi^T du + \frac{1}{2}\psi^T (\rho_*(\zeta))\, \psi\Big) \mathcal{D}\psi.$$

The values of f are forms on V whose components relative to $du_1, \ldots, du_{2k}$ are polynomials in $\zeta_1, \ldots, \zeta_n$. These polynomials are not in the Schwartz space $\mathcal{S}(\mathfrak{g})$, but we nevertheless propose to apply the Fourier formula (42) componentwise to $f(\zeta)$. This is rather sloppy, of course, and could be made more precise by inserting a rapidly decaying factor $e^{-\epsilon(\zeta,\zeta)}$ and taking the limit as $\epsilon \to 0$. However, since our objective is a formula to be applied formally to an infinite-dimensional situation in which even k is infinite and complete rigor is (for the time being at least) out of the question anyway, we will not be scrupulous about such matters. Thus, we let $\lambda = (\lambda_1, \ldots, \lambda_n)$ be another Lie algebra variable in $\mathfrak{g}$ and apply (42) with $w = dC^*$, that is, with $w = dC^*(\chi_1, \chi_2)$ for each fixed pair (χ_1, χ_2) of tangent vectors to P. Evaluating this on horizontal parts then gives, by (41), our next expression for $\tilde{U}$. The result is as follows.

$$\tilde{U} = (2\pi)^{-n}(2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \int_{\mathfrak{g}} \int_{\mathfrak{g}} \int \exp\Big(i\psi^T du + \frac{1}{2}\psi^T (\rho_*(\zeta))\psi + i(dC^*, \lambda) - i(\zeta, R\lambda)\Big) \det R\, \mathcal{D}\psi\, d\lambda\, d\zeta$$

$$\text{(evaluated on horizontal parts)} \tag{43}$$

Notice that this expression contains one Berezin ("fermionic") and two ordinary ("bosonic") integrations.

Next we would like to include the parenthetical remark "evaluated on horizontal parts" in (43) directly into the integral expression for $\tilde{U}$. For this we need the notion of a *normalized vertical volume form* for a principal bundle, which is essentially an analogue of the Thom form for a vector bundle. Let $G \hookrightarrow Q \overset{\pi_Q}{\to} M$ be a principal G-bundle with M compact and orientable and $\dim G = n$. We assume that the bundle is orientable in the sense that there exists a smooth n-form Ψ on Q such that, if $m \in M$ and $\iota_m : \pi_Q^{-1}(m) \hookrightarrow Q$

is the inclusion map of the fiber above m into Q, then $\iota_m^*\Psi$ is an orientation for the submanifold $\pi_Q^{-1}(m) \cong G$ of Q. It follows that Q is orientable and one can assume that the action of G on Q is orientation preserving. Then a *normalized vertical volume form* is an n-form Θ on Q such that the integral of Θ over each fiber is one, that is, if $m \in M$ and $\iota_m : \pi_Q^{-1}(m) \hookrightarrow Q$ is the inclusion map, then

$$\int_{\pi_Q^{-1}(m)} \iota_m^*\Theta = 1.$$

If β is any top rank form on M, then

$$\int_M \beta = \int_Q \pi_Q^*\beta \wedge \Theta. \tag{44}$$

One can produce a normalized vertical volume form from a connection ω on the bundle as follows. Assume that $\{\zeta_1, \ldots, \zeta_n\}$ is an orthonormal basis for $\mathfrak{g}$ relative to the normalized, ad-invariant inner product $(\ ,\ )$ and consistent with the orientation G inherits as a fiber in Q. Write $\omega = \omega^j\zeta_j$, where $\omega^j \in \Omega^1(Q)$ for $j = 1, \ldots, n$. Then we claim that

$$\Theta = \omega^1 \wedge \cdots \wedge \omega^n \tag{45}$$

is a normalized vertical volume form on Q. To see this note that, for each $m \in M$,

$$\iota_m^*\Theta = (\iota_m^*\omega^1) \wedge \cdots \wedge (\iota_m^*\omega^n).$$

Now notice that, on $\pi_Q^{-1}(m) \cong G$, the form $\iota_m^*\omega^j$ is dual to the vector field $\zeta_j^\#$ for $j = 1, \ldots, n$. Indeed, a defining property of a connection is that $\omega(A^\#) = A$ for every $A \in \mathfrak{g}$ so, on $\pi_Q^{-1}(m)$, $(\iota_m^*\omega)(\zeta_k^\#) = \zeta_k$ for each $k = 1, \ldots, n$. Thus,

$$\zeta_k = (\iota_m^*\omega)(\zeta_k^\#) = (\iota_m^*\omega^j)(\zeta_k^\#)\,\zeta_j$$

so $(\iota_m^*\omega^j)(\zeta_k^\#) = \delta_{jk} = (\zeta_j^\#, \zeta_k^\#)$. Consequently, the integral of $\iota_m^*\Theta$ over $\pi_Q^{-1}(m)$ agrees with the volume of G with respect to the normalized, bi-invariant Riemannian metric on G corresponding to $(\ ,\ )$ and this is one.

Now, notice that wedging the normalized vertical volume form Θ given by (45) with the horizontal projection $\pi_{\mathrm{Hor}}(\omega)(\varphi)$ given by (38) gives

$$\pi_{\mathrm{Hor}}(\omega)(\varphi) \wedge \Theta = \varphi \wedge \Theta. \tag{46}$$

Integrating the left-hand side over a fiber gives

$$\pi_{\mathrm{Hor}}(\omega)(\varphi) \int_{\pi_Q^{-1}(m)} \iota_m^*\Theta = \pi_{\mathrm{Hor}}(\omega)(\varphi).$$

We conclude that computing the ω-horizontal projection of a form φ on Q can be accomplished by wedging φ with the normalized vertical volume form Θ and then integrating over a fiber.

We would like to apply this to the form $\tilde{U}$ in (43), where the principal bundle is $G \hookrightarrow P \times V \xrightarrow{Q} P \times_\rho V$. However, we would like to include the procedure as part of the integration formula so we first note that Θ can be written as a Berezin integral. For this we let $\{\eta_1, \ldots, \eta_n\}$ denote either the oriented, orthonormal basis $\{\zeta_1, \ldots, \zeta_n\}$ if $n(n-1)/2$ is even or an odd permutation of $\{\zeta_1, \ldots, \zeta_n\}$ if $n(n-1)/2$ is odd. Regard these as odd generators of the exterior algebra $\bigwedge \mathfrak{g}$ and consider the following element of $\Omega^*(Q) \otimes \bigwedge \mathfrak{g}$.

$$e^{\sum_{a=1}^n \omega_a\eta_a} = e^{\omega_1\eta_1} \cdots e^{\omega_n\eta_n} = (1 + \omega_1\eta_1) \cdots (1 + \omega_n\eta_n)$$

Performing the Berezin integration gives

$$\int e^{\sum_{a=1}^{n}\omega_a\eta_a}\mathcal{D}\eta = (-1)^{n(n-1)/2}\int \omega_1\cdots\omega_n\eta_1\cdots\eta_n\mathcal{D}\eta = \omega_1\cdots\omega_n = \omega_1\wedge\cdots\wedge\omega_n = \Theta. \tag{47}$$

Now we would like to express the Berezin integral representation (47) of Θ without explicit reference to the connection forms ω_a. For this we assume, as we did earlier for $G \hookrightarrow P \overset{\pi_P}{\to} X$, that Q has an invariant Riemannian metric and that ω is the connection on Q whose horizontal spaces are the orthogonal complements of the G-orbits. Then we have available the maps C, C^*, and R. Now consider the element of $\mathbb{C}^1[\mathfrak{g}]\,\hat{\otimes}\,\Omega^1(Q)$ whose value at any $\eta\in\mathfrak{g}$ is the 1-form on Q defined by

$$(C^*,\eta)(\chi) = (C^*\chi,\eta) = (\chi, C\eta)$$

for any vector field χ on Q. One can exponentiate (C^*,η) in $\mathbb{C}[\mathfrak{g}]\,\hat{\otimes}\,\Omega^*(Q)$ and compute the Berezin integral of the result. If one writes out everything explicitly and has a little patience one finds that

$$\int e^{(C^*,\eta)}\,\mathcal{D}\eta = \det R\int e^{\sum_{a=1}^{n}\omega_a\eta_a}\,\mathcal{D}\eta = (\det R)\,\Theta$$

(see pages 106-109 of [Nab2]). Thus,

$$\Theta = (\det R)^{-1}\int e^{(C^*,\eta)}\,\mathcal{D}\eta. \tag{48}$$

We can now ensure the horizontal projection in (43) by multiplying through by the normalized vertical volume form Θ of $G \hookrightarrow P\times V \overset{Q}{\to} P\times_\rho V$ as given by (48), thus adding one more term in the exponential and canceling the determinant.

$$\tilde{U} = (2\pi)^{-n}(2\pi)^{-k}e^{-\frac{1}{2}\|u\|^2}\int_{\mathfrak{g}}\int_{\mathfrak{g}}\int\int \exp\Big(i\psi^T du + \frac{1}{2}\psi^T(\rho_*(\zeta))\psi + i(dC^*,\lambda) - i(\zeta,R\lambda) + (C^*,\eta)\Big)\mathcal{D}\eta\,\mathcal{D}\psi\,d\lambda\,d\zeta \tag{49}$$

Notice that, in addition to the two Berezin integrations and the two ordinary integrations, there is an implicit integration over the fibers of $P\times V\to P\times_\rho V$ to "integrate out" the extra factor of Θ in (46); it is customary to leave this understood rather than explicit. Then (49) gives a basic form on $P\times V$ which descends to a representative U of the Thom class on $P\times_\rho V$. Pulling U back by a section of $P\times_\rho V$ gives a representative of the Euler class of $P\times_\rho V$ which, when integrated over X, gives the Euler number. We have already seen that pulling back by such a section amounts to first pulling the V-parts of $\tilde{U}$ back by an equivariant map $S : P\to V$ and then pulling the resulting form on P back by a section s of the principal bundle $\pi_P : P\to X$. The first of these steps simply replaces u by $S^*u = u\circ S$, that is, by the component functions of $S : P\to V$. We will adhere to the usual convention of simply writing this as

$$(2\pi)^{-n}(2\pi)^{-k}e^{-\frac{1}{2}\|S\|^2}\int_{\mathfrak{g}}\int_{\mathfrak{g}}\int\int \exp\Big(i\psi^T dS + \frac{1}{2}\psi^T(\rho_*(\zeta))\psi + i(dC^*,\lambda) - i(\zeta,R\lambda) + (C^*,\eta)\Big)\mathcal{D}\eta\,\mathcal{D}\psi\,d\lambda\,d\zeta \tag{50}$$

Pulling back this form on P by the section $s : X\to P$ of $\pi_P : P\to X$ gives a representative of the Euler class of $P\times_\rho V$ so its integral over X is the Euler number of $P\times_\rho V$.

Remark 1.18. One can look at this calculation of the Euler number from a somewhat different point of view. Notice that if e is any representative of the Euler class of $P \times_\rho V$ and Θ is the normalized vertical volume form for the principal bundle $\pi_P : P \to X$, then (44) gives

$$\int_X e = \int_P \pi_P^* e \wedge \Theta$$

so the Euler number can also be computed as an integral over P.

We have one last bit of cosmetic surgery to perform. There is a common notational device in (supersymmetric) physics whereby the integral of a top rank form over a manifold is written as two successive integrations, one fermionic and one bosonic. As motivation let us first suppose P is a d-dimensional, oriented, real vector space with a positive definite inner product. Let $u_1, \dots, u_d$ be coordinate functions on P corresponding to some oriented, orthonormal basis for P and let $du = du_1 \wedge \cdots \wedge du_d$ be the volume form for P. The integral of any (properly decaying) function φ on P is then defined to be the integral of the d-form $\varphi\, du$ over P.

$$\int_P \varphi\, du$$

Introduce generators $\chi_1, \dots \chi_d$ for the exterior algebra $\bigwedge P$. Now let $\alpha = \alpha(u_j, du_j)$ be a d-form on P written in terms of the coordinates $u_1, \dots, u_d$ and their differentials. Formally substituting $du_j \mapsto \chi_j$ for $j = 1, \dots, d$ we find that the Berezin integral

$$\int \alpha(u_j, \chi_j)\, \mathcal{D}\chi$$

is precisely the function one integrates next to du to get the integral of α over P.

$$\int_P \alpha = \int_P \int \alpha(u_j, \chi_j)\, \mathcal{D}\chi\, du$$

Physicists employ precisely the same notational device when P is a manifold by thinking of $u_1, \dots, u_d$ as local coordinates and $\chi_1, \dots, \chi_d$ as generators for the exterior algebra of the tangent space. Since it *is* just a notational device one does not worry too much about making this precise. Applying this convention to the expression for the Euler number obtained by integrating over P gives the final formula toward which all of this has been leading us. This introduces one more fermionic integration $\mathcal{D}\chi$ and one more bosonic integration du. Since there is now an integration over $u \in P$ we will include the exponential factor $e^{-\frac{1}{2}\|S\|^2}$ in the integrand. Also, since the proliferation of integral signs is becoming rather tiresome we will now suppress all but one of them and leave it to the "differentials" $\mathcal{D}\chi\, \mathcal{D}\eta\, \mathcal{D}\psi\, d\lambda\, d\zeta\, du$ to indicate the integration intended. Finally, we will, for the sake of clarity, indicate explicitly the variables on which each term depends. The result of all this is

$$(2\pi)^{-n}(2\pi)^{-k} \int \exp\Big(-\frac{1}{2}\|S(u)\|^2 + i\psi^T dS_u(\chi) + \frac{1}{2}\psi^T(\rho_*(\zeta))\psi + i(dC_u^*(\chi), \lambda) - i(\zeta, R_u\lambda) + (C_u^*\chi, \eta)\Big) \mathcal{D}\chi \mathcal{D}\eta\, \mathcal{D}\psi\, d\lambda\, d\zeta\, du. \tag{51}$$

We will refer to (51) as the *Atiyah-Jeffrey formula* for the Euler number of $P \times_\rho V$. Our objective is to formally apply it to the infinite-dimensional vector bundle (26) associated with Donaldson theory. The result will be, formally at least, an expression for the "Euler number" of the bundle, that is, the 0-dimensional Donaldson invariant (see page 40) and also, as it so happens, the partition function for Witten's topological quantum field theory. We should emphasize at the outset that what we intend to do is not mathematics (and certainly not physics). We proceed by analogy, isolating formal field-theoretic analogues of the various fermionic and bosonic variables in (51) and natural identifications of the terms in the exponent with functions of these variables. In the process the perfectly well-defined fermionic and bosonic integrals in (51) metamorphose into Feynman integrals over spaces of fields with all of their attendant mathematical difficulties. The purist will argue that this is meaningless manipulation of symbols and we can offer no credible defense against the charge. The only mitigating circumstance is that such formal manipulations have proved extraordinarily productive for both physics and mathematics and so should be viewed with some tolerance.

Throughout the remainder of this section X will denote a compact, connected, simply connected, oriented, smooth 4-manifold with $b_2^+(X) > 1$ and odd, g is a generic Riemannian metric on X and we will use the notation established on pages 39-40. It will be more convenient to deal with matrices rather than quaternions so we will identify Sp(1) with SU(2) and consider a smooth, principal SU(2)-bundle $\mathrm{SU}(2) \hookrightarrow P \xrightarrow{\pi} X$ over X whose Chern number k has been fixed so that the dimension of the moduli space $\hat{\mathcal{M}} = \hat{\mathcal{M}}(P, g)$ of irreducible g-ASD connections is zero. We consider the principal $\hat{\mathcal{G}}$-bundle

$$\hat{\mathcal{G}} \hookrightarrow \hat{\mathcal{A}} \to \hat{\mathcal{A}}/\hat{\mathcal{G}} = \hat{\mathcal{B}},$$

the vector bundle (26)

$$\hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega^2_+(X, \mathrm{ad}\, P)$$

associated to it by the action of $\hat{\mathcal{G}}$ on $\Omega^2_+(X, \mathrm{ad}\, P)$ and the section

$$s_+ : \hat{\mathcal{B}} \to \hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega^2_+(X, \mathrm{ad}\, P),$$

of it determined by the equivariant map $S = F^+ : \hat{\mathcal{A}} \to \Omega^2_+(X, \mathrm{ad}\, P)$, $S(\omega) = F^+(\omega) = F^+_\omega = \frac{1}{2}(F_\omega +^* F_\omega)$, that is,

$$s_+([\omega]) = [\omega, F^+_\omega].$$

The application of the Atiyah-Jeffrey formula (51) requires the existence of a Riemannian metric on the principal bundle space $\hat{\mathcal{A}}$ for which the group $\hat{\mathcal{G}}$ acts by isometries. To produce such a metric we recall that $\hat{\mathcal{A}}$ is an open subset of $\mathcal{A}$ and that $\mathcal{A}$ is an affine space modeled on $\Omega^1(X, \mathrm{ad}\, P)$ so that

$$T_\omega(\hat{\mathcal{A}}) \cong \Omega^1(X, \mathrm{ad}\, P)$$

for each $\omega \in \hat{\mathcal{A}}$. Now, each of the spaces $\Omega^p(X, \mathrm{ad}\, P)$ has a natural L^2-inner product arising from the metric g on X, the corresponding Hodge star operator * and an ad-invariant inner product $(\ ,\ )$ on the Lie algebra $\mathfrak{su}(2)$. Taking $(A, B) = -\mathrm{tr}\,(AB)$ for all $A, B \in \mathfrak{su}(2)$, this is given by

$$\langle \alpha, \beta \rangle_p = -\int_X \mathrm{tr}\,(\alpha \wedge^* \beta).$$

In particular, this is true for $T_\omega(\hat{\mathcal{A}})$ and this defines a metric on $\hat{\mathcal{A}}$. Since $(\ ,\ )$ is ad-invariant, $\langle\ ,\ \rangle$ is invariant under the action of $\hat{\mathcal{G}}$ (pointwise conjugation by elements of $P \times_{\mathrm{Ad}} \mathrm{SU}(2)$), so $\hat{\mathcal{G}}$ acts by isometries on $\hat{\mathcal{A}}$. This metric, in turn, determines a connection on $\hat{\mathcal{G}} \hookrightarrow \hat{\mathcal{A}} \to \hat{\mathcal{A}}/\hat{\mathcal{G}} = \hat{\mathcal{B}}$ whose horizontal spaces are the orthogonal complements to the gauge orbits. Using the Hodge Decomposition Theorem for elliptic complexes one can show that, for every $\omega \in \hat{\mathcal{A}}$,

$$T_\omega(\hat{\mathcal{A}}) \cong T_\omega(\omega \cdot \hat{\mathcal{G}}) \oplus \text{Kernel}\,(\delta^\omega) = \text{Image}\,(d^\omega) \oplus \text{Kernel}\,(\delta^\omega),$$

where $d^\omega : \Omega^0(X, \text{ad}\,P) \to \Omega^1(X, \text{ad}\,P)$ is the covariant exterior derivative and $\delta^\omega : \Omega^1(X, \text{ad}\,P) \to \Omega^0(X, \text{ad}\,P)$ is its formal adjoint with respect to the inner products we have just introduced (see pages 73-74 of [Nab2]).

Now we will begin to determine the appropriate analogue for each of the terms in the exponent of the Atiyah-Jeffrey formula (51). First consider $-\frac{1}{2}\|S(u)\|^2$. This is a function defined on $\hat{\mathcal{A}}$ so we will write ω rather than u. The norm is computed from the inner product $\langle\, ,\, \rangle_2$ on $\Omega^2(X, \text{ad}\,P)$ introduced above. Thus,

$$-\frac{1}{2}\|S(\omega)\|^2 = -\frac{1}{2}\|F_\omega^+\|^2 = \frac{1}{2}\int_X \text{tr}\,(F_\omega^+ \wedge {}^*F_\omega^+) = \frac{1}{2}\int_X \text{tr}\,(F_\omega^+ \wedge F_\omega^+)$$

Using the orthogonality of the Hodge decomposition one finds that this can be written

$$-\frac{1}{2}\|S(\omega)\|^2 = \frac{1}{4}\int_X \text{tr}\,(F_\omega \wedge {}^*F_\omega) + \frac{1}{4}\int_X \text{tr}\,(F_\omega \wedge F_\omega) \tag{52}$$

The first term is of the typical Yang-Mills variety for a classical gauge theory, whereas the second term Witten [Witt2] calls a *topological term* since it is, up to a constant, the Chern number of the underlying SU(2)-bundle.

Remark 1.19. Witten [Witt2] employs the notation more common in physics whereby everything is written in such a way as to appear local. We will not attempt to translate all that we do into this language, but will illustrate with (52). Let $\{T_a\}$ be an orthonormal basis for $\mathfrak{su}(2)$ relative to $(A, B) = -\text{tr}\,(AB)$. Write $F_\omega = \frac{1}{2}F_{\alpha\beta}dx^\alpha \wedge dx^\beta$, where $F_{\alpha\beta} = F^a_{\alpha\beta}T_a$ and ${}^*F_\omega = \frac{1}{2}\tilde{F}_{\alpha\beta}dx^\alpha \wedge dx^\beta$, where $\tilde{F}_{\alpha\beta} = \tilde{F}^a_{\alpha\beta}T_a$. Raise indices with the metric g to get $F^{\alpha\beta} = g^{\alpha\alpha'}g^{\beta\beta'}F_{\alpha'\beta'}$ and $\tilde{F}^{\alpha\beta} = g^{\alpha\alpha'}g^{\beta\beta'}\tilde{F}_{\alpha'\beta'}$. Writing the volume element of g as $d^4x\,\sqrt{g}$ one finds that $\frac{1}{4}\text{tr}\,(F_\omega \wedge {}^*F_\omega) = \frac{1}{4}\text{tr}\,(F_{\alpha\beta}F^{\alpha\beta})\,\sqrt{g}\,d^4x$ and $\frac{1}{4}\text{tr}\,(F_\omega \wedge F_\omega) = \frac{1}{4}\text{tr}\,(F_{\alpha\beta}\tilde{F}^{\alpha\beta})\,\sqrt{g}\,d^4x$. Thus,

$$-\frac{1}{2}\|S(\omega)\|^2 = \int_X d^4x\,\sqrt{g}\,\text{tr}\left(\frac{1}{4}F_{\alpha\beta}F^{\alpha\beta} + \frac{1}{4}F_{\alpha\beta}\tilde{F}^{\alpha\beta}\right)$$

These are the first two terms in the Witten's TQFT Lagrangian (see (2.13) and (2.41) of [Witt2]).

To proceed we must sort out the appropriate analogues, in the Donaldson theory context, of the maps C, C^*, and R. At each point ω in the principal bundle space $\hat{\mathcal{A}}$, C_ω is the map from the Lie algebra of $\hat{\mathcal{G}}$, which can be identified with $\Omega^0(X, \text{ad}\,P)$, to the tangent space $T_\omega(\hat{\mathcal{A}}) \cong \Omega^1(X, \text{ad}\,P)$, defined, for each $\zeta \in \Omega^0(X, \text{ad}\,P)$ by

$$C_\omega(\zeta) = \frac{d}{dt}\,(\omega \cdot \exp\,(t\zeta))\,\Big|_{t=0}.$$

Computing this locally by expanding the exponential gives the covariant exterior derivative of ζ with respect to ω, that is,

$$C_\omega = d^\omega : \Omega^0(X, \text{ad}\,P) \to \Omega^1(X, \text{ad}\,P). \tag{53}$$

Consequently, C^*_ω is the formal adjoint

$$C^*_\omega = \delta^\omega : \Omega^1(X, \mathrm{ad}\, P) \to \Omega^0(X, \mathrm{ad}\, P) \tag{54}$$

of d^ω relative to the inner products on the spaces of $\mathrm{ad}\, P$-valued forms introduced above and

$$R_\omega = C^*_\omega \circ C_\omega = \Delta^0_\omega : \Omega^0(X, \mathrm{ad}\, P) \to \Omega^0(X, \mathrm{ad}\, P) \tag{55}$$

is the scalar Laplacian corresponding to ω.

With this information in hand we consider the term $-i(\zeta, R_\omega\lambda)$ in (51). Both ζ and λ are in the Lie algebra of the gauge group so we introduce two "bosonic" fields

$$\zeta, \lambda \in \Omega^0(X, \mathrm{ad}\, P)$$

and interpret $(\ ,\ )$ as the natural inner product $\langle\ ,\ \rangle_0$ on $\Omega^0(X, \mathrm{ad}\, P)$.

Remark 1.20. We apply the adjectives "bosonic" and "fermionic" to the fields we introduce only because of the type of integral these variables correspond to in (51). The terms have physical significance in QFT, but we do not claim to have justified any such physical connotations.

Thus, the term $-i(\zeta, R_\omega\lambda)$ is interpreted as

$$-i(\zeta, R_\omega\lambda) = -i\langle\zeta, \Delta^0_\omega\lambda\rangle_0 = i\int_X \mathrm{tr}\,(\zeta \wedge {}^*(\Delta^0_\omega\lambda)) = i\int_X \mathrm{tr}\,({}^*(\zeta\Delta^0_\omega\lambda)). \tag{56}$$

Remark 1.21. Atiyah and Jeffrey [AJ] point out that, to conform with Witten's notation in [Witt2], ζ and λ must be replaced by $i\zeta$ and $\frac{1}{2}\lambda$, respectively. In the notation used in physics the corresponding term in [Witt2] is

$$\int_X d^4x\ \sqrt{g}\,\mathrm{tr}\Big(\frac{1}{2}\zeta\, D_\alpha D^\alpha\lambda\Big).$$

Next we consider the term $(C^*_\omega\chi, \eta)$ in (51). Since C^*_ω maps $\Omega^1(X, \mathrm{ad}\, P)$ to $\Omega^0(X, \mathrm{ad}\, P)$ we will need two fermionic fields

$$\eta \in \Omega^0(X, \mathrm{ad}\, P) \quad \text{and} \quad \chi \in \Omega^1(X, \mathrm{ad}\, P)$$

and, as above, $(\ ,\ ) = \langle\ ,\ \rangle_0$. Thus, we find that

$$(C^*_\omega\chi, \eta) = \langle\delta^\omega\chi, \eta\rangle_0 = \langle\chi, d^\omega\eta\rangle_1 = -\int_X \mathrm{tr}\,(\chi \wedge {}^*d^\omega\eta). \tag{57}$$

The fermionic variable ψ in the Mathai-Quillen formalism arises from the odd generators of the exterior algebra of the fiber V in the vector bundle. In our case this fiber is $\Omega^2_+(X, \mathrm{ad}\, P)$ so we introduce a fermionic field

$$\psi \in \Omega^2_+(X, \mathrm{ad}\, P).$$

Now we consider the term $i\psi^T dS_\omega(\chi)$ in (51). $S = F^+ : \hat{\mathcal{A}} \to \Omega^2_+(X, \mathrm{ad}\, P)$ is the self-dual curvature map. The derivative of this map at $\omega \in \hat{\mathcal{A}}$ can be identified with the self-dual projection of the covariant exterior derivative associated with ω which we will denote

$$d^\omega_+ : \Omega^1(X, \mathrm{ad}\, P) \to \Omega^2_+(X, \mathrm{ad}\, P).$$

Thus, $dS_\omega(\chi) = d^\omega_+ \chi$. We will interpret finite-dimensional expressions such as

$$A^T B = (A^1 \cdots A^r) \begin{pmatrix} B^1 \\ \vdots \\ B^r \end{pmatrix} = A^1 B^1 + \cdots A^r B^r$$

in terms of the appropriate field-theoretic inner products. Thus, we obtain, using the self-duality of ψ and the orthogonality of the Hodge decomposition,

$$i\psi^T dS_\omega(\chi) = i\langle \psi, d^\omega_+ \chi\rangle_2 = i\langle \psi, d^\omega \chi\rangle_2 = i\langle d^\omega \chi, \psi\rangle_2 = -i \int_X \mathrm{tr}\,(d^\omega \chi \wedge \psi). \tag{58}$$

Next consider the term $\frac{1}{2}\psi^T(\rho_*(\zeta))\psi$. We know that $\zeta \in \Omega^0(X, \mathrm{ad}\, P)$ and $\psi \in \Omega^2_+(X, \mathrm{ad}\, P)$. In the Mathai-Quillen formalism, ρ corresponds to the action of G on V that gives rise to the associated vector bundle. In our case, $\hat{\mathcal{G}}$, regarded as sections of the nonlinear adjoint bundle, acts on $\Omega^2_+(X, \mathrm{ad}\, P)$ pointwise by conjugation. The infinitesimal action is therefore bracket

$$\rho_*(\zeta)\psi = [\zeta, \psi]$$

so $\frac{1}{2}\psi^T(\rho_*(\zeta))\psi$ is interpreted as the inner product

$$\frac{1}{2}\psi^T(\rho_*(\zeta))\psi = \frac{1}{2}\langle \psi, [\zeta, \psi]\rangle_2 = \frac{1}{2}\langle [\zeta, \psi], \psi\rangle_2 = -\frac{1}{2}\int_X \mathrm{tr}\,([\zeta, \psi] \wedge \psi).$$

Now, any ad-invariant inner product (,) on any Lie algebra satisfies $(x, [y, z]) = ([x, y], z)$ so we can write this term as

$$\frac{1}{2}\psi^T(\rho_*(\zeta))\psi = -\frac{1}{2}\int_X \mathrm{tr}\,(\,\zeta\,[\psi, \psi]\,) \tag{59}$$

The only remaining term is $i\,(dC^*_\omega(\chi), \lambda)$ and this requires a bit more work. C^* is a 1-form on $\hat{\mathcal{A}}$ with values in $\Omega^0(X, \mathrm{ad}\, P)$. We compute dC^* at $\omega \in \hat{\mathcal{A}}$ as follows. Fix $\chi_1, \chi_2 \in T_\omega\hat{\mathcal{A}} \cong \Omega^1(X, \mathrm{ad}\, P)$. Since $\mathcal{A}$ is an affine space and $\hat{\mathcal{A}}$ is open in $\mathcal{A}$, we can regard χ_1 and χ_2 as constant vector fields on $\hat{\mathcal{A}}$. Thus,

$$dC^*_\omega(\chi_1, \chi_2) = \chi_1(C^*\chi_2) - \chi_2(C^*\chi_1) - C^*([\chi_1, \chi_2]) = \chi_1(C^*\chi_2) - \chi_2(C^*\chi_1),$$

where $C^*\chi_j$ is the function $\theta \mapsto C^*_\theta\chi_j$ on $\hat{\mathcal{A}}$ for $j = 1, 2$ and $[\chi_1, \chi_2] = 0$ because the vector fields are constant. Now,

$$(\chi_1(C^*\chi_2))(\omega) = \chi_1(\omega)(C^*\chi_2) = \chi_1(C^*\chi_2) = \frac{d}{dt}\, C^*_{\omega+t\chi_1}(\chi_2)\Big|_{t=0} = \frac{d}{dt}\, \delta^{\omega+t\chi_1}(\chi_2)\Big|_{t=0}$$

Notice that, for any $\lambda \in \Omega^0(X, \mathrm{ad}\, P)$,

$$d^{\omega+t\chi_1}(\lambda) = d^\omega\lambda + t\,[\chi_1, \lambda] = d^\omega\lambda + tB_{\chi_1}(\lambda)$$

where $B_{\chi_1} : \Omega^0(X, \mathrm{ad}\, P) \to \Omega^1(X, \mathrm{ad}\, P)$ is given by $B_{\chi_1}(\lambda) = [\chi_1, \lambda]$. Thus,

$$\delta^{\omega+t\chi_1}(\chi_2) = \delta^\omega\chi_2 + tB^*_{\chi_1}(\chi_2),$$

where $B^*_{\chi_1} : \Omega^1(X, \operatorname{ad} P) \to \Omega^0(X, \operatorname{ad} P)$ is the adjoint of B_{χ_1}. We claim that

$$B^*_{\chi_1}(\chi_2) = -^*[\chi_1, {}^*\chi_2].$$

Indeed, for any $\lambda \in \Omega^0(X, \operatorname{ad} P)$,

$$\begin{aligned} \langle B_{\chi_1}(\lambda), \chi_2 \rangle_1 = \langle [\chi_1, \lambda], \chi_2 \rangle_1 = -\int_X \operatorname{tr}([\chi_1, \lambda] \wedge {}^*\chi_2) &= \int_X \operatorname{tr}([\lambda, \chi_1] \wedge {}^*\chi_2) \\ &= \int_X \operatorname{tr}(\lambda \wedge [\chi_1, {}^*\chi_2]) = \int_X \operatorname{tr}(\lambda \wedge {}^{**}[\chi_1, {}^*\chi_2]) = -\langle \lambda, {}^*[\chi_1, {}^*\chi_2] \rangle_0 \\ &= \langle \lambda, -^*[\chi_1, {}^*\chi_2] \rangle_0 \end{aligned}$$

which shows that $-^*[\chi_1, {}^*\chi_2] = B^*_{\chi_1}(\chi_2)$. Thus,

$$\delta^{\omega + t\chi_1}(\chi_2) = \delta^\omega \chi_2 - t\,^*[\chi_1, {}^*\chi_2]$$

so computing the t-derivative at $t = 0$ gives

$$(\chi_1(C^*\chi_2))(\omega) = -^*[\chi_1, {}^*\chi_2]$$

Interchanging χ_1 and χ_2 gives

$$(\chi_2(C^*\chi_1))(\omega) = -^*[\chi_2, {}^*\chi_1] = {}^*[\chi_1, {}^*\chi_2]$$

Notice that these are independent of ω so $\chi_1(C^*\chi_2) = -^*[\chi_1, {}^*\chi_2]$ and $\chi_2(C^*\chi_1) = {}^*[\chi_1, {}^*\chi_2]$. We conclude that

$$dC^*_\omega(\chi_1, \chi_2) = -2\,^*[\chi_1, {}^*\chi_2]$$

for any $\omega \in \hat{\mathcal{A}}$. Putting all of this together we find that

$$i\,(dC^*_\omega(\chi_1, \chi_2), \lambda) = 2i \int_X \operatorname{tr}(\lambda\,[\chi_1, {}^*\chi_2]). \tag{60}$$

Using the metric on P we identify vector fields with 1-forms and these, in turn, with elements of the exterior algebra and thereby arrive at the following identification of the final term in (51).

$$i\,(dC^*_\omega(\chi), \lambda) = 2i \int_X \operatorname{tr}(\lambda\,[\chi, {}^*\chi]) \tag{61}$$

With this we have identified all of the terms in the exponent of (51). Each is the integral over X of a trace so we can collect them altogether into

$$\int_X \operatorname{tr}\Big(\frac{1}{4} F_\omega \wedge {}^*F_\omega + \frac{1}{4} F_\omega \wedge F_\omega - \frac{1}{2}\zeta\,[\psi, \psi] - i\,d^\omega\chi \wedge \psi + 2i\,\lambda\,[\chi, {}^*\chi] + i\,^*(\zeta\,\Delta^0_\omega \lambda) - \chi \wedge {}^*d^\omega \eta \Big) \tag{62}$$

Now we will introduce some of the terminology used in physics. Each fixed choice of the three bosonic $(\omega \in \hat{\mathcal{A}}, \zeta, \lambda \in \Omega^0(X, \operatorname{ad} P))$ and three fermionic $(\eta \in \Omega^0(X, \operatorname{ad} P), \chi \in \Omega^1(X, \operatorname{ad} P), \psi \in \Omega^2_+(X, \operatorname{ad} P))$ fields will be called a *field configuration* and will be denoted

$$\Phi = (\chi, \eta, \psi, \lambda, \zeta, \omega).$$

We will also write

$$\mathcal{D}\Phi = \mathcal{D}\chi\,\mathcal{D}\chi\,\mathcal{D}\psi\,d\lambda\,d\zeta\,d\omega$$

Changing all of the signs in (62) we define the *Donaldson-Witten action functional* $S_{DW}[\Phi]$ on the space of field configurations by

$$S_{DW}[\Phi] = \int_X \operatorname{tr}\Big(-\frac{1}{4}F_\omega \wedge {}^*F_\omega - \frac{1}{4}F_\omega \wedge F_\omega + \frac{1}{2}\zeta\,[\psi,\psi] + i\,d^\omega\chi \wedge \psi - 2i\,\lambda\,[\chi, {}^*\chi] - i\,{}^*(\zeta\,\Delta^0_\omega\lambda) + \chi \wedge {}^*d^\omega\eta \Big)$$

Now, the integral in (51) can be written

$$\int e^{-S_{DW}[\Phi]}\,\mathcal{D}\Phi. \tag{63}$$

Whereas $S_{DW}[\Phi]$ is a perfectly well-defined mathematical object, (63) can only be thought of as a Feynman integral over the space of field configurations.

We will conclude this section with a few remarks. Notice that in (63) we have omitted any constants that might correspond to $(2\pi)^{-n}(2\pi)^{-k}$ in (51) since both n and k are infinite in our present circumstances. Physicists generally include in the exponent in (63), or directly in the action $S_{DW}[\Phi]$, a factor of the form $1/e^2$, where e is a so-called *coupling constant*. Mathematically, one can view the inclusion of such a factor in $S_{DW}[\Phi]$ as simply a different choice of invariant inner product on the Lie algebra $\mathfrak{su}(2)$, which is determined only up to a positive constant multiple. Classically, one can rescale and give this factor any convenient value. However, upon quantization with the Feynman integral (63) the different values of the coupling constant give rise to an entire 1-parameter family of quantum field theories and the computability of the theory generally, but not always depends on this value. Since this will be relevant to some other comments we would like to make we will record the following alternative to (63).

$$Z_{DW} = \int e^{-S_{DW}[\Phi]/e^2}\,\mathcal{D}\Phi \tag{64}$$

The integral (64) represents the *partition function* of the quantum field theory constructed by Witten in [Witt2]. It is (formally) a function of the coupling constant e. Of course, Witten arrived at the action, and therefore the partition function, by quite a different route than the one we have followed. We began with the 0-dimensional Donaldson invariant, regarded it as an "Euler number" and massaged the Mathai-Quillen integral representation of the Euler number until it could be formally applied in the infinite-dimensional context of Donaldson theory. Witten's arguments in [Witt2] were physical, but the objective was to describe a quantum field theory in which the Donaldson invariants appeared as expectation values of certain observables and, in particular, the 0-dimensional invariant appeared as the partition function (the observables whose expectation values give rise to other Donaldson invariants are discussed in Section 3 of [Witt2]). He chose the field content Φ and the action in order to ensure that $S_{DW}[\Phi]$ exhibited certain symmetries which, together with various formal manipulations of Feynman integrals, implied that the theory was independent of both the Riemannian metric g and the coupling constant e in the sense that the infinitesimal variation of the partition function with respect to either g or e is zero.

The g-independence has led physicists to refer to such field theories as *topological quantum field theories* and to the expectation values we have mentioned as *topological invariants*. One should be aware, however, that this conflicts rather strikingly with the terminology that one would encounter in mathematics. Indeed, we have seen in Section 1.4 that the Donaldson invariants are only independent of a *generic* choice of Riemannian metric g and even this is true only if $b_2^+(X) > 1$. Moreover, even granting this, the Donaldson invariants are invariants, not of the topology of X, but rather of its smooth structure and this is quite a different thing.

The e-independence of the theory is particularly significant since one is then free to compute in either the *weak coupling* ($e \to 0$) or *strong coupling* ($e \to \infty$) regime. For small values of e physicists employ various approximation techniques such as the semi-classical approximation which, again because of the symmetries of the action $S_{DW}[\Phi]$, Witten was able to show (formally) must be exact. This fortuitous circumstance is an analogue of a well-known finite-dimensional theorem on the exactness of the stationary phase approximation due to Duistermaat and Heckman [DH]. This result is most properly understood within the context of equivariant cohomology and the localization of certain integrals of equivariant differential forms to the fixed point set of the group action (see Section 7.2 of [BGV] or Section 6 of [Nab3]). For the particular case we have under consideration, the partition function, being gauge invariant, descends to the moduli space of field configurations and localizes to a sum over the 0-dimensional moduli space of ASD connections thus yielding the 0-dimensional Donaldson invariant.

Remark 1.22. This localization is, intuitively, not unlike the Residue Theorem which localizes a contour integral around a closed path to a sum of contributions from the singularities of the integrand inside (although there are no symmetry considerations in this case). Much closer analogues are the Equivariant Localization Theorems of Berline and Vergne which extract and generalize the essential content of the Duistermaat-Heckman Theorem (see Section 7.2 of [BGV] or Section 6 of [Nab3]).

There is only one question remaining. Witten found the Donaldson invariants by studying his TQFT in the weak coupling regime ($e \to 0$) by performing a semi-classical approximation and arguing that it must be exact. In the strong coupling regime such arguments are unavailable and the theory can be expected to look quite different, but, because of the e-independence, it is physically equivalent and so, one might hope, mathematically equivalent in some sense as well. What can be learned about Donaldson theory in the $e \to \infty$ limit of Witten's TQFT? This is the question to which we turn in the next section.

1.6. **The Seiberg-Witten Invariants for Spin^c-Manifolds.** Remarkable as it is that the subtle differential-topological invariants of Donaldson can be recast in the quantum field-theoretic terms described in the previous section, the most remarkable part of the story is yet to come. Witten unearthed the Donaldson invariants in 1988 by analyzing his TQFT perturbatively in the weak coupling regime. He was, of course, well-aware of the fact that the e-independence of the theory suggested the possibility of an alternative description of the invariants that lay buried in an analysis of the strong coupling regime. However, in 1988, there were no tools available for studying the strongly coupled, nonperturbative behavior of quantum field theories. This all changed in 1994 when Seiberg and Witten [SW] developed techniques for performing exact calculations in strongly coupled $N = 2$ supersymmetric Yang-Mills theories, of which Witten's 1988 TQFT was an example. How they managed to do this is a story that lies in very deep waters indeed and, alas, we cannot presume to offer even an exegesis (one might consult Section 5 of [Witt3], or Lectures 17, 18, and 19, pages 1351-1401, of [Delig2]). The end result, however, when applied to Witten's TQFT gave rise to a new set of 4-manifold invariants which, from the perspective of physics at least, were "dual" to the Donaldson invariants and so, conjectured Witten, must contain the same topological information. These new invariants arose in much the same way as the Donaldson invariants from a moduli space of solutions to certain partial differential equations. These *Seiberg-Witten equations*, however, were vastly simpler than the anti-self-dual equations so Witten's conjecture was received with considerable skepticism. Very soon, however, it became clear that the skepticism was unwarranted and that the new and much simpler *Seiberg-Witten invariants* provided a much more felicitous tool for the study of smooth 4-manifolds. The story of how the Witten conjecture was sprung on the mathematical community and the pandemonium that ensued has been told many times, but is best heard from one who was there. For this as well as a lovely introduction to what is to come and a very pleasant afternoon's entertainment we heartily recommend [Taub3]. We will now simply get on with the business of describing the equations and the invariants (references for this material are [Mor1], [Nicol], [Sal], Appendix A of [Nab5], and Section 7 of [Nab3]).

1.6.1. *Real and Complex Clifford Algebras of* $\mathbb{R}^4$. The price one must pay for the eventual simplicity of the Seiberg-Witten invariants is an initial expenditure of time and energy to assemble the rather considerable algebraic machinery required to simply write down the relevant equations. This is most conveniently phrased in the language of Clifford algebras so we will begin with a brief synopsis of just what we need and no more (the standard reference for Clifford algebras and spin geometry is [LM]).

Any finite-dimensional, real vector space V with an inner product $\langle\ ,\ \rangle$ has a *Clifford algebra* $Cl(V)$ that can be realized concretely as follows. If $\{e_1, \ldots, e_n\}$ is any orthonormal basis for V, then $Cl(V)$ is the real, associative algebra with unit 1 generated by $\{e_1, \ldots, e_n\}$ and subject to the relations

$$e_i e_j + e_j e_i = -2\langle e_i, e_j\rangle\, 1, \quad i, j = 1, \ldots, n. \tag{65}$$

Since we will need only the Clifford algebra $Cl(4) = Cl(\mathbb{R}^4)$ of $\mathbb{R}^4$ with its usual positive definite inner product, we will construct an explicit matrix model. The procedure will be to identify $\mathbb{R}^4$ with a real linear subspace of a matrix algebra, find an orthonormal basis for this copy of $\mathbb{R}^4$ satisfying the defining relations (65), where the product is matrix multiplication and 1 is the identity matrix, and then form the subalgebra it generates. First, identify $\mathbb{R}^4$ with the algebra $\mathbb{H}$ of quaternions $q = q^1 + q^2\mathbf{i} + q^3\mathbf{j} + q^4\mathbf{k}$. Next we embed $\mathbb{H}$ in the real, associative algebra $\mathbb{H}^{2\times 2}$ of 2×2 quaternionic matrices with the product given by matrix multiplication.

$$\mathbb{H}^{2\times 2} = \left\{ \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} : q_{ij} \in \mathbb{H}, \quad i, j = 1, 2 \right\}$$

Specifically, we will identify $\mathbb{H}$ with the real, linear subspace of $\mathbb{H}^{2\times 2}$ consisting of all elements of the form

$$x = \begin{pmatrix} 0 & q \\ -\bar{q} & 0 \end{pmatrix}, \quad q \in \mathbb{H}. \tag{66}$$

Notice that $\det x = \|q\|^2$ so defining a norm on the set of all such x by $\|x\|^2 = \det x$ and an inner product by polarization $\langle x, y\rangle = \frac{1}{4}(\,\|x+y\|^2 - \|x-y\|^2\,)$, we find that the real linear subspace of $\mathbb{H}^{2\times 2}$ consisting of all x of the form (66) is isomorphic to $\mathbb{R}^4$ as an inner product space. One easily checks that

$$e_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & \mathbf{i} \\ \mathbf{i} & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & \mathbf{j} \\ \mathbf{j} & 0 \end{pmatrix}, \quad e_4 = \begin{pmatrix} 0 & \mathbf{k} \\ \mathbf{k} & 0 \end{pmatrix} \tag{67}$$

form an orthonormal basis and satisfy (65). The real subalgebra of $\mathbb{H}^{2\times 2}$ generated by $\{e_1, e_2, e_3, e_4\}$ is the *real Clifford algebra* of $\mathbb{R}^4$ and will be denoted $Cl(4)$. Writing out the matrix products of these basis elements and using (65) to eliminate linear dependencies gives the following basis for $Cl(4)$.

$$
\begin{aligned}
e_0 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1, \\
e_1 &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & \mathbf{i} \\ \mathbf{i} & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & \mathbf{j} \\ \mathbf{j} & 0 \end{pmatrix}, \quad e_4 = \begin{pmatrix} 0 & \mathbf{k} \\ \mathbf{k} & 0 \end{pmatrix}, \\
e_1e_2 &= \begin{pmatrix} \mathbf{i} & 0 \\ 0 & -\mathbf{i} \end{pmatrix}, \quad e_1e_3 = \begin{pmatrix} \mathbf{j} & 0 \\ 0 & -\mathbf{j} \end{pmatrix}, \quad e_1e_4 = \begin{pmatrix} \mathbf{k} & 0 \\ 0 & -\mathbf{k} \end{pmatrix}, \\
e_2e_3 &= \begin{pmatrix} \mathbf{k} & 0 \\ 0 & \mathbf{k} \end{pmatrix}, \quad e_2e_4 = \begin{pmatrix} -\mathbf{j} & 0 \\ 0 & -\mathbf{j} \end{pmatrix}, \quad e_3e_4 = \begin{pmatrix} \mathbf{i} & 0 \\ 0 & \mathbf{i} \end{pmatrix}, \\
e_1e_2e_3 &= \begin{pmatrix} 0 & \mathbf{k} \\ -\mathbf{k} & 0 \end{pmatrix}, \quad e_1e_2e_4 = \begin{pmatrix} 0 & -\mathbf{j} \\ \mathbf{j} & 0 \end{pmatrix}, \\
e_1e_3e_4 &= \begin{pmatrix} 0 & \mathbf{i} \\ -\mathbf{i} & 0 \end{pmatrix}, \quad e_2e_3e_4 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \\
e_1e_2e_3e_4 &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}
\end{aligned} \tag{68}
$$

Thus, $\dim\, Cl(4) = 16$. Since the real dimension of $\mathbb{H}^{2\times 2}$ is also 16, we conclude that

$$Cl(4) = \mathbb{H}^{2\times 2}.$$

Notice that this basis provides $Cl(4)$ with a natural $\mathbb{Z}_2$-grading

$$Cl(4) = Cl_{\mathbf{0}}(4) \oplus Cl_{\mathbf{1}}(4),$$

where $Cl_{\mathbf{0}}(4)$ is spanned by $\{e_0, e_1e_2, e_1e_3, e_1e_4, e_2e_3, e_2e_4, e_1e_2e_3e_4\}$ and $Cl_{\mathbf{1}}(4)$ is spanned by $\{e_1, e_2, e_3, e_4, e_1e_2e_3, e_1e_2e_4, e_1e_3e_4, e_2e_3e_4\}$. The elements of $Cl_{\mathbf{0}}(4)$ are said to be *even*, while those of $Cl_{\mathbf{1}}(4)$ are *odd*. The decomposition corresponds to splitting an element of $Cl(4)$ into its diagonal and anti-diagonal parts. Since

$$(Cl_{\mathbf{i}}(4))\,(Cl_{\mathbf{j}}(4)) \subseteq Cl_{\mathbf{i}+\mathbf{j}}(4), \quad \mathbf{i}, \mathbf{j} \in \mathbb{Z}_2,$$

$Cl(4)$ is a $\mathbb{Z}_2$-graded algebra, that is, a superalgebra.

Lemma 1.21. *The center $Z(Cl(4))$ of $Cl(4)$ is Span $\{e_0\} \cong \mathbb{R}$.*

Proof. Span $\{e_0\} \subseteq Z(Cl(4))$ is clear. To complete the proof it will suffice to show that every

$$e_I = e_{i_1} \cdots e_{i_k}, \quad 1 \le k \le 4, \quad 1 \le i_1 < \cdots < i_k \le 4$$

fails to commute with something in $Cl(4)$. For $k = 1$ this is clear since $e_ie_j = -e_je_i$ for $i \ne j$. For $k = 4$, $e_I = e_1e_2e_3e_4$ so $e_1e_I = (e_1e_1)e_2e_3e_4 = -e_2e_3e_4$, whereas $e_Ie_1 = (-1)^3(e_1e_1)e_2e_3e_4 = e_2e_3e_4$. Now suppose $1 < k < 4$. Then $e_Ie_{i_1} = (-1)^{k-1}e_{i_1}e_I$, but if e_l is not among $e_{i_1}, \ldots, e_{i_k}$, then $e_Ie_l = (-1)^k e_le_I$ so e_I cannot commute with both e_{i_1} and e_l. □

Notice that if $x, y \in \mathbb{R}^4 \subseteq Cl(4)$, then it follows from (65) and the bilinearity of the multiplication in $Cl(4)$ that

$$xy + yx = -2\langle x, y\rangle\, 1.$$

If $x \in \mathbb{R}^4 \subseteq Cl(4)$ and $\|x\| = 1$, then $\langle x, x\rangle = 1$ and $xx + xx = -2\langle x, x\rangle\, 1$ imply $xx = -1$ so x is invertible in $Cl(4)$ and $x^{-1} = -x$. We will denote by $Cl^{\times}(4)$ the multiplicative group of units (invertible elements) in $Cl(4)$ and by Pin(4) the subgroup of $Cl^{\times}(4)$ generated by the set of all $x \in \mathbb{R}^4$ with $\|x\| = 1$, that is, by the unit sphere $\mathbb{S}^3 \subseteq \mathbb{R}^4 \subseteq Cl(4)$. Now, notice that an x of the form (66) has $\|x\| = 1$ if and only if q is a unit quaternion, that is, if and only if $q \in \mathrm{Sp}(1)$ and we have just seen that the set of all such x is closed under inversion ($x^{-1} = -x$). Consequently, Pin(4) is just the set of products of such elements. The even elements of Pin(4) are just its diagonal elements and these form a subgroup denoted

$$\mathrm{Spin}(4) = \mathrm{Pin}(4) \cap Cl_{\mathbf{0}}(4) = \left\{ \begin{pmatrix} u_1 & 0 \\ 0 & u_2 \end{pmatrix} : u_1, u_2 \in \mathrm{Sp}(1) \right\} \cong \mathrm{Sp}(1) \times \mathrm{Sp}(1). \tag{69}$$

The topology and differentiable structures that Spin (4) inherits from $\mathbb{H}^{2\times 2} \cong \mathbb{H}^4 \cong \mathbb{R}^{16}$ are the same as the product structures from Sp(1) so Spin (4) is a compact, simply connected Lie group. Since the Lie algebra of Sp(1) can be identified with the pure imaginary quaternions, the Lie algebra of Spin(4) can be identified with

$$\mathfrak{spin}(4) = \left\{ \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} : Q_1, Q_2 \in \mathrm{Im}\,\mathbb{H} \right\} \cong \mathrm{Im}\,\mathbb{H} \oplus \mathrm{Im}\,\mathbb{H}. \tag{70}$$

For any $u \in Cl^{\times}(4)$ we define an isomorphism $\mathrm{ad}_u : Cl(4) \to Cl(4)$ by

$$\mathrm{ad}_u(p) = upu^{-1}.$$

Each ad_u clearly preserves the grading of $Cl(4)$. For proofs of the following results, leading up to Theorem 1.22, one can consult Theorem 7.3 of [Nab3] or Theorem A.2.3 of [Nab5]. If $u \in \mathrm{Pin}(4)$ one can show that ad_u carries $\mathbb{R}^4$ onto itself and is, in fact, an orthogonal transformation of $\mathbb{R}^4$. If $u \in \mathrm{Spin}(4)$, then the determinant of ad_u is 1 so ad_u is a rotation of $\mathbb{R}^4$.

$$u \in \mathrm{Spin}(4) \Rightarrow \mathrm{ad}_u \in \mathrm{SO}(\mathbb{R}^4).$$

The orthonormal basis $\{e_1, e_2, e_3, e_4\}$ for $\mathbb{R}^4$ identifies $\mathrm{SO}(\mathbb{R}^4)$ with SO(4) so we have a map

$$\mathrm{Spin} : \mathrm{Spin}(4) \to \mathrm{SO}(\mathbb{R}^4) \cong \mathrm{SO}(4)$$
$$\mathrm{Spin}(u) = \mathrm{ad}_u$$

This is called the *spinor map*. It is a surjective Lie group homomorphism with kernel $\mathbb{Z}_2$ so we arrive at the following.

Theorem 1.22. *Spin(4) is the universal double cover of SO(4).*

Remark 1.23. In Section 1.6.2 we will see how globalizing these constructions leads to the notion of a *Spin structure* on a manifold and how this additional structure gives rise to the existence of *spinor fields* on the manifold. There is, however, an obstruction to the existence of a Spin structure called the *second Stiefel-Whitney class* (see Section 6.5 of [Nab5]) and many interesting 4-manifolds ($\mathbb{CP}^2$, for example) fail to admit such a structure. Since a spinor field is an essential ingredient of Seiberg-Witten theory and since we would like the theory to be applicable to as large a class of 4-manifolds as possible it will be necessary to generalize these ideas. As it happens there is a very natural generalization obtained by complexifying our previous algebraic considerations and this is what we will turn to next.

In order to define complex analogues of the algebraic objects we have introduced we will embed $Cl(4)$ into a complex algebra of matrices and form the complex subalgebra it generates. The basic tool we use is the usual matrix model of the quaternions. Specifically, we consider the map $\gamma : \mathbb{H} \to \mathbb{C}^{2\times 2}$ from the quaternions to the 2×2 complex matrices given by

$$\gamma(q) = \gamma(\alpha + \beta\mathbf{j}) = \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}, \tag{71}$$

where we have written $q = q^1 + q^2\mathbf{i} + q^3\mathbf{j} + q^4\mathbf{k} = (q^1 + q^2\mathbf{i}) + (q^3 + q^4\mathbf{i})\mathbf{j} = \alpha + \beta\mathbf{j}$. One easily checks that γ is real linear, injective, preserves products, carries $\overline{q}$ to $\overline{\gamma(q)}^T$ and satisfies $\det(\gamma(q)) = \|q\|^2$ so that we can identify $\mathbb{H}$ with the set of all 2×2 complex matrices of the form $\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}$. Specifically, if we let

$$\gamma(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{1}, \quad \gamma(\mathbf{i}) = \begin{pmatrix} \mathbf{i} & 0 \\ 0 & -\mathbf{i} \end{pmatrix} = \mathbf{I},$$
$$\gamma(\mathbf{j}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \mathbf{J}, \quad \gamma(\mathbf{k}) = \begin{pmatrix} 0 & \mathbf{i} \\ \mathbf{i} & 0 \end{pmatrix} = \mathbf{K} \tag{72}$$

then we identify $q = q^1 + q^2\mathbf{i} + q^3\mathbf{j} + q^4\mathbf{k}$ with $q^1\mathbf{1} + q^2\mathbf{I} + q^3\mathbf{J} + q^4\mathbf{K}$. Next we identify $Cl(4) = \mathbb{H}^{2\times 2}$ with a subset of $\mathbb{C}^{4\times 4}$. Define $\Gamma : \mathbb{H}^{2\times 2} \to \mathbb{C}^{4\times 4}$ by

$$\Gamma\begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = \begin{pmatrix} \gamma(q_{11}) & \gamma(q_{12}) \\ \gamma(q_{21}) & \gamma(q_{22}) \end{pmatrix}. \tag{73}$$

This map Γ is also real linear, injective and preserves products so we can identify $Cl(4)$ with its image

$$Cl(4) = \Gamma(\mathbb{H}^{2\times 2}) \subseteq \mathbb{C}^{4\times 4}.$$

The restriction of Γ to $\mathbb{R}^4 \subseteq Cl(4)$ is

$$x = \begin{pmatrix} 0 & q \\ -\overline{q} & 0 \end{pmatrix} \mapsto \Gamma(x) = \begin{pmatrix} 0 & \gamma(q) \\ -\overline{\gamma(q)}^T & 0 \end{pmatrix}.$$

Since $\det \Gamma(x) = \det x = \|x\|^2 = \|q\|^2$ we can define an inner product on the copy $\Gamma(\mathbb{R}^4)$ of $\mathbb{R}^4$ via polarization from $\|\Gamma(x)\|^2 = \det \Gamma(x)$ and then $\gamma \,|\, \mathbb{R}^4$ is an isometry. With this we fully identify $\mathbb{R}^4$ with $\Gamma(\mathbb{R}^4)$ and obtain an orthonormal basis $E_j = \Gamma(e_j)$, $j = 1, 2, 3, 4$, satisfying

$$E_iE_j + E_jE_i = -2\langle E_i, E_j\rangle\mathbf{1}, \quad i, j = 1, 2, 3, 4.$$

In this context, $Cl\,(4)$ is the *real* subalgebra of $\mathbb{C}^{4\times 4}$ generated by $\{E_1, E_2, E_3, E_4\}$ and a basis for $Cl\,(4)$ is as in (68), but with everything capitalized and 1 changed to $\mathbf{1}$. Moreover, under γ, Sp(1) is mapped to SU(2) so

$$\mathrm{Spin}(4) = \left\{ \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix} : U_1, U_2 \in \mathrm{SU}(2) \right\} \cong \mathrm{SU}(2) \times \mathrm{SU}(2) \tag{74}$$

and

$$\mathfrak{spin}(4) = \left\{ \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} : A_1, A_2 \in \mathfrak{su}(2) \right\} \cong \mathfrak{su}(2) \oplus \mathfrak{su}(2). \tag{75}$$

Now we regard $\mathbb{C}^{4\times 4}$ as a complex algebra. The complex linear subspace spanned by $\{E_1, E_2, E_3, E_4\}$ is a copy of $\mathbb{C}^4$. We define the *complexified Clifford algebra* $Cl\,(4) \otimes \mathbb{C}$ *of* $\mathbb{R}^4$ to be the complex subalgebra of $\mathbb{C}^{4\times 4}$ generated by $\{E_1, E_2, E_3, E_4\}$, that is, by $Cl\,(4)$. A basis for $Cl\,(4) \otimes \mathbb{C}$ over $\mathbb{C}$ is given by (68) with all of the e_i capitalized and 1 changed to $\mathbf{1}$. Thus, the dimension of $Cl\,(4)\otimes\mathbb{C}$ over $\mathbb{C}$ is 16 so $Cl\,(4)\otimes\mathbb{C} \cong \mathbb{C}^{4\times 4}$. We will let

$$S_{\mathbb{C}} = \mathbb{C}^4$$

with its usual Hermitian inner product ($\langle z, w\rangle = \bar{z}^1 w^1 + \bar{z}^2 w^2 + \bar{z}^3 w^3 + \bar{z}^4 w^4$) and use the basis $\{E_1, E_2, E_3, E_4\}$ for $\mathbb{C}^4$ to identify

$$Cl\,(4) \otimes \mathbb{C} = \mathrm{End}_{\mathbb{C}}(S_{\mathbb{C}}).$$

Thus, the elements of $Cl\,(4) \otimes \mathbb{C}$ (and therefore also $Cl\,(4), \mathbb{R}^4$ and Spin(4)) act as endomorphisms of $S_{\mathbb{C}}$. This action is called *Clifford multiplication* and will be indicated with a dot $\cdot$. Restricting the Clifford action to $\mathrm{Spin}(4) \subseteq Cl\,(4) \otimes \mathbb{C}$ gives a *group* representation of Spin(4) on $S_{\mathbb{C}}$ that we will denote

$$\Delta_{\mathbb{C}} : \mathrm{Spin}(4) \to \mathrm{GL}_{\mathbb{C}}(S_{\mathbb{C}})$$

and call the *complex spin representation* This is *not* irreducible. Indeed, if we write

$$S_{\mathbb{C}} = S_{\mathbb{C}}^{+} \oplus S_{\mathbb{C}}^{-}$$
$$\begin{pmatrix} z^1 \\ z^2 \\ z^3 \\ z^4 \end{pmatrix} = \begin{pmatrix} z^1 \\ z^2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ z^3 \\ z^4 \end{pmatrix},$$

then Clifford multiplication by elements of $\mathrm{Spin}(4) \subseteq Cl_{\mathbf{0}}(4)\otimes\mathbb{C}$, because they are block diagonal, preserves $S_{\mathbb{C}}^{+}$ and $S_{\mathbb{C}}^{-}$, whereas Clifford multiplication by $Cl_{\mathbf{1}}(4)\otimes\mathbb{C}$ interchanges $S_{\mathbb{C}}^{+}$ and $S_{\mathbb{C}}^{-}$. In particular, $\Delta_{\mathbb{C}}$ resolves into a direct sum

$$\Delta_{\mathbb{C}} = \Delta_{\mathbb{C}}^{+} \oplus \Delta_{\mathbb{C}}^{-},$$

where

$$\Delta_{\mathbb{C}}^{\pm} : \mathrm{Spin}(4) \to \mathrm{SU}(S_{\mathbb{C}}^{\pm}).$$

These are called the *positive* and *negative complex spin representations*. Notice, however, that Clifford multiplication by elements of $\mathbb{R}^4 \subseteq Cl\,(4) \otimes \mathbb{C}$, being odd, interchanges $S_{\mathbb{C}}^{+}$ and $S_{\mathbb{C}}^{-}$.

Recall that Spin(4) is the set of all even elements in the subgroup of multiplicative units in $Cl(4)$ generated by the unit sphere in $\mathbb{R}^4 \subseteq Cl(4)$. For the complex analogue we add to the generators the unit circle in $\mathbb{C}$. More precisely, we identify U(1) with the subset

$$\mathrm{U}(1) = \{ e^{\theta \mathbf{i}}\mathbf{1} : \theta \in \mathbb{R} \}$$

of $Cl(4) \otimes \mathbb{C}$ (often dropping the $\mathbf{1}$ and thinking of $e^{\theta\mathbf{i}}$ as an element of $Cl(4) \otimes \mathbb{C}$). Then

$$\mathrm{Spin}^c(4)$$

is defined to be the subgroup of the group of multiplicative units in $Cl(4) \otimes \mathbb{C}$ generated by Spin(4) and U(1). Since U(1) is in the center of $Cl(4) \otimes \mathbb{C}$ we have

$$\begin{aligned}\mathrm{Spin}^c(4) = \{ e^{\theta\mathbf{i}}u : \theta \in \mathbb{R}, u \in \mathrm{Spin}(4) \} &= \left\{ e^{\theta\mathbf{i}} \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix} : \theta \in \mathbb{R}, U_1, U_2 \in \mathrm{SU}(2) \right\} \\ &= \left\{ \begin{pmatrix} U_+ & 0 \\ 0 & U_- \end{pmatrix} : U_+, U_- \in \mathrm{U}(2), \det\, U_+ = \det\, U_- \right\}.\end{aligned}$$

From Lemma 1.21 and the fact that U(1) is Abelian, it follows that the center of $\mathrm{Spin}^c(4)$ is

$$Z(\mathrm{Spin}^c(4)) = \mathrm{U}(1).$$

There is another useful way of looking at $\mathrm{Spin}^c(4)$. The mapping

$$\begin{aligned}\mathrm{Spin}(4) \times \mathrm{U}(1) &\to \mathrm{Spin}^c(4) \\ (u, e^{\theta\mathbf{i}}) &\to e^{\theta\mathbf{i}}u\end{aligned}$$

is a surjective homomorphism. Its kernel is the set of all $(\alpha, \alpha^{-1}) \in \mathrm{Spin}(4) \cap \mathrm{U}(1)$. But Spin(4) intersects the scalars only in $\pm\mathbf{1}$ so the kernel is $\mathbb{Z}_2 = \pm(\mathbf{1}, \mathbf{1})$. Thus,

$$\mathrm{Spin}^c(4) \cong \mathrm{Spin}(4) \times \mathrm{U}(1)/\mathbb{Z}_2.$$

Now we will define a few mappings that will be required in the next section. The first is a map

$$\delta : \mathrm{Spin}^c(4) \to \mathrm{U}(1) \tag{76}$$

defined as follows. For $\xi = \begin{pmatrix} U_+ & 0 \\ 0 & U_- \end{pmatrix} = \begin{pmatrix} e^{\theta\mathbf{i}}U_1 & 0 \\ 0 & e^{\theta\mathbf{i}}U_2 \end{pmatrix} \in \mathrm{Spin}^c(4)$,

$$\delta(\xi) = \det\, U_+ = \det\, U_- = e^{2\theta\mathbf{i}}.$$

Then δ is a surjective homomorphism with kernel Spin(4). Next define

$$\pi : \mathrm{Spin}^c(4) \to \mathrm{SO}(4) \tag{77}$$

as follows. For each $u \in \mathrm{Spin}(4)$, $\mathrm{ad}_u : \mathbb{R}^4 \to \mathbb{R}^4$ is a rotation of $\mathbb{R}^4$ defined by $\mathrm{ad}_u(x) = uxu^{-1}$ for every $x \in \mathbb{R}^4$. Notice, however, that if $\xi = e^{\theta\mathbf{i}}u \in \mathrm{Spin}^c(4)$, then $\xi x \xi^{-1} = uxu^{-1} = \mathrm{ad}_u(x)$ so we can extend the adjoint action of Spin(4) on $\mathbb{R}^4$ to $\mathrm{Spin}^c(4)$ by setting $\mathrm{ad}_\xi = \mathrm{ad}_u$. We define

$$\pi(\xi) = \mathrm{ad}_\xi = \mathrm{ad}_u = \mathrm{Spin}(u),$$

where $\xi = e^{\theta \mathbf{i}} u$ for some $\theta \in \mathbb{R}$ and some $u \in \mathrm{Spin}(4)$. Putting δ and π together we obtain

$$\mathrm{Spin}^c : \mathrm{Spin}^c(4) \to \mathrm{SO}(4) \times \mathrm{U}(1)$$
$$\mathrm{Spin}^c(\xi) = \mathrm{Spin}^c(e^{\theta \mathbf{i}} u) = (\pi(\xi), \delta(\xi)) = (\mathrm{Spin}(u), e^{2\theta \mathbf{i}}).$$

Spin^c is a surjective homomorphism with kernel $\mathbb{Z}_2 = \pm \mathbf{1}$ so $\mathrm{Spin}^c(4)$ is a double cover of $\mathrm{SO}(4) \times \mathrm{U}(1)$ (not the universal cover, however, since its fundamental group is $\mathbb{Z}$). The Lie algebra is therefore $\mathfrak{so}(4) \oplus \mathfrak{u}(1) \cong \mathfrak{spin}(4) \oplus \mathfrak{u}(1)$ which can be identified with the following subset of $Cl(4) \otimes \mathbb{C}$.

$$\mathfrak{spin}^c(4) = \left\{ \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} + t\,\mathbf{i} \begin{pmatrix} \mathbf{1} & 0 \\ 0 & \mathbf{1} \end{pmatrix} : t \in \mathbb{R}, A_1, A_2 \in \mathfrak{su}(2) \right\}$$

In particular, $\mathfrak{spin}^c(4)$ also acts by Clifford multiplication on $S_{\mathbb{C}}$, preserving both $S^{\pm}_{\mathbb{C}}$.

The complex spin representation $\Delta_{\mathbb{C}} : \mathrm{Spin}(4) \to \mathrm{GL}_{\mathbb{C}}(S_{\mathbb{C}})$ satisfies $\Delta_{\mathbb{C}}(-\mathbf{1}) = -\mathbf{1}$ and so it extends to a representation

$$\hat{\Delta}_{\mathbb{C}} : \mathrm{Spin}^c(4) \to \mathrm{GL}_{\mathbb{C}}(S_{\mathbb{C}})$$

of $\mathrm{Spin}^c(4)$. Since the elements of $\mathrm{Spin}^c(4)$ are all block diagonal, this representation also splits into

$$\hat{\Delta}_{\mathbb{C}} = \hat{\Delta}^+_{\mathbb{C}} \oplus \hat{\Delta}^-_{\mathbb{C}},$$

where

$$\hat{\Delta}^{\pm}_{\mathbb{C}} : \mathrm{Spin}^c(4) \to \mathrm{U}(S^{\pm}_{\mathbb{C}}).$$

One of the Seiberg-Witten equations relates the self-dual part of the curvature of a U(1)-connection to a certain trace-free endomorphism of a positive spinor bundle. The last of our algebraic preliminaries describes the relationship between 2-forms and endomorphisms. Note first that there is a linear isomorphism from the complex-valued 2-forms $\Omega^2(\mathbb{R}^4; \mathbb{C})$ on $\mathbb{R}^4$ into $Cl_{\mathbf{0}}(4) \otimes \mathbb{C}$. Indeed, if $\{e_1, e_2, e_3, e_4\}$ is the basis (67) for $\mathbb{R}^4$ and $\{e^1, e^2, e^3, e^4\}$ is its dual basis and if $\{E_1, E_2, E_3, E_4\}$ is the image of $\{e_1, e_2, e_3, e_4\}$ under Γ, then we define

$$\rho : \Omega^2(\mathbb{R}^4; \mathbb{C}) \to Cl_{\mathbf{0}}(4) \otimes \mathbb{C} \tag{78}$$

by

$$\rho(\eta) = \rho\Big(\sum_{i<j} \eta_{ij} e^i \wedge e^j\Big) = \sum_{i<j} \eta_{ij} E_i E_j$$

Notice that, although ρ is clearly a linear isomorphism, it is not multiplicative. For example, $e^1 \wedge e^1 = 0$, but $E_1 E_1 = -\mathbf{1}$. There is, of course, an analogous map in any rank. Also notice that, if η real-valued (respectively, $\mathrm{Im}\,\mathbb{C}$-valued), then $\rho(\eta)$ is skew-Hermitian (respectively, Hermitian). For example, if η is real-valued,

$$\overline{\rho(\eta)}^T = \sum_{i<j} \overline{\eta}_{ij} \overline{E_i E_j}^T = \sum_{i<j} \eta_{ij} \overline{E}_j^T \overline{E}_i^T = \sum_{i<j} \eta_{ij} (-E_j)(-E_i) = \sum_{i<j} \eta_{ij} E_j E_i = \sum_{i<j} \eta_{ij} (-E_i E_j) = -\rho(\eta).$$

Moreover, in the definition of η, $\{e_1, e_2, e_3, e_4\}$ can be replaced by any oriented, orthonormal basis provided $\{E_1, E_2, E_3, E_4\}$ is replaced by the image of the basis under Γ. Being even, that is, block diagonal, any $\rho(\eta)$ preserves the subspaces $S^{\pm}_{\mathbb{C}}$ of $S_{\mathbb{C}}$ so we obtain endomorphisms of $S^{\pm}_{\mathbb{C}}$ by setting

$$\rho^{\pm}(\eta) = \rho(\eta)\,|\,S^{\pm}_{\mathbb{C}}.$$

For example,

$$\rho^{+}(\eta) = (\eta_{12} + \eta_{34})\,\mathbf{I} + (\eta_{13} + \eta_{42})\,\mathbf{J} + (\eta_{14} + \eta_{23})\,\mathbf{K}.$$

Thus, we have two maps

$$\rho^{\pm} : \Omega^2(\mathbb{R}^4; \mathbb{C}) \to \mathrm{End}_{\mathbb{C}}(S^{\pm}_{\mathbb{C}}).$$

The usual orientation and inner product on $\mathbb{R}^4$ determine a Hodge star operator * and therefore a decomposition

$$\Omega^2(\mathbb{R}^4; \mathbb{C}) = \Omega^2_+(\mathbb{R}^4; \mathbb{C}) \oplus \Omega^2_-(\mathbb{R}^4; \mathbb{C})$$

of $\Omega^2(\mathbb{R}^4; \mathbb{C})$ into self-dual and anti-self-dual 2-forms.

Lemma 1.23. $\rho^{\pm}\,|\,\Omega^2_{\pm}(\mathbb{R}^4; \mathbb{C})$ *is a complex-linear isomorphism onto the space* $End_0(S^{\pm}_{\mathbb{C}})$ *of trace-free endomorphisms of* $S^{\pm}_{\mathbb{C}}$.

Proof. We give the argument for $\rho^{+}\,|\,\Omega^2_+(\mathbb{R}^4; \mathbb{C})$. The $\rho^{-}\,|\,\Omega^2_-(\mathbb{R}^4; \mathbb{C})$ case is analogous. A simple computation shows that

$$\rho(e^1 \wedge e^2 + e^3 \wedge e^4) = 2\begin{pmatrix} \mathbf{I} & 0 \\ 0 & 0 \end{pmatrix}$$

$$\rho(e^1 \wedge e^3 + e^4 \wedge e^2) = 2\begin{pmatrix} \mathbf{J} & 0 \\ 0 & 0 \end{pmatrix}$$

$$\rho(e^1 \wedge e^4 + e^2 \wedge e^3) = 2\begin{pmatrix} \mathbf{K} & 0 \\ 0 & 0 \end{pmatrix}$$

Since $\{e^1 \wedge e^2 + e^3 \wedge e^4,\ e^1 \wedge e^3 + e^4 \wedge e^2,\ e^1 \wedge e^4 + e^2 \wedge e^3\}$ spans the set of self-dual 2-forms on $\mathbb{R}^4$, it is clear that $\rho^{+}\,|\,\Omega^2_+(\mathbb{R}^4; \mathbb{C})$ is a complex-linear, injective map into the endomorphisms of $S^{+}_{\mathbb{C}}$. Because $\mathbf{I}, \mathbf{J}$, and $\mathbf{K}$ are trace-free, so is everything in the image of $\rho^{+}\,|\,\Omega^2_+(\mathbb{R}^4; \mathbb{C})$. Furthermore, every 2×2 complex, trace-free matrix is a complex linear combination of $\mathbf{I}, \mathbf{J}$, and $\mathbf{K}$ so $\rho^{+}\,|\,\Omega^2_+(\mathbb{R}^4; \mathbb{C})$ maps onto $\mathrm{End}_0(S^{+}_{\mathbb{C}})$. ◻

It follows, in particular, that the map $\rho^{+}\,|\,\Omega^2_+(\mathbb{R}^4; \mathbb{C}) : \Omega^2_+(\mathbb{R}^4; \mathbb{C}) \to \mathrm{End}_0(S^{+}_{\mathbb{C}})$ has an inverse that we will denote simply

$$\sigma^{+} : \mathrm{End}_0(S^{+}_{\mathbb{C}}) \to \Omega^2_+(\mathbb{R}^4; \mathbb{C}). \tag{79}$$

We would like to describe the image under σ^{+} of a particular type of trace-free endomorphism of $S^{+}_{\mathbb{C}}$. For this we consider an arbitrary element ψ of $S^{+}_{\mathbb{C}}$ and suppress its two zero components to write

$$\psi = \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix}.$$

Define an endomorphism of $S^+_{\mathbb{C}}$ by the matrix

$$\psi \otimes \psi^* = \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix} \begin{pmatrix} \overline{\psi}^1 & \overline{\psi}^2 \end{pmatrix} = \begin{pmatrix} |\psi^1|^2 & \psi^1 \overline{\psi}^2 \\ \overline{\psi}^1 \psi^2 & |\psi^2|^2 \end{pmatrix}.$$

The trace-free part of this endomorphism is

$$(\psi \otimes \psi^*)_0 = \psi \otimes \psi^* - \frac{1}{2} \operatorname{tr}(\psi \otimes \psi^*) \, \mathbf{1} = \begin{pmatrix} \frac{1}{2}(|\psi^1|^2 - |\psi^2|^2) & \psi^1 \overline{\psi}^2 \\ \overline{\psi}^1 \psi^2 & \frac{1}{2}(|\psi^2|^2 - |\psi^1|^2) \end{pmatrix} \tag{80}$$

The following expression for $\sigma^+((\psi \otimes \psi^*)_0)$ can be checked by simply applying ρ^+ to the right-hand side.

$$\begin{aligned} \sigma^+((\psi \otimes \psi^*)_0) = -\frac{1}{4} \mathbf{i} \, [\, & (|\psi^1|^2 - |\psi^2|^2)(e^1 \wedge e^2 + e^3 \wedge e^4) \\ & - 2 \operatorname{Im}(\psi^1 \overline{\psi}^2)(e^1 \wedge e^3 + e^4 \wedge e^2) \\ & + 2 \operatorname{Re}(\psi^1 \overline{\psi}^2)(e^1 \wedge e^4 + e^2 \wedge e^3) \,] \end{aligned} \tag{81}$$

1.6.2. *Spin and Spin^c Structures.* Our task now is to globalize the algebraic constructions of the preceding section. Throughout this section and the next, X will denote an oriented, smooth 4-manifold.

Remark 1.24. Compactness, connectedness, and simple connectivity are generally not required for the constructions we wish to describe in this section and the next, but only for one of the results we wish to state. The situation changes in Sections 1.6.4 and 1.6.5, where we will impose these conditions (and more) on all of the oriented, smooth 4-manifolds we consider.

Choose a Riemannian metric g on X. The orientation and the Riemannian metric determine an oriented, orthonormal frame bundle

$$SO(4) \hookrightarrow F_{SO}(X) \stackrel{\pi_{SO}}{\to} X.$$

The fiber above each point in X is a copy of SO(4). We now know that SO(4) is double covered by Spin(4).

$$\text{Spin} : \text{Spin}(4) \to SO(4)$$

What we would like to do is piece together these double covers over the points of X and build a principal Spin(4)-bundle over X. More precisely, a *Spin structure* for X consists of a principal Spin(4)-bundle

$$\text{Spin}(4) \hookrightarrow S(X) \stackrel{\pi_S}{\to} X$$

over X and a smooth map

$$\lambda : S(X) \to F_{SO}(X)$$

of $S(X)$ onto $F_{SO}(X)$ satisfying

$$\pi_{SO} \circ \lambda = \pi_S \tag{82}$$

and

$$\lambda(p \cdot u) = \lambda(p) \cdot \text{Spin}(u) \tag{83}$$

for each $p \in S(X)$ and every $u \in \mathrm{Spin}(4)$. Condition (82) says that the fiber $\pi_S^{-1}(x)$ of π_S above $x \in X$ is λ^{-1} of the fiber $\pi_{\mathrm{SO}}^{-1}(x)$ of π_{SO} above x. On each fiber $\pi_S^{-1}(x)$ we can fix p and regard λ as a function of $u \in \mathrm{Spin}(4)$ and (83) asserts that this is just the double cover $\mathrm{Spin} : \mathrm{Spin}(4) \to \mathrm{SO}(4)$ on this fiber. In this sense, λ is just the spinor map on each fiber of π_S.

$$\left.\begin{array}{cl} S(X) & \\ \downarrow & \lambda \\ F_{\mathrm{SO}}(X) & \\ \downarrow & \pi_{\mathrm{SO}} \\ X & \end{array}\right\} \pi_S$$

Two Spin structures

$$\mathrm{Spin}(4) \hookrightarrow S(X) \stackrel{\pi_S}{\to} X$$
$$\lambda : S(X) \to F_{\mathrm{SO}}(X)$$

and

$$\mathrm{Spin}(4) \hookrightarrow S'(X) \stackrel{\pi_{S'}}{\to} X$$
$$\lambda' : S'(X) \to F_{\mathrm{SO}}(X)$$

for X are said to be *equivalent* if there exists a bundle isomorphism $f : S'(X) \to S(X)$ for which $\lambda \circ f = \lambda'$, that is, $f^*\lambda = \lambda'$. Note that it is *not* enough to assume that the the Spin(4)-bundles are isomorphic (see the Example, page 200, of [Miln2]).

Unlike the frame bundle $F_{\mathrm{SO}}(X)$, which exists for any oriented, Riemannian manifold, there is an obstruction to the existence of a Spin structure. We should briefly indicate how this comes about. Let $\{U_\alpha\}$ be any trivializing good cover for $\mathrm{SO}(4) \hookrightarrow F_{\mathrm{SO}}(X) \stackrel{\pi_{\mathrm{SO}}}{\to} X$ so that each intersection $U_\alpha \cap U_\beta$ is contractible (see Theorem 3.7 of [KN1]). Denote the transition functions by $\tau_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{SO}(4)$. Because Spin(4) is the universal cover of SO(4), each of the maps $\tau_{\alpha\beta}$ lifts to a map $\tilde{\tau}_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{Spin}(4)$ satisfying

$$\tau_{\alpha\beta} = \mathrm{Spin} \circ \tilde{\tau}_{\alpha\beta}$$

(see Theorem 6.1 of [Green]). If these lifts satisfy the cocycle condition ($\tilde{\tau}_{\alpha\beta}\tilde{\tau}_{\beta\gamma} = \tilde{\tau}_{\alpha\gamma}$ on $U_\alpha \cap U_\beta \cap U_\gamma$ whenever this is nonempty), then they determine a principal Spin(4)-bundle over X that is unique up to equivalence and is easily seen to give rise to a Spin structure for X (see Proposition 5.2 of [KN1] or Theorem 4.3.4 of [Nab4]). It might well happen that no such family of lifts satisfying the cocycle condition exists and, in this case, no Spin structure exists. In terms of the Čech cohomology of X with coefficients in $\mathbb{Z}_2$ it is easy to isolate a class $w_2(X) \in \check{H}(X; \mathbb{Z}_2)$, called the *second Stiefel-Whitney class*, the vanishing of which is equivalent to the existence of the required lifts (see Section 6.5 of [Nab5]). Thus, an oriented, Riemannian manifold X has a Spin structure if and only if the second Stiefel-Whitney class is trivial. If X has a Spin structure, then it is called a *Spin manifold*. The following result, which describes the situation of most interest to us, is Corollary 2.12 of [LM].

Theorem 1.24. *Let X be a compact, connected, simply connected, oriented, smooth 4-manifold. Then the second Stiefel-Whitney class $w_2(X)$ is trivial if and only if the intersection form Q_X is even.*

Given a Spin structure on X and a representation ρ of Spin(4) on some vector space V there is an associated vector bundle $S(X) \times_\rho V$. Sections of such a vector bundle will be referred to as *spinor fields* on X. On the space $\Gamma(S(X) \times_\rho V)$ of all such sections one can define a canonical first order differential operator $D : \Gamma(S(X) \times_\rho V) \to \Gamma(S(X) \times_\rho V)$ called the *Dirac operator* (see Chapter II, Section 5, pages 112-134, of [LM]). The study of Dirac operators on Spin manifolds has led to some of the deepest mathematics of the 20^{th} century, the most renowned example being the *Atiyah-Singer Index Theorem* (see Chapter III, Section 13, pages 243-259, of [LM]). For our purposes, however, that is, for the Seiberg-Witten equations, we will require a complex analogue of Spin structures and Dirac operators and it is to this that we turn now.

A *Spin^{c}structure* $\mathcal{L}$ on X consists of a principal Spin^c(4)-bundle

$$\mathrm{Spin}^c(4) \hookrightarrow S^c(X) \overset{\pi_{S^c}}{\to} X$$

over X and a smooth map of $S^c(X)$ onto $F_{\mathrm{SO}}(X)$

$$\Lambda : S^c(X) \to F_{\mathrm{SO}}(X)$$

satisfying

$$\pi_{\mathrm{SO}} \circ \Lambda = \pi_{S^c}$$

and

$$\Lambda(p \cdot \xi) = \Lambda(p) \cdot \pi(\xi).$$

The picture is essentially the same as for a Spin structure.

$$\left.\begin{array}{ll} S^c(X) & \\ \downarrow & \Lambda \\ F_{\mathrm{SO}}(X) & \\ \downarrow & \pi_{\mathrm{SO}} \\ X & \end{array}\right\} \pi_{S^c}$$

Two Spin^c structures

$$\mathrm{Spin}^c(4) \hookrightarrow S^c(X) \overset{\pi_{S^c}}{\to} X$$
$$\Lambda : S^c(X) \to F_{\mathrm{SO}}(X)$$

and

$$\mathrm{Spin}^c(4) \hookrightarrow (S^c)'(X) \overset{\pi_{(S^c)'}}{\to} X$$
$$\Lambda' : S'(X) \to F_{\mathrm{SO}}(X)$$

for X are said to be *equivalent* if there exists a bundle isomorphism $F : (S^c)'(X) \to S^c(X)$ for which $\Lambda \circ F = \Lambda'$, that is, $F^*\Lambda = \Lambda'$.

The reason we work with Spin^c structures rather that Spin structures is that, for oriented, Riemannian 4-manifolds, they always exist (see Lemma 3.1.2, page 25, of [Mor1]). Each of the representations δ :

$\text{Spin}^c(4) \to \text{U}(1)$, $\hat{\Delta}_{\mathbb{C}} : \text{Spin}^c(4) \to \text{GL}_{\mathbb{C}}(S_{\mathbb{C}})$, and $\hat{\Delta}^{\pm}_{\mathbb{C}} : \text{Spin}^c(4) \to \text{U}(S^{\pm}_{\mathbb{C}})$ gives rise to a vector bundle associated to $\text{Spin}^c(4) \hookrightarrow S^c(X) \stackrel{\pi_{S^c}}{\to} X$ and we will write these as follows.

$$L(\mathcal{L}) = S^c(X) \times_{\delta} \mathbb{C}$$
$$\mathbb{S}(\mathcal{L}) = S^c(X) \times_{\hat{\Delta}_{\mathbb{C}}} S_{\mathbb{C}}$$
$$\mathbb{S}^{\pm}(\mathcal{L}) = S^c(X) \times_{\hat{\Delta}^{\pm}_{\mathbb{C}}} S^{\pm}_{\mathbb{C}}$$

We refer to $L(\mathcal{L})$ as the *determinant line bundle* of $\mathcal{L}$, $\mathbb{S}(\mathcal{L})$ as the *spinor bundle* of $\mathcal{L}$, and $\mathbb{S}^{\pm}(\mathcal{L})$ as the *positive* and *negative spinor bundles* of $\mathcal{L}$. The algebraic decomposition $\hat{\Delta}_{\mathbb{C}} = \hat{\Delta}^{+}_{\mathbb{C}} \oplus \hat{\Delta}^{-}_{\mathbb{C}}$ of $\hat{\Delta}_{\mathbb{C}}$ gives rise to a Whitney sum decomposition

$$\mathbb{S}(\mathcal{L}) = \mathbb{S}^{+}(\mathcal{L}) \oplus \mathbb{S}^{-}(\mathcal{L})$$

of $\mathbb{S}(\mathcal{L})$. Sections of $\mathbb{S}(\mathcal{L})$, $\mathbb{S}^{+}(\mathcal{L})$, and $\mathbb{S}^{-}(\mathcal{L})$ are referred to as *spinor fields*, *positive spinor fields* and *negative spinor fields*, respectively. The spaces of such sections will be denoted $\Gamma(\mathbb{S}(\mathcal{L}))$, $\Gamma(\mathbb{S}^{+}(\mathcal{L}))$, and $\Gamma(\mathbb{S}^{-}(\mathcal{L}))$, respectively.

The determinant line bundle $L(\mathcal{L})$ is, in particular, a complex line bundle and so has an associated principal U(1)-bundle

$$\text{U}(1) \hookrightarrow L^0(\mathcal{L}) \stackrel{\pi_{L^0}}{\to} X$$

that can be described as follows. Choose a Hermitian fiber metric on $L(\mathcal{L})$. Then $L^0(\mathcal{L})$ is the unit circle bundle of $L(\mathcal{L})$, that is, the corresponding orthonormal frame bundle. One can retrieve $L(\mathcal{L})$ from $L^0(\mathcal{L})$ as the vector bundle associated to $L^0(\mathcal{L})$ by complex multiplication. $L^0(\mathcal{L})$ has a first Chern class $c_1(L^0(\mathcal{L}))$ and one can show that the second Stiefel-Whitney class of X is the mod 2 reduction of $c_1(L^0(\mathcal{L}))$

$$w_2(X) = c_1(L^0(\mathcal{L})) \bmod 2$$

and that, conversely, given a U(1)-bundle L^0 over X with $w_2(X) = c_1(L^0) \bmod 2$, there is a Spin^c structure $\mathcal{L}$ on X with $L^0(\mathcal{L}) = L^0$ (see Section 1.3, pages 42-43, of [Nicol]).

Recall that $\text{Spin}^c(4)$ double covers $\text{Spin}(4)\times\text{U}(1)$ by the map Spin^c. It follows that $S^c(X)$ double covers the fiber product $F_{\text{SO}}(X) \dot{\times} L^0(\mathcal{L})$ (this is just that part of the product bundle $\text{SO}(4)\times\text{U}(1) \hookrightarrow F_{\text{SO}}(X)\times L^0(\mathcal{L}) \to X \times X$ above the diagonal in $X \times X$ with this diagonal identified with X in the obvious way). We will use the symbol Spin^c also for this double cover.

$$\begin{array}{ccccc} \text{Spin}^c(4) & \hookrightarrow & S^c(X) & & \longrightarrow X \\ & & \downarrow & \text{Spin}^c & \\ \text{SO}(4) \times \text{U}(1) & \hookrightarrow & F_{\text{SO}}(X) \dot{\times} L^0(\mathcal{L}) & & \longrightarrow X \end{array}$$

Locally, the map $\text{Spin}^c : S^c(X) \to F_{\text{SO}}(X) \dot{\times} L^0(\mathcal{L})$ is given by

$$(x, \xi) \mapsto (x, \text{Spin}^c(\xi)) = (x,\ (\pi(\xi), \delta(\xi))\,) = (x,\ (\text{Spin}(u), e^{2\theta \mathbf{i}})\,),$$

where $\xi = e^{\theta \mathbf{i}} u$ with $\theta \in \mathbb{R}$ and $u \in \text{Spin}(4)$.

We will also require two bundles associated to the oriented, orthonormal frame bundle of X that do *not* require the existence of a Spin or Spin^c structure. Notice that Spin(4), being contained in $Cl^{\times}(4)$, acts on $Cl(4)$ by conjugation and, since $(-u)p(-u)^{-1} = upu^{-1}$, this descends to an action of $\text{SO}(4)=\text{Spin}(4)/\mathbb{Z}_2$ on

$Cl(4)$ which clearly preserves products since $u(pq)u^{-1} = (upu^{-1})(uqu^{-1})$. The *Clifford bundle* is the bundle of algebras with typical fiber $Cl(4)$ over X associated to $SO(4) \hookrightarrow F_{SO}(X) \overset{\pi_{SO}}{\to} X$ by this action. It is denoted

$$Cl(X) = F_{SO}(X) \times_{SO(4)} Cl(4).$$

In the same way, SO(4) acts on $Cl(4) \otimes \mathbb{C}$ and one defines the *complexified Clifford bundle*

$$Cl(X) \otimes \mathbb{C} = F_{SO}(X) \times_{SO(4)} (Cl(4) \otimes \mathbb{C}).$$

These decompose into even and odd summands $Cl(X) = Cl_{\mathbf{0}}(X) \oplus Cl_{\mathbf{1}}(X)$ and $Cl(X) \otimes \mathbb{C} = Cl_{\mathbf{0}}(X) \otimes \mathbb{C} \oplus Cl_{\mathbf{1}}(X) \otimes \mathbb{C}$. Moreover, pointwise multiplication provides the spaces $\Gamma(Cl(X))$ and $\Gamma(Cl(X) \otimes \mathbb{C})$ of sections of these bundles with algebra structures and such sections act on sections of the spinor bundles by pointwise Clifford multiplication.

Finally, we will need to globalize the isomorphism $\sigma^+ : \mathrm{End}_0(S^+_{\mathbb{C}}) \to \Omega^2_+(\mathbb{R}^4; \mathbb{C})$ of (84). The map $\rho : \Omega^2(\mathbb{R}^4; \mathbb{C}) \to Cl_0(4) \otimes \mathbb{C}$ of (78) is independent of the choice of oriented, orthonormal basis for $\mathbb{R}^4$ so, using local, oriented, orthonormal frame fields, that is, sections of $F_{SO}(X)$, it gives a map from 2-forms on X to sections of $Cl_0(X) \otimes \mathbb{C}$. At each point these sections, in turn, act on the spinor bundle $\mathcal{S}(\mathcal{L}) = \mathcal{S}^+(\mathcal{L}) \oplus \mathcal{S}^-(\mathcal{L})$ of any Spin^c structure for X. Since this action is fiberwise, the image of a self-dual 2-form on X preserves $\mathcal{S}^+(\mathcal{L})$ and is, at each point, a trace-free endomorphism of $S^+_{\mathbb{C}}$. Thus, a self-dual 2-form on X gives rise to a section of the trace-free endomorphism bundle $\mathrm{End}_0(\mathcal{S}^+(\mathcal{L}))$ of the positive spinor bundle. This is an isomorphism of bundles and we will continue to write the inverse as

$$\sigma^+ : \Gamma(\,\mathrm{End}_0(\mathcal{S}^+(\mathcal{L}))\,) \to \Omega^2_+(X; \mathbb{C}). \tag{84}$$

1.6.3. *Seiberg-Witten Equations.* To formulate the Seiberg-Witten equations for X one chooses a Riemannian metric g and an associated Spin^c structure $\mathcal{L}$.The field content of the theory consists of a connection A on the U(1)-principal bundle $L^0(\mathcal{L})$, the *gauge field*, and a positive spinor field $\psi \in \Gamma(\mathcal{S}^+(\mathcal{L}))$, the *matter field*. These two fields are coupled by two equations, the first of which, called the *curvature equation*, we now describe.

The positive spinor field determines a section $(\psi \otimes \psi^*)_0$ of the trace-free endomorphism bundle $\mathrm{End}_0(\mathcal{S}^+(\mathcal{L}))$ as in (80). This corresponds, via (84), to a self-dual 2-form on X. The first of the Seiberg-Witten equations requires that this coincide with the self-dual part F_A^+ of the curvature of A.

$$F_A^+ = \sigma^+(\,(\psi \otimes \psi^*)_0\,) \tag{85}$$

Remark 1.25. We should be clear on what we mean by the curvature of A. This is defined to be a Lie algebra-valued 2-form on the bundle space $L^0(\mathcal{L})$. Since the structure group is U(1), this is a $\mathrm{Im}\,\mathbb{C}$-valued 2-form. Pulling this curvature back to X by the sections corresponding to some trivializing cover one obtains a family of locally defined, $\mathrm{Im}\,\mathbb{C}$-valued 2-forms on X. These are related by the adjoint action which, for the Abelian group U(1), is trivial. Consequently, these locally defined pullbacks agree on the intersections of their domains and so determine a globally defined 2-form on X. This is what we have called the curvature and denoted F_A.

The second Seiberg-Witten equation, called a *Dirac equation*, still requires a bit more preparation. The oriented, orthonormal frame bundle $SO(4) \hookrightarrow F_{SO}(X) \stackrel{\pi_{SO}}{\to} X$ has a distinguished (Levi-Civita) connection which we will denote ω_{LC}. This can be characterized locally as follows. If $\{e_1, e_2, e_3, e_4\}$ is a local, oriented, orthonormal frame field on X, that is, a section of $F_{SO}(X)$, with dual 1-form field $\{e^1, e^2, e^3, e^4\}$, then ω_{LC} is represented by a skew-symmetric matrix (ω^i_j) of real-valued 1-forms satisfying $de^i = -\omega^i_j \wedge e^j$, $i = 1, 2, 3, 4$ (see Chapter IV, Section 2, of [KN1]). Now, notice that, if X has a Spin structure $\text{Spin}(4) \hookrightarrow S(X) \stackrel{\pi_S}{\to} X$, then the map $\lambda : S(X) \to F_{SO}(X)$ is a double cover that respects the group actions so ω_{LC} automatically lifts to a connection on $S(X)$ (think of the connection as a distribution of horizontal spaces). However, if X has only a Spin^c structure $\text{Spin}^c(4) \hookrightarrow S^c(X) \stackrel{\pi_{S^c}}{\to} X$, then the map $\Lambda : S^c(X) \to F_{SO}(X)$ is not a finite covering so ω_{LC} alone will not determine a connection on $S^c(X)$. However, $\text{Spin}^c : S^c(X) \to F_{SO}(X) \dot{\times} L^0(\mathcal{L})$ is a double cover so, if A is any connection on $L^0(\mathcal{L})$, then ω_{LC} and A together determine a connection on $F_{SO}(X) \dot{\times} L^0(\mathcal{L})$ which will then lift to a connection on $S^c(X)$. Specifically, if we denote by π_1 and π_2 the restrictions to $F_{SO}(X) \dot{\times} L^0(\mathcal{L})$ of the projections of $F_{SO}(X) \times L^0(\mathcal{L})$ onto $F_{SO}(X)$ and $L^0(\mathcal{L})$, respectively, then $\pi_1^*\omega_{LC} \oplus \pi_2^* A$ is a connection on the fiber product and

$$\omega_A = (\text{Spin}^c)^*(\pi_1^*\omega_{LC} \oplus \pi_2^* A) \tag{86}$$

is a connection on $S^c(X)$. Any such ω_A is called a *Spin^{c}connection* on $S^c(X)$. Notice that the Levi-Civita connection ω_{LC} on the frame bundle is fixed, but the U(1)-connection A will play the role of the gauge field in Seiberg-Witten theory and is constrained only by the Seiberg-Witten equations, that is, by (85) and the Dirac equation that we are in the process of introducing.

A Spin^c connection ω_A determines a covariant derivative on sections of all of the associated spinor bundles $\mathcal{S}(\mathcal{L})$, $\mathcal{S}^+(\mathcal{L})$, and $\mathcal{S}^-(\mathcal{L})$. We will denote all of these by ∇_A. It is in terms of these covariant derivatives that we will now build a "Dirac operator" on spinor fields. Begin with

$$\nabla_A : \Gamma(\mathcal{S}(\mathcal{L})) \to \Omega^1(X; \mathbb{C}) \otimes \Gamma(\mathcal{S}(\mathcal{L})).$$

Define

$$\tilde{\not{D}}_A : \Gamma(\mathcal{S}(\mathcal{L})) \to \Gamma(\mathcal{S}(\mathcal{L}))$$

as follows. Let $\{e_1, e_2, e_3, e_4\}$ be an oriented, orthonormal frame field on $U \subseteq X$. Each e_i can be thought of either as a vector field on U or as a section of the Clifford bundle which therefore acts by Clifford multiplication on sections of $\mathcal{S}(\mathcal{L})$ defined on U. Thus, for each $\Psi \in \Gamma(\mathcal{S}(\mathcal{L}))$ we can define $\tilde{\not{D}}_A$ on Ψ by

$$\tilde{\not{D}}_A \Psi = \sum_{i=1}^{4} e_i \cdot \nabla_A \Psi(e_i). \tag{87}$$

One shows that this is independent of the choice of $\{e_1, e_2, e_3, e_4\}$ and so defines $\tilde{\not{D}}_A \Psi$ globally (see Lemma 3.3.1 of [Mor1]). Since $\mathcal{S}(\mathcal{L}) = \mathcal{S}^+(\mathcal{L}) \oplus \mathcal{S}^-(\mathcal{L})$, we can restrict $\tilde{\not{D}}_A$ to sections of either $\mathcal{S}^+(\mathcal{L})$ or $\mathcal{S}^-(\mathcal{L})$. Since Clifford multiplication by e_i switches $\mathcal{S}^\pm(\mathcal{L})$, so will these restrictions. We will write these as

$$\not{D}_A : \Gamma(\mathcal{S}^+(\mathcal{L})) \to \Gamma(\mathcal{S}^-(\mathcal{L})) \tag{88}$$

and

$$\not{D}_A^* : \Gamma(\mathcal{S}^-(\mathcal{L})) \to \Gamma(\mathcal{S}^+(\mathcal{L})) \tag{89}$$

The operators $\not{D}_A$ and $\not{D}_A^*$ are, in fact, formal adjoints of each other with respect to the L^2-inner product on sections induced by the pointwise Hermitian inner product on fibers (see Lemma 3.3.3 of [Mor1]). They are, moreover, elliptic operators (see pages 43-44 of [Mor1]). We will refer to $\not{D}_A$ as the *Dirac operator* of Seiberg-Witten theory. The second of the Seiberg-Witten equations requires that the positive spinor field ψ be annihilated by $\not{D}_A$.

$$\not{D}_A \psi = 0. \tag{90}$$

We will collect together all of this information in the form of a few definitions and then try to gain some sense of what the Seiberg-Witten equations look like locally. Thus, we let X denote an oriented, smooth 4-manifold. Choose a Riemannian metric g for X and a Spin^c structure $\mathcal{L}$ for the corresponding oriented, orthonormal frame bundle $SO(4) \hookrightarrow F_{SO}(X) \xrightarrow{\pi_{SO}} X$. The *Seiberg-Witten configuration space* $\mathcal{A}(\mathcal{L})$ consists of all pairs (A, ψ), where A is a connection on the principal U(1)-bundle $L^0(\mathcal{L})$ and $\psi \in \Gamma(\mathcal{S}^+(\mathcal{L}))$ is a positive spinor field. An $(A, \psi) \in \mathcal{A}(\mathcal{L})$ is called a *Seiberg-Witten monopole* if it satisfies the *Seiberg-Witten Equations*.

$$\begin{aligned} F_A^+ &= \sigma^+(\,(\psi \otimes \psi^*)_0\,) \\ \not{D}_A \psi &= 0 \end{aligned} \tag{91}$$

Example 1.9. To gain some sense of what the Seiberg-Witten equations actually look like we will write them out explicitly when $X = \mathbb{R}^4$ with its standard orientation and Riemannian metric. Since $\mathbb{R}^4$ is contractible all of the relevant bundles over it are trivial and we will work with explicit trivializations (see Corollary 11.6, page 53, of [Steen]). Thus, the oriented, orthonormal frame bundle is

$$SO(4) \hookrightarrow \mathbb{R}^4 \times SO(4) \xrightarrow{\pi_{SO}} \mathbb{R}^4,$$

where π_{SO} is the projection onto the first factor. We introduce a Spin^c structure $\mathcal{L}$ as follows. Begin with the trivial Spin$^c(4)$-bundle over $\mathbb{R}^4$

$$\mathrm{Spin}^c(4) \hookrightarrow \mathbb{R}^4 \times \mathrm{Spin}^c(4) \xrightarrow{\pi_{S^c}} \mathbb{R}^4,$$

where π_{S^c} is the projection onto the first factor. Now define

$$\Lambda : \mathbb{R}^4 \times \mathrm{Spin}^c(4) \to \mathbb{R}^4 \times SO(4)$$

by

$$\Lambda(x, \xi) = (x, \pi(\xi)) = (x, \mathrm{Spin}(u)),$$

where $\xi = e^{\theta \mathbf{i}} u \in \mathrm{Spin}^c(4)$. The associated spinor bundles are therefore also trivial so their sections can be identified with globally defined functions on $\mathbb{R}^4$ which we will write as

$$\Psi = \begin{pmatrix} \psi^1 \\ \psi^2 \\ \psi^3 \\ \psi^4 \end{pmatrix} : \mathbb{R}^4 \to S_{\mathbb{C}} \cong \mathbb{C}^4$$

$$\psi = \begin{pmatrix} \psi^1 \\ \psi^2 \\ 0 \\ 0 \end{pmatrix} : \mathbb{R}^4 \to S^+_{\mathbb{C}} \cong \mathbb{C}^2 \qquad \phi = \begin{pmatrix} 0 \\ 0 \\ \psi^3 \\ \psi^4 \end{pmatrix} : \mathbb{R}^4 \to S^-_{\mathbb{C}} \cong \mathbb{C}^2.$$

For convenience we will often abuse the notation a bit and write $\Psi = \begin{pmatrix} \psi \\ \phi \end{pmatrix}$ by suppressing the zero components. We will use x^1, x^2, x^3, x^4 for the standard coordinates on $\mathbb{R}^4$ and write ∂_i for $\frac{\partial}{\partial x^i}$, $i = 1, 2, 3, 4$, these being applied componentwise to the spinor fields. The determinant line bundle is likewise trivial, as is the corresponding principal U(1)-bundle $L^0(\mathcal{L})$

$$\mathrm{U}(1) \hookrightarrow \mathbb{R}^4 \times \mathrm{U}(1) \xrightarrow{\pi_{L^0}} \mathbb{R}^4,$$

where, once again, π_{L^0} is the projection onto the first factor. A connection on this U(1)-bundle is then uniquely determined by an $\mathrm{Im}\,\mathbb{C}$-valued 1-form

$$A = A_i dx^i, \qquad A_i : \mathbb{R}^4 \to \mathrm{Im}\,\mathbb{C}, \quad i = 1, 2, 3, 4.$$

In orthonormal coordinates, the covariant exterior derivative induced by the Levi-Civita connection on $\mathbb{R}^4$ is just ordinary componentwise exterior differentiation. Consequently, the covariant derivative ∇_A induced by it and the U(1)-connection A takes the form

$$\nabla_A = \nabla_i \, dx^i = (\partial_i + A_i) \, dx^i.$$

Thus, for any spinor field Ψ,

$$\nabla_A \Psi = \begin{pmatrix} (\partial_i \psi^1 + A_i \psi^1) \, dx^i \\ (\partial_i \psi^2 + A_i \psi^2) \, dx^i \\ (\partial_i \psi^3 + A_i \psi^3) \, dx^i \\ (\partial_i \psi^4 + A_i \psi^4) \, dx^i \end{pmatrix}.$$

Taking $\{e_i\} = \{\partial_i\}$ to be the standard oriented, orthonormal frame field on $\mathbb{R}^4$ we have

$$\nabla_A \Psi(e_i) = \nabla_i \Psi = (\partial_i + A_i)\Psi = \begin{pmatrix} \partial_i \psi^1 + A_i \psi^1 \\ \partial_i \psi^2 + A_i \psi^2 \\ \partial_i \psi^3 + A_i \psi^3 \\ \partial_i \psi^4 + A_i \psi^4 \end{pmatrix}.$$

The Dirac operator (87) requires that we Clifford multiply by the basis elements e_i, that is, matrix multiply by $E_i = \Gamma(e_i) \in Cl(4) \otimes \mathbb{C}$, and add. Writing this out explicitly with $\Psi = \begin{pmatrix} \psi \\ \phi \end{pmatrix}$ we have

$$\tilde{\not{D}}_A \Psi = \sum_{i=1}^{4} e_i \cdot \nabla_A \Psi(e_i) = \sum_{i=1}^{4} E_i \nabla_A \Psi(e_i) = \begin{pmatrix} \nabla_1 \phi + \mathbf{I}\,\nabla_2 \phi + \mathbf{J}\,\nabla_3 \phi + \mathbf{K}\,\nabla_4 \phi \\ -\nabla_1 \psi + \mathbf{I}\,\nabla_2 \psi + \mathbf{J}\,\nabla_3 \psi + \mathbf{K}\,\nabla_4 \psi \end{pmatrix}.$$

The restriction of $\tilde{\not{D}}_A$ to a positive spinor field ψ is therefore given by

$$\not{D}_A \psi = -\nabla_1 \psi + \mathbf{I}\,\nabla_2 \psi + \mathbf{J}\,\nabla_3 \psi + \mathbf{K}\,\nabla_4 \psi.$$

The Seiberg-Witten Dirac equation (90) therefore becomes

$$\nabla_1 \psi = \mathbf{I}\,\nabla_2 \psi + \mathbf{J}\,\nabla_3 \psi + \mathbf{K}\,\nabla_4 \psi,$$

or, in complete detail,

$$\begin{pmatrix} -(\partial_1 + A_1) + \mathbf{i}\,(\partial_2 + A_2) & (\partial_3 + A_3) + \mathbf{i}\,(\partial_4 + A_4) \\ -(\partial_3 + A_3) + \mathbf{i}\,(\partial_4 + A_4) & -(\partial_1 + A_1) - \mathbf{i}\,(\partial_2 + A_2) \end{pmatrix} \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{92}$$

Notice that these are *first order, linear* equations.

For the Seiberg-Witten curvature equation (85) we will use the local expressions (81) for $\sigma^+((\psi \otimes \psi^*)_0)$ and the following local description of F_A^+. If $A = A_i dx^i$, then, since U(1) is Abelian, $F_A = dA = \sum_{i<j} F_{ij} dx^i \wedge dx^j$, where $F_{ij} = \partial_i A_j - \partial_j A_i$. A basis for the self-dual 2-forms on $\mathbb{R}^4$ is given by

$$\{\, dx^1 \wedge dx^2 + dx^3 \wedge dx^4,\ dx^1 \wedge dx^3 + dx^4 \wedge dx^2,\ dx^1 \wedge dx^4 + dx^2 \wedge dx^3 \,\}$$

so

$$\begin{aligned} F_A^+ = \frac{1}{2}(F_A + {}^*F_A) = \frac{1}{2}&(F_{12} + F_{34})(dx^1 \wedge dx^2 + dx^3 \wedge dx^4) \\ + \frac{1}{2}&(F_{13} + F_{42})(dx^1 \wedge dx^3 + dx^4 \wedge dx^2) \\ + \frac{1}{2}&(F_{14} + F_{23})(dx^1 \wedge dx^4 + dx^2 \wedge dx^3). \end{aligned}$$

With this and (81) the curvature equation (85) becomes the following system of three equations.

$$\begin{aligned} F_{12} + F_{34} &= -\frac{1}{2}\,\mathbf{i}\,(|\psi^1|^2 - |\psi^2|^2) \\ F_{13} + F_{42} &= \ \ \mathbf{i}\,\mathrm{Im}\,(\psi^1 \overline{\psi}^2) \\ F_{14} + F_{23} &= -\mathbf{i}\,\mathrm{Re}\,(\psi^1 \overline{\psi}^2) \end{aligned}$$

or, in complete detail,

$$\begin{aligned} (\partial_1 A_2 - \partial_2 A_1) + (\partial_3 A_4 - \partial_4 A_3) &= -\frac{1}{2}\,\mathbf{i}\,(|\psi^1|^2 - |\psi^2|^2) \\ (\partial_1 A_3 - \partial_3 A_1) + (\partial_4 A_2 - \partial_2 A_4) &= \ \ \mathbf{i}\,\mathrm{Im}\,(\psi^1 \overline{\psi}^2) \\ (\partial_1 A_4 - \partial_4 A_1) + (\partial_2 A_3 - \partial_3 A_2) &= -\mathbf{i}\,\mathrm{Re}\,(\psi^1 \overline{\psi}^2) \end{aligned} \tag{93}$$

These are again first order equations, but this time (mildly) nonlinear.

It is perhaps worth pointing out that the Seiberg-Witten equations (92) and (93) on $\mathbb{R}^4$ *do* have nontrivial solutions. Indeed, if $\psi = 0$, then the Dirac equation is satisfied identically and A can be taken to be any anti-self-dual connection on the trivial U(1)-bundle $U(1) \hookrightarrow \mathbb{R}^4 \times U(1) \to \mathbb{R}^4$ and these are abundant (see pages 384-385 of [Nab5] for an explicit construction).

Remark 1.26. Our discussion of the Donaldson invariants began with the Yang-Mills action functional. The corresponding Euler-Lagrange equations are the Yang-Mills equations. These are second order partial differential equations. The absolute minima of the Yang-Mills action are solutions to the first order anti-self-dual equations and it is the moduli space of these that give rise to the Donaldson invariants. The pattern is the same for the Seiberg-Witten invariants. One can write down a Seiberg-Witten action functional and the corresponding second order Euler-Lagrange equations . Once again, however, the absolute minima of this action functional are the objects that give rise to the Seiberg-Witten invariants and these are just the solutions to the first order Seiberg-Witten equations (see Section 7.1 of [Sal] or Section 7.2 of [Jost]). Specifically, the Seiberg-Witten action functional/energy has the following form.

$$E(A,\psi) = \int_X \Big(|\nabla_A \psi|^2 + \frac{s}{4}|\phi|^2 + \frac{1}{4}|\phi|^4 + |F_A|^2 \Big) d\,\mathrm{vol}_g,$$

where s is the scalar curvature of X. This can be rewritten in the form

$$E(A,\psi) = \int_X \Big(|\not{D}_A \psi|^2 + 2\,|F_A^+ - \sigma^+((\psi \otimes \psi^*)_0)|^2 \Big) d\,\mathrm{vol}_g - \pi \langle c_1(L^0(\mathcal{L}))^2, [X] \rangle, \tag{94}$$

where $c_1(L^0(\mathcal{L}))$ is the first Chern class of the U(1)-bundle $L^0(\mathcal{L})$ and $[X]$ is the fundamental class of X (see Proposition 7.3 of [Sal]). From this it is clear that the absolute minima occur for configurations that satisfy the Seiberg-Witten equations. In the next two sections we will sketch how these invariants come about.

1.6.4. *Seiberg-Witten Moduli Space.* Throughout this section and the next X will denote a compact, connected, simply connected, oriented, smooth 4-manifold (simple connectivity is not really necessary here, but will streamline the exposition and facilitate comparison with Donaldson theory). Choosing a Riemannian metric g on X gives an oriented, orthonormal frame bundle and one can then select a Spin^c structure $\mathcal{L}$. Recall that an element (A,ψ) of the Seiberg-Witten configuration space $\mathcal{A}(\mathcal{L})$ is called a Seiberg-Witten monopole if it satisfies the Seiberg-Witten equations (91). As in the case of Donaldson theory we are interested in a moduli space of such Seiberg-Witten monopoles so we must begin by isolating the appropriate gauge group. This can be defined in a manner entirely analogous to Donaldson theory, but in this case it admits a much simpler formulation and we will begin with this. We let $C^\infty(X, U(1))$ denote the group of all smooth maps of X into U(1) under pointwise multiplication in U(1). Let $\gamma \in C^\infty(X, U(1))$ and identify U(1) with a subgroup of Spin$^c(4)$. Now define the map

$$\sigma_\gamma : S^c(X) \to S^c(X)$$

by

$$\sigma_\gamma(p) = p \cdot \gamma(\pi_{S^c}(p))$$

where the dot $\cdot$ indicates the right action of Spin$^c(4)$ on $S^c(X)$. Then each σ_γ is an automorphism of the Spin^c-bundle, that is, a diffeomorphism of $S^c(X)$ onto itself satisfying $\sigma_\gamma(p\cdot\xi) = \sigma_\gamma(p)\cdot\xi$ for each $p \in S^c(X)$

and each $\xi \in \mathrm{Spin}^c(4)$, and which covers the identity on X in the sense that $\pi_{S^c} \circ \sigma_\gamma = \pi_{S^c}$. Moreover, each σ_γ covers the identity on $F_{\mathrm{SO}}(X)$ in the following sense. Let $\mathrm{Spin}^c : S^c(X) \to F_{\mathrm{SO}}(X) \dot{\times} L^0(\mathcal{L})$ be the double cover of the fiber product and $\pi_1 : F_{\mathrm{SO}}(X) \dot{\times} L^0(\mathcal{L}) \to F_{\mathrm{SO}}(X)$ the restriction to $F_{\mathrm{SO}}(X) \dot{\times} L^0(\mathcal{L})$ of the projection of $F_{\mathrm{SO}}(X) \times L^0(\mathcal{L})$ onto the first factor. Then

$$\pi_1 \circ \mathrm{Spin}^c \circ \sigma_\gamma = \pi_1 \circ \mathrm{Spin}^c.$$

Remark 1.27. One can show that, conversely, any automorphism of the Spin^c-bundle that covers the identity on $F_{\mathrm{SO}}(X)$ is of this form (see Lemma 7.5, page 55, of [Nab3]).

The group $C^\infty(X, \mathrm{U}(1))$ under pointwise multiplication is isomorphic to the group of all σ_γ under composition and to the group of automorphisms of the Spin^c-bundle that cover the identity on $F_{\mathrm{SO}}(X)$ under composition. This group is called the *Seiberg-Witten gauge group* and denoted $\mathcal{G}(\mathcal{L})$. We will use whichever of these three descriptions is most convenient in any particular context.

To define an action of $\mathcal{G}(\mathcal{L})$ on the configuration space $\mathcal{A}(\mathcal{L})$ we proceed as follows. First consider the positive spinor field $\psi \in \Gamma(\mathcal{S}^+(\mathcal{L}))$. Since $\mathrm{U}(1)$ acts by scalar multiplication on the fibers $S^+_{\mathbb{C}}$ of $\mathcal{S}^+(\mathcal{L})$ we can define a right action of $\gamma \in C^\infty(X, \mathrm{U}(1))$ on $\Gamma(\mathcal{S}^+(\mathcal{L}))$ by

$$(\psi \cdot \gamma)(x) = \gamma(x)^{-1}\psi(x)$$

for every $x \in X$. The same formula defines a right action of $C^\infty(X, \mathrm{U}(1))$ on $\Gamma(\mathcal{S}^-(\mathcal{L}))$.

Next we turn to the connection A on $L^0(\mathcal{L})$. First note that each of the automorphisms $\sigma_\gamma : S^c(X) \to S^c(X)$ induces an automorphism $\sigma'_\gamma : L^0(\mathcal{L}) \to L^0(\mathcal{L})$ defined by

$$\sigma'_\gamma \circ (\pi_2 \circ \mathrm{Spin}^c) = (\pi_2 \circ \mathrm{Spin}^c) \circ \sigma_\gamma.$$

We define the action of $\mathcal{G}(\mathcal{L})$ on A by

$$A \cdot \gamma = A \cdot \sigma_\gamma = (\sigma'_\gamma)^* A.$$

One checks that the Spin^c-connection corresponding to $A \cdot \gamma$ is the pullback by σ_γ of that corresponding to A, that is,

$$\omega_{A \cdot \gamma} = \sigma^*_\gamma\, \omega_A \tag{95}$$

(see page 58 of [Nab3]). It is useful to have the action on A written locally on X. Let s be a local section of $L^0(\mathcal{L})$ and consider the 1-form s^*A on X. Then define

$$(s^*A) \cdot \gamma = s^*(A \cdot \gamma) = s^*(\,(\sigma'_\gamma)^* A\,) = (\sigma'_\gamma \circ s)^* A.$$

A computation using the local expression for Spin^c gives

$$(s^*A) \cdot \gamma = s^*A + 2\gamma^{-1} d\gamma$$

(see pages 57-58 of [Nab3]).

Putting the actions of $\mathcal{G}(\mathcal{L})$ on ψ and A together we arrive at a right action of $\mathcal{G}(\mathcal{L})$ on the Seiberg-Witten configuration space $\mathcal{A}(\mathcal{L})$

$$(A, \psi) \cdot \gamma = (A, \psi) \cdot \sigma_\gamma = (\,(\sigma'_\gamma)^* A,\ \gamma^{-1}\psi\,)$$

or, locally on X,

$$(s^*A, \psi) \cdot \gamma = s^*A + 2\gamma^{-1} d\gamma.$$

One finds that this action carries solutions to the Seiberg-Witten equations onto other solutions to the Seiberg-Witten equations. More precisely, one has the following (see Theorem 7.6 of [Nab3]).

Theorem 1.25. *Suppose* $(A, \psi) \in \mathcal{A}(\mathcal{L})$ *satisfies*

$$\begin{aligned} \not{D}_A \psi &= 0 \\ F_A^+ &= \sigma^+(\,(\psi \otimes \psi^*)_0\,). \end{aligned}$$

Then, for any $\gamma \in C^\infty(X, U(1))$, $(A, \psi) \cdot \gamma = (A \cdot \gamma, \psi \cdot \gamma)$ *satisfies*

$$\begin{aligned} \not{D}_{A\cdot\gamma}(\psi \cdot \gamma) &= 0 \\ F_{A\cdot\gamma}^+ &= \sigma^+(\,((\psi \cdot \gamma) \otimes (\psi \cdot \gamma)^*)_0\,). \end{aligned}$$

Thus, the space of solutions to the Seiberg-Witten equations is invariant under the action of the gauge group $\mathcal{G}(\mathcal{L})$ and we may, as for the anti-self-dual equations, consider the *moduli space* $\mathcal{M}(\mathcal{L})$ of gauge equivalence classes of Seiberg-Witten monopoles.

$$\mathcal{M}(\mathcal{L}) = \{\,(A, \psi) \in \mathcal{A}(\mathcal{L}) : \not{D}_A \psi = 0, F_A^+ = \sigma^+(\,(\psi \otimes \psi^*)_0\,)\,\}/\mathcal{G}(\mathcal{L})$$

Remark 1.28. The ensuing analysis required to manufacture a differential-topological invariant of X from $\mathcal{M}(\mathcal{L})$ is in many ways analogous to that of Donaldson theory. In particular, this analysis cannot be carried out in the smooth category in which we currently find ourselves and one must work with appropriate Sobolev completions. As we did in the anti-self-dual case (see Remarks 1.5 and 1.8) we will forgo this here and simply record the results we will need. We refer those interested in the analytical details to [Mor1], [Nicol], and [Sal].

In Donaldson theory the reducible connections are those left fixed by some nontrivial element of the gauge group and these give rise to singularities in the moduli space. In Seiberg-Witten theory the situation is slightly different, but the reducible configurations are particularly easy to describe.

Lemma 1.26. *An element* (A, ψ) *of* $\mathcal{A}(\mathcal{L})$ *is left fixed by some nontrivial element* γ *of* $\mathcal{G}(\mathcal{L})$ *if and only if* $\psi = 0$ *and, in this case,* $\gamma : X \to U(1)$ *is a constant map.*

Proof. It will suffice to argue locally so we let s be a section of $L^0(\mathcal{L})$ and consider s^*A. Then s^*A is left fixed by $\gamma \in \mathcal{G}(\mathcal{L})$ if and only if

$$(s^*A + 2\gamma^{-1} d\gamma,\ \gamma^{-1}\psi) = (s^*A, \psi)$$

and this, in turn, is the case if and only if

$$\gamma^{-1}\psi = \psi \quad \text{and} \quad 2\gamma^{-1} d\gamma = 0.$$

Since $\gamma \neq 1$ the first of these is true if and only if $\psi = 0$. The second is true if and only if $d\gamma = 0$ and, since X is connected, γ must be constant. □

Thus, we will say that a configuration $(A, \psi) \in \mathcal{A}(\mathcal{L})$ is *reducible* if and only if $\gamma = 0$ and *irreducible* otherwise. In Donaldson theory, Theorem 1.12 guarantees the existence of a dense subspace of the space $\mathcal{R}(X)$ of Riemannian metrics for X for which reducible anti-self-dual connections do not exist and the moduli space is a smooth manifold. One cannot do quite so well for the Seiberg-Witten equations. We will first describe the part of this scenario that is the same and then what must be done to get the full strength of the theorem (see Chapter 4 of [Mor1] or Sections 7.1 and 7.2 of [Sal]).

Theorem 1.27. *Let X denote a compact, connected, simply connected, oriented, smooth 4-manifold with $b_2^+(X) > 0$. Then there is a dense G_δ-subset $\mathcal{R}_G^{SW}(X)$ of the space $\mathcal{R}(X)$ of Riemannian metrics on X with the following property. For any $g \in \mathcal{R}_G^{SW}(X)$ and any corresponding Spin^c structure $\mathcal{L}$, every solution (A, ψ) to the Seiberg-Witten equations is irreducible. If $b_2^+(X) > 1$, then for any generic path $g(t)$, $0 \le t \le 1$, of Riemannian metrics in $\mathcal{R}(X)$ there are no reducible solutions to the Seiberg-Witten equations for any Spin^c structure corresponding to any of the metrics $g(t)$, $0 \le t \le 1$.*

Missing here is any statement about the manifold structure of the moduli space. For this one must perturb, not only the metric, but the equations themselves. Specifically, if we fix a Riemannian metric g and a Spin^c structure $\mathcal{L}$, then we can consider, for any $\eta \in \Omega_+^2(X, \operatorname{Im}\mathbb{C})$, the *perturbed Seiberg-Witten equations*

$$\begin{aligned} \not{D}_A \psi &= 0 \\ F_A^+ &= \sigma^+((\psi \otimes \psi^*)_0) + \eta. \end{aligned} \tag{96}$$

Remark 1.29. The motivation here is not so hard to discern. If we define a map

$$F : \mathcal{A}(\mathcal{L}) \to \Omega_+^2(X, \operatorname{Im}\mathbb{C}) \oplus \Gamma(\mathcal{S}^-(\mathcal{L}))$$

by

$$F(A, \psi) = (F_A^+ - \sigma^+((\psi \otimes \psi^*)_0, \not{D}_A \psi),$$

then (A, ψ) satisfies the Seiberg-Witten equations if and only if it is in the zero set $F^{-1}(0, 0)$. With the appropriate analytical structures one can show that F is smooth and this solution space is a smooth manifold if and only if $(0, 0)$ is a regular value of F. If it is not, then the Sard-Smale Theorem (see Section 3.6 of [AMR]) suggests that a small perturbation of $(0, 0)$ of the form $(\eta, 0)$ will be a regular value so that $F(A, \psi) = (\eta, 0)$ will define a smooth manifold of configurations.

The gauge group $\mathcal{G}(\mathcal{L})$ acts on solutions to the perturbed equations in the same way so one can once again consider the moduli space $\mathcal{M}(\mathcal{L}, \eta)$ of gauge equivalence classes of solutions. One then has the following *Generic Perturbations Theorem* (see Theorem 7.16 of [Sal] or Corollary 6.3.2 of [Mor1]).

Theorem 1.28. *Let X denote a compact, connected, simply connected, oriented, smooth 4-manifold with* $b_2^+(X) > 0$. *Fix a* $g \in \mathcal{R}_G^{SW}(X)$ *and a corresponding Spin*^{c} *structure* $\mathcal{L}$. *Then there is a dense* G_δ*-set* $\Omega_+^2(X, Im\,\mathbb{C})_G$ *in* $\Omega_+^2(X, Im\,\mathbb{C})$ *with the following property. For any* $\eta \in \Omega_+^2(X, Im\,\mathbb{C})_G$, *the moduli space* $\mathcal{M}(\mathcal{L}, \eta)$ *of solutions* (A, ψ) *to the perturbed Seiberg-Witten equations (96) is either empty or a smooth manifold of dimension*

$$\frac{1}{4}\Big[\,\langle\, c_1(L^0(\mathcal{L}))^2, [X]\,\rangle - 2\chi(X) - 3\sigma(X)\,\Big],$$

where $c_1(L^0(\mathcal{L}))$ *is the first Chern class of the U(1)-bundle* $L^0(\mathcal{L})$, $[X]$ *is the fundamental class of X,* $\chi(X)$ *is the Euler characteristic of X and* $\sigma(X)$ *is its signature.*

Much of the analysis of the Seiberg-Witten moduli space (orientability, cobordism, calculation of the dimension, etc.) proceeds in a manner entirely analogous to that of the anti-self-dual moduli space (see Sections 7.2 and 7.3 of [Sal]). There is one very substantial difference, however, and it accounts for the relative simplicity of Seiberg-Witten theory. There is, in this case, no need for an "Uhlenbeck-type" compactification since the Seiberg-Witten moduli spaces are *always compact*. The proof of this rests on the fact that solutions to the (perturbed) Seiberg-Witten equations satisfy certain uniform, *a priori* bounds. This is something that is categorically false for the anti-self-dual equations because these are conformally invariant in dimension four. The arguments are quite long and technical and we will simply refer those interested in the details to either Section 7.2 and Chapter 8 of [Sal] or Sections 5.3 and 6.4 of [Mor1], where the results proved are much more general than the special case we will now record in the final theorem of this section.

Theorem 1.29. *Let X denote a compact, connected, simply connected, oriented, smooth 4-manifold with* $b_2^+(X) > 0$. *Fix a generic Riemannian metric* $g \in \mathcal{R}_G^{SW}(X)$, *a corresponding Spin*^{c} *structure* $\mathcal{L}$, *a generic perturbation* $\eta \in \Omega_+^2(X, Im\,\mathbb{C})_G$, *and an orientation for* $H_+^2(X;\mathbb{R})$. *Then the moduli space* $\mathcal{M}(\mathcal{L}, \eta)$ *of solutions* (A, ψ) *to the perturbed Seiberg-Witten equations (96) is either empty or a smooth, compact, oriented manifold of dimension*

$$\frac{1}{4}\Big[\,\langle\, c_1(L^0(\mathcal{L}))^2, [X]\,\rangle - 2\chi(X) - 3\sigma(X)\,\Big],$$

where $c_1(L^0(\mathcal{L}))$ *is the first Chern class of the U(1)-bundle* $L^0(\mathcal{L})$, $[X]$ *is the fundamental class of X,* $\chi(X)$ *is the Euler characteristic of X and* $\sigma(X)$ *is its signature.*

1.6.5. *Zero-Dimensional Seiberg-Witten Invariant and Witten's Conjecture.* Throughout this section X will denote a compact, connected, simply connected, oriented, smooth 4-manifold and we will assume that $b_2^+(X)$ is greater than 1 and odd (the reason for strengthening the assumption $b_2^+(X) > 0$ in Theorem 1.29 will be described in a moment). We now know that, for such a 4-manifold, we have available some very nice moduli spaces $\mathcal{M}(\mathcal{L}, \eta)$ and one would like to use these to produce differential-topological invariants of X *à la* Donaldson. The game plan is the same: integrate selected cohomology classes over the moduli space to produce numbers and show that orientation preserving diffeomorphisms preserve these numbers. In the case of Donaldson theory we focused on 0-dimensional moduli spaces because this was the simplest case and we will do the same for Seiberg-Witten, but now the reasons are even more compelling, as we shall see quite soon.

Thus, we fix a generic metric g and perturbation η and suppose that there exists a Spin^c structure $\mathcal{L}$ for which

$$\langle\, c_1(L^0(\mathcal{L}))^2, [X]\,\rangle = 2\chi(X) + 3\sigma(X). \tag{97}$$

This can be written as

$$\frac{\langle\, c_1(L^0(\mathcal{L}))^2, [X]\,\rangle - \sigma(X)}{4} = \frac{\chi(X) + \sigma(X)}{2} = 1 + b_2^+(X)$$

and the index theory of the Dirac operator implies that the leftmost quantity is even. Consequently, $b_2^+(X)$ must be odd. We assume that $b_2^+(X) > 1$ for the cobordism result suggested in Theorem 1.27 which is responsible for the metric independence of the invariants. Also choose an orientation for the vector space $H_+^2(X; \mathbb{R})$ and thereby orient all of the moduli spaces. Then $\mathcal{M}(\mathcal{L}, \eta)$ is either empty or a smooth, compact, oriented manifold of *dimension zero*, that is, a finite set of isolated points each of which is equipped with a sign ± 1 (there is an explicit description of the sign on page 246 of [Sal]). The *0-dimensional Seiberg-Witten invariant* , denoted

$$\mathrm{SW}_0(X, \mathcal{L}),$$

is defined to be zero if the moduli space is empty and, otherwise, it is the sum of these signs. This is, indeed, an invariant in the sense that it is independent of the choice of g and η and depends only on the isomorphism class of the Spin^c structure $\mathcal{L}$ (see Theorem 7.22, page 247, of [Sal]). Consequently, if $f : X' \to X$ is an orientation preserving diffeomorphism for which the induced map $f^* : H_+^2(X; \mathbb{R}) \to H_+^2(X'; \mathbb{R})$ preserves the chosen orientations, then the induced Spin^c structure $f^*\mathcal{L}$ obtained by pulling back the Spin$^c(4)$-bundle and the map Λ by f also satisfies (97) and, moreover, $\mathrm{SW}_0(X', f^*\mathcal{L}) = \mathrm{SW}_0(X, \mathcal{L})$.

Seiberg-Witten invariants can be defined even when the moduli space is not 0-dimensional (see Section 7.4 of [Sal] or Section 6.7 of [Mor1]), but we will not do so here and we would like to provide a few words of explanation. The empirical evidence, based on the calculation of specific examples, suggests that when $b_2^+(X) > 1$ nonzero Seiberg-Witten invariants occur *only* when the moduli space is 0-dimensional. We will say that an X with $b_2^+(X) > 1$ is of *SW-simple type* if nonzero Seiberg-Witten invariants of X occur only for 0-dimensional moduli spaces. The conjecture then is that every compact, connected, simply connected, oriented, smooth 4-manifold X with $b_2^+(X) > 1$ and odd is of SW-simple type. Any element of $H^2(X; \mathbb{Z})$ that arises as $c_1(L^0(X))$ for some Spin^c structure $\mathcal{L}$ satisfying (97) is referred to as a *SW-basic class*. Thus, SW-basic classes are just the 1^{st} Chern classes of the U(1)-bundles corresponding to Spin^c structures for which the Seiberg-Witten moduli space is 0-dimensional. With this terminology we can describe more precisely our motivation for restricting attention to $\mathrm{SW}_0(\mathcal{L}, \eta)$.

We have already mentioned that Witten was led to the Seiberg-Witten equations by a deep physical "duality" between the weak and strong coupling regimes of the TQFT constructed in [Witt2]. This duality symmetry led Witten not just to some vague sense that the Seiberg-Witten invariants should contain the same topological information as the Donaldson invariants, but rather to a very explicit formula relating the two for manifolds of KM-simple type (see Theorem 1.19). This conjecture, called *Witten's Magical Formula* by

Taubes [Taub3], essentially rewrites the Donaldson series $\mathcal{D}_X(\alpha)$ in terms of *0-dimensional* Seiberg-Witten invariants. In the following statement of the conjecture we use the notation of Theorem 1.19.

Witten's Conjecture: Let X be a compact, connected, simply connected, oriented, smooth 4-manifold with $b_2^+(X) > 1$ and odd. Then

(1) X is of SW-simple type if and only if it is of KM-simple type and, in this case,
(2) KM-basic classes coincide with the SW-basic classes, and
(3) for each basic class K_i, $i = 1, \ldots, s$, the corresponding rational number a_i, $i = 1, \ldots, s$, is given by
$$a_i = 2^{2+\frac{1}{4}(7\chi(X)+11\,\sigma(X))}\,\mathrm{SW}_0(\mathcal{L}_i),$$
where $\mathcal{L}_i$ is a Spin^c structure for which $c_1(L^0(\mathcal{L}_i)) = K_i$.

The Donaldson series can then be written
$$\mathcal{D}_X(\alpha) = e^{Q_X(\alpha,\alpha)/2}\sum_{i=1}^{s} a_i e^{K_i(\alpha)} = e^{Q_X(\alpha,\alpha)/2}\sum_{i=1}^{s} 2^{2+\frac{1}{4}(7\chi(X)+11\,\sigma(X))}\,\mathrm{SW}_0(\mathcal{L}_i)\,e^{c_1(L^0(\mathcal{L}_i))(\alpha)}.$$

Witten's Conjecture has been verified for all of the examples for which both the Donaldson and Seiberg-Witten invariants are known, but a completely general proof is still lacking. A proposal for constructing such a proof was outlined shortly after the appearance of the conjecture by Pidstrigatch and Tyurin [PT] and this proposal was taken up in a long series of technically very demanding papers by Feehan and Leness (the current state of affairs is described in [FL2]). A special case of the conjecture proved by Feehan and Leness in [FL1] was used by Kronheimer and Mrowka [KM2] to resolve a long-standing problem in knot theory.

References

[AMR] Abraham, R., J.E. Marsden, and T. Ratiu, *Manifolds, Tensor Analysis, and Applications*, Second Edition, Springer, New York, NY, 2002.

[Atiyah] Atiyah, M.F., *Geometry of Yang-Mills Fields*, Accademia Nazionale Dei Lincei, Scuola Normale Superiore, Pisa, Italy, 1979.

[AHS] Atiyah, M.F., N.J. Hitchin and I.M. Singer, Self-Duality in Four-Dimensional Riemannian Geometry, Proc. Roy. Soc. Lond. **A 362**(1978), 425-461.

[AJ] Atiyah, M.F. and L. Jeffrey, Topological Lagrangians and Cohomology, J. Geo. Phys. **7**(1990), 119-136.

[Beau] Beauville, A., *Complex Algebraic Surfaces*, Second Edition, Cambridge University Press, Cambridge, England, 1996.

[BPST] Belavin, A., A. Polyakov, A. Schwarz and Y. Tyupkin, Pseudoparticle Solutions of the Yang-Mills Equations, Phys. Lett. **59 B**(1975), 85-87.

[BGV] Berline, N., E. Getzler and M. Vergne, *Heat Kernels and Dirac Operators*, Springer, New York, NY, 2004.

[Blau] Blau, M., The Mathai-Quillen Formalism and Topological Field Theory, J. Geo. Phys. **11**(1993), 95-127.

[BT] Bott, R. and L.W. Tu, *Differential Forms in Algebraic Topology*, Springer, New York, NY, 1982.

[Delig1] Deligne, P., *et al* (Editors), *Quantum Fields and Strings: A Course for Mathematicians*, Volume 1, American Mathematical Society, Providence, RI, 1999.

[Delig2] Deligne, P., *et al* (Editors), *Quantum Fields and Strings: A Course for Mathematicians*, Volume 2, American Mathematical Society, Providence, RI, 1999.

[Don1] Donaldson, S.K., An Application of Gauge Theory to Four-Dimensional Manifolds, J. Diff. Geo. **18**(1983), 279-315.

[Don2] Donaldson, S.K., Irrationality and the h-Cobordism Conjecture, J. Diff. Geo. **26**(1987), 141-168.

[Don3] Donaldson, S.K., Polynomial Invariants for Smooth 4-Manifolds, Topology **29**(1990), 257-315.

[Don4] Donaldson, S.K., Yang-Mills Invariants of Four-Manifolds, in *Geometry of Low-Dimensional Manifolds 1: Gauge Theory and Algebraic Surfaces*, 5-40, Cambridge University Press, Cambridge, England, 1992.

[DK] Donaldson, S.K. and P. Kronheimer, *The Geometry of Four-Manifolds*, Oxford University Press, Oxford, England, 1990.

[DH] Duistermaat, J.J. and G. Heckman, On the Variation in the Cohomology of the Symplectic Form of the Reduced Phase Space, Invent. Math. **69**(1982), 250-268. Addendum, **72**(1983), 153-158.

[FL1] Feehan, P.M.N. and T.G. Leness, An SO(3)-Monopole Cobordism Formula Relating Donaldson and Seiberg-Witten Invariants, Available at https://arxiv.org/pdf/math/0203047.pdf

[FL2] Feehan, P.M.N. and T.G. Leness, Superconformal Simple Type and Witten's Conjecture, Available at https://arxiv.org/pdf/1408.5085.pdf

[FS1] Fintushel, R. and R.J. Stern, SO(3)-Connections and the Topology of 4-Manifolds, J. Diff. Geo. **20**(1984), 523-539.

[FS2] Fintushel, R. and R.J. Stern, Donaldson Invariants of 4-Manifolds with Simple Type, J. Diff. Geo. **42**(1995), 577-633.

[FU1] Freed, D.S. and K.K. Uhlenbeck, *Instantons and Four-Manifolds*, Second Edition, Springer, New York, NY, 1991.

[FU2] Freed, D.S. and K.K. Uhlenbeck, Editors, *Geometry and Quantum Field Theory*, American Mathematical Society, Providence, RI, 1991.

[Freed] Freedman, M., The Topology of Four-Dimensional Manifolds, J. Diff. Geo. **17**(1982), 357-454.

[FQ] Freedman, M. and F. Quinn, *Topology of 4-Manifolds*, Princeton University Press, Princeton, NJ, 1990.

[FM] Friedman, R. and J.W. Morgan, *Smooth Four-Manifolds and Complex Surfaces*, Springer, New York, NY, 1994.

[GJ] Glimm, J. and A. Jaffe, *Quantum Physics: A Functional Integral Point of View*, Second Edition, Springer, New York, NY, 1987.

[GS] Gompf, R.E. and A.I. Stipsicz, *4-Manifolds and Kirby Calculus*, American Mathematical Society, Providence, RI, 1999.

[Green] Greenberg, M., *Lectures on Algebraic Topology*, W.A. Benjamin, Inc., New York, NY, 1967.

[GuSt] Guillemin, V. and S. Sternberg, *Supersymmetry and Equivariant De Rham Theory*, Springer, New York, NY, 1999

[Hirsch] Hirsch, M.W., *Differential Topology*, Springer, New York, NY, 1994.

[Hitch] Hitchin, N.J., The Geometry and Topology of Moduli Spaces, in *Global Geometry and Mathematical Physics*, Lecture Notes in Mathematics 1451, 1-48, Springer, New York, NY, 1990.

[Jost] Jost, J., *Riemannian Geometry and Geometric Analysis*, Third Edition, Springer, New York, NY, 2002.

[Kirby] Kirby, R.C., *Topology of 4-Manifolds*, Lecture Notes in Mathematics 1374, Springer, New York, NY, 1989.

[KN1] Kobayashi, S. and K. Nomizu, *Foundations of Differential Geometry*, Volume 1, Wiley-Interscience, New York, NY, 1963.

[KN2] Kobayashi, S. and K. Nomizu, *Foundations of Differential Geometry*, Volume 2, Wiley-Interscience, New York, NY, 1969.

[KM1] Kronheimer, P.B. and T.S. Mrowka, Embedded Surfaces and the Structure of Donaldson's Polynomial Invariants, J. Diff. Geo. **41**(1995), 573-734.

[KM2] Kronheimer, P.B. and T.S. Mrowka, Witten's Conjecture and Property P, Geometry and Topology **8**(2004), 295-310.

[Law] Lawson, H.B., *The Theory of Gauge Fields in Four Dimensions*, American Mathematical Society, Providence, RI, 1985.

[LM] Lawson, H.B. and M-L Michelsohn, *Spin Geometry*, Princeton University Press, Princeton, NJ, 1989.

[MT] Madsen, I. and J. Thornehave, *From Calculus to Cohomology: De Rham Cohomology and Characteristic Classes*, Cambridge University Press, Cambridge, England, 1997.

[MQ] Mathai, V. and D. Quillen, Superconnections, Thom Classes and Equivariant Differential Forms, Topology **25**(1986), 85-110.

[Miln1] Milnor, J.W., On Manifolds Homeomorphic the 7-Sphere, Ann. Math. **64**(1956), 399-405.

[Miln2] Milnor, J.W., Spin Structures on Manifolds, L'Enseignement Mathématique **9**(1963), 198-203.

[Miln3] Milnor, J.W., *Topology from the Differentiable Viewpoint*, Princeton University Press, Princeton, NJ, 1997.

[MH] Milnor, J. and D. Husemoller, *Symmetric Bilinear Forms*, Springer, New York, NY, 1973.

[Moi] Moise, E., Affine Structures in 3-Manifolds V: The Triangulation Theorem and the Hauptvermutung, Ann. Math. **56**(1952), 96-114.

[Mor1] Morgan, J.W., *The Seiberg-Witten Equations and Applications to the Topology of Smooth Four-Manifolds*, Princeton University Press, Princeton, NJ, 1996.

[Mor2] Morgan, J.W., An Introduction to Gauge Theory, in *Gauge Theory and the Topology of Four-Manifolds*, 53-143, American Mathematical Society, Providence, RI, 1998.

[Nab1] Naber, G., A Survey of Donaldson Theory, in the *Proceedings of the Second International Conference on Geometry, Integrability and Quantization*, 33-71, Coral Press, Sofia, Bulgaria, 2001.

[Nab2] Naber, G., Topology, Geometry and Physics: Background for the Witten Conjecture I, J. Geo. and Sym. Phys., **2**(2004), 27-123.

[Nab3] Naber, G., Topology, Geometry and Physics: Background for the Witten Conjecture II, J. Geo. and Sym. Phys., **3**(2005), 1-83 .

[Nab4] Naber, G., *Topology, Geometry and Gauge Fields: Foundations*, Second Edition, Springer, New York, NY, 2011.

[Nab5] Naber, G., *Topology, Geometry and Gauge Fields: Interactions*, Second Edition, Springer, New York, NY, 2012.

[Nash] Nash, C., *Differential Topology and Quantum Field Theory*, Academic Press (Elsevier), San Diego, CA, 2003.

[Nicol] Nicolaescu, L.I., *Notes on Seiberg-Witten Theory*, American Mathematical Society, Providence, RI, 2000.

[O'G] O'Grady, K.G., Donaldson's Polynomials for K3 Surfaces, J. Diff. Geo. **35**(1992), 415-427.

[Park] Parker, T.H., Gauge Theories on Four-Dimensional Riemannian Manifolds, Comm. Math. Phys. **85**(1982), 563-602.

[PT] Pidstrigatch, V. and A. Tyurin, Localization of Donaldson Invariants along Seiberg-Witten Classes, Available at https://arxiv.org/pdf/dg-ga/9507004.pdf.

[Rado] Rado, T. Über den Begriff der Riemannschen Fläche, Acta Litt. Scient. Univ. Szegd. **2**(1925), 101-121.

[RS2] Reed, M. and B. Simon, *Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness*, Academic Press, New York, NY, 1975.

[Rokh] Rokhlin, V., New Results in the Theory of Four Dimensional Manifolds, Dokl. Akad. Nauk. USSR **84**(1952), 221-224.

[Sal] Salamon, D., *Spin Geometry and Seiberg-Witten Invariants*, ETH-Zürich, 1999, Available at https://people.math.ethz.ch/ salamon/PREPRINTS/witsei.pdf

[Scor] Scorpan, A., *The Wild World of 4-Manifolds*, American Mathematical Society, Providence, RI, 2005.

[SW] Seiberg, N. and E. Witten, Monopoles, Duality and Chiral Symmetry Breaking in $N = 2$ Supersymmetric QCD, Nucl. Phys. **B 431**(1994), 581-640.

[Steen] Steenrod, N., *The Topology of Fibre Bundles*, Princeton University Press, Princeton, NJ, 1951.

[Stern] Stern, R.J., Computing Donaldson Invariants, in *Gauge Theory and the Topology of Four-Manifolds*, 147-194, American Mathematical Society, Providence, RI, 1998.

[Taub1] Taubes, C.H., Self-Dual Connections on Non-Self-Dual 4-Manifolds, J. Diff. Geo. **17**(1982), 139-170.

[Taub2] Taubes, C.H., Gauge Theory on Asymptotically Periodic 4-Manifolds, J. Diff. Geo. **25**(1987), 363-430.

[Taub3] Taubes, C.H., The Seiberg-Witten Invariants (Videotape), AMS Selected Lectures in Mathematics, American Mathematical Society, Providence, RI, 1995.

[Uhl] Uhlenbeck, K.K., Removable Singularities in Yang-Mills Fields, Comm. Math. Phys. **83**(1982), 11-30.

[Warn] Warner, F.W., *Foundations of Differentiable Manifolds and Lie Groups*, Springer, New York, NY, 1983.

[White] Whitehead, J.H.C., On Simply Connected 4-Dimensional Polyhedra, Comment. Math. Helv. **22**(1949), 48-92.

[Witt1] Witten, E., Supersymmetry and Morse Theory, J. Diff. Geo. **17**(1982), 661-692.

[Witt2] Witten, E., Topological Quantum Field Theory, Comm. Math. Phys. **117**(1988), 353-386.

[Witt3] Witten, E., Monopoles and Four-Manifolds, Math. Res. Lett. **1**(1994), 769-796.

[YM] Yang, C.N. and R.L. Mills, Conservation of Isotopic Spin and Isotopic Gauge Invariance, Phys. Rev. **96**(1954), 191-195.

Index