



ELSEVIER

Nuclear Physics B 431 (1994) 3–77

NUCLEAR
PHYSICS B

A strong coupling test of S -duality

Cumrun Vafa^{a,b}, Edward Witten^b

^a *Lyman Laboratory of Physics, Harvard University, Cambridge, MA 02138, USA*

^b *School of Natural Sciences, Institute for Advanced Study, Olden Lane, Princeton, NJ 08540, USA*

Received 15 August 1994; revised 2 September 1994; accepted 5 September 1994

Abstract

By studying the partition function of $N = 4$ topologically twisted supersymmetric Yang–Mills on four-manifolds, we make an exact strong coupling test of the Montonen–Olive strong-weak duality conjecture. Unexpected and exciting links are found with two-dimensional rational conformal field theory.

1. Introduction

One of the most remarkable known quantum field theories in four dimensions is the $N = 4$ supersymmetric Yang–Mills theory. This theory has the largest possible number of supersymmetries for a four-dimensional theory without gravity. It is believed to be exactly finite and conformally invariant.

A long-standing conjecture asserts that this theory has a symmetry exchanging strong and weak coupling and exchanging electric and magnetic fields. This conjecture originated with work of Montonen and Olive, who [1] proposed a symmetry with the above properties and also exchanging the gauge group G with the dual group \hat{G} (whose weight lattice is the dual of that of G). It was soon realized that this duality was more likely to hold supersymmetrically [2] and in fact the $N = 4$ theory was seen to be the most likely candidate [3] since only in that case the elementary electrons and monopoles have the same quantum numbers. (It has recently been argued that an analog of Montonen–Olive duality does hold for a certain $N = 2$ theory with matter hypermultiplets [4]).

While Montonen–Olive duality was originally proposed as a \mathbb{Z}_2 symmetry involving the coupling constant only, the $N = 4$ theory has one more parameter that should be included, namely the θ angle. As was originally recognized in lattice models [5,6] and string theory [7,8], when the θ angle is included, it is natural to combine it with the gauge coupling constant g in a complex parameter

$$\tau = \frac{\theta}{2\pi} + \frac{4\pi i}{g^2}. \quad (1.1)$$

Then the \mathbb{Z}_2 originally proposed by Olive and Montonen can be extended to a full $\text{SL}(2, \mathbb{Z})$ symmetry acting on τ in the familiar fashion

$$\tau \rightarrow \frac{a\tau + b}{c\tau + d}; \quad (1.2)$$

where a, b, c , and d are integers with $ad - bc = 1$, so that the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (1.3)$$

has determinant 1. Indeed, $\text{SL}(2, \mathbb{Z})$ is generated by the transformations

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (1.4)$$

and

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \quad (1.5)$$

Invariance under T is the assertion that physics is periodic in θ with period 2π , and S is equivalent at $\theta = 0$ to the transformation $g^2/4\pi \rightarrow (g^2/4\pi)^{-1}$ originally proposed by Montonen and Olive.

The difficulty in testing a conjecture that relates weak coupling to strong coupling is, of course, that it is difficult to know what happens for strong coupling. Until now, tests of this conjecture have involved quantities that have no quantum corrections and can be determined exactly at the semiclassical level; one then checks that the semiclassical results or formulas are invariant under $\text{SL}(2, \mathbb{Z})$. For instance, the masses of stable particles that saturate the Bogomol'nyi–Prasad–Sommerfeld (BPS) bound are given exactly by the semiclassical result, as one can deduce [2] from the structure of the supersymmetry algebra. As explained by Sen in a recent survey [9], the $\text{SL}(2, \mathbb{Z})$ symmetry predicts the existence of BPS-saturated multimonopole bound states. Sen verified this [10] for the case of magnetic charge two in an elegant calculation that gave some of the most striking new evidence in many years for the strong-weak duality conjecture. The topological aspects of the generalization to arbitrary magnetic charge have been demonstrated by Segal [11].

Relation to string theory

The conjectured $\text{SL}(2, \mathbb{Z})$ symmetry of the $N = 4$ theory gets further appeal from the proposal that this is actually a low energy manifestation of a similar symmetry in string theory. To be precise, the conjecture [12] involves the compactification of the heterotic string theory on a six-torus, which gives a four-dimensional theory with $N = 4$ supersymmetry; the expectation value of the dilaton multiplet determines the low energy parameters. This theory is conjectured to have $\text{SL}(2, \mathbb{Z})$ symmetry acting on the dilaton multiplet. In this context, $\text{SL}(2, \mathbb{Z})$ symmetry has been called *S*-duality; it is

strikingly similar to the usual $R \leftrightarrow 1/R$ symmetry of toroidal compactification of string theory known as T -duality. The analogy helped motivate the original speculations about S -duality in string theory. T -duality has a conjectured generalization, mirror symmetry [13,14], for which there is ample evidence [15–17]. In this generalized sense T -duality relates compactification on one manifold to compactification on another manifold. It has been suggested that string-fivebrane duality [18,19] could exchange S and T -duality, perhaps shedding light on the former; see again Ref. [9].

The existing evidence for S -duality in string theory has been surveyed in Ref. [9]. It largely concerns the existence of quantities (like the low energy effective action for the axion-dilaton system) that are unaffected by quantum corrections and whose $SL(2, \mathbb{Z})$ symmetry can be verified at tree level. Also, there is the fascinating occurrence of duality of root systems both in the heterotic string and in the Montonen–Olive conjecture.

The extension of $SL(2, \mathbb{Z})$ from field theory to string theory greatly increases its potential significance. In field theory, this strong-weak duality would appear to be, at most, a curious phenomenon applying to very special field theories. In string theory, however, if this symmetry appears under toroidal compactification, then – as the radius of the torus is arbitrary – on thinking about the large volume limit it would appear that the $SL(2, \mathbb{Z})$ must be a manifestation of some property of the uncompactified theory (just as the same is believed to be true for T -duality, which appears after compactification). And if S -duality comes from a property of the uncompactified theory, it must have some significance after any compactification. Thus, if valid in string theory, S -duality should have some implications not just for special models but for the real world.

In field theory, one has to be careful in calling $SL(2, \mathbb{Z})$ a “symmetry”; it not only changes the coupling parameters but exchanges the gauge group G with the dual group \hat{G} . Only a subgroup of $SL(2, \mathbb{Z})$ maps G to itself. In toroidal compactifications of string theory, $SL(2, \mathbb{Z})$ is (if valid at all) a normal, albeit for the most part spontaneously broken, symmetry (in fact a gauge symmetry [9]) with the unusual property of being valid only quantum mechanically. That last fact suggests that S -duality may have a message that is now hard to perceive, especially since upon compactification to three dimensions, S and T duality are apparently combined into a bigger symmetry group [20], perhaps giving a unification of \hbar and α' – the parameters that control quantum mechanical and stringy corrections.

Perhaps, as in the case of T -duality and mirror symmetry, examples will be found of string theories in which the strong coupling limit of one compactification is the weak coupling limit of another. A proper general statement may well be simply that dilaton moduli space is compact and smooth except for singularities that correspond to weak coupling limits¹; in particular, the same string theory might have several very different-looking weak coupling limits.

Testing S -duality for strong coupling

Despite the substantial evidence for S -duality, it is frustrating that none of the existing tests of this strong coupling symmetry really involve the strong coupling behavior. The

¹ And perhaps also orbifold singularities where some discrete symmetries are unbroken.

purpose of the present paper is to fill this gap by developing a true strong coupling test of S -duality.

At the same time, the test we carry out will involve computations in the vacuum of the $N = 4$ theory in which the non-abelian gauge symmetry is unbroken. Existing tests have generally involved computations in vacua in which the gauge symmetry is spontaneously broken to an abelian subgroup. To test S -duality for strong coupling involves finding quantities that we can calculate for strong coupling. Most strong coupling calculations are out of reach, but exceptions are sometimes provided by quantities that can be interpreted as correlation functions in a topological field theory. $N = 2$ supersymmetric Yang–Mills theory has a twisted version that is a topological field theory [21] and the same is true for $N = 4$ [22]. The twisted theories coincide with the physical theory on a flat manifold or more generally on a hyper-Kähler manifold, but not in general.

In this paper, we will consider just one of the twisted theories, and show that its partition function for gauge fields in a given topological class is the Euler characteristic of instanton moduli space. Thus, to simplify a bit, if a_k is the Euler characteristic of the moduli space of k -instantons, and $q = \exp(2\pi i \tau)$, then the partition function is $Z(q) = \sum_k a_k q^k$. We will test S -duality by examining the modular properties of the function $Z(q)$, on various manifolds, mainly for gauge groups $SU(2)$ and $SO(3)$. For computing $Z(q)$ on various four-manifolds, we rely almost entirely on known mathematical results. For $K3$, we rely on constructions by Mukai and others [23–25]; for CP^2 we use formulas of Yoshioka and Klyachko [26–28]; for the case of blowing up a point in a Kähler manifold we use a formula of Yoshioka [26]; and for ALE spaces we use formulas of Nakajima [29]. All of these formulas give functions with modular properties.

The organization of this paper is as follows. In Section 2 we discuss the topologically twisted $N = 4$ Yang–Mills theory and explain why, in favorable conditions, its partition function is the Euler characteristic of instanton moduli space. (More generally, one needs to consider a more complicated equation that we describe in Section 2.) This section includes a self-contained review of the relevant aspects of topological theories.

In Section 3 we formulate more precisely the predictions of S -duality that we are going to test. We discuss some subtleties that may arise when discussing S -duality on a compact four-manifold; they have the effect that the partition function can transform as a modular form rather than a modular function and that it can differ by an overall factor of q^{-s} from the generating function of instanton Euler characteristics. We then go on to sharpen the S -duality conjecture to allow for non-abelian electric and magnetic flux [30]. (While this paper was in gestation, we received a paper by Girardello, Giverson, Porrati, and Zaffaroni, who similarly incorporated the discrete flux on the four-torus in the context of S -duality [31].) We propose a transformation law for the partition function computed with fixed electric and magnetic fluxes. We also determine some general constraints on exponents and singularities.

In Section 4 we begin testing the predictions. We first consider the case in which space-time is a $K3$ surface; this is a particularly nice case, as it is a hyper-Kähler

manifold so the physical and twisted models coincide. Enough is known about instanton moduli spaces on $K3$ for sufficiently many instanton numbers to allow a strong test of the S -duality. As most of the instanton moduli spaces on $K3$ can be explicitly exhibited as orbifolds, one can use techniques from orbifold constructions to compute its Euler characteristic. We not only find that the partition function of $N = 4$ super Yang–Mills theory on $K3$ is a modular object, as expected from S -duality, but in the midst of this calculation we unexpectedly encounter the partition function of bosonic strings!

We next consider \mathbf{CP}^2 using the formulas of Yoshioka and Klyachko. We find that the topological partition function has interesting modular properties but does not quite converge well enough to be modular. A natural non-holomorphic modification makes it modular [32,33]. Presumably, what is going on is that there is a holomorphic anomaly somewhat analogous to the one that arises [34] in certain two dimensional models. The two examples of $K3$ and \mathbf{CP}^2 enable us to fix some important unknown constants that appeared in Section 3.

Next we consider the case of the blow-up of a point on a four-manifold (i.e. gluing in a copy of \mathbf{CP}^2 with opposite orientation) using the formula of Yoshioka [26]; it turns out that under the blow-up the partition function essentially is multiplied by a character of the two-dimensional WZW model of $SU(2)$ at level 1! We have no idea why two-dimensional field theory makes this appearance, but at any rate the function involved is certainly modular. (We propose a natural generalization of the blowing-up formula for groups of rank bigger than one, again involving WZW characters and satisfying some non-trivial checks.)

Finally, following results of Nakajima, we consider the $U(k)$ and $SU(k)$ theories on ALE spaces, which are non-compact hyper-Kähler manifolds. Nakajima's formulas once again involve two-dimensional current algebra in a beautiful and unexpected way. His results appear not just to incorporate S -duality but to go beyond what one would expect from *field theoretic* S -duality; unfortunately we do not understand the predictions of S -duality on noncompact manifolds precisely enough to fully exploit them.

In Section 5 we attempt to generalize these results using physical arguments. Imitating a strategy used recently in Donaldson theory [35], we restrict ourselves to Kähler manifolds with $h^{2,0} \neq 0$, and make a massive perturbation of the twisted $N = 4$ theory that preserves part of the topological symmetry. We propose a formula for the partition function of the $N = 4$ theory on these manifolds which beautifully obeys the various constraints.

We conclude in Section 6 with some comments on the relation to string theory. On the one hand, we summarize the facts concerning the odd appearances in Section 4 of formulas from two-dimensional rational conformal field theory. And we pursue further the peculiar appearance of the (left-moving) bosonic string partition function in Section 4; we explain that a similar computation using the four-torus instead of $K3$ gives the (left-moving) oscillator states of a fermionic string. In a way these mysterious observations generalize an observation by Gauntlett and Harvey for the string ground state [36].

2. Twistings of supersymmetric Yang–Mills theory

Before describing how topological field theories can be constructed by twisting of $N = 4$ super Yang–Mills theory, let us recall the situation for $N = 2$ [21]. $N = 2$ super Yang–Mills has a global symmetry group $SU(2)_I$. The supercharges Q_{ai} and $\bar{Q}_{\dot{a}j}$ transform in the two-dimensional representation of this group. Working on a flat \mathbb{R}^4 , the rotation group $K = SO(4)$ is locally $SU(2)_L \times SU(2)_R$. Under $SU(2)_L \times SU(2)_R \times SU(2)_I$, the supercharges transform as $(2, 1, 2) \oplus (1, 2, 2)$.

Now, as long as we are on flat \mathbb{R}^4 , we could find an alternative embedding of K in $SU(2)_L \times SU(2)_R \times SU(2)_I$ and declare this to be the rotation group. This can be done by leaving $SU(2)_L$ undisturbed and replacing $SU(2)_R$ by a diagonal combination of $SU(2)_R \times SU(2)_I$; we will call this diagonal combination $SU(2)'_R$. The modified rotation group is hence $K' = SU(2)_L \times SU(2)'_R$. When one departs from flat \mathbb{R}^4 , either by considering a curved metric or by working on a different four-manifold altogether, one uses not the usual stress tensor T_{ij} but a modified stress tensor T'_{ij} chosen so that the corresponding rotation operators are in fact K' . The “twisted” theory so defined therefore coincides with the physical theory only when the metric is flat.

Under K' , the supercharges transform as $(2, 2) \oplus (1, 3) \oplus (1, 1)$. Let us call the $(1, 1)$ element Q . Its claim to fame is that it obeys $Q^2 = 0$, and (roughly because it has spin zero in the sense of K'), under the twisted coupling to gravity, it is conserved on an arbitrary four-manifold M . Moreover, one finds that $T'_{ij} = \{Q, \Lambda_{ij}\}$ for some Λ ; this means that if Q is interpreted as a BRST-like operator, only Q -invariant observables being considered, then the coupling to the gravitational field is a BRST commutator and the theory is a topological field theory.

The topological field theory constructed this way is quite interesting, being equivalent to Donaldson theory.

Generalization to $N = 4$

Now we come to the generalization to $N = 4$. From the above, it is clear that the main point is to pick an homomorphism of K into the global symmetry group of the theory to get a twisted Lorentz group K' . $N = 4$ supersymmetric Yang–Mills theory in four dimensions has global symmetry group $SU(4)$. The possible homomorphisms of K in $SU(4)$ can be described by telling how the **4** of $SU(4)$ transforms under K . Thus they correspond simply to four-dimensional representations of K . To get a topological field theory, we need a representation such that at least one component of the supercharge is a K' singlet. Up to an exchange of left and right, there are three four-dimensional representations with this property: (i) $(2, 2)$; (ii) $(1, 2) \oplus (1, 2)$; (iii) $(1, 2) \oplus (1, 1) \oplus (1, 1)$. Correspondingly, there are three topological field theories that one can consider. The theories corresponding to the last two of those three representations were discussed some years ago by Yamron [22]. Our interest in the present paper will be in the theory determined by the representation $(1, 2) \oplus (1, 2)$. Note that this embedding of K in $SU(4)$ commutes with a subgroup $F \cong SU(2)$ of $SU(4)$ that transforms the two copies of $(1, 2)$. This becomes a global symmetry of the twisted theory.

The supercharges, which under $K \times SU(4)$ transform as $(2, 1, 4) \oplus (1, 2, \bar{4})$, transform

under $K' \otimes F$ as $(2, 2, 2) \oplus (1, 3, 2) \oplus (1, 1, 2)$. Thus there are two K' singlets, say Q and Q' . They obey $Q^2 = (Q')^2 = \{Q, Q'\} = 0$. They transform as a doublet of F .

The gauge bosons of the $N = 4$ theory are, of course, scalars under $SU(4)$. The left and right handed fermions transform under $K \times SU(4)$ like the Q 's – so under $K' \otimes F$, they transform as $(2, 2, 2) \oplus (1, 3, 2) \oplus (1, 1, 2)$. If we ignore F , this is just two copies of the K' representation that appears in Donaldson theory. The scalars of the $N = 4$ theory transform in the six dimensional representation of K ; under $K' \times F$ they transform as $(1, 1, 3) \oplus (1, 3, 1)$. In other words, in the twisted theory these fields turn into three scalars (a triplet of F) and a self-dual antisymmetric tensor (a singlet of F). Of course, the fermions and scalars all take values in the Lie algebra of the gauge group.

With the scalars denoted as v_y , $y = 1, \dots, 6$, the bosonic part of the Lagrangian of the $N = 4$ theory is

$$L = \frac{1}{2e^2} \int d^4x \operatorname{Tr} \left(\frac{1}{2} F_{ij} F^{ij} + \sum_{i=1}^6 (D_i v_y)^2 + \sum_{1 \leq y < z \leq 6} [v_y, v_z]^2 \right), \quad (2.1)$$

plus a possible theta term

$$\frac{i\theta}{8\pi^2} \int \operatorname{Tr} F \wedge F. \quad (2.2)$$

After twisting, there is an important curvature coupling term that will emerge later.

If the twisted theory is formulated on a Kähler manifold X , some special features arise. The holonomy of the Riemannian connection on a Kähler manifold is $SU(2)_L \times U(1)_R$ (for a generic oriented Riemannian four-manifold it is $SU(2)_L \times SU(2)_R$). The twisting involves the embedding in the global symmetry group $SU(4)$ of only $U(1)_R$, not $SU(2)_R$. The global symmetry group is the subgroup of $SU(4)$ that commutes with the embedding of $U(1)_R$; it is isomorphic to $SU(2) \times SU(2) \times U(1)$. Four supercharges instead of two are invariant under the twisted holonomy group $SU(2)_L \times U(1)'_R$, so the twisted $N = 4$ theory on a Kähler manifold has four fermionic symmetries instead of two. Indeed, one of these originates from each of the four underlying supersymmetries of the $N = 4$ model.

2.1. The Euler class

The goal in the rest of this section is to explain why the partition function of the twisted $N = 4$ theory that was just described is, under suitable conditions, the Euler characteristic of instanton moduli space. This will also help us understand the deviations from this formula that occur under certain conditions. Actually, the general background is not new [37–40] and the specific issues that lead to the Euler characteristic have also been discussed previously [41–44], but we will attempt to develop the subject here in such a way as to make this paper as self-contained and readable as possible. To do so, we will first explain some very simple models, beginning with finite dimensional systems.

To begin with, we consider a compact² oriented manifold M of dimensions $d = 2n$, endowed with a real oriented vector bundle V of rank d . We choose on V a metric g_{ab} (we write $(v, w) = g_{ab}v^a w^b$) and an $\text{SO}(d)$ connection A . We will consider a system with a topological symmetry Q ($Q^2 = 0$) that carries charge one with respect to a “ghost number” operator U . There will be two multiplets. The first consists of local coordinates u^i on M (of $U = 0$) together with fermions ψ^i tangent to M , of $U = 1$. The transformation laws are

$$\begin{aligned}\delta u^i &= i\epsilon\psi^i, \\ \delta\psi^i &= 0.\end{aligned}\tag{2.3}$$

Here ϵ is an anticommuting parameter. We also define δ_0 to be the variation with ϵ removed, so for instance $\delta_0\phi^i = i\psi^i$. The second multiplet consists of an anticommuting section χ^a of V of $U = -1$, and a commuting section H^a of V ; H has $U = 0$. The transformation laws are

$$\begin{aligned}\delta\chi^a &= \epsilon H^a - \epsilon\delta_0 u^i A_i^a{}_b \chi^b, \\ \delta H^a &= \epsilon\delta_0 u^i A_i^a{}_b H^b - \frac{\epsilon}{2}\delta_0 u^i \delta_0 u^j F_{ij}{}^a{}_b \chi^b.\end{aligned}\tag{2.4}$$

Of course, one could here substitute for $\delta_0 u^i$ from (2.3); we have written the formula in this way to indicate that it is a covariantized version of the more naive $\delta\chi^a = \epsilon H^a$, $\delta H^a = 0$.

The Lagrangian is to be $L = \delta_0 W$ for a suitable W . Such a W is

$$W = \frac{1}{2\lambda}(\chi, H + 2is)\tag{2.5}$$

with s an arbitrary c -number section of V and λ a small positive real number. We get

$$L = \frac{1}{2\lambda}(H, H - 2is) + \frac{1}{\lambda}g_{ab}\chi^a \frac{\partial s^b}{\partial u^i} \psi^i - \frac{1}{2\lambda}F_{ijab}\psi^i \psi^j \chi^a \chi^b.\tag{2.6}$$

Now we want to do the integral

$$Z = \left(\frac{1}{2\pi}\right)^d \int du d\psi d\chi dH e^{-L}.\tag{2.7}$$

(The factors of 2π correspond to the standard factor of $1/\sqrt{2\pi}$ for every bosonic variable in the Feynman path integral.) This integral is guaranteed to be a topological invariant – that is, to depend only on M and V – since the derivative of L with respect to any of the other data (λ , g , A , and s) is of the form $\{Q, \dots\}$. We will call it the partition invariant of the system.

As a first step to evaluate the integral, we integrate over H , getting

² Compactness is not necessary if the section s , introduced presently, has a suitable behavior at infinity. A non-compact manifold with a suitable s will, eventually, be the main situation we study.

$$Z = \left(\frac{\lambda}{2\pi} \right)^{d/2} \int du d\psi d\chi \exp \left(-\frac{(s, s)}{2\lambda} - \frac{1}{\lambda} g_{ab} \chi^a \frac{\partial s^b}{\partial u^i} \psi^i + \frac{1}{2\lambda} F_{ijab} \psi^i \psi^j \chi^a \chi^b \right). \quad (2.8)$$

To proceed, we first consider the case $s = 0$. The integral is then done by expanding the four fermi interaction, giving

$$Z = \int_M \frac{\text{Pf}(F \wedge F \wedge \dots \wedge F)}{(2\pi)^{d/2} \cdot d!}, \quad (2.9)$$

with $\text{Pf}(F \wedge \dots \wedge F)$ the Pfaffian on the a, b indices. The curvature integral in (2.9) is a standard integral representation for a topological invariant that is known as the Euler class of V (integrated over M); we will denote it as $\chi(V)$. In case $V = TM$ is the tangent bundle of M , $\chi(V)$ coincides with the Euler characteristic $\chi(M)$ of M .

Now we consider the case of $s \neq 0$, which is much closer to our general interests in this paper. The main idea is to consider the behavior for $\lambda \rightarrow 0$. In this limit, the integral is dominated by contributions from infinitesimal neighborhoods of zeroes of s . For the first basic case, we suppose that s has only isolated and non-degenerate zeroes P_α . In that case, near each zero one can choose local coordinates on M and a trivialization of s so that $s^a = f_a u^a$ (no sum over a here and in similar formulas below), with some real numbers f_a . Higher order terms are irrelevant for small λ . Then the contribution of a particular zero is

$$\left(\frac{\lambda}{2\pi} \right)^{d/2} \prod_{a=1}^d \int du^a d\psi^a d\chi^a \exp \left(-\frac{(f_a u^a)^2}{2\lambda} + \frac{1}{\lambda} f_a \psi^a \chi^a \right) = \prod_{a=1}^d \frac{f_a}{|f_a|} = \pm 1. \quad (2.10)$$

The answer is thus

$$Z = \sum_{P_\alpha} \epsilon_\alpha, \quad (2.11)$$

where ³

$$\epsilon_\alpha = \text{sign} \left(\det \left(\frac{\partial s^a}{\partial u^i} \right) \right). \quad (2.12)$$

Since the integral is independent of s , we learn by comparing to the result for $s = 0$ that the Euler class of a bundle can be computed by counting the zeroes of a section, weighted with signs; this is a standard theorem.

In our applications, we will actually need a hybrid of the two cases considered above. Suppose that $s = 0$ on a union of submanifolds M_α of M , of dimensions d_α . We assume that the behavior of s in the normal directions to M_α is non-degenerate; this means that

³ Another way to describe this is that ϵ_α measures the relative orientation of $TM|_{P_\alpha}$ and $V|_{P_\alpha}$; regarding $ds|_{P_\alpha}$ as an isomorphism between those spaces, $\epsilon_\alpha = \pm 1$ depending on whether the orientations of the two spaces agree or disagree under this isomorphism.

locally one can pick coordinates u^i , $i = 1, \dots, d - d_\alpha$ in the directions normal to M_α and a trivialization of V such that

$$\begin{aligned} s^a &= f^a_i u^i, & \text{for } i, a = 1, \dots, d - d_\alpha \\ s^a &= 0, & \text{for } a > d - d_\alpha \end{aligned} \quad (2.13)$$

where f^a_i (with $a, i = 1, \dots, d - d_\alpha$) is an invertible $(d - d_\alpha) \times (d - d_\alpha)$ matrix. It is convenient to regard f as a $d \times d$ matrix whose other components are zero. Looking at the Lagrangian (2.6), we see that this f is the “mass matrix” for the fermions near M_α . The “massless components” of ψ are those that are tangent to M_α . The massless components of χ are in the above trivialization the χ^a for $a > d - d_\alpha$. That trivialization is valid only locally; globally the massless components of χ are sections of a vector bundle V_α over M_α .

We will adopt the following terminology: we call the χ^a (which have ghost number -1) “antighosts,” and we refer to V_α as the vector bundle of antighost zero modes.

Now we will evaluate the integral for Z , in the limit of small λ . The integral will be a sum of contributions from the various M_α . These contributions can be evaluated as follows. Fixing a particular M_α , the integral over the “massive modes,” which roughly are those “normal” to M_α , proceeds precisely as in the derivation of (2.10). One gets a Gaussian integral with bosons and fermions canceling up to sign, giving a factor of $\epsilon_\alpha = \pm 1$.⁴ Then one has the integral “tangent” to M_α . The Lagrangian (2.6) has the property that if one sets all of the “massive” fields to 0, one gets a Lagrangian of the same type, but with M replaced by M_α , V by V_α , and s by 0. The integral over the “massless” fields is thus an integral of the type that we have already seen in getting (2.9), and so equals $\chi(V_\alpha)$, the Euler class of V_α (integrated over M_α). The final result is then

$$Z = \sum_\alpha \epsilon_\alpha \chi(V_\alpha). \quad (2.14)$$

Another way to obtain the same result is to perturb s to a nearby section \tilde{s} that has isolated zeroes. For instance, on each M_α , pick a section s_α of V_α with only isolated zeroes. Regard s_α as a section of $V|_{M_\alpha}$; extend it in an arbitrary fashion to a section of V that vanishes outside a small tubular neighborhood of M_α (a neighborhood disjoint from M_β for $\beta \neq \alpha$). Then $\tilde{s} = s + \epsilon \sum_\alpha s_\alpha$ for sufficiently small ϵ vanishes precisely on points on M_α on which $s_\alpha = 0$. This gives a check on (2.14) in the following sense: evaluating (2.14) by using the zeroes of s_α to compute $\chi(V_\alpha)$ one gets the same result that one gets by evaluating (2.11) for the section \tilde{s} .

2.2. Counting solutions of an equation

It is illuminating to consider a very special case of this: counting the solutions of an equation. We will start with an elementary example. Consider a single real variable u and an equation

⁴ In the spirit of the last footnote, one can describe ϵ as a factor which, given orientations of TM and V , produces a relative orientation of TM_α and V_α . This is the data needed to fix the sign of $\chi(V_\alpha)$ below.

$$u^2 - a = 0, \quad (2.15)$$

with real a . Obviously, the number of solutions is not a topological invariant; there are two for positive a , and none for negative a .

To put this in the above format, take M to be the u -axis, V to be a one-dimensional trivial bundle, and s to be the section of V

$$s(u) = u^2 - a. \quad (2.16)$$

We want to compute the partition invariant of this system. According to (2.10), the result is

$$Z = \sum_{\alpha} \epsilon_{\alpha}, \quad (2.17)$$

with α running over the zeroes of s , and ϵ the sign of df/du at a given zero. For $a < 0$, $Z = 0$ since there are no zeroes. For $a > 0$, there are two zeroes, at $u = \pm\sqrt{a}$. Z still vanishes since $\epsilon = \pm 1$ for $u = \pm\sqrt{a}$.

The moral is, of course, that while the total number of solutions of an equation is not a topological invariant, the number of solutions weighted with signs (or in general, with multiplicities, if one encounters degenerate solutions) is such an invariant.

Suppose, however, that we want to find an integral formula that counts *without signs* the total number of solutions of an equation. This cannot really be done, as is clear from the above example. But there is a partial substitute which we will explain first in the above special case. Double the degrees of freedom, adding a new variable y and replacing the u axis by the u - y plane. Take V to be a two dimensional trivial bundle, and let s be the section of V given by the two functions

$$\begin{aligned} s_1 &= u^2 - a - y^2 \\ s_2 &= 2uy. \end{aligned} \quad (2.18)$$

Now we consider the system of equations $s = 0$, that is $s_1 = s_2 = 0$. The partition invariant Z is a topological invariant of this system which is the number of solutions of the equations *weighted by sign*. Let us compute this number – which of course can be defined as above by an integral formula – for various a .

Suppose first that $a > 0$. The equations $s_1 = s_2 = 0$ have the two solutions $u = \pm\sqrt{a}$, $y = 0$. The result of including y is that $\epsilon = 1$ for both solutions. Indeed, the determinant in (2.12) is always positive (as $\partial s_1/\partial u$ and $\partial s_2/\partial y$ have the same signs at each root of the equations). So $Z = 2$.

What about $a < 0$? There are again two solutions, now at $u = 0$, $y = \pm\sqrt{-a}$. One can verify that again the contributions are $+1$ and so $Z = 2$.

The reason that, for $y = 0$, $\epsilon = +1$ at each zero is that

$$s_2 = y \frac{\partial s_1}{\partial u} + O(y^2). \quad (2.19)$$

This ensures that for zeroes with $y = 0$ the determinant in (2.12) is positive. The conclusion is that *not for all sections s , but for those sections that obey (2.19) and vanish*

only at $y = 0$, Z is equal to the total number of solutions of the “original equations” $s_1(u, y = 0) = 0$.⁵

The general story

The generalization of this is as follows. Suppose that one is interested in counting *without signs* the solutions of some equations

$$s^a(u^i) = 0, \quad a = 1, \dots, n \quad (2.20)$$

in n variables u^1, \dots, u^n . The u^i are, in general, coordinates on some manifold M and, globally, the s^a define a section of a vector bundle V . Introduce another set of variables y_a , $a = 1, \dots, n$, and extend s^a to arbitrary functions $s^a(u^i, y_b)$ such that⁶

$$s^a(u^i, 0) = s^a(u^i) \quad (2.21)$$

Let $h_i(u^j, y_b)$ be any n additional functions such that

$$h_i(u^j, y_b) = \sum_a y_a \frac{\partial s^a}{\partial u^i} + O(y^2). \quad (2.22)$$

Consider the system of equations

$$s^a = h_i = 0. \quad (2.23)$$

Every solution of the original system (2.20) gives by taking $y = 0$ a solution of the extended system (2.23). Solutions of this kind have $\epsilon = +1$ (since $\partial h_i / \partial y_a = \partial s^a / \partial u^i$ and the determinant in (2.12) is positive). If these are the only solutions, then (2.11) reduces to

$$Z = \sum_{P_a} 1 = N, \quad (2.24)$$

with N the total number of solutions of the original equation. Thus *if all solutions of the extended system are at $y^i = 0$, then the number of solutions of the extended system, weighted by sign, is the same as the total number of solutions of the original system.* Under this restriction, therefore, we do get an integral representation for the *unweighted* number of solutions of the original equation.

Of course, the utility of all this depends on finding an interesting situation in which there is a suitable vanishing theorem. This paper will be based on such a situation.

Generalization

More generally, we will need the following variant of the above construction. Consider a system of d' equations

$$s^a = 0, \quad a = 1, \dots, d' \quad (2.25)$$

⁵ In this statement and similar statements below, we assume the zeroes are nondegenerate; otherwise one must count the multiplicities.

⁶ Geometrically, let \widehat{M} be the total space of the bundle $V \rightarrow M$, and \widehat{V} the pullback of V to \widehat{M} . The extended functions $s^a(y^i, y_b)$ define a section of $\widehat{V} \rightarrow \widehat{M}$. Other formulas below have analogous interpretations.

for d variables u^1, \dots, u^d , with $d' < d$. If everything is sufficiently generic, the solutions will consist of disjoint, smooth, compact manifolds M_α of dimension $d - d'$ and nondegenerate in the sense of (2.13). Introduce d' new variables y_b , and extend the s^a to functions $s^a(u^i, y_b)$. Introduce d additional functions h_i such that $h_i = \sum_b y_b (\partial s^b / \partial u^i) + O(y^2)$. Consider the system

$$s^a = h_i = 0 \quad (2.26)$$

of $d + d'$ equations for $d + d'$ unknowns u^i, y_a .

The partition invariant Z is the number of solutions of this system weighted by signs. The solutions that have $y = 0$ are simply the solutions of the original system $s^a = 0$, and thus consist of the union of manifolds M_α . Suppose that there is a vanishing theorem that ensures that all solutions of (2.26) are at $y = 0$. Then Z can be evaluated using (2.14), and is

$$Z = \sum_\alpha \chi(V_\alpha), \quad (2.27)$$

where V_α is the bundle of antighost zero modes along M_α . The signs ϵ_α are all +1 because of cancellation between u and y .

Moreover, in this situation, V_α has a special interpretation. Since there are two sets of equations s^a and h_i , there are two sets of antighosts, say χ^a and $\tilde{\chi}_i$. The assumption that the M_α have the expected dimension $d - d'$ and are nondegenerate means that the original antighosts χ^a have no zero modes. As for the $\tilde{\chi}_i$, they can be analyzed as follows: they are cotangent to the original target space M , and their zero modes are cotangent to the space M_α of classical solutions. So V_α is the cotangent bundle of M_α , and the Euler class $\chi(V_\alpha)$ is the same as the Euler characteristic $\chi(M_\alpha)$. Hence we can rewrite (2.27) as

$$Z = \sum_\alpha \chi(M_\alpha) = \chi(\mathcal{W}), \quad (2.28)$$

with $\mathcal{W} = \cup_\alpha M_\alpha$ the space of solutions of the original equations $F(u) = 0$. So under the above-stated restrictions, the Euler characteristic of the space of solutions of a system of equations can be given an integral representation.

2.3. Gauge invariance

The situation that we really want is a gauge invariant version of the above. So let us explain how to incorporate a symmetry group in the finite dimensional model. First we consider the general construction that counts solutions weighted by signs and then the more special construction that (given a vanishing theorem) eliminates the signs.

To incorporate a group action in these models, we assume that a compact Lie group G acts on M and V (preserving all the data such as the metric g on V and the section s). If G has dimension t , we take the dimension of M to be $d = 2n + t$ and the rank of V to be $\tilde{d} = 2n$. Thus

$$d - \tilde{d} - t = 0. \quad (2.29)$$

We introduce a field ϕ , in the adjoint representation of G , with ghost number $U = 2$. We write $\phi = \sum_{x=1}^l \phi^x T_x$ with T_x a basis of the Lie algebra of G . The action of T_x on the manifold M is described by a vector field U_x^i , and the lifting of T_x to act on the bundle V is described by the action on a section $\chi : \delta\chi^a = U_x^i D_i \chi^a + Y_x^a{}_b \chi^b$, with some Y . (Of course D_i is the covariant derivative with respect to the connection A_i on V .)

For the fermionic symmetry, we take

$$\delta\phi = 0. \quad (2.30)$$

The transformation laws of other fields are as follows. (2.3) is modified to

$$\begin{aligned} \delta u^i &= i\epsilon\psi^i \\ \delta\psi^i &= \epsilon\phi^x U_x^i \end{aligned} \quad (2.31)$$

while (2.4) is replaced by

$$\begin{aligned} \delta\chi^a &= \epsilon H^a - \epsilon\delta_0 u^i A_i^a{}_b \chi^b \\ \delta H^a &= \epsilon\delta_0 u^i A_i^a{}_b H^b - \frac{\epsilon}{2}\delta_0 u^i \delta_0 u^j F_{ij}^a{}_b \chi^b + i\epsilon\phi^x Y_x^a{}_b \chi^b. \end{aligned} \quad (2.32)$$

(This is just the gauge-covariant version of $\delta\chi = \epsilon H$, $\delta H = -\epsilon[\phi, \chi]$, which is analogous to (2.31).) It is no longer the case that $Q^2 = 0$; rather, Q^2 is equal to a gauge transformation with generator $\phi^x T_x$. (This structure gives a model of what mathematically is called equivariant cohomology; see Refs. [37,38].)

To make it possible to write a Lagrangian, we introduce another multiplet $(\bar{\phi}, \eta)$, in the adjoint representation of G , with ghost number $U = (-2, -1)$, and a transformation law analogous to the above:

$$\begin{aligned} \delta\bar{\phi} &= i\epsilon\eta \\ \delta\eta &= i\epsilon[\phi, \bar{\phi}]. \end{aligned} \quad (2.33)$$

We also pick a G -invariant metric g_{ij} on M .

Now, set

$$W = \frac{1}{2\lambda}(\chi, H + 2is) + \frac{1}{\lambda'}\bar{\phi}^x g_{ij} U_x^i \psi^j + W', \quad (2.34)$$

with λ' a new small parameter and W' consisting of possible non-minimal terms. Then define the Lagrangian

$$\begin{aligned} L = \delta_0 W &= \frac{1}{2\lambda}(H, H - 2is) + \frac{1}{\lambda}g_{ab}\chi^a \frac{\partial S^b}{\partial u^i} \psi^i - \frac{1}{2\lambda}F_{ijab}\psi^i \psi^j \chi^a \chi^b \\ &\quad - \frac{i}{2\lambda}\chi^a \chi^b \phi^x Y_{xab} + \frac{i}{\lambda'}\bar{\eta}^x g_{ij} U_x^i \psi^j + \frac{1}{\lambda'}\bar{\phi}^x g_{ij} U_x^i U_y^j \phi^y \\ &\quad + \frac{i}{2\lambda'}\bar{\phi}^x (\partial_k U_{xi} - \partial_i U_{xk}) \psi^k \psi^i + \delta_0 W'. \end{aligned} \quad (2.35)$$

Notice that $\delta L = \delta^2 W = \phi^x T_x(W) = 0$, as W is gauge-invariant.

Now we wish to study an integral

$$Z = \frac{1}{\text{Vol}(G) \cdot (2\pi)^d (-i)^t} \int d\phi d\bar{\phi} d\eta du d\psi d\chi dH e^{-L}. \quad (2.36)$$

Notice that, while all other degrees of freedom are in bose-fermi pairs and so have a natural measure, ϕ is unpaired. To make sense of the integration measure in (2.36), we pick a translation-invariant measure $d\phi$ on the Lie algebra of G . This determines a measure on the group manifold, and by $\text{Vol}(G)$ we mean the volume of G with that measure. The choice of measure therefore cancels out of the ratio $d\phi/\text{Vol}(G)$, so (2.36) has no unspecified or arbitrary normalization. It is convenient to use the chosen measure on the Lie algebra to define separately $d\bar{\phi}$ and $d\eta$ (whose product, in any case, is naturally defined without any choices). Similarly, we take the Riemannian metric g_{ij} on M to define the separate u and ψ measures – and measures on any subspaces of u 's or ψ 's.

The standard BRST argument shows that the integral in (2.36), if sufficiently well convergent, is a deformation invariant and depends only on the manifold M and bundle V , and the G action on them. There is one basic case in which this invariant is easy to determine.

That is the case in which G acts freely on M . When that happens, it is possible to reduce the gauge invariant problem on M to an ordinary problem, without gauge invariance, on the quotient $M' = M/G$. The bundle V and section s will be replaced by the objects V' and s' over M' which pull back to V, s over M . The integral (2.36) will reduce to the one on M' that counts – with signs – the number of solutions of $s' = 0$ on M' . Or equivalently, it counts with signs the number of solutions of $s = 0$ on M , up to gauge equivalence.

To justify these claims, note first that the statement that G acts freely on M implies that for any complex-valued $\phi^x \neq 0$, the vector field $\theta = \phi^x U_x^i$ has no zeroes anywhere on M and therefore the ϕ kinetic energy in (2.35), which is $|\theta|^2$, is strictly positive. Thus there are no ϕ or $\bar{\phi}$ zero modes. (If G does not act freely at least locally, there are instead ϕ and $\bar{\phi}$ zero modes, at some points on M .)

Turning this around, the expression $g_{xy} = U_x^i U_y^j g_{ij}$ is for each G orbit in M a positive definite metric on the Lie algebra of G ; it also determines a metric on the G manifold. This is simply the metric that comes from the fact that (i) G acts freely on M , so the orbits are copies of G ; (ii) the orbits are embedded in M and get an induced metric from the Riemannian metric on M . The ratio of the induced measure to the one that was chosen in defining (2.36) will be called $\sqrt{\det(g_{xy})}$. (That is really the definition of $\det(g_{xy})$; as g_{xy} is a quadratic form rather than a matrix, its determinant only makes sense as a number if there is a pre-existing measure to compare to.) If we call the volume of G with the metric induced from the embedding $\text{Vol}'(G)$, then of course

$$\text{Vol}'(G) = \sqrt{\det(g_{xy})} \cdot \text{Vol}(G). \quad (2.37)$$

Since the $\phi, \bar{\phi}$ kinetic energy is nondegenerate, one can perform the Gaussian integral over ϕ and $\bar{\phi}$. It gives a factor of

$$(2\pi\lambda')^t \det(g_{xy})^{-1}. \quad (2.38)$$

Similarly, one can integrate over η , which appears linearly in the Lagrangian. This gives a factor of

$$\left(\frac{-i}{\lambda'}\right)^t \delta(g_{ij} U_x^i \psi^j). \quad (2.39)$$

The delta function in (2.39) has the following interpretation. The vectors U_x^i generate the G action and so are tangent to the G orbits on M . The delta function in (2.39) thus projects onto the components of ψ that are normal to the group orbits. The surviving components can be interpreted as giving a section of the pullback to M of the tangent bundle of $M' = M/G$.

In fact, we can divide by the free action of G and reduce what is left of (2.36) to an integral on M' . From integrating over the G orbits, we get a factor of $\text{Vol}'(G)$, which is given in (2.37). The delta function in (2.39) is $\sqrt{\det g_{xy}} \cdot \delta(\tilde{\psi})$ where $\tilde{\psi}$ are orthonormal components of ψ tangent to the G orbits. With the Riemannian measure for $\tilde{\psi}$,

$$\int d\tilde{\psi} \delta(\tilde{\psi}) = 1. \quad (2.40)$$

In this process of eliminating components of u tangent to the G orbits by dividing by G , and eliminating components of ψ tangent to the orbits by using the delta function, the factors of λ' and $\det(g_{xy})$ and extra factors of 2π cancel out. What remains is a Q -invariant integral on M' of the standard type, with the standard measure. V and s “go along for the ride” in the above manipulations, and so are simply replaced on M' by the objects V' and s' that pull back to V and s on M .

So we can carry over all of our analysis of (2.7). When G acts freely on M , the invariant Z defined by the integral in (2.36) simply counts, with signs, the solutions of $s' = 0$ on M' , or, equivalently, the gauge orbits of solutions of $s = 0$ on M .

Counting solutions without signs

If we want to find a way to count gauge orbits of solutions *without signs*, we must imitate the special construction that led to (2.24) and (2.28).

We recall that in that discussion, we started with fields u^i , $i = 1, \dots, d$ and equations $s^a(u) = 0$, $a = 1, \dots, d'$, with $d' < d$. Then we added dual variables y_a , extended the s^a to possibly depend on y , and added dual equations $h_i = 0$. Among other things, these steps gave a system with equally many fields and equations. When the appropriate vanishing theorem holds, the partition function was the Euler characteristic of the space \tilde{M} of solutions of the original equations $s^a(u) = 0$.

In the G -invariant case, that is not quite right, because (assuming G has dimension t and acts freely) by dividing by G one could remove t degrees of freedom from u , so that there are effectively only $d - t$ fields to begin with. So to balance the fields and equations, one would need not d but $d - t$ dual equations h . Moreover, instead of the h 's being dual to the tangent space of M , they should be dual to the pullback to M of the tangent space of $M' = M/G$. The latter condition will ensure that – after we descend to M' by dividing by G – we will arrive at the construction that we have already analyzed.

A suitable set of $d - t$ dual equations can be constructed as follows. Start with any G -invariant set of d functions h_i such that

$$h_i = y_a \frac{\partial s^a}{\partial u^i} + O(y^2). \quad (2.41)$$

Let

$$L_x = U_x^i h_i. \quad (2.42)$$

Let Π be the projection operator (using the metric on M) onto the subset of h 's for which $L = 0$, and let

$$\tilde{h} = \Pi(h). \quad (2.43)$$

The desired set of $d - t$ equations is $\tilde{h} = 0$.

Suppose that a vanishing theorem ensures that the solutions of $s = \tilde{h} = 0$ are all nondegenerate and have $y = 0$. The partition invariant for the system consisting of fields u^i, y_a , equations $s = \tilde{h} = 0$, and symmetry group G can in that case be evaluated by dividing by G and using (2.28). It equals the Euler characteristic of \mathcal{W}/G , where \mathcal{W} is the space of solutions of the original equations $F(u) = 0$ for $u \in M$.

Locality

In field theory, however, the projection operator Π may be nonlocal and for that reason its use is best avoided. Instead of using this projection operator to reduce the number of equations, one can increase the number of fields, as follows. The assertion that $\Pi(h) = 0$ is equivalent to the assertion that there exists an adjoint-valued function C^x on M such that

$$h_j + C^x U_x^i g_{ij} = 0. \quad (2.44)$$

Let us include the C^x as additional fields of $U = 0$. Their ghost number one partners will be called ζ^x . The multiplet is the standard one $\delta C = i\epsilon\zeta$, $\delta\zeta = i\epsilon[\phi, C]$. Let k_i be any functions such that

$$k_j = h_j + C^x U_x^i g_{ij} + O(y^2, Cy, C^2). \quad (2.45)$$

Consider the system of equations $s^a = k_i = 0$ for fields u, y, C , with group action G . Suppose it is the case that there is a vanishing theorem ensuring that the solutions are all at $y = C = 0$. Then the partition invariant Z for this system, upon integrating out the (C, ζ) multiplet, reduces to the partition function for the system with fields u, y and equations $s^a = \tilde{h}_i = 0$. This reflects the fact that the equations $k = 0$ with C present are equivalent to the equations $\tilde{h} = 0$ with C absent. Under these conditions, we can again invoke (2.28) and conclude

$$Z = \chi(\mathcal{W}/G), \quad (2.46)$$

with, again, \mathcal{W} the space of solutions of the original equations.

We should stress that the derivation of this formula has assumed that G acts freely on M and that \mathcal{W} is a smooth nondegenerate compact manifold of the expected dimension. When these assumptions fail, the integral must be examined more closely.

2.4. Gauge theories in four dimensions

At last we have assembled the needed tools, and we turn to our real interest – four dimensional gauge theories.

We take X to be an oriented four-manifold with local coordinates x^i , $i = 1, \dots, 4$. We pick a finite dimensional gauge group G_0 , and a G_0 bundle E with a connection A and curvature $F_{ij} = \partial_i A_j - \partial_j A_i + [A_i, A_j]$. The curvature can be decomposed in self-dual and anti-self-dual pieces,

$$F_{ij}^{\pm} = \frac{1}{2} \left(F_{ij} \pm \frac{1}{2} \epsilon_{ijkl} F^{kl} \right). \quad (2.47)$$

We take G to be the group of all gauge transformations of this bundle, acting on A in the standard fashion, $D_i \rightarrow h^{-1} D_i h$, where $D_i = \partial_i + A_i$.

If G_0 is connected and simply-connected, the bundle E is determined topologically by a single integer called the instanton number. For $G_0 = \text{SU}(N)$ this is

$$k = \frac{1}{8\pi^2} \int_M \text{Tr} F \wedge F, \quad (2.48)$$

with Tr the trace in the N -dimensional representation. k can be an arbitrary integer.⁷ If G_0 is not simply connected, k is still defined, but may not be integral; for instance, for $G_0 = \text{SO}(3)$, $k \in \frac{1}{4}\mathbb{Z}$.

To apply the above constructions, we take M to be the space of connections on E . We take V to be the bundle of self-dual two forms with values in the adjoint-representation of E . A natural section of V is given by the $+$ part of the curvature:

$$s(A) = F^+(A). \quad (2.49)$$

A zero of this section is called an instanton. The space \mathcal{W} of instantons – solutions of $s = 0$ – is infinite dimensional because of gauge equivalence. However [45], the moduli space of instantons $\mathcal{M} = \mathcal{W}/G$ is finite dimensional. Under assumptions analogous to the ones that we have made in discussing the finite dimensional examples, the dimension of the moduli space is [45]

$$\dim(\mathcal{M}) = 4kh(G_0) - \dim(G_0)(1 + b_2^+). \quad (2.50)$$

Here b_2^+ is the dimension of the space of self-dual harmonic two forms, and $h(G_0)$ is the dual Coxeter number of G_0 (equal to N for $G_0 = \text{SU}(N)$).

In finite dimensions, under our usual assumptions, $\dim(\mathcal{M})$ is equal to $d - \tilde{d} - t$, with d and t the dimensions of M and G and \tilde{d} the rank of V . In the gauge theory problem, d , \tilde{d} , and t are all infinite, but the difference $d - \tilde{d} - t$ makes sense as the index of a certain elliptic operator that appears in the moduli problem [45], and still equals $\dim(\mathcal{M})$.

⁷ For any compact simple G_0 , one defines k by a curvature integral analogous to (2.48), normalized so that if G_0 is replaced by the universal cover \widehat{G}_0 of its identity component, then k is an arbitrary integer.

If E is chosen so that $\dim(\mathcal{M}) = 0$, we are in the situation in which one can try to count, with signs, the number of points in \mathcal{M} . From our finite dimensional discussion, we know exactly how to formulate an integral that will compute this quantity. We need a system with the following supermultiplets.

At ghost numbers 0, 1 we have the gauge fields A_i and ghosts ψ_i . The ghosts have the gauge and Lorentz quantum numbers of the fields and so are a one-form with values in the adjoint representation. At ghost numbers $-1, 0$, we have antighosts χ and auxiliary fields H ; they have the quantum numbers of the equations, and so are self-dual two-forms with values in the adjoint representation. At ghost number 2, we have a field ϕ with the quantum numbers of a generator of gauge transformations – that is, ϕ is a scalar field with values in the adjoint representation. At ghost numbers $-2, -1$ are the conjugate $\bar{\phi}$ of ϕ and its fermionic partner η .

The propagating bosonic fields are thus $A, \phi, \bar{\phi}$ – a gauge field and a complex scalar in the adjoint representation. This is precisely the bosonic part of the field content of $N = 2$ super Yang–Mills theory. The fermionic fields ψ, χ , and η , transforming as $(2, 2)$, $(1, 3)$, and $(1, 1)$ under $SU(2)_L \times SU(2)_R$, similarly coincide with the fermions of the topologically twisted $N = 2$ theory, as described at the outset of this section.

The bosonic part of the action can be read off from (2.35). In doing so, we set $\lambda = \lambda' = e^2$ with e the gauge coupling. We also note that $U_y^j \phi^y$ is the change in the field under a gauge transformation generated by ϕ so in the gauge theory case is determined by the formula $\delta A_i = -D_i \phi$. Finally, after eliminating the auxiliary field H , the bosonic part of the action is

$$L = \frac{1}{2e^2} \int_X \text{Tr} (|F^+|^2 + |D_i \phi|^2). \quad (2.51)$$

If we bear in mind that

$$\int_X |F^+|^2 = \frac{1}{2} \int_X \text{Tr} F_{ij} F^{ij} + \frac{1}{4} \int_X \epsilon^{ijkl} \text{Tr} F_{ij} F_{kl}, \quad (2.52)$$

we see that – except for the last term, which is a topological invariant, a multiple of the instanton number k – the bosonic part of the action (2.51) is just the standard kinetic energy. From $N = 2$ super Yang–Mills theory, we are still missing a bosonic interaction

$$\frac{1}{2e^2} \text{Tr} [\phi, \bar{\phi}]^2. \quad (2.53)$$

This will appear if for W' in (2.34) we take

$$W' = \frac{1}{2e^2} \text{Tr} \eta [\phi, \bar{\phi}]. \quad (2.54)$$

The whole topological Lagrangian (2.35) is indeed, in this situation, the twisted version of $N = 2$ super Yang–Mills theory that we recalled at the beginning of this section. The topological meaning of this theory is now clear, at least when the bundle E is such that the expected dimension of instanton moduli space is zero: this theory computes with signs the number of instantons, up to gauge transformation.

For other E , the partition function of the theory vanishes because of ghost counting. However, for $\dim(\mathcal{M}) > 0$, there are interesting BRST-invariant operators that can have non-trivial correlation functions. They are described in Ref. [21]. Their correlation functions are in fact the celebrated Donaldson invariants of four-manifolds.

Euler characteristic of moduli space

Closer to our interests in this paper is to find a way to eliminate the minus signs and compute, when $\dim(\mathcal{M}) = 0$, the total number of instanton solutions, up to gauge transformations. More generally, when $\dim(\mathcal{M}) > 0$, we want to compute the Euler characteristic of instanton moduli space.

From our finite dimensional discussion, we know how to do this as well:

(1) We introduce conjugate fields with the quantum numbers of the equations. In the present problem, the conjugate fields are a self-dual two-form B^+ with values in the adjoint representation. They come with ghosts $\tilde{\psi}^+$ with the same gauge and Lorentz quantum numbers.

(2) We introduce additional fields C with the quantum numbers of the gauge generators. In the present problem, C is a scalar field with values in the adjoint representation of G . C comes with a ghost ζ with similar quantum numbers.

(3) We extend the original equations $F^+ = 0$ to have possible dependence on the new fields. Success depends on whether the extension can be chosen so that, eventually, a suitable vanishing theorem will hold. In the present case, a suitable choice is to modify the section s to

$$s_{ij} = F^+_{ij} + \frac{1}{2}[C, B^+_{ij}] + \frac{1}{4}[B^+_{ik}, B^+_{jl}]g^{kl}. \quad (2.55)$$

The additions are needed to ultimately get a vanishing theorem – and compare to twisted $N = 4$ super Yang–Mills theory – as we will see.

(4) Finally, one needs conjugate equations. Their general structure is given in (2.45); up to first order in B and C , they are uniquely determined, but the higher order terms are arbitrary. In the present problem, the higher order terms are best set to zero. So the conjugate equations, found by interpreting (2.45) in the present situation, are $k = 0$ with

$$k_j = D^i B^+_{ij} + D_j C. \quad (2.56)$$

Associated with the conjugate equations, one adds new antighosts $\tilde{\chi}_j$, and new auxiliary fields.

Let us examine the field content of the theory. The bosonic fields are the gauge field A , a self-dual two-form B , and three scalars $C, \phi, \bar{\phi}$. The fermionic fields are two self-dual two-forms, χ and $\tilde{\psi}$, two vectors ψ and $\tilde{\chi}$, and two scalars, η and ζ . This is precisely the field content of a certain topologically twisted version of $N = 4$ super Yang–Mills theory, as described at the outset of this section.

Now let us work out the bosonic part of the action (2.35). Apart from comparing to the $N = 4$ theory, this will enable us to see the conditions for a vanishing theorem. The crucial terms are the square of the section $(|s|^2 + |k|^2)/2e^2$. Integration by parts and use of the Jacobi identity with some slightly delicate cancellations leads to the following identity:

$$\begin{aligned}
\frac{|s|^2 + |k|^2}{2e^2} &= \frac{1}{2e^2} \int_X d^4x \sqrt{g} \text{Tr} \left(\left(F^+_{ij} + \frac{1}{4} [B_{ik}, B_{jl}] g^{kl} + \frac{1}{2} [C, B_{ij}] \right)^2 \right. \\
&\quad \left. + (D^j B_{ij} + D_i C)^2 \right) \\
&= \frac{1}{2e^2} \int_X d^4x \sqrt{g} \text{Tr} \left(F^+_{ij}{}^2 + \frac{1}{4} (D_l B_{ij})^2 + (D_i C)^2 \right. \\
&\quad \left. + \frac{1}{16} [B_{ik}, B_{jk}] [B_{ir}, B_{jr}] + \frac{1}{4} [C, B_{ij}]^2 \right. \\
&\quad \left. + \frac{1}{4} B_{ij} \left(\frac{1}{6} (g_{ik} g_{jl} - g_{il} g_{jk}) R + W^+_{ijkl} \right) B_{kl} \right), \quad (2.57)
\end{aligned}$$

with R the scalar curvature of X and W^+ the self-dual part of the Weyl tensor.

The vanishing theorem

Now let us look for a suitable vanishing theorem. Of course, if there is no vanishing theorem, the theory still exists. It is just harder to study, though conceivably richer. But to study the theory, it is certainly important to understand whatever vanishing theorems do exist.

According to our general discussion, the appropriate vanishing theorem would assert that the solutions of $s = k = 0$ all have $B = C = 0$; then it follows that the topological partition function associated with the equations $s = k = 0$ has for its partition function the Euler characteristic of instanton moduli space.

The most obvious inference from (2.57) is that if the metric is such that

$$\sum_{ijkl} B_{ij} \left(\frac{1}{6} (g_{ik} g_{jl} - g_{il} g_{jk}) R + W^+_{ijkl} \right) B_{kl} > 0 \quad (2.58)$$

for any non-zero B , then any solution of $s = k = 0$ has

$$B = D_i C = F^+ = 0. \quad (2.59)$$

This is slightly less than we hoped for because we learn only that C is covariantly constant, not zero. However, the condition $D_i C = 0$ has the following significance: it means that C is covariantly constant and generates a gauge transformation that leaves the gauge field invariant, so that the gauge group G does not act freely on the space of solutions of $s = k = 0$.⁸ Instanton moduli space \mathcal{M} is singular at such points and our general assumptions fail there. Moreover, when \mathcal{M} is singular, one would want to specify exactly what one means by the Euler characteristic. Our general formal arguments really need an extension (which we do not know how to give) when such singularities occur.

At least informally, though, (2.59) can be described by saying that when (2.58) is positive definite, the argument identifying the partition function with $\chi(\mathcal{M})$ is valid if

⁸ A gauge field that admits a nonzero solution C of $D_i C = 0$ is called reducible. Such a gauge field can be interpreted as a connection with values in the subgroup G'_0 of the gauge group G_0 that leaves C invariant. For instance, for $G_0 = \text{SU}(2)$, G'_0 will be the abelian group $\text{U}(1)$.

one treats singularities of \mathcal{M} properly. This is to be contrasted with the generic situation that would prevail in the absence of a vanishing theorem: then \mathcal{M} might be perfectly smooth, yet there might be solutions of $s = k = 0$ with $B, C \neq 0$ having nothing to do with instantons.

The vanishing theorem that we have just stated, which is a nonlinear version of the one in Ref. [45], applies, for instance, to the four-sphere with its standard metric (for which $W = 0$ and $R > 0$). However, there is a severe topological limitation on its applicability, namely b_2^+ (the dimension of the space of self-dual harmonic two forms) must vanish. To see this, note the following identity for a (neutral) self-dual two-form w :

$$\int_X (D^i w_{ij})^2 = \frac{1}{4} \int_X \left((D_i w_{jk})^2 + w_{ij} \left(\frac{1}{6} (g_{ik} g_{jl} - g_{il} g_{jk}) R + W^+_{ijkl} \right) w_{kl} \right). \quad (2.60)$$

(This identity was part of the derivation of (2.57).) If w is harmonic, $D^i w_{ij} = 0$, so if the quadratic form in (2.58) is strictly positive, then $w = 0$. So if that quadratic form is positive, then $b_2^+ = 0$.

Examples with $b_2^+ = 0$ are very restricted and will not be useful in this paper because our computations will all involve gauge groups that are locally isomorphic to a product of $SU(2)$'s. For $SU(2)$ the dimension of instanton moduli space is $8k - 3(1 + b_2^+)$, and so is odd if $b_2^+ = 0$.⁹ The partition function is therefore zero for $SU(2)$ if $b_2^+ = 0$.¹⁰ Manifolds with $b_2^+ = 0$ might be of interest with other gauge groups such as $SU(3)$.

We will therefore need some further vanishing theorems, and we will discuss several variants. As a preliminary, let us consider the important case in which the quadratic form in (2.58) is positive *semi*-definite. Then from (2.60) we learn that a harmonic self-dual two-form w is covariantly constant. The existence of such a nonzero w reduces the holonomy group of X , and there are two possibilities: (1) If $b_2^+ = 1$, so there is essentially a single w , the holonomy group is reduced to $U(2)$; then X is Kähler and w is the Kähler form. (2) If $b_2^+ > 1$, the holonomy is reduced still further. The only possibility is that X is hyper-Kähler and $b_2^+ = 3$. We will later discuss more precise vanishing theorems for such manifolds.

The following simple fact will be useful. Though the right-hand side of (2.57) is not manifestly positive (unless one is given some information about the curvature of X), the equation itself shows that the right-hand side is non-negative and vanishes when and only when $s = k = 0$. Now the right-hand side of (2.57) is invariant under

$$\tau : C \rightarrow -C. \quad (2.61)$$

(This is part of the $SU(2)$ global symmetry of the $N = 4$ super Yang–Mills theory mentioned in the introduction.) Hence, given any solution of $s = k = 0$, we get a new solution by replacing C by $-C$. It follows that any solution of $s = k = 0$ has

⁹ For $SO(3)$, $4k$ is integral and the same statement holds.

¹⁰ If instanton moduli space is an odd dimensional smooth compact manifold (without boundary), its Euler characteristic vanishes. Normally, those assumptions are too optimistic. However, the Euler characteristic appears in our formulas as a curvature integral – this is explicit in (2.9) – and so the Euler characteristic in the sense we want vanishes when \mathcal{M} is of odd dimension.

$$D_i C = [C, B] = 0, \quad (2.62)$$

without any assumption on the curvature of X . Thus, either $C = 0$ or C generates a gauge transformation that acts trivially on both the gauge connection and B .

If the gauge group is (locally) $SU(2)$ (or a product of $SU(2)$'s) we can make a more precise statement since if $C \neq 0$, it breaks $SU(2)$ to an abelian subgroup $U(1)$. If $C \neq 0$, then $[C, B] = 0$ implies B lies in the same $U(1)$ so $[B_{ij}, B_{kl}] = 0$, and hence $s = 0$ implies that $F^+ = 0$. But if $b_2^+ > 0$ (which we may as well assume, if the gauge group is $SU(2)$), then for a generic metric on X , there are no abelian instantons.¹¹ So for such gauge groups, we can assume that $C = 0$.

Comparison to $N = 4$

Before resuming the discussion of vanishing theorems, it will be helpful to make a more precise comparison of the topological theory that we have been considering so far to $N = 4$ super Yang–Mills theory.

Apart from terms involving ϕ and $\bar{\phi}$, the bosonic part of the action in the topological theory is simply the right-hand side of (2.57). If we work on flat \mathbb{R}^4 and set $B_{0i} = B_i$, the right-hand side of (2.57) is

$$\frac{1}{2e^2} \int_X \text{Tr} \left(F^+_{ij}{}^2 + (D_i B_j)^2 + (D_i C)^2 + \sum_{i < j} [B_i, B_j]^2 + \sum_j [C, B_j]^2 \right). \quad (2.63)$$

This has an $O(4)$ symmetry rotating C, B_i ; it is the subgroup of the underlying $O(6)$ symmetry of the $N = 4$ theory that does not act on $\phi, \bar{\phi}$. In fact, (2.63) is precisely the bosonic part (2.1) of the action of the $N = 4$ theory with two of the scalars – ϕ and $\bar{\phi}$ – set to zero, and a θ term added.

The bosonic terms involving ϕ can be found in a fashion similar to our discussion of the $N = 2$ theory. The ϕ kinetic energy arises as in the discussion of (2.51). The other bosonic interactions, namely

$$\frac{1}{2e^2} \int_X d^4x \sqrt{g} \text{Tr} \left([C, \bar{\phi}][C, \phi] + [B_i, \bar{\phi}][B_i, \phi] - \frac{1}{4} [\phi, \bar{\phi}]^2 \right) \quad (2.64)$$

originate as in (2.53) by adding to W' some terms with the structure $\text{Tr } \eta_T [T, \bar{\phi}]$ where T ($= \bar{\phi}, C$, or B) is a bosonic field that transforms into η_T under the fermionic symmetry. In this way one gets a topological theory whose bosonic part on flat \mathbb{R}^4 agrees precisely with $N = 4$ super Yang–Mills theory. The fermionic part of the action similarly coincides on flat \mathbb{R}^4 with the $N = 4$ theory; it could hardly be otherwise given the supersymmetry.

¹¹ The first Chern class of an abelian instanton is a two-dimensional cohomology class of X that (i) is integral, and (ii) is an eigenstate of the Hodge $*$ operator with eigenvalue -1 . (The minus sign is because we take the instanton equations to be that $F^+ = 0$, so that the non-zero part of F is F^- .) If $b_2^+ > 0$ (which means that the subspace of $H^2(X)$ with eigenvalue $+1$ of $*$ is non-empty) then for a generic metric on X , there are no non-zero cohomology classes obeying conditions (i) and (ii).

Vanishing theorems on Kähler manifolds.

On a general four-manifold X , the topological theory differs from the $N = 4$ theory by the twisting that shifts the spins and by the curvature coupling on the right-hand side of (2.57). Generally, these couplings break the $O(4)$ symmetry that we noted in (2.63), leaving only the \mathbb{Z}_2 in (2.61) and a similar \mathbb{Z}_2 acting on B . There is an important case in which a larger subgroup survives. This is the case that X is a Kähler manifold. In that case, one can naturally decompose the self-dual two-form B into components of type $(2, 0)$, $(1, 1)$, and $(0, 2)$. We will write the $(1, 1)$ piece as $b\omega$, with ω the Kähler form and b a scalar field, and call the $(2, 0)$ and $(0, 2)$ pieces β and $\bar{\beta}$.

Of the $SO(4)$ symmetry of (2.63) on a flat manifold, an $O(2)$ that rotates b and C survives on a Kähler manifold. In fact, b and C are both scalars and have kinetic energy of the same form. Also, the $(1, 1)$ piece of B is in the kernel of (2.58) (this is more or less obvious from (2.60)) so the curvature term does not spoil the symmetry between b and C .

Hence the arguments that we gave above for C carry over to b , and for instance, on a Kähler manifold, with gauge group locally a product of $SU(2)$'s, we can assume that $b = 0$ in a solution of $s = k = 0$, since we have proved that assertion for C .

Now let us analyze the situation for the $(2, 0)$ and $(0, 2)$ part of B . For these components, (2.58) collapses to a positive multiple of $R \operatorname{Tr} \beta \bar{\beta}$, and so is positive if the scalar curvature R is positive. Thus, in that case $\beta = 0$. Even if R is zero rather than positive, (2.57) implies that $D_i B_{jk} = 0$, so that if not zero B is covariantly constant. For gauge group locally a product of $SU(2)$'s, it follows (unless A is gauge equivalent to the trivial connection) that $[B, B] = 0$ and hence (if s vanishes) $F^+ = 0$.¹² But for a generic Kähler metric, there are no abelian instantons (Kähler manifolds always have $b_2^+ > 0$ since the Kähler form is self-dual), contradicting the fact that $B \neq 0$ forces the connection to be abelian.

So we conclude that, for a Kähler metric with $R \geq 0$ and gauge group locally a product of $SU(2)$'s, the desired vanishing theorem holds and the partition function of the topological theory is the Euler characteristic of instanton moduli space. This important vanishing theorem applies to examples such as $K3$, \mathbb{CP}^2 , and blowups of \mathbb{CP}^2 at a small number of points.

More general Kähler manifolds

What about more general Kähler manifolds? We still have $b = C = 0$. The curvature F^+ can be usefully decomposed on a Kähler manifold in pieces $F^{p,q}$ of types (p, q) , with $(p, q) = (2, 0)$, $(1, 1)$, or $(0, 2)$. The equations $s = k = 0$ give first

$$F^{2,0} = F^{0,2} = 0, \quad (2.65)$$

that is, the connection A endows the bundle E with a holomorphic structure; second

$$\bar{D}\beta = 0, \quad (2.66)$$

that is, β is a holomorphic section of $\operatorname{End}(E) \otimes K$ (with K the canonical bundle of X), and finally

¹² If A is trivial, the stated conclusions still hold since then $F^+ = 0$ and $s = 0$ implies that $[B, B] = 0$.

$$\omega \wedge F + [\beta, \bar{\beta}] = 0 \quad (2.67)$$

(ω is the Kähler form, of type $(1,1)$; only the $(1,1)$ part of F contributes in the equation. Interpreting β as $(2,0)$ form, both terms in the equation are $(2,2)$ forms.) Analogy with other somewhat similar problems (such as the “Higgs bundle” equations [46]) suggests that the last equation can be interpreted holomorphically as a kind of stability condition for the pair (E, β) . If so, a determination of contributions – if any – to the partition function from solutions with $\beta \neq 0$ should be quite accessible.

The following is a severe constraint. The above equations have the obvious $U(1)$ symmetry

$$\beta \rightarrow e^{i\theta} \beta \quad (2.68)$$

(which is, again, a survivor of the $SO(4)$ symmetry of (2.63)). The contributions of solutions with $\beta \neq 0$ to the topological partition function would equal the number of gauge orbits of such solutions, weighted by signs, if the number is finite. If there is instead a manifold \mathcal{W} of such solutions, the contribution (according to Eq. (2.14)) is $\pm \chi(V)$ with V the bundle of antighost zero modes. The Euler class $\chi(V)$ of a $U(1)$ -equivariant bundle V can be computed by summing over fixed points of the $U(1)$ action. Thus, the only solutions with $\beta \neq 0$ that really have to be considered are those that are invariant under (2.68), up to a gauge transformation.

That is only possible if the gauge connection is reducible. For gauge group $SU(2)$, for instance, the only fixed points with $\beta \neq 0$ are abelian configurations, with a connection of the form

$$A = \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix} \quad (2.69)$$

and β of the form

$$\beta = \begin{pmatrix} 0 & 0 \\ * & 0 \end{pmatrix}. \quad (2.70)$$

Thus, the bundle E is $E \cong L \oplus L^{-1}$ with L a holomorphic line bundle; β is a holomorphic section of $K \otimes L^{-2}$. Equation (2.67) reduces in this situation to

$$\omega \wedge F = \beta \wedge \bar{\beta} \quad (2.71)$$

with now F the curvature of the connection on L . Since the right-hand side is positive, this equation requires that $[\omega] \cdot c_1(L) > 0$, where $[\omega]$ is the Kähler class and $c_1(L)$ the first Chern class of L . Conversely, if L is a line bundle with $[\omega] \cdot c_1(L) > 0$ and β is a holomorphic section of $K \otimes L^{-2}$, then a standard convexity argument shows formally that there is a unique metric on L giving a solution of (2.71).

We can now uncover a qualitative consequence of the failure of the vanishing theorem on some Kähler manifolds. Of course, when the vanishing theorem holds, solutions of the equations are instantons and necessarily have $c_2(E) > 0$. But connections of the form (2.69) have

$$c_2(E) = -c_1(L)^2, \quad (2.72)$$

and this can be negative. For instance, on a minimal surface of general type,¹³ we can obey the conditions if we take $L = K^{1/2}$ ¹⁴ and $\beta = 1$. (In fact, in that case one can pick a Kähler metric such that $[\omega] = c_1(K)$ and solve (2.71) very explicitly with F a positive multiple of ω .) This solution has instanton number

$$c_2(E) = -\frac{1}{4}c_1(K)^2. \quad (2.73)$$

Conversely, on a minimal surface of general type, if we pick a Kähler metric with $[\omega] = c_1(L)$, the instanton number of a solution of (2.71) is bounded below by (2.73).¹⁵ The Hodge index theorem states that the intersection pairing on $H^{1,1}(X)$ is of “Lorentz signature” $(+ - -, \dots, -)$. On a minimal surface of general type, $c_1(K)^2 > 0$ and hence $c_1(L) = \lambda c_1(K) + \alpha$, where λ is a real number, $\alpha \cdot c_1(K) = 0$, $\alpha \cdot \alpha < 0$. The fact that $K \otimes L^{-2}$ has a nonzero holomorphic section β implies that $\lambda \leq 1/2$; the fact that $[\omega] \cdot c_1(L) > 0$ implies $\lambda > 0$. These conditions together imply $c_1(L)^2 \leq c_1(K)^2/4$. This will be useful in Section 5.

One more small extension of the vanishing theorems will be helpful in Section 4. $K3$ with a generic complex structure has no non-trivial holomorphic line bundles. Let X be such a generic $K3$ surface with one point blown up. On X there is an exceptional divisor D produced by the blow up; the canonical bundle is $K = \mathcal{O}(D)$, with $c_1(K)^2 = -1$. Any line bundle on X is of the form $K^{\otimes n}$ for some integer n . Any Kähler form ω on X has

$$[\omega] \cdot c_1(K) = \omega \cdot [D] > 0; \quad (2.74)$$

the right-hand side is just the area of D in the Kähler metric. Let us now show that (2.71) has no non-trivial solutions on X . The line bundle L would have to be of the form $K^{\otimes n}$ as those are the only line bundles. For $K \otimes L^{-2}$ to have a non-zero holomorphic section, one needs $n \leq 0$. But in view of (2.74), $[\omega] \cdot c_1(L) > 0$ requires $n > 0$. So we get a vanishing theorem on X : one can compute via instantons. Essentially the same argument holds for $K3$ with any number of points blown up.

2.5. Singularities of instanton moduli space

In at least one respect, the above discussion is misleading. We have constantly assumed that the moduli space \mathcal{M} of solutions of the original equations is compact and non-singular. For moduli spaces of instantons, those assumptions are unrealistic.

Compactness fails because instantons can shrink to zero size. The Euler characteristic entered our problem as a kind of curvature integral, beginning with (2.9). Only for a compact manifold (or for connections of very special type) does such an integral

¹³ The condition roughly means that the line bundle K is very positive. “Most” two dimensional compact complex manifolds are of this type or obtained from such manifolds by blowing up points.

¹⁴ That is, L is a line bundle such that $L \otimes L \cong K$. Such a line bundle only exists globally if X is a spin manifold, so only if X is spin does the bundle $E = L \oplus L^{-1}$ exist and contribute for the $SU(2)$ theory. The corresponding $SO(3)$ bundle $\text{ad}(E) \cong L^2 \oplus \mathcal{O} \oplus L^{-2} \cong K \oplus \mathcal{O} \oplus K^{-1}$ (here \mathcal{O} is a trivial line bundle) always exists and contributes to the $SO(3)$ theory.

¹⁵ The argument was explained to us by D. Morrison.

reproduce the Euler characteristic as defined topologically. To facilitate computations, one would hope that the curvature integral can be interpreted as the Euler characteristic of some compactification of \mathcal{M} . At least for X a Kähler manifold, there is a natural compactification of \mathcal{M} by stable sheaves in algebraic geometry; we will very optimistically use this compactification, since our computations will be based on results borrowed from mathematicians who used it. We do not know how to justify this assumption.

Physically, the only obvious place that a compactification of \mathcal{M} would come from is string theory. To the extent that \mathcal{M} can be interpreted as a space of classical solutions of string theory, the good ultraviolet behavior of string theory should lead to a natural compactification.

Instanton moduli space may also have singularities; these arise at points in \mathcal{M} where C or B has a zero mode. For a generic metric on X , \mathcal{M} is smooth. Nevertheless, the zero modes and singularities that occur in a one-parameter family of metrics are important in verifying the formal arguments for topological invariance of the theory.

2.6. Self-conjugacy

One point that may puzzle the reader is that the twisted $N = 4$ system has $N = 2$ topological symmetry and an $SU(2)$ global symmetry; but we have so far discussed it as an $N = 1$ topological system, and (in discussing the vanishing theorems) we exhibited only a small piece of the global symmetry. Here we will fill this gap in generality. It is convenient to do so in the general context of the whole class of models that we have discussed in this section.

First we consider the general construction that counts solutions of a system of equations weighted by signs. The fields at various ghost numbers are as follows:

$$\begin{aligned}
 U=2 : & \quad \phi^x \\
 U=1 : & \quad \psi^i \\
 U=0 : & \quad u^i, H^a \\
 U=-1 : & \quad \chi^a, \eta^x \\
 U=-2 : & \quad \bar{\phi}^x.
 \end{aligned} \tag{2.75}$$

Here (u^i, ψ^i) are a multiplet of fields and ghosts; (χ^a, H^a) are a multiplet associated with the equations; and $\phi, \bar{\phi}$, and η are fields associated with the symmetry group G .

Now we consider the more detailed construction that eliminates signs when a vanishing theorem holds. In this case, there are three sets of multiplets containing fields: the “original” multiplets (u^i, ψ^i) , the “dual” multiplets $(y_a, \tilde{\psi}_a)$, and the multiplet (C^x, ζ^x) associated with the symmetries. In addition, there are auxiliary multiplets (χ^a, H^a) associated with the original equations and $(\tilde{\chi}_i, \tilde{H}_i)$ associated with the dual equations. The other fields $\phi^x, \bar{\phi}^x, \eta^x$ are unchanged from the general picture in (2.75). So we get this setup:

$$\begin{aligned}
 U=2 : & \quad \phi^x \\
 U=1 : & \quad \psi^i, \tilde{\psi}_a, \zeta^x \\
 U=0 : & \quad u^i, y_a, C^x, H^a, \tilde{H}_i
 \end{aligned}$$

$$\begin{aligned} U = -1 : & \quad \chi^a, \tilde{\chi}_i, \eta^x \\ U = -2 : & \quad \bar{\phi}^x. \end{aligned} \quad (2.76)$$

Now, (2.76) differs from the more general structure (2.75) in being self-conjugate in the following sense. The fields at ghost number $U = -1$ have the same quantum numbers as the fields at ghost number 1 (if we bear in mind that there are metrics g_{ij} and g_{ab} that can be used to raise and lower indices), and likewise the quantum numbers are the same for ghost number 2 and -2 . In this self-conjugate case, instead of our usual BRST-like operator Q of $U = 1$, we obviously could define a similar operator Q' of $U = -1$. It will soon be clear that we can take $Q^2 = (Q')^2 = \{Q, Q'\} = 0$, up to gauge transformation.

One can actually ask for more. Define an $SU(2)$ action on the fields in (2.76) such that $\phi, C, \bar{\phi}$ make a three-dimensional representation, $\psi, \tilde{\chi}$ and $\tilde{\psi}, \chi$ make up two different two-dimensional representations, and the other fields are invariant. One can arrange so that the pair Q, Q' transform in a two-dimensional representation of this $SU(2)$.

To implement this, even in a superfield language, is really quite easy. Introduce a doublet of anticommuting variables θ^A , $A = 1, 2$, transforming in a two-dimensional representation of $SU(2)$. Arrange $\phi, C, \bar{\phi}$ into an $SU(2)$ triplet $\phi_{AB} = \phi_{BA}$. The supersymmetry transformations are to be

$$Q_A = i \frac{\partial}{\partial \theta^A} - \theta^B [\phi_{AB}, \cdot], \quad (2.77)$$

with $[\phi_{AB}, \cdot]$ denoting the infinitesimal G transformation generated by the Lie algebra element ϕ_{AB} . Obviously, $\{Q_A, Q_B\} = 0$ up to a gauge transformation.

Form superfields

$$T^i = u^i + i\theta^A \psi_A^i + \frac{i\epsilon_{AB}\theta^A\theta^B}{2} \tilde{H}^i, \quad (2.78)$$

with ψ^i and $\tilde{\chi}^i$ being the components of ψ_A^i , and

$$Y^a = y^a + i\theta^A \tilde{\psi}_A^a + \frac{i\epsilon_{AB}\theta^A\theta^B}{2} H^a, \quad (2.79)$$

with $\tilde{\psi}^a$ and χ^a being the components of $\tilde{\psi}_A^a$. The transformation laws given earlier for the fields in those multiplets can be summarized by

$$\begin{aligned} \delta T^i &= -i\epsilon^A \{Q_A, T^i\} \\ \delta Y^a - \delta T^i A_i^a{}_b Y^b &= -i\epsilon^A \{Q_A, Y^a\}. \end{aligned} \quad (2.80)$$

If we combine η, ζ as an $SU(2)$ doublet η_A , then the transformation laws for the fields ϕ_{AB} and η_A associated with the gauge symmetry are

$$\begin{aligned} \delta \phi_{AB} &= i(\epsilon_A \eta_B + \epsilon_B \eta_A) \\ \delta \eta_B &= -\frac{1}{2} \epsilon_A [\phi_{BC}, \phi^{CA}]. \end{aligned} \quad (2.81)$$

Unfortunately, we do not know of a general description of the possible Q_A -invariant Lagrangians. In the case of gauge theory, of course, the standard $N = 4$ Lagrangian is one.

3. Predictions of strong-weak duality

In this section we formulate precisely the predictions of strong weak duality that we are going to test. Consider $N = 4$ super Yang–Mills theory with gauge group G ¹⁶ on flat Euclidean four-dimensional space. Then the fact that the energy-momentum tensor $T_{\mu\nu}$ is invariant under S -duality implies that if we consider the same theory on a curved background it should still respect S -duality. The simplest object to compute is the partition function of the theory. This will in general depend on the manifold, its metric $g_{\mu\nu}$, and the gauge coupling constant and θ angle which we combine as $\tau = (\theta/2\pi) + (4\pi i/g^2)$. One would expect from S -duality that (up to some universal factors) the partition function $Z_M(\tau, \bar{\tau}, g_{\mu\nu}, G)$ should transform under $\tau \rightarrow -1/\tau$ to the same object for the dual group \hat{G} ¹⁷; the transformation from G to \hat{G} is part of the original Montonen–Olive conjecture.

To test any prediction of S -duality, we need to be able to compute exact (or at least strong coupling) quantities in the theory, as S -duality relates weak to strong coupling. Unfortunately the partition function Z_M in general is too difficult to compute.¹⁸ Here is where topological twisting becomes helpful. First of all, the twisted theory should still be S -dual since twisting, as discussed in the last section, basically means introducing background fields which couple to the $SU(4)$ global symmetry current of the theory, and those currents are S -dual. The theory being topological means in particular that, barring anomalies,¹⁹ the partition function is a holomorphic function of τ and is independent of the metric on M . Thus we would formally expect the partition function on M to depend only on τ and the group chosen; we write it therefore as $Z_M(\tau, G)$. As discussed in the previous section, Z_M is determined (when the appropriate vanishing theorem holds) by the Euler characteristics of instanton moduli spaces. Suppose for simplicity that G is connected and simply-connected. Then G -bundles on M are classified by a single integer, the instanton number. If \mathcal{M}_n is the moduli space of instantons of instanton number n , χ denotes the Euler characteristic, and $q = \exp(2\pi i\tau)$, then the partition function would be²⁰

¹⁶ We are making a small change of notation relative to Section 2, where the finite dimensional gauge group was called G_0 , and the name G was reserved for the group of local gauge transformations, that is (roughly) the group of maps of space-time to G_0 .

¹⁷ Recall that \hat{G} is the group whose weight lattice is dual to that of G .

¹⁸ The case that M is a hyper-Kähler manifold is an exception as we see later.

¹⁹ As we will find later, such anomalies do occur in certain cases, as in Donaldson theory. However the discussion of modular properties of the partition function in the rest of this section still holds.

²⁰ If G is not connected and simply-connected, the classification of bundles is finer, as we discuss later, and the instanton number n may not be an integer. However, we still write the following formula, with the sum now running over all topological types of G -bundles. The discussion below is valid with obvious modifications. In any case, we systematically examine these issues later.

$$Z_M = \frac{1}{\#Z(G)} \sum_n \chi(\mathcal{M}_n) q^n \quad (3.1)$$

($\#Z(G)$ is the number of elements of the center $Z(G)$ of G ; this factor is present because $Z(G)$ acts trivially on the space of connections and one divides by it in performing the Feynman path integral.) Under the S transformation, $\tau \rightarrow -1/\tau$, how should Z_M transform? The most naive guess is that simply

$$Z_M(-1/\tau, G) = Z_M(\tau, \hat{G}),$$

in other words that Z_M is strictly invariant under S -duality.

However, two natural generalizations of this come to mind. One is the possibility that instead of being modular invariant, Z_M might transform like a “modular form,”

$$Z_M(-1/\tau, G) = \pm \left(\frac{\tau}{i}\right)^{w/2} Z_M(\tau, \hat{G}) \quad (3.2)$$

for some w (the factor of i in the denominator is there to guarantee $S^2 = 1$; this leaves room for an extra overall \pm sign). One other familiar fact suggests a modification of S -duality in curved space: the leading power of q in a modular object is not always 0. For example the Dedekind η -function has an integral expansion multiplied by $q^{1/24}$. In string theory this comes from a shift in the zero point of the energy. Similarly here we might expect a shift in the zero point of the instanton number, as a result of which the formula for the partition function should be modified by an overall multiplicative factor to be

$$Z_M = \frac{q^{-s}}{\#Z(G)} \sum_n \chi(\mathcal{M}_n) q^n \quad (3.3)$$

for some s .

To give at least some explanation of how such subtleties could arise, note that even if a Lagrangian L is S -dual in flat space and has an S -dual extension to curved space, we have to ask precisely what extension of L to curved space is S -dual. For instance, as we saw in Section 2, the BRST symmetry of the twisted theory requires the presence of some curvature couplings that one might not have guessed. Even if all q -number terms are known in the extension of L to curved space (in the twisted theory they are all determined by BRST invariance, modulo BRST commutators), one can still add c -number terms. Thus, if L_1 is one extension of the theory to curved space, another is

$$L'_1 = L_1 + \int_M d^4x \sqrt{g} (e(\tau) + f(\tau)R + f(\tau)R^2 + \dots), \quad (3.4)$$

where the terms are local operators constructed from the metric, and we have to ask whether it is L_1 or L'_1 that is S -dual. We have here made an expansion in local operators because we assume that the statement “the theory is S -dual in curved space” means that there is a *local* Lagrangian which is S -dual in curved space.

Taking L_1 to be the twisted $N = 4$ theory as formulated in Section 2, topological invariance means that the c -number terms must themselves be topological invariants.

The only topological invariants of a four-manifold that can be written as the integral of a local operator are the Euler characteristic χ and the signature σ . Thus, the unknown c -number terms must be of the form $e(\tau)\chi + f(\tau)\sigma$, with e and f being unknown functions of τ . If this is so, then Z_M defined as in (3.3) would fail to be $SL(2, \mathbb{Z})$ -invariant by a universal χ and σ -dependent factor, and with suitable e and f , this could lead to the subtleties suggested above. We now however have the additional information that we should expect the modular weight $w/2$ and the instanton shift s to be linear functions of χ and σ :

$$\begin{aligned} w &= a\chi + b\sigma \\ s &= \alpha\chi + \beta\sigma, \end{aligned} \quad (3.5)$$

with universal constants a, b, α , and β . For the physical $N = 4$ theory, the coefficients b and β would have to be zero as σ is odd under parity (which is a symmetry of the physical model when $\theta = 0$). For the twisted theory, parity is violated explicitly and it is not clear that b and β should vanish. However, it will turn out that they do.

A further subtlety arises if one does not require that s be integral. Then, although (3.1) is strictly invariant under

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad (3.6)$$

which corresponds to $\theta \rightarrow \theta + 2\pi$ or $\tau \rightarrow \tau + 1$, (3.3) would change under τ by an overall phase. One could interpret this as a kind of global gravitational anomaly in the 2π periodicity in θ – or perhaps better, as a clash between S -duality and the 2π periodicity. If this possibility is realized, then Z_M is an object somewhat like the Dedekind η function – transforming under $SL(2, \mathbb{Z})$ almost like a modular form of some particular weight, but with some additional phase factors. We will see that all of these possibilities are realized; the zero of instanton number is shifted by a multiple of χ , which is not always integral so one gets the phases just mentioned, and there is also a modular weight that is a multiple of χ .

It is convenient sometimes to get rid of the modular weight. This can be done by multiplying Z_M by η^{-w} to get

$$\hat{Z}_M = \eta^{-w} Z_M.$$

Under S -duality we have

$$\hat{Z}_M \rightarrow \pm \hat{Z}_M.$$

We denote the shift from integer power q expansion in \hat{Z}_M by $q^{-c/24}$. This definition is motivated by a similar appearance of the central charge c in two-dimensional conformal field theory. In principle, c depends on both M and G . Note that c is related to s by

$$c = 24s + w = a'\chi + b'\sigma, \quad (3.7)$$

where a', b' can be written in terms of a, b, α, β .

3.1. Fractional instanton numbers

The original Montonen–Olive conjecture states that under $S : \tau \rightarrow -1/\tau$, the gauge group G is replaced by the dual group \widehat{G} . With the exception of $G = E_8$, it is impossible for both G and \widehat{G} to be simply-connected, and therefore we are all but forced to discuss the phenomena that arise for non-simply-connected groups.

Before being general, let us discuss the important special case that $G = \mathrm{SU}(2)$; then the dual group is $\widehat{G} = \mathrm{SU}(2)/\mathbb{Z}_2 = \mathrm{SO}(3)$. Of course, $\pi_1(\mathrm{SO}(3)) = \mathbb{Z}_2$.

$\mathrm{SU}(2)$ or $\mathrm{SO}(3)$ bundles on the four-sphere are both classified by a single integer, the instanton number. That is not so on a more general four-manifold. The basic difference between $\mathrm{SU}(2)$ and $\mathrm{SO}(3)$ bundles arises first in two dimensions. On the two-sphere \mathbb{S}^2 , bundles with any gauge group G are classified by $\pi_1(G)$, so $\mathrm{SU}(2)$ bundles are trivial, but there are two types of $\mathrm{SO}(3)$ bundle, labeled by \mathbb{Z}_2 . The non-trivial $\mathrm{SO}(3)$ bundle can be described explicitly as follows. Let L be the basic $\mathrm{U}(1)$ magnetic monopole bundle of magnetic charge one. Then the $\mathrm{SU}(2)$ bundle $F = L^{1/2} \oplus L^{-1/2}$ does not exist, since $L^{1/2}$ – which would be a line bundle of magnetic charge $1/2$ – does not exist. However, the $\mathrm{SO}(3)$ bundle that would be derived from F (for instance, by taking the symmetric part of $F \otimes F$ to construct spin one from spin $1/2$) is

$$E = L \oplus \mathcal{O} \oplus L^{-1} \quad (3.8)$$

(\mathcal{O} is a trivial bundle), and does exist. The fact that E exists but there is no associated $\mathrm{SU}(2)$ bundle makes it clear that E is non-trivial.

Now, if E is an $\mathrm{SO}(3)$ bundle on a four-manifold M , then for every two sphere S in M , we can assign an element $\alpha(S) \in \mathbb{Z}_2$ that measures whether the restriction of E to S is trivial or not. $\alpha(S)$ is what 't Hooft [30] called the non-abelian magnetic flux through S . One justification for the terminology is that if S is the boundary of a three-manifold, then automatically $\alpha(S) = 0$, as one would expect for magnetic flux. Mathematically, the association $S \rightarrow \alpha(S)$ (or a slight elaboration of it if M is not simply-connected), is an element of $H^2(M, \mathbb{Z}_2)$ which is called the second Stiefel–Whitney class of E , $w_2(E)$. The $\mathrm{SO}(3)$ bundles that are associated with $\mathrm{SU}(2)$ bundles are precisely those for which $w_2 = 0$.

For instance, if E is of the form (3.8) for some line bundle L , then $\alpha(S)$ is the mod two reduction of $\langle S, c_1(L) \rangle$ (the latter is the pairing of S with the first Chern class $c_1(L)$; it vanishes if S is a boundary). Therefore, $w_2(E)$ in this situation is the mod two reduction of $c_1(L)$, a fact that will be used in Section 5.

Since $\pi_1(\mathrm{SO}(3)) = \mathbb{Z}_2$ and $\pi_3(\mathrm{SO}(3)) = \mathbb{Z}$ are the only non-trivial homotopy groups of $\mathrm{SO}(3)$ in dimensions low enough to matter, an $\mathrm{SO}(3)$ bundle E on a four-manifold M is classified by two invariants, which are $v = w_2(E)$ and the instanton number k . However, these are correlated in a perhaps surprising fashion; k is not an integer in general, but rather

$$k = -\frac{v \cdot v}{4} \pmod{1}. \quad (3.9)$$

Here $v \cdot v$ is defined as follows: one lifts v to an integral cohomology class v' and then interprets $v \cdot v$ as the usual cup product $v' \cdot v'$; it is evident that $v \cdot v/4$ is independent of

the lifting modulo 1.

At least for simply connected M , (3.9) can be proved as follows. After lifting v as above, pick a line bundle L with $c_1(L) = v'$ and let $E' = L \oplus \mathcal{O} \oplus L^{-1}$. Thus $w_2(E') = w_2(E)$. The $SU(2)$ bundle associated with E' would be $L^{1/2} \oplus L^{-1/2}$ and its instanton number would be $-c_1(L)^2/4 = -v \cdot v/4$.²¹ Thus we have proved (3.9) for a bundle with the same w_2 as E . However, any two $SO(3)$ bundles E and E' on M with the same w_2 become isomorphic if a point p is deleted from M (since then the classification of bundles involves the homotopy groups $\pi_k(SO(3))$ for $k \leq 2$, and π_1 is the only non-trivial one). They differ therefore by a topological twist that is localized near p , but the localized topological twists are the ones that can be defined on the four-sphere, and shift the instanton number by an integer.

3.2. The group and the dual group

Eq. (3.9) implies that the instanton numbers in the $SO(3)$ theory, for which we sum over all bundles with arbitrary w_2 , take values in $\mathbb{Z}/4$ but not in general in \mathbb{Z} . Therefore, while the $SU(2)$ theory on a four-manifold is invariant under $\tau \rightarrow \tau + 1$, the $SO(3)$ theory is in general only invariant under $\tau \rightarrow \tau + 4$. There is one situation in which this is improved slightly: if M is a spin manifold, then $v \cdot v$ is even for all v ; and hence the $SO(3)$ instanton numbers take values in $\mathbb{Z}/2$, so the $SO(3)$ theory is invariant under $\tau \rightarrow \tau + 2$.

Now we can understand better the precise implications of S -duality. The modular transformation $T : \tau \rightarrow \tau + 1$ maps the $SU(2)$ theory to itself (perhaps with an anomalous phase, as discussed above). The transformation $S : \tau \rightarrow -1/\tau$, according to Montonen and Olive, maps $SU(2)$ to $SO(3)$. A transformation by T^4 then maps the $SO(3)$ theory to itself, and subsequent transformation by S will map us back to $SU(2)$. So the $SU(2)$ theory will be mapped to itself by the operation ST^4S . The $SU(2)$ theory should therefore be transformed to itself by the subgroup of $SL(2, \mathbb{Z})$ generated by the matrices

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad ST^4S = \begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix}. \quad (3.10)$$

These matrices generate the subgroup of $SL(2, \mathbb{Z})$ consisting of matrices whose lower left entry is congruent to 0 modulo 4. This subgroup is known as $\Gamma_0(4)$. In the case of a spin manifold, one would get the subgroup of $SL(2, \mathbb{Z})$ generated by S and ST^2S ; this is the group $\Gamma_0(2)$ of matrices whose lower left entry is congruent to 0 modulo 2.

Therefore, the prediction of S -duality is that the $SU(2)$ (or $SO(3)$) partition function Z_M is modular for $\Gamma_0(4)$, or $\Gamma_0(2)$ for spin manifolds. Z_M is not quite a modular form

²¹ The fact that this $SU(2)$ bundle may not really exist is immaterial; the instanton number is defined by a curvature integral which can be taken either in the spin one-half representation or – with an extra factor of $1/4$ – in the spin one representation. Writing down the formal expression for E' simply lets us use a short-cut: for an $SU(2)$ bundle $N \oplus N^{-1}$ with N a line bundle, the instanton number is $-c_1(N)^2$.

for these groups, but transforms with extra phases, because we have allowed a shift in the instanton number in multiplying by q^{-s} and moreover because (3.2) is not quite the conventional definition of the transformation of a modular form. We can get rid of most of the phases by studying the function

$$f_M = \eta^{-w+c} Z = \eta^c \widehat{Z} \quad (3.11)$$

We see that f_M has weight $c/2$ and is invariant under $\tau \rightarrow \tau + 1$. It is still not quite a standard modular form unless c is an integer, but this will turn out to be the case in our examples. If c is an odd integer, we get what is called a modular form of half-integral weight.²²

Let us note some aspects of the fundamental domains of each of these two groups. These fundamental domains parametrize the inequivalent values of τ if S -duality is valid. The fundamental domain of $\Gamma_0(2)$ has two cusps and a \mathbb{Z}_2 orbifold point.²³ One cusp (say at $\tau = \infty$) corresponds to the instanton expansion for $SU(2)$ and one (say at $\tau = 0$) corresponds to the instanton expansion for $SO(3)$. The \mathbb{Z}_2 orbifold point has $\theta = \pi$ from the $SU(2)$ viewpoint. The fundamental domain of $\Gamma_0(4)$ is a two-fold cover of that of $\Gamma_0(2)$; this eliminates the orbifold point and creates a third cusp. While again two of the cusps correspond to the instanton expansions of the $SU(2)$ and $SO(3)$ theories, the meaning of the third cusp is more mysterious. We will make a proposal later: the expansion near the third cusp is the instanton expansion of the $SO(3)$ theory restricted to bundles E for which $w_2(E) = w_2(M)$. Here $w_2(M)$ is the second Stiefel–Whitney class of the tangent bundle of M and vanishes precisely if M is a spin manifold; when that occurs two of the three cusps of $\Gamma_0(4)$ are equivalent and we get back to the $\Gamma_0(2)$ picture.

Generalization to $SU(N)$

For $SU(N)$, the story is very similar. The dual group is now $SU(N)/\mathbb{Z}_N$. As $\pi_1(SU(N)/\mathbb{Z}_N) = \mathbb{Z}_N$, an $SU(N)/\mathbb{Z}_N$ bundle has in addition to the instanton number an additional invariant v taking values in $H^2(M, \mathbb{Z}_N)$. One can think of this as a \mathbb{Z}_N -valued magnetic flux. The instanton number h_v for a bundle characterized by a magnetic flux v is not necessarily an integer but obeys

$$h_v = \frac{v \cdot v}{2N} - \frac{v \cdot v}{2} \pmod{1}. \quad (3.12)$$

Note that this is invariant under $v \rightarrow v + Nv'$ and so is independent of the choice of lifting of v to an integral class.

To justify (3.12), we follow the same steps as in the $SO(3)$ case; it is enough to prove the result for some bundle with given v . Picking a line bundle L with $c_1(L) = v \pmod{N}$, we let F be the “ $SU(N)$ bundle” $F = \bigoplus_{i=1}^N L^{a_i}$, where (for instance) $a_i = 1/N$

²² These are objects that transform as θ^c under $\Gamma_0(4)$, where θ is the usual theta function. To check that f_M has the appropriate q expansion near the third cusp of $\Gamma_0(4)$ one needs the sharpened version discussed later of the S -duality conjecture. f_M may still fail to be a modular form in the usual sense if it has poles at some cusps, which may occur – as we discussed in connection with (2.71) – if bundles of negative instanton number contribute due to a failure of the vanishing theorem.

²³ See Ref. [47], p. 231, solutions to exercises 12 and 14, for this and subsequent assertions.

for $i < N$ and $a_N = -(N-1)/N$. Though F may not exist as an $SU(N)$ bundle, the corresponding $SU(N)/\mathbb{Z}_N$ bundle E' does. Its “magnetic flux” is v . The instanton number of E' (which can be computed as if F exists) is

$$v^2 \cdot \left(\frac{1}{2N} - \frac{1}{2} \right), \quad (3.13)$$

and this gives (3.12).

The vector $\vec{a} = (a_1, \dots, a_N)$ used above, with $a_i \in \mathbb{Z}_N$, $a_i - a_j \in \mathbb{Z}$, and $\sum_i a_i = 0$, can be interpreted as a vector in the weight lattice Γ_w of $SU(N)$. Γ_w contains the root lattice Γ as a sublattice of index N with $\Gamma_w/\Gamma \cong \mathbb{Z}_N$. If we replace \vec{a} by an arbitrary vector in Γ_w , the flux v of the bundle E would be $[\vec{a}] \cdot c_1(L)$, and the instanton number would be $-c_1(L)^2 \vec{a}^2/2$ modulo 1 (with $[\vec{a}]$ the coset of \vec{a} in $\Gamma_w/\Gamma \cong \mathbb{Z}_N$, and \vec{a}^2 the length squared of \vec{a}). This is similar to conformal field theory, where $SU(N)$ characters at level one are theta functions of the Γ cosets in Γ_w , and for the character derived from a given coset, $L_0 = \vec{a}^2/2$ modulo 1. The analogy would be perfect for a one-dimensional $H^2(M, \mathbb{Z})$ generated by e with $e^2 = -1$ (see the discussion below).

Just as in conformal field theory, this way of writing things leads at once to the generalization for arbitrary simply laced groups. The magnetic flux v takes values in $H^2(M, \Gamma_w/\Gamma)$, and the instanton numbers are equal modulo 1 to $-v \cdot v/2$ where the $v \cdot v$ is computed using the tensor product of the natural inner product on $H^2(M, \mathbb{Z})$ and the inner product on the weight lattice Γ_w .

3.3. Sharpening the S-duality conjecture

Let us consider the theory with $SU(N)/\mathbb{Z}_N$ gauge group. As we have discussed, an $SU(N)/\mathbb{Z}_N$ bundle on M has a “magnetic” invariant $v \in H^2(M, \mathbb{Z}_N)$. Let b_2 be the second Betti number of M . If for simplicity we suppose that M is simply-connected (or at least that there is no N -torsion in $H^2(M, \mathbb{Z})$), then the magnetic flux takes b_2^N distinct possible values. In the $SU(N)/\mathbb{Z}_N$ theory we sum over all these possibilities. We will consider the partition function of the theory with a fixed magnetic flux vector v , that is summing over all values of the instanton number with a fixed v . Let us denote this partition function by Z_v where we hide the M dependence to avoid too many labels. It is natural to ask how the individual Z_v ’s transform under $SL(2, \mathbb{Z})$. It will also be natural to study how $\widehat{Z}_v = \eta^{-w} Z_v$ transforms. The T -transformation is particularly easy. From (3.12) and from the fact that we have shifted by an overall $q^{-s} = q^{-\frac{s}{24} + \frac{w}{24}}$ we have

$$\begin{aligned} Z_v(\tau + 1) &= \exp(2\pi i(-s + h_v)) Z_v(\tau), \\ \widehat{Z}_v(\tau + 1) &= \exp(2\pi i(-\frac{c}{24} + h_v)) \widehat{Z}_v(\tau), \end{aligned} \quad (3.14)$$

where h_v is given by (3.12). More subtle is of course the S transformation.

It has been shown by ’t Hooft [30] that in the Hamiltonian formulation, analogous to the \mathbb{Z}_N -valued magnetic flux, it is natural to introduce also \mathbb{Z}_N -valued electrical fluxes. Moreover he showed that they are Fourier transforms of one another. In our setting this means that the path integral with electrical flux w is

$$Z_w^{\text{elec.}} = \text{const.} \sum_v \exp \left(\frac{2\pi i v \cdot w}{N} \right) Z_v,$$

where again the $v \cdot w$ is the inner product on $H^2(M)$. It is natural to extend the conjecture of strong/weak duality to include not only the exchange of ordinary electric and magnetic flux but also the statement that 't Hooft's electric and magnetic fluxes get exchanged under S transformation. This means in particular that the transformation laws are

$$\begin{aligned} Z_u(-1/\tau) &= \pm N^{-b_2/2} \left(\frac{\tau}{i} \right)^{w/2} \sum_v \exp \left(\frac{2\pi i v \cdot u}{N} \right) Z_v(\tau), \\ \widehat{Z}_u(-1/\tau) &= \pm N^{-b_2/2} \sum_v \exp \left(\frac{2\pi i v \cdot u}{N} \right) \widehat{Z}_v(\tau). \end{aligned} \quad (3.15)$$

The proportionality constant was fixed by requiring $S^2 = 1$, which leaves a \pm sign ambiguity in fixing S . The proof that $S^2 = 1$ requires using the fact that by Poincaré duality the cup product on the integral lattice $H^2(M, \mathbb{Z})$ is self-dual. In particular one uses

$$\sum_v \exp \left(\frac{2\pi i v \cdot u}{N} \right) = N^{b_2} \delta_{u,0} \quad (3.16)$$

The partition function for the $SU(N)$ theory is the same as the contribution of zero magnetic flux to the $SU(N)/\mathbb{Z}_N$ partition function, times an elementary factor:

$$Z_{SU(N)} = N^{-1+b_1} Z_0. \quad (3.17)$$

Here b_1 is the first Betti number of M . The prefactor in (3.17) arises as follows. The volume of the group of $SU(N)$ gauge transformations is (i) bigger than that of $SU(N)/\mathbb{Z}_N$ by a factor of N because constant gauge transformations by an element of the center of $SU(N)$ are non-trivial in $SU(N)$ but trivial in $SU(N)/\mathbb{Z}_N$, but (ii) smaller by a factor of N^{b_1} because an $SU(N)/\mathbb{Z}_N$ gauge transformation, in traversing a loop in M , may make a non-trivial loop in $\pi_1(SU(N)/\mathbb{Z}_N)$, a possibility that is absent for $SU(N)$. (For simplicity we assume that $H_1(M, \mathbb{Z})$ is torsion-free or at least that the torsion is prime to N , so that the ordinary first Betti number b_1 enters this assertion.) So the volume of the gauge group for $SU(N)$ is N^{1-b_1} times that of $SU(N)/\mathbb{Z}_N$, and the $SU(N)$ partition function is obtained by dividing by this factor the contribution Z_0 of bundles with zero flux to the $SU(N)/\mathbb{Z}_N$ partition function.

In the $SU(N)/\mathbb{Z}_N$ theory we have to sum over all allowed bundles with equal weight, which means summing over all allowed magnetic flux vectors. So

$$Z_{SU(N)/\mathbb{Z}_N} = \sum_v Z_v.$$

Thus, the original Montonen–Olive relation between $SU(N)$ and $SU(N)/\mathbb{Z}_N$, with some slight correction factors, is a consequence of (3.15):

$$\begin{aligned}
Z_{\mathrm{SU}(N)}(-1/\tau) &= \pm N^{-1+b_1-b_2/2} \left(\frac{\tau}{i}\right)^{w/2} Z_{\mathrm{SU}(N)/\mathbb{Z}_N}(\tau) \\
&= \pm N^{-\chi/2} \left(\frac{\tau}{i}\right)^{w/2} Z_{\mathrm{SU}(N)/\mathbb{Z}_N}(\tau).
\end{aligned} \tag{3.18}$$

The third cusp

Using (3.15), we can determine the meaning of the third cusp for the $\mathrm{SU}(2)$ or $\mathrm{SO}(3)$ theory. The $\mathrm{SL}(2, \mathbb{Z})$ transformation ST^2S takes the $\mathrm{SU}(2)$ cusp to the other cusp. Since the partition function at the $\mathrm{SU}(2)$ cusp is proportional to Z_0 , applying ST^2S to it with the help of (3.15) and (3.14) (and setting $N = 2$) we get

$$Z_0 \rightarrow \frac{1}{2^{b_2}} \sum_{v,u} (-1)^{v \cdot (v+u)} Z_u = \frac{1}{2^{b_2}} \sum_{u,v} (-1)^{v \cdot (w_2(M)+u)} Z_u = Z_{w_2(M)}. \tag{3.19}$$

Here we have used the Wu formula (see Ref. [48] for a quick proof in the simply-connected case) which asserts that $v^2 = v \cdot w_2(M) \bmod 2$ for every vector $v \in H^2(M, \mathbb{Z}_2)$, where $w_2(M)$ is the second Stiefel–Whitney class of M . Thus the expansion at the third cusp gives the partition function with $v = w_2(M)$.

An interesting self-dual group

The transformation property (3.15) can be in part tested (and was originally guessed) by the following considerations. Consider the group

$$\frac{\mathrm{SU}(N) \times \mathrm{SU}(N)}{\mathbb{Z}_N},$$

where the \mathbb{Z}_N is embedded diagonally in the product of the centers of the two $\mathrm{SU}(N)$'s. This is a self-dual group. Its weight lattice is the sublattice of the product of the two $\mathrm{SU}(N)$ weight lattices given by the condition that the difference of weights is in the root lattice; this is a self-dual lattice. Note that since the \mathbb{Z}_N is embedded diagonally in $\mathrm{SU}(N) \times \mathrm{SU}(N)$, an $(\mathrm{SU}(N) \times \mathrm{SU}(N))/\mathbb{Z}_N$ bundle is an $(\mathrm{SU}(N)/\mathbb{Z}_N)^2$ bundle such that the two magnetic flux vectors are equal. Allowing for the possibility of different coupling constants for the two $\mathrm{SU}(N)$'s, the partition function of the $\mathrm{SU}(N) \times \mathrm{SU}(N)/\mathbb{Z}_N$ theory is

$$Z_{\mathrm{SU}(N) \times \mathrm{SU}(N)/\mathbb{Z}_N} = N^{-1+b_1} \sum_v Z_v(q_1) Z_v(q_2),$$

where we have denoted the instanton counting parameters as q_1 and q_2 . (The prefactor has the same origin as in (3.17).) Since the group is self-dual, S -duality says that Z should be invariant (up to a factor associated with the modular weight) under simultaneous transformations $\tau_i \rightarrow -1/\tau_i$ for $i = 1, 2$. In order to have a nice action also for T , it is convenient to choose $\tau_2 = \bar{\tau}_1$. This in particular implies that the θ angles have opposite signs and therefore the fractionality of the instanton number (3.9) cancels between the two groups; so the partition function is invariant under $\tau \rightarrow \tau + 1$. With this choice the partition function is therefore $\mathrm{SL}(2, \mathbb{Z})$ -invariant up to a factor involving the modular weight; in what follows we divide by a power of $|\eta|^{2w}$ to remove this factor. The partition function so corrected is

$$\widehat{Z}_{\mathrm{SU}(N) \times \mathrm{SU}(N)/\mathbb{Z}_N} = N^{-1+b_1} \sum_v |\widehat{Z}_v(q)|^2.$$

So we learn that the “partition functions” \widehat{Z}_v transform as a unitary representation of $\mathrm{SL}(2, \mathbb{Z})$. This is a consequence of Eqs. (3.14) and (3.15), as we will verify below; that is an important check on our ansatz.

Analogy with rational conformal field theory

This structure is reminiscent of rational conformal field theory, with the \widehat{Z}_v playing the role of the conformal blocks. To pursue the analogy, consider for simplicity the case that $H^2(M, \mathbb{Z})$ is one dimensional with the lattice generated by a vector e with $e^2 = -1$. (This situation arises for \mathbf{CP}^2 with the opposite of the usual complex orientation or in connection with the blow-up of a point on a complex surface.) Then the partition functions are Z_{re} for $r = 0, \dots, N-1$.

This is reminiscent of the partition function of a two-dimensional rational conformal theory with N blocks. In fact the T transformation (3.14) in this case is exactly the same as for the $\mathrm{SU}(N)$ level 1 WZW model; this model has one block for each conjugacy class of $\mathrm{SU}(N)$ and the conformal weight for each one is given (mod 1) by $w^2/2$ where w is a weight vector in that conjugacy class – precisely the quantity appearing in (3.12). Therefore they have the same T matrix (up to at most an overall phase coming from c). Meanwhile, S is strongly constrained from the condition that $(ST)^3 = 1$ and in fact the transformation law under S in (3.15) coincides (up to an overall sign) with that of the $\mathrm{SU}(N)$ WZW theory at level 1.

This does not necessarily imply that the \widehat{Z}_{re} ’s equal the characters of $\mathrm{SU}(N)$ at level 1, but only that they transform the same way. In Section 4, we will see that in a particular case – involving blow-ups and $G = \mathrm{SU}(2)$ – these functions do coincide.

Verification of modular behavior

We now return to the general case of an arbitrary four-manifold M (with no N -torsion in $H^2(M, \mathbb{Z})$, for simplicity). We wish to verify that the formulas (3.14) and (3.15) give a unitary representation of $\mathrm{SL}(2, \mathbb{Z})$. The non-trivial point is to verify that $(ST)^3 = 1$. (Given the known structure of T , this equation highly constrains S and under some assumptions on M uniquely determines it; in that sense our ansatz in (3.15) can nearly be derived from conventional S -duality.)

In verifying that $(ST)^3 = 1$, we will also find some interesting restrictions on c . In particular we will find that $c = (N-1)\chi \bmod 4$ for $\mathrm{SU}(N)$. Later by studying certain examples we will prove that for $\mathrm{SU}(2)$, $c = \chi$ thus suggesting that the $\mathrm{SU}(N)$ answer is also $c = (N-1)\chi$. The same mod 4 condition can be shown for simply laced groups with $(N-1)$ replaced by the rank of the group. For non-simply laced group G , it is tempting given the analogy with conformal field theory to conjecture that $c = c_1(G)\chi$ where $c_1(G) = \dim G / (1 + h(G))$ is the central charge of the WZW theory with target G at level 1 ($h(G)$ is the dual Coxeter number of G).

The $\{\widehat{Z}_v\}$, form an N^{b_2} dimensional representation of $\mathrm{SL}(2, \mathbb{Z})$ that is described explicitly in (3.14) and (3.15). One can use those formulas to check the statement $(ST)^3 = 1$. When $H^2(M, \mathbb{Z})$ is one-dimensional the fact that the partition functions

transform the same way as the characters of $SU(N)$ at level 1 implies that $(ST)^3 = 1$ (with a suitable choice of c). It is easy to generalize this to the arbitrary case. Using the self-duality of $H^2(M, \mathbb{Z})$ (and in particular using (3.16)) we find that $(ST)^3 = 1$ holds up to an overall factor, which disappears if

$$(\exp(2\pi ic/24))^3 = \pm N^{-b_2/2} \sum_v (-1)^{v^2} \omega^{v^2/2}, \quad (3.20)$$

where $\omega = \exp(2\pi i/N)$. We will refer to the right hand side of (3.20) (with the + sign) as $A_N(L)$. $A_N(L)$ depends on N and on the lattice $L = H^2(M, \mathbb{Z})$ (with its canonical quadratic form). We would like to compute $A_N(L)$ for an arbitrary self-dual lattice L and verify (3.20).²⁴ It is easy to prove the following properties for $A_N(L)$:

$$\begin{aligned} |A_N(L)| &= 1, \\ A_N(-L) &= \overline{A_N(L)}, \\ A_N(L_1 \oplus L_2) &= A_N(L_1) A_N(L_2). \end{aligned} \quad (3.21)$$

Here $-L$ denotes the same lattice as L except with the opposite sign for the quadratic form. Let I^\pm denote the one-dimensional lattices generated by a vector v with $v^2 = \pm 1$. Then [49] every indefinite odd self-dual lattice²⁵ is a direct sum of the form

$$r_+ I^+ \oplus r_- I^-.$$

So we learn from (3.21) that for such lattices

$$A_N(L) = A_N(I^+)^{r_+} A_N(I^-)^{r_-} = A_N(I^+)^{r_+ - r_-} = A_N(I^+)^{\sigma(L)}, \quad (3.22)$$

where $\sigma = r_+ - r_-$ is the signature of the lattice. Moreover, it is known [49] that every self-dual lattice can be transformed to an odd, indefinite self-dual lattice by taking a direct sum, if necessary, with I^+ or I^- . Therefore using (3.21) we learn that (3.22) is true for all lattices, and we are left just with computing $A_N(I^+)$:

$$A_N(I^+) = \frac{1}{\sqrt{N}} \sum_{s=0}^{N-1} (-1)^s \exp(2\pi i s^2/2N).$$

This sum can be computed using the Poisson resummation technique (the sum is a simple generalization of so-called “Gauss sums” (see appendix 4 of Ref. [50])). We learn

$$A_N(I^+) = \exp \frac{-2i\pi(N-1)}{8}, \quad (3.23)$$

which implies

$$\exp \frac{2\pi ic}{8} = \pm \exp \frac{-2i\pi(N-1)\sigma}{8} \quad (3.24)$$

²⁴ How to compute such sums is explained in Ref. [49], as was pointed out to us by Dick Gross.

²⁵ An odd lattice is simply a lattice such that $v \cdot v$ is odd for some v .

and so

$$c = -(N-1)\sigma + 4\epsilon \quad \text{mod } 8,$$

where $\epsilon = 0, 1$ depending on the \pm in (3.24). We can rewrite c as

$$c = (N-1)\chi - (N-1)(\chi + \sigma) + 4\epsilon \quad \text{mod } 8.$$

Note that if N is odd, then $N-1$ is even; and since $\chi + \sigma = 0 \text{ mod } 2$, $(N-1)(\chi + \sigma) = 0 \text{ mod } 4$. For N even, the partition function vanishes unless $\chi + \sigma = 0 \text{ mod } 4$ (otherwise instanton moduli space is odd dimensional and its Euler characteristic in the relevant sense is zero). Thus in either case for a non-trivial partition function we have $(N-1)(\chi + \sigma) = 0 \text{ mod } 4$, and so

$$c = (N-1)\chi \quad \text{mod } 4. \quad (3.25)$$

We will see in Section 4 that at least for $N = 2$, this equation for c is true identically and not just mod 4. This will in particular imply

$$\epsilon = \frac{(N-1)(\chi + \sigma)}{4} \quad \text{mod } 2. \quad (3.26)$$

Note that we have thus determined the \pm sign in (3.15) to be $(-1)^\epsilon$. In the case of $SU(2)$ note that $(-1)^\epsilon = (-1)^\nu$ where $\nu = (\chi + \sigma)/4$. We will use this fact later on.

4. Testing the predictions of S -duality

In this section we will test the predictions made in the previous section in a small but satisfying set of examples. The examples we consider, mostly with gauge group $SU(2)$, are, in turn, $K3$, \mathbf{CP}^2 , blow ups of Kähler manifolds, and ALE spaces. These examples not only provide strong coupling tests of S -duality but also fix the quantities c and w which were not completely fixed in Section 3 to be for $SU(2)$ $c = \chi$ and $w = -\chi$. We will also find a little bit of a surprise in the context of our discussion of \mathbf{CP}^2 . We find that the S -duality predicts that there are holomorphic anomalies (a dependence of the partition function on topologically trivial observables) in the topologically twisted $N = 4$ super Yang–Mills theory. These holomorphic anomalies are somewhat reminiscent of holomorphic anomalies that arise for certain two-dimensional topological theories [34]. There also may be a relation to a sort of anomaly that arises in Donaldson theory on certain manifolds. Both in the case of blowups and for ALE spaces, the analogy with rational conformal field theory that we have seen already will reappear.

4.1. $N = 4$ Yang–Mills on $K3$

For our first example, we study the partition function of $N = 4$ Yang–Mills theory on $K3$. This turns out to be computable because of the existence of nice constructions of instantons on $K3$. It is also a particularly nice example because – as it is hyper-Kähler – the physical model coincides with the topological model if one orients $K3$ correctly (so

that holomorphic vector bundles are instantons). Furthermore, the vanishing theorem of Section 2 applies to $K3$, so we can expect to compute just in terms of instantons.

In general, the dimension of the $SU(2)$ instanton moduli space \mathcal{M}_k of instanton number k is

$$\dim \mathcal{M}_k = 8k - \frac{3}{2}(\chi + \sigma), \quad (4.1)$$

where χ and σ are the Euler characteristic and the signature of the manifold respectively. For $K3$, one has $\chi = 24$ and $\sigma = -16$, so

$$\dim \mathcal{M}_k = 8k - 12.$$

If E is an $SU(2)$ instanton bundle on $M = K3$ with instanton number k , one can seek [23,24] a convenient description of E by finding a line bundle L on $K3$ such that the index of the $\bar{\partial}$ operator coupled to $L^{-1} \otimes E$ is 1.²⁶ It will then be generically true that $H^0(M, L^{-1} \otimes E)$ has (up to scalar multiple) a single holomorphic section s . The number of zeroes of s , counted with multiplicity, is equal to the second Chern class of $L^{-1} \otimes E$, which (with L and E as stated) turns out to be $2k - 3$. Thus, if L exists, one has a natural way to extract from E a configuration of $2k - 3$ points on $K3$, namely the zeroes of s . Conversely (when an appropriate L exists), given a configuration of $2k - 3$ points on $K3$, one can reconstruct a unique E , using a process of extension of sheaves due originally to Serre.

A configuration of $2k - 3$ points on the four-manifold M depends on $8k - 12$ real parameters. This number equals the dimension of instanton moduli space as given above. In fact it is true in this situation that the instanton moduli space can be identified with the space of configurations of $2k - 3$ distinct (but unordered) points on $K3$, and is, in particular, hyper-Kähler. If the points are permitted to coincide, one gets a compact hyper-Kähler manifold with orbifold singularities. A resolution of the singularities of that moduli space, preserving the hyper-Kähler structure, gives the algebrogeometric compactification of instanton moduli space. Such resolutions all have the same Betti numbers and Euler characteristic and can be studied by standard orbifold methods that we use below.

A line bundle L with the properties needed for the above construction exists if and only if k is odd and a suitable complex structure is picked on $K3$. (For different k 's one needs different complex structures.) The restriction on the complex structure will not cause difficulties, because the partition function we are trying to compute is independent of the complex structure. The restriction to odd k will result in the appearance of an extra function below.

The construction just sketched can also be carried out for $SO(3)$ bundles with non-zero w_2 . In this case, a suitable L always exists (if the complex structure on $K3$ is suitably chosen). The instanton number k might be half-integral ($K3$ is a spin manifold so a denominator of 4 cannot arise), but $2k - 3$ is always integral and the moduli spaces with $w_2 \neq 0$ have the Euler characteristic of (a hyper-Kähler resolution of) the symmetric product of $2k - 3$ copies of $K3$.

²⁶ These matters were explained to us by P. Kronheimer.

Even though we do not have any convenient description of the moduli spaces with $w_2 = 0$ and even k , we will be able from the facts stated above both to test S -duality and to use it to predict the Euler characteristics in the missing cases.

SO(3) bundles on $K3$

We will need to know some facts about $SO(3)$ bundles on $K3$. Of course, in addition to the instanton number, such a bundle E is classified topologically by the “magnetic flux” $v = w_2(E)$. As $H^2(K3, \mathbb{Z})$ is 22 dimensional (and torsion-free), v can take 2^{22} values. In the $SO(3)$ theory, we must sum over them.

There is no need to study separately 2^{22} possibilities because $K3$ has a very large diffeomorphism group which permutes the possible values of v . One obvious diffeomorphism invariant of v is the value of v^2 modulo 4; if it is 0 we call v even and if it is 2 we call v odd. If v is odd, it is certainly non-zero, but for v even there is one more obvious invariant: whether v is zero or not. It turns out that up to diffeomorphism, the invariants just stated are the only invariants of v . So on $K3$ there are really three partition functions to compute, namely the partition functions for $v = 0$, v even but non-zero, and v odd. We call these Z_0 , Z_{even} , and Z_{odd} . Similarly, we write n_0 , n_{even} , and n_{odd} for the number of values of v that are, respectively, trivial, even but non-trivial, and odd.

Let us count how many v ’s of each type there are. It turns out that for this purpose, the intersection form on $H^2(K3, \mathbb{Z})$ can be replaced by the sum of 11 copies of

$$H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (4.2)$$

This makes the combinatorics of counting the different kinds of bundle straightforward – in fact, equivalent to the combinatorics of counting the number of even and odd spin structure on a Riemann surface of genus 11, except that one must separate out the case of zero flux. The result is

$$\begin{aligned} n_0 &= 1 && \text{trivial type} \\ n_{\text{even}} &= \frac{2^{22} + 2^{11}}{2} - 1 && \text{even type} \\ n_{\text{odd}} &= \frac{2^{22} - 2^{11}}{2} && \text{odd type} \end{aligned}$$

So in particular the $SU(2)$ and $SO(3)$ answers are

$$\begin{aligned} Z_{SU(2)} &= \frac{1}{2} Z_0, \\ Z_{SO(3)} &= Z_0 + n_{\text{even}} Z_{\text{even}} + n_{\text{odd}} Z_{\text{odd}}. \end{aligned}$$

(The factor of $1/2$ in the first equation comes from the factor of $1/\#Z(G)$ in (3.3).)

Euler characteristic of a symmetric product

According to the description sketched above of the instanton moduli spaces on $K3$, we need to calculate the Euler characteristic of a symmetric product of $K3$ ’s, that is, of the quotient $(K3 \times K3 \times \dots \times K3)/S_n$ of the product of n copies of $K3$ by the group S_n of permutations of n objects; S_n acts by permuting the factors. The hyper-Kähler

condition means that one can use the orbifold description [51] of the cohomology of the quotient of a manifold by a finite group. The result we need has been computed before [52] and has even been obtained using the orbifold formula [53]. However, we will include a derivation for completeness. It is actually convenient, and no more difficult, to apply the orbifold formula to the symmetric product of an arbitrary manifold M (the orbifold formula is only expected to agree with the cohomology of a resolution of the symmetric product if M is hyper-Kähler or at least Calabi–Yau). In the computation, we will aim to describe the cohomology of the symmetric product, and not just compute the Euler characteristic. This will make the meaning of the result clearer.

First we recall the general construction of the cohomology, in the orbifold sense, of a quotient X/Γ , with X a manifold acted on by a discrete group Γ . It is constructed as

$$H^*(X/\Gamma) = \bigoplus_{\gamma} H_{\gamma}, \quad (4.3)$$

where γ runs over conjugacy classes in Γ (for each conjugacy class we pick a representative that we also call γ) and H_{γ} is the cohomology in the sector “twisted” by γ . This is obtained as follows. Let X^{γ} be the subset of X left fixed by γ . Let N_{γ} be the subgroup of Γ that commutes with γ . Let $H^*(X^{\gamma})$ be the cohomology of X^{γ} and let $H^*(X^{\gamma})^{N_{\gamma}}$ be the part of $H^*(X^{\gamma})$ that is invariant under N_{γ} . Then

$$H_{\gamma} = H^*(X^{\gamma})^{N_{\gamma}}. \quad (4.4)$$

In the case at hand, X is the product $M^n = M \times M \times \dots \times M$ of n copies of M , and Γ is the group S_n of permutations of the factors. We also want to take the sum over n . It is convenient, as we will see, to introduce an operator L_0 that acts on $H^*(M^n/S_n)$ with eigenvalue n , and to write the sum over n as

$$\mathcal{H} = \bigoplus_{n=0}^{\infty} q^n H^*(M^n/S_n), \quad (4.5)$$

where the powers of q are a formal way to keep track of the L_0 action. \mathcal{H} will turn out to be a kind of Fock space.

First let us work out using the definition in (4.4) the contribution of the untwisted sector. For $\gamma = 1$, $M^{\gamma} = M^n$. The cohomology of M^n is $H^*(M^n) = H^*(M)^{\otimes n}$, the tensor product of n copies of the cohomology of M . Also, $N_1 = S_n$, the full permutation group. So H_1 is the S_n -invariant part of $(H^*(M))^{\otimes n}$. This can be interpreted as follows. If we think of an element of $H^*(M)$ as a “one particle state” and an element of $H^*(M)^{\otimes n}$ as an “ n particle state,” then S_n invariance means that we should think of the n particles as being identical bosons or fermions²⁷ and impose bose and fermi statistics. If therefore we pick a basis w^a of the cohomology of M , and introduce a corresponding set of “creation operators” α_{-1}^a acting on a “Fock vacuum” $|\Omega\rangle$, then the contribution of the untwisted sectors to (4.5) is a Fock space generated by the α_{-1}^a .

We now want to show that the inclusion of twisted sectors results instead in a Fock space generated by a “stringy” set of oscillators α_{-n}^a $n = 1, 2, 3, \dots$, which have $L_0 = n$ just as one would expect in string theory. We have to remember that every element of S_n can be decomposed in disjoint cycles, for instance a permutation of six objects

²⁷ That is, the cohomology classes of even and odd dimension correspond to bosons and fermions respectively.

might take the form of a one-cycle, a two-cycle, and a three-cycle, often denoted (1)(23)(456). The conjugacy classes in S_n are labeled by giving the number n_l of l -cycles for $l = 1, 2, 3, \dots$. The n_l are arbitrary non-negative integers except for the obvious restriction

$$n = \sum_l l n_l. \quad (4.6)$$

N_γ is of the form

$$N_\gamma = \prod_{l=1}^{\infty} N_\gamma(l), \quad (4.7)$$

where $N_\gamma(l)$ is the subgroup of N_γ consisting of permutations that act non-trivially only on objects in the l -cycles of γ . This factorization, which holds because the order of an object under γ is invariant under N_γ , is reflected in a similar factorization of \mathcal{H} .

It is convenient to consider first the contributions of l -cycles of a fixed l . Suppose that γ consists only of l -cycles, say k of them. Then of course $n = kl$, and M^n is a product of n copies of M , divided into k sets of l such copies; each set is permuted cyclically by γ . The fixed point set $(M^n)_\gamma$ is a product of k copies of M , one for each l -cycle, so its cohomology is $H^*(M)^{\otimes k}$. N_γ again acts on $H^*(M)^{\otimes k}$ by permuting the factors, and again this means that if we think of the elements of $H^*(M)$ as “one particle states,” then we should impose bose and fermi statistics on the “ k particle states” in $H^*(M)^{\otimes k}$. Also, we want to sum over $k = 0, 1, 2, \dots$. So if we introduce a new set of creation operators α_{-l}^a , then the contribution to \mathcal{H} from conjugacy classes consisting of l -cycles only is the Fock space generated by the α_{-l}^a acting on the Fock vacuum. There is one novelty: α_{-l}^a has $L_0 = l$ because of the relation $n = kl$, or if you will because taking the fixed points of γ collapsed all the copies of M in an l cycle into a single copy, the diagonal.

Now it is easy to put the results together. Once we sum over n the restriction in (4.6) loses its force and the n_l are independent. The fixed point set of a general permutation γ contains a factor of M for each l -cycle in γ of any l . So the cohomology $H^*(M_\gamma^n)$ has a factor of $H^*(M)$ for each cycle. The symmetry group N_γ acts separately, according to (4.7), on the cycles of different order. So bose and fermi statistics are only imposed on “particles” of the same l . The net effect of this is that if we introduce the creation operators α_{-n}^a for all positive integers n , then \mathcal{H} is the Fock space generated by the α_{-n}^a , as claimed above.

For instance, if b_+ and b_- are the dimensions of the bosonic and fermionic subspaces of $H^*(M)$, then we can write a simple result for the generating functional of the number of states,

$$\sum q^n \dim(H^*(M^n/S_n)) = \frac{\prod_{n=1}^{\infty} (1 + q^n)^{b_-}}{\prod_{n=1}^{\infty} (1 - q^n)^{b_+}}. \quad (4.8)$$

We can also write a simple result for the generating function of the Euler characteristics,

$$\sum q^n \chi(M^n/S_n) = \frac{\prod_{n=1}^{\infty} (1 - q^n)^{b_-}}{\prod_{n=1}^{\infty} (1 - q^n)^{b_+}} = \frac{1}{\prod_{n=1}^{\infty} (1 - q^n)^{\chi(M)}}. \quad (4.9)$$

Happily, up to elementary factors, those formulas are modular. In particular the generating function for the Euler characteristic after multiplying by $q^{-\chi/24}$ becomes

$$G(q) = q^{-\chi/24} \sum q^n \chi(M^n/S_n) = \frac{1}{\eta^\chi}, \quad (4.10)$$

where η is the Dedekind eta-function.

Instantons on $K3$

Now we return to the computation of Euler characteristic of instantons on $K3$. For $K3$, Eq. (4.10) gives

$$G(q) = \frac{1}{\eta^{24}},$$

which is none other than the partition function of left-moving modes of the bosonic string!

We can now calculate the Euler characteristic of the moduli spaces of instantons on $K3$. It is easier to consider first the $SO(3)$ bundles with nonzero w_2 because then the description by a symmetric product of $K3$'s work for all values of the instanton number. We recall that if the instanton number is k , then the number of points is $n = 2k - 3$, so for even (or odd) bundles only odd (or even) n contributes. Also the formula $n = 2k - 3$ shows that adding another copy of $K3$ adds only $1/2$ to the instanton number, so if we want to count instantons with powers of q , we must count points on $K3$ with powers of $q^{1/2}$. So we can evaluate the partition functions

$$Z = q^{-s} \sum_n q^n \chi(\mathcal{M}_n). \quad (4.11)$$

If we pick $s = 2$, we get

$$\begin{aligned} Z_{\text{even}}(q) &= \frac{1}{2} \left(G(q^{1/2}) + G(-q^{1/2}) \right), \\ Z_{\text{odd}}(q) &= \frac{1}{2} \left(G(q^{1/2}) - G(-q^{1/2}) \right). \end{aligned} \quad (4.12)$$

These formulas are modular (of weight -12), giving finally our first strong coupling test of S -duality. In writing (4.12), we added or subtracted $G(q^{1/2})$ and $G(-q^{1/2})$ to project onto terms with an even or odd number of copies of $K3$.

The formula $\dim \mathcal{M}_k = 8k - 12$ for the dimension of instanton moduli space shows that the smallest value of k for which \mathcal{M}_k is non-empty²⁸ is $k = 3/2$. The leading behavior $Z_{\text{odd}} \sim q^{-1/2}$ of Z_{odd} comes from an instanton of $k = 3/2$; its contribution is shifted from $q^{3/2}$ to $q^{-1/2}$ by the overall factor of q^{-2} that comes from picking $s = 2$. Similarly, the leading behavior $Z_{\text{even}} \sim 1$ of Z_{even} is the shifted contribution from $k = 2$.

²⁸ Generically, and with an exception discussed below.

The contribution of the trivial $SO(3)$ bundle on $K3$ can be treated similarly, with one difference: since the description of the moduli spaces by symmetric products of $K3$ is in this case only valid for odd instanton number, we have to add an unknown function $F(q^2)$ contributing to the terms with even instanton number. The partition function is thus of the form

$$Z_0(q) = F(q^2) + \frac{1}{2}(G(q^{1/2}) + G(-q^{1/2})), \quad (4.13)$$

where we have multiplied by the same factor of q^{-2} .

Now we will make a precise test of S -duality in its sharpened form of Eq. (3.15). This predicts the transformation law under $\tau \rightarrow -1/\tau$ to be

$$\begin{aligned} Z_0 &\rightarrow 2^{-11} \left(\frac{\tau}{i}\right)^{w/2} Z_{SO(3)}, \\ Z_{\text{even}} &\rightarrow 2^{-11} \left(\frac{\tau}{i}\right)^{w/2} [Z_0 + (2^{10} - 1)Z_{\text{even}} - 2^{10}Z_{\text{odd}}], \\ Z_{\text{odd}} &\rightarrow 2^{-11} \left(\frac{\tau}{i}\right)^{w/2} [Z_0 + (-2^{10} - 1)Z_{\text{even}} + 2^{10}Z_{\text{odd}}]. \end{aligned} \quad (4.14)$$

To obtain these formulas from (3.15), one needs to evaluate, for given v , the sums

$$\sum_{u \text{ odd}} (-1)^{v \cdot u}, \quad \sum_{u \text{ even}} (-1)^{v \cdot u}. \quad (4.15)$$

This is easily done using the fact that modulo two the intersection form on $H^2(K3, \mathbb{Z})$ is equivalent to 11 copies of H , and picking a convenient choice of v .

Since we have incomplete information about Z_0 , we can eliminate it from the above to get the prediction that under $\tau \rightarrow -1/\tau$,

$$Z_{\text{even}} - Z_{\text{odd}} \rightarrow \left(\frac{\tau}{i}\right)^{w/2} (Z_{\text{even}} - Z_{\text{odd}})$$

which is indeed the case

$$Z_{\text{even}} - Z_{\text{odd}} = G(-q^{1/2}) \rightarrow \tau^{-12} G(-q^{1/2}).$$

In particular, the modular weight is -12 , and $w = -24$. But we can go further and determine $F(q^2)$. Using the modular transformation properties

$$\begin{aligned} G(-q^{1/2}) &\rightarrow \tau^{-12} G(-q^{1/2}), \\ G(q^{1/2}) &\rightarrow 2^{-12} \tau^{-12} G(q^2), \\ G(q^2) &\rightarrow 2^{12} \tau^{-12} G(q^{1/2}), \end{aligned}$$

we find that *all* of the equations (4.14) are satisfied if and only if

$$F(q^2) = \frac{1}{4} G(q^2).$$

That there exists an F which satisfies all three equations at the same time is a very precise further test of S -duality.

Note that the leading behavior of $F(q^2)$ for small q is $F(q^2) \sim q^{-2} + \text{constant}$. Thus the partition function Z_0 for zero flux has an expansion

$$Z_0 \sim \frac{1}{4q^2} + O(1). \quad (4.16)$$

The contribution of order q^{-2} must be interpreted as the contribution of the trivial connection, shifted from q^0 to q^{-2} because we have multiplied the whole series by q^{-2} . There is indeed one exception to the statement that instantons generically do not exist for $8k - 12 < 0$: the trivial connection always exists, even though one might think that generically it shouldn't. Evidently, it contributes to the twisted $N = 4$ theory, and (for $SO(3)$) its contribution has the somewhat mysterious value $1/4$. On the other hand, $8k - 12$ is negative for $k = 1$, and for $k = 1$ there are in fact generically no instantons. That explains the absence of a term of order q^{-1} in (4.16).

Of course, we can reassemble our results into formulas for the $SU(2)$ and $SO(3)$ partition functions. The $SU(2)$ partition function, allowing for the usual factor of $1/2$, is

$$Z_{SU(2)} = \frac{1}{8}G(q^2) + \frac{1}{4}G(q^{1/2}) + \frac{1}{4}G(-q^{1/2}). \quad (4.17)$$

This transforms as a modular form of $\Gamma_0(2)$, as it should, since it is invariant under T and transforms correctly under ST^2S . This formula makes predictions for the Euler characteristics of the moduli spaces with $w_2 = 0$ and even instanton number; the predictions contain mysterious denominators of $1/4$ which must somehow reflect the singularities of these moduli spaces. The $SO(3)$ formula is similarly

$$Z_{SO(3)} = \frac{1}{4}G(q^2) + 2^{21}G(q^{1/2}) + 2^{10}G(-q^{1/2}) \quad (4.18)$$

and again has modular properties for $\Gamma_0(2)$.

Since $w = -24$ and $s = -2$, the “central charge” defined in (3.7) is $c = 24$. This agrees with the general arguments discussed in the last section (3.25) which gave $c = \chi = 24 \bmod 4$. We see that the equality is true even if we remove the mod 4 condition. We also see that at least in this case $w = -\chi = -24$. Since in general c and w are linear combinations of χ and σ we need one other example with a different ratio of χ and σ to completely fix c and w in general; we will find that the coefficients of σ vanish (as one might have expected from parity conservation were it not that parity is explicitly broken by the twisting used to construct a topological field theory).

Before leaving $K3$, we will point out a fairly natural guess for the partition function for the $SU(N)$ model on $K3$. We will state the guess for the partition function Z_0 of bundles with zero magnetic flux. Contributions of the other bundles are determined by modular transformations. Our guess is

$$Z_0 = \frac{1}{N^2}G(q^N) + \frac{1}{N} \left[G(q^{1/N}) + G(\omega q^{1/N}) + \dots + G(\omega^{N-1} q^{1/N}) \right],$$

where $\omega = \exp(2\pi i/N)$. The formula has some attractive properties (but considerations of its modular properties as well as some considerations explained at the end of Section

5.3 suggest that it may be valid, if at all, only for N prime). For $SU(N)$ the dimension of instanton moduli space on $K3$ is $4kN - 4(N^2 - 1)$, so after the trivial instanton at instanton number 0, instantons appear first at $k = N$. That agrees with the fact that in the above formula $Z_0 \sim q^{-N}/N^2 + O(1)$ with a gap between q^{-N} and q^0 . The modular transform of the formula for Z_0 can also be seen to give the right gaps for bundles with non-zero flux.

The above guess suggests that the instanton moduli spaces for integral instanton numbers that are not 0 mod N (or perhaps only those prime to N , in view of what we find in Section 5.3), can be identified with appropriate symmetric products of $K3$'s. At least for instanton numbers 1 mod N there are some indications of this [54]. Our guess is also compatible with $c = (N - 1)\chi$ in accord with (3.25), and with $w = -\chi$ (in other words this guess suggests that the modular weight is independent of the gauge group for a simple group; this is also supported by the analysis discussed below on ALE spaces).

4.2. \mathbf{CP}^2

We now consider the case of \mathbf{CP}^2 , again with gauge group $SU(2)$ or $SO(3)$. \mathbf{CP}^2 is another case in which the vanishing theorems of Section 2 apply, so it would seem that we can compute from instantons only. We will actually run into a surprise, a kind of anomaly that affects holomorphy.

As $H^2(\mathbf{CP}^2)$ is one-dimensional, an $SO(3)$ bundle E has two possible values of $v = w_2(E)$. There is $v = 0$, with $v^2 = 0$; and there is $v \neq 0$, with $v^2 = 1$ modulo 4. Accordingly, there are two partition functions,

$$\begin{aligned} Z_0 &= q^{-s} \sum_n \chi(\mathcal{M}_{0,n}) q^n \\ Z_1 &= q^{-s} \sum_n \chi(\mathcal{M}_{1,n}) q^n, \end{aligned} \quad (4.19)$$

where $\mathcal{M}_{0,n}$ and $\mathcal{M}_{1,n}$ are, respectively, moduli spaces of bundles with $v = 0$ and $v \neq 0$, and instanton number n . To study these functions, we rely on formulas of Yoshioka and Klyachko [26,28]. Yoshioka determined a formula for Z_1 , while Klyachko determined formulas for the related functions

$$Y_i(q) = q^{-s} \sum_{n=0}^{\infty} \chi(\mathcal{M}^{(0)}_{i,n}) q^n, \quad (4.20)$$

where $\mathcal{M}^{(0)}_{i,n}$ is the *uncompactified* moduli space of instantons of instanton number n and magnetic flux determined by i .

Klyachko's formula for Y_1 is

$$Y_1(q) = 3 \sum_{n=1}^{\infty} H(4n - 1) q^{n - \frac{1}{4}}, \quad (4.21)$$

where $H(m)$ is the Hurwitz function that equals the number of equivalence classes of integral binary quadratic forms of discriminant m , weighted by their number of

automorphisms.²⁹ A formula of Yoshioka³⁰ shows that for any Kähler manifold of Euler characteristic χ , if we take $s = -\chi/12$ then

$$Z_v(q) = \frac{Y_v(q)}{\eta(q)^{2\chi}} \quad (4.22)$$

for appropriate non-zero v ; the formula applies for $v \neq 0$ on \mathbf{CP}^2 . (The fact that the modular function η appears here should hopefully be related to S -duality, though we do not know how.) Combining these formulas, one has

$$Z_1(q) = \frac{3}{\eta(q)^6} \sum_{n=1}^{\infty} H(4n-1)q^{n-\frac{1}{4}}. \quad (4.23)$$

(Yoskioka also obtained directly [26] a formula for Z_1 which must coincide with this formula – we checked this for the first few terms – but the modular properties are easier to see in (4.23).) Since the modular properties of (4.23), which will be discussed presently, would be ruined by multiplying by a power of q , the value of s is the one required to make the formula (4.22) modular, or

$$s = -\frac{\chi}{12}, \quad (4.24)$$

which is the same result that we found for $K3$. Similarly, as the modular weight of Z_1 is $-3/2$ (in a suitable sense that will be explained) we can determine in general (using also the $K3$ result) that the modular weight is $-\chi/2$ or that

$$w = -\chi. \quad (4.25)$$

We will have a further check on (4.24) and (4.25) when we consider blowups.

Klyachko's formula for Y_0 has an extra term relative to (4.21), but Yoskioka's analog of (4.22) for $v = 0$ is also more complicated³¹; combining them, it appears that Z_0 is the obvious analog of (4.23):

$$Z_0 = \frac{3}{\eta^6} \sum_{n=0}^{\infty} H(4n)q^n. \quad (4.26)$$

The series in (4.23) and (4.26) are known as Eisenstein series of weight $3/2$. For a relatively elementary introduction to such series see Ref. [47], section IV.2, especially p. 194; see also [55]. The simplest Eisenstein series of half-integral weight for $\Gamma_0(4)$ are defined by series such as

$$E_{k/2}(z) = \sum_{\gamma} j(\gamma, z)^{-k}. \quad (4.27)$$

²⁹ $H(m)$ vanishes unless m is congruent to 0 or -1 modulo 4.

³⁰ This follows from the Weil conjectures and Theorem 0.4 of [26].

³¹ This formula is unpublished and we are grateful to Yoshioka for providing it to us.

Here the sum runs over certain cosets of $\Gamma_0(4)$, and $j(\gamma, z)^{-k}$ generalizes the factor $(mz + n)^{-k}$ that appears in the conventional Eisenstein series

$$G_k(z) = \sum_{m,n} (mz + n)^{-k}. \quad (4.28)$$

For k an odd integer ≥ 5 , (4.27) defines a modular form of weight $k/2$ for $\Gamma_0(4)$. However, for $k = 3$, the sum in (4.27) does not converge quite well enough to define a holomorphic modular form. To define a modular object, one regularizes the sum, replacing (4.27) with

$$E_{3/2}(z, s) = \sum_{\gamma} j(\gamma, z)^{-3} |j|^{-2s}. \quad (4.29)$$

With this regularization, $E_{3/2}(z, s)$ transforms as a modular form of weight $3/2$ (for any fixed s) but is not holomorphic in z . One can actually take the limit $s \rightarrow 0$; the function $E_{3/2}(z) = E_{3/2}(z, 0)$ that one obtains this way is modular but not holomorphic. The story is just analogous for the ordinary Eisenstein series (4.29) for $k = 2$, which similarly does not quite converge well enough to define a holomorphic modular form.

As was discovered twenty years ago by Zagier [32] (see also Ref. [33]), the process just described, starting with a slightly different Eisenstein series of weight $3/2$, gives such functions as

$$\begin{aligned} f_0 &= \sum_{n \geq 0} 3H(4n)q^n + 6\tau_2^{-1/2} \sum_{n \in \mathbb{Z}} \beta(4\pi n^2 \tau_2) q^{-n^2}, \\ f_1 &= \sum_{n > 0} 3H(4n-1)q^{n-\frac{1}{4}} + 6\tau_2^{-1/2} \sum_{n \in \mathbb{Z}} \beta(4\pi(n + \frac{1}{2})^2 \tau_2) q^{-(n+\frac{1}{2})^2}, \end{aligned} \quad (4.30)$$

where $q = e^{2\pi i \tau}$, $\tau_2 = \text{Im}(\tau)$, and

$$\beta(t) = \frac{1}{16\pi} \int_1^\infty u^{-3/2} \exp(-ut) \, du.$$

Thus, the modular status of (4.23) and (4.26) is clear: these functions are the “holomorphic part” of the non-holomorphic modular functions in (4.30) (divided by η^6). The “holomorphic part” can be more formally defined by considering the limit of $\bar{\tau} \rightarrow \infty$ while keeping τ fixed. It is easy to see that the additional non-holomorphic terms vanish in this limit.

Moreover³² under $\tau \rightarrow -1/\tau$, these functions transform as

$$\begin{pmatrix} f_0(-1/\tau) \\ f_1(-1/\tau) \end{pmatrix} = \begin{pmatrix} \tau \\ i \end{pmatrix}^{3/2} \cdot \begin{pmatrix} -1/\sqrt{2} \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} f_0(\tau) \\ f_1(\tau) \end{pmatrix}. \quad (4.31)$$

This is exactly the transformation law predicted in (3.15). Since f_0 is invariant under T and f_1 under T^4 , it follows, for instance, that f_0 has the expected modular transfor-

³² According to an unpublished argument kindly explained to us by D. Zagier.

mation law under the group $\Gamma_0(4)$ generated by T and ST^4S .

Holomorphic anomaly?

Thus, all is well except that the Z_v are not holomorphic. This is reminiscent of the holomorphic anomaly in certain topological two-dimensional models [34]. The situation there is as follows: One has an untwisted physical theory with coupling constants τ and $\bar{\tau}$, parametrizing some moduli space, appearing in the action. In the physical theory $\bar{\tau}$ is the complex conjugate of τ . One then considers the twisted version and shows that the $\bar{\tau}$ dependence in the action is of the form $\bar{\tau}\{Q, \dots\}$, and so formally the path integral is independent of $\bar{\tau}$. One then takes the unphysical limit $\bar{\tau} \rightarrow \infty$ keeping τ fixed to argue that the twisted theory computes the Euler characteristic of a certain moduli space. However, under certain circumstances [34], the formal independence of the path integral from $\bar{\tau}$ is anomalous and the twisted theory has nice modular properties only when $\bar{\tau}$ is chosen to be the complex conjugate of τ . Nevertheless the topological answer is recovered, and modularity lost, by taking $\bar{\tau} \rightarrow \infty$ and keeping τ fixed. This is very similar to the above situation for the $N = 4$ twisted Yang–Mills on \mathbf{CP}^2 , with the obvious substitutions. In particular the S-duality conjecture seems to suggest that the twisted physical theory on \mathbf{CP}^2 has $\bar{\tau}$ dependence.

Another precedent, perhaps equally relevant, is Donaldson theory which (as Donaldson discovered in one of his early papers [56]) fails to produce topological invariants on manifolds with $b_2^+ = 1$. \mathbf{CP}^2 has $b_2^+ = 1$ (though in Donaldson theory it is not a typical example with $b_2^+ = 1$ since also $b_2^- = 0$).

The anomaly in Donaldson theory comes from abelian connections where zero modes appear for the fields called $\phi, \bar{\phi}$ in Section 2. Though this anomaly has not been given its proper physical expression, it very plausibly involves the essential failure of compactness of field space due to the flat directions in the ϕ potential. (Those flat directions are suppressed when the gauge field has $SU(2)$ holonomy.) The condition $b_2^+ = 1$ for the anomaly looks like it might be natural from this point of view, since for $b_2^+ > 1$ there are extra fermion zero modes along the flat directions.

The behavior of the twisted $N = 4$ theory on \mathbf{CP}^2 suggests that it too may have an anomaly associated with abelian configurations when $b_2^+ = 1$. In fact, the instanton numbers of abelian configurations on \mathbf{CP}^2 are $-n^2$ for $v = 0$ and $-(n + 1/2)^2$ for $v = 1$. These are precisely the exponents that appear in the anomalous, non-holomorphic sum in (4.30). The failure of holomorphy of the f_i can be summarized in the elegant equations

$$\begin{aligned} \frac{\partial}{\partial \bar{\tau}} f_0 &= \frac{3}{16\pi i} \tau_2^{-3/2} \sum_{n \in \mathbb{Z}} \bar{q}^{n^2}, \\ \frac{\partial}{\partial \bar{\tau}} f_1 &= \frac{3}{16\pi i} \tau_2^{-3/2} \sum_{n \in \mathbb{Z}} \bar{q}^{(n+\frac{1}{2})^2}, \end{aligned} \quad (4.32)$$

which give further encouragement for seeking a field theoretic explanation of the anomaly.

The formulas of Yoskioka [27] for instantons on rational ruled surfaces (which all have $b_2^+ = 1$) strongly suggest that on this whole class of manifolds there is some sort

of anomaly generalizing the one that we have found experimentally on \mathbf{CP}^2 .

4.3. Blow ups

Our next example will involve the behavior under blowing up a point on a Kähler manifold. Thus, we let M be a Kähler manifold, and \hat{M} the new manifold obtained by blowing up a point $p \in M$. Topologically, the effect of the blow-up is to glue in into M a copy of $\overline{\mathbf{CP}}^2$ (that is, \mathbf{CP}^2 with the opposite of the usual complex orientation).

The effect on the second homology is that the lattice L of M is replaced by the lattice

$$\hat{L} = L \oplus I^- \quad (4.33)$$

for \hat{M} (where I^- , as in Section 3, is the one-dimensional lattice with quadratic form $-x^2$). If E is an $\mathrm{SO}(3)$ bundle on \hat{M} , then $\hat{v} = w_2(E)$ can be decomposed as

$$\hat{v} = v \oplus r \quad (4.34)$$

under the decomposition (4.33).

In Eq. (3.15), we proposed a precise formula for how the partition functions transform under S -duality. Since \hat{L} is a direct sum, the formula implies that the action of $\mathrm{SL}(2, \mathbb{Z})$ on $Z_{\hat{v}} = Z_{v,r}$ is the product of a representation acting on the v label and a representation acting on the r label. Thus, the action of $\mathrm{SL}(2, \mathbb{Z})$ is compatible with the possibility that the Z 's have a factorization as

$$Z_{\hat{M},v} = Z_{M,v} Z_r \quad (4.35)$$

where for clarity we write the manifold \hat{M} or M explicitly, but the function Z_r could be “universal,” independent of M .

The possible Z_r 's can be analyzed just as in Section 3.3, where we considered a manifold with the intersection form I^- ; if a factorization such as (4.35) holds, that discussion can be applied to the blowup of an arbitrary M . In that discussion, we pointed out that (for $\mathrm{SU}(N)$, not just $\mathrm{SU}(2)$) the modular transformations of the Z_r are those of the level one characters of the $\mathrm{SU}(N)$ WZW model. However, the modular transformations alone do not imply that a factorization such as (4.35) holds, or if it does that the Z_r are the WZW characters.

These assertions are, however, true, at least in one important case, according to another formula of Yoshioka ([26], proposition (0.3)). He demonstrates that under blowup of a Kähler manifold, the factorization of (4.35) holds for $\mathrm{SU}(2)$ and $r = 0$, with

$$Z_0 = \frac{\theta_0(q)}{\eta(q)^2}, \quad (4.36)$$

where

$$\theta_0(q) = \sum_{n \in \mathbb{Z}} q^{n^2}. \quad (4.37)$$

This is indeed $1/\eta$ times the appropriate $\mathrm{SU}(2)$ character at level 1.

Under blowup, the Euler characteristic of a manifold is increased by 1. On the other hand, the function in (4.36) behaves for small q as $q^{-1/12}$ and transforms as a modular form of weight $-1/2$. This gives a further check on our earlier finding that the “zero point” of the instanton number is $-\chi/12$ and the modular weight of the partition function is $-\chi/2$.

S-duality implies that the factorization (4.35) must also hold for $r = 1$ and that

$$Z_1 = \frac{\theta_1(q)}{\eta(q)^2}, \quad (4.38)$$

where

$$\theta_1(q) = \sum_{n \in \mathbb{Z} + 1/2} q^{n^2}. \quad (4.39)$$

4.4. ALE spaces

For our last example, following Nakajima [29], we consider instantons on ALE (asymptotically locally Euclidean) spaces. In many ways this is the richest example, as precise information is available for $U(k)$ or $SU(k)$ gauge group, not just $SU(2)$,³³ and one encounters WZW models at arbitrary levels, not just level 1. Unfortunately, it is hard to exploit the examples fully as one does not know enough about the implications of S-duality on non-compact manifolds.

The ALE spaces X_Γ we have in mind arise as hyper-Kähler resolutions of hyper-Kähler orbifolds of the form \mathbb{C}^2/Γ , where Γ is a finite subgroup of $SU(2)$ acting linearly on \mathbb{C}^2 . These spaces are noncompact relatives of $K3$. Indeed, $K3$ has a somewhat analogous construction beginning with the quotient of a four-torus by a finite group, and the singularities so obtained have the local structure of \mathbb{C}^2/Γ .

Every finite subgroup Γ of $SU(2)$ has, up to isomorphism, a finite set of irreducible complex representations ρ_i . If ρ is the two-dimensional representation associated with the embedding of Γ in $SU(2)$, one can decompose the tensor product $\rho \otimes \rho_i$ as the direct sum of n_{ij} copies of ρ_j ; the n_{ij} are all 0 or 1. Form a graph with the ρ_i for vertices, with two vertices ρ_i, ρ_j connected by a line if and only if $n_{ij} \neq 0$. It turns out that the graph so obtained is the extended Dynkin diagram of a simply laced Lie group H_Γ . The correspondence $\Gamma \leftrightarrow H_\Gamma$ is the McKay correspondence between finite subgroups of $SU(2)$ and simply laced Lie groups. (The subgroup \mathbb{Z}_N of $SU(2)$ corresponds to A_{N-1} , the dihedral groups correspond to the D_N , and the symmetries of the three regular solids correspond to the E series.)

The Dynkin diagram of H_Γ appears in the geometry of X_Γ as the intersection form of two-spheres S_i created in blowing up the singularities of X_Γ . These two-spheres give a basis of $H_2(X_\Gamma, \mathbb{Z})$, and the first Chern class of a vector bundle over X_Γ can be interpreted as a weight of H_Γ .

³³ The partition function of the $U(k)$ theory would vanish on a compact four-manifold because of fermion zero modes in the $U(1)$ factor, but that is not so on the ALE spaces, where there are no normalizable zero modes. Nakajima's results are most elegantly stated for $U(k)$.

The hyper-Kähler metrics on X_Γ are asymptotic at infinity to \mathbb{R}^4/Γ . The fundamental group at infinity is non-trivial, and admits non-trivial homomorphisms

$$\phi : \Gamma \rightarrow G \quad (4.40)$$

to the gauge group G . These homomorphisms enter because requiring that the action of an instanton should be finite does not imply that the instanton approaches the trivial connection at infinity, but only that it should be flat at infinity, that is, given by a homomorphism from Γ to G .

Nakajima analyzes instantons with gauge group $U(k)$ on X_Γ . Such an instanton determines at infinity a k dimensional representation (4.40) of Γ , which decomposes as a sum of n_i copies of the irreducible representation ρ_i , with

$$k = \sum_i n_i \dim(\rho_i). \quad (4.41)$$

Since the ρ_i can be interpreted as points on the extended Dynkin diagram of H_Γ , one gets a natural assignment of integers n_i to the points of the Dynkin diagram. Those points also correspond to simple roots e_i of H_Γ , so we get an assignment of integers to simple roots. This enables us to form the sum $w_\phi = \sum_i n_i e_i$, which is a positive weight of H_Γ , so it determines a representation so determined is a highest weight of an integrable representation of H_Γ current algebra at level k . The instanton number of these instantons is shifted from being an integer by an amount that depends on ϕ .

Nakajima now considers the middle dimensional cohomology $H_{\phi, c_1, n}$ of the moduli space of $U(k)$ instantons with a representation ϕ at infinity, arbitrary first Chern class c_1 , and instanton number n . (n is shifted from an integer by an amount that depends on ϕ .) More precisely, he considers the cohomology of some complete hyper-Kähler manifolds obtained by a particular partial compactification and deformation of singularities of the moduli spaces.

To describe the results, introduce an operator L_0 that has eigenvalue n on $H_{\phi, c_1, n}$, and introduce the sum

$$\mathcal{H}_\phi = \bigoplus_{c_1, n} q^n H_{\phi, c_1, n}, \quad (4.42)$$

where the variable q is formally included to keep track of the L_0 action. Nakajima's remarkable result is that \mathcal{H} has a natural structure of irreducible representation of H_Γ current algebra at level k , with the highest weight being w_ϕ . L_0 acts by multiplication by n , while c_1 , interpreted as we mentioned above as a weight of H_Γ , determines the action of the Cartan subalgebra of H_Γ . The rest of the action of the affine Lie algebra is defined through some operations of twisting vector bundles along the S_i ; these operations conceivably should be regarded as analogs of vertex operators in conformal field theory.

At any rate, the functions

$$Z_\phi(q) = \sum_{c_1, n} q^{n-c/24} \dim H_{\phi, c_1, n} \quad (4.43)$$

are therefore the characters of integrable representations of the loop group of H_Γ . (c is the central charge of H_Γ current algebra at level k .) So they have modular

properties and in fact transform in a unitary representation of $SL(2, \mathbb{Z})$. Because of not understanding the implications of S -duality on manifolds with boundary, and for other reasons mentioned presently, we do not know how to explain the particular representation that arises.

To interpret the facts just sketched in terms of S -duality, one might hope that the middle dimensional cohomology of the moduli spaces coincides with the L^2 cohomology, in which case $\dim H_{\phi, c_1, n}$ can perhaps be interpreted as a kind of Euler characteristic. Then one would hope to deduce the modularity of (4.43) from S -duality. Unfortunately, there are difficulties with this interpretation. In particular the partition function of the twisted $N = 4$ theory gives the Euler characteristic defined as a kind of curvature integral, not the Euler characteristic of L^2 cohomology of the moduli space; therefore it is not obvious how to compare Nakajima's results to S -duality.

Even if one could predict from S -duality the modular properties of the ALE partition function, that would not do justice to Nakajima's story, which involves construction of a natural Hilbert space with group action and not just a partition function. It does not seem that a natural origin of this would be in four-dimensional field theory – which associates Hilbert spaces with quantization on a three-manifold, and a partition function with the path integral on a four-manifold, but does not normally produce a Hilbert space associated with a four-manifold. It would seem that the full story here should have its origins in a supersymmetric theory of dimension ≥ 5 , which, in quantization on $X_F \times \mathbb{R}$ (where \mathbb{R} is the “time” axis) would produce a Hilbert space of physical states. The middle dimensional cohomology of the instanton moduli spaces would appear as a set of BPS-saturated states, perhaps making it possible to do justice to the results of Ref. [29]. Note that the dimension of the cohomology of the instanton number k moduli space (on $K3$ or an ALE space) grows exponentially like $\exp(a\sqrt{k})$ for some constant a . This implies that in a five dimensional theory that would be relevant to Nakajima's results, the multiplicity of particles of mass m goes as $\exp(a\sqrt{m})$ for some constant a . This growth is huge for a field theory though it is somewhat less than what one encounters in string theory (whose spectrum grows as $\exp(am)$ because $m^2 = k$). A natural origin for such a huge BPS-saturated spectrum may well require string theory. Anyway, the only sensible theory that we know of in dimension ≥ 5 is string theory, so some of Nakajima's results may have their natural origin in S -duality of string theory.

$SU(k)$

It is also of interest to extract from Nakajima's formulas the results for gauge group $SU(k)$. For $SU(k)$ on compact manifolds, we worked out precise predictions of S -duality in Section 3. We would like to see what happens in the ALE case.

If one wishes to consider $SU(k)$ rather than $U(k)$ on X_F , one must restrict to bundles with $c_1 = 0$. As the action of the Cartan subalgebra of H_F is proportional to c_1 , this entails restriction to states annihilated by the maximal torus T . This suggests that the “partition functions,” in the above sense, of the $SU(k)$ theory would be the characters of the coset conformal field theory H_F/T , times the contribution of extra oscillators coming from T (which give a non-zero modular weight). This appears to be so, though a verification requires a careful study of the (apparently non-standard) way that Nakajima's partial compactification affects the $U(1)$ factor.

Let us specialize to the the ALE space X_Γ with $\Gamma = \mathbb{Z}_N$, so $H_\Gamma = \mathrm{SU}(N)$. A representation ϕ of the fundamental group at infinity determines a highest weight w_ϕ of $\mathrm{SU}(N)$ and also a representation Λ of the loop group of $\mathrm{SU}(N)$ at level k . Nakajima's result for $\mathrm{U}(k)$ means that (with the partition function Z interpreted in his sense)

$$Z_{H_\Gamma}(\tau, \mathrm{U}(k), \phi) = \chi^\Lambda(q),$$

where χ^Λ is the character of the representation Λ . χ^Λ can be written [57]

$$\chi^\Lambda = \sum_{\lambda} c_{\lambda}^{\Lambda} \Theta_{\lambda, k},$$

where the c_{λ}^{Λ} , known as “string functions,” are characters of the $\mathrm{SU}(N)/T$ coset model times the partition function of oscillators of T , and the $\Theta_{\lambda, k}$ are certain theta functions. The λ 's are as follows: λ is an $\mathrm{SU}(N)$ weight that is congruent to w_ϕ modulo the root lattice L ; also one identifies $\lambda \cong \lambda'$ if $\lambda - \lambda' \in kL$. Thus λ takes $(N-1)^k$ values. Since L can be identified with $H_2(X_\Gamma)$, λ can be identified with the \mathbb{Z}_k -valued magnetic flux. The \mathbb{Z}_k magnetic flux and c_1 (which is the magnetic flux for the center $\mathrm{U}(1)$ of $\mathrm{U}(k)$) are correlated in this way simply because

$$\mathrm{U}(k) = \frac{\mathrm{SU}(k) \times \mathrm{U}(1)}{\mathbb{Z}_k},$$

where \mathbb{Z}_k is a diagonal subgroup of the center of $\mathrm{SU}(k)$ and a \mathbb{Z}_k subgroup of $\mathrm{U}(1)$.

How does c_{λ}^{Λ} transforms under $\tau \rightarrow -1/\tau$? As is well known in the study of coset models, the answer is the product of a matrix acting on the Λ index and a matrix acting on λ :

$$c_{\lambda}^{\Lambda}(-1/\tau) = \left(\frac{\tau}{i}\right)^{-(N-1)/2} \mathrm{const.} \sum_{\Lambda', \lambda'} b(\Lambda, \Lambda', \lambda, \lambda') c_{\lambda'}^{\Lambda'}(\tau), \quad (4.44)$$

where

$$b(\Lambda, \Lambda', \lambda, \lambda') = \exp(2\pi i \lambda \cdot \lambda' / k) \sum_{w \in W} \epsilon(w) \times \exp(2\pi i (\Lambda + \rho) w (\Lambda' + \rho) / k + N - 1), \quad (4.45)$$

where W is the Weyl group of $\mathrm{SU}(N)$ and ρ is half the sum of positive roots of $\mathrm{SU}(N)$. The factor in (4.45) that acts on the the magnetic flux λ is consistent with what one expects in (3.15). The factor acting on Λ , which is determined by the boundary conditions, is more mysterious; we do not understand why the modular transformations of the $\mathrm{SU}(N)$ current algebra appear here. The formula can actually be motivated to a certain extent by considering S -duality for (purely bosonic) $\mathrm{U}(1)$ gauge theory on X_Γ ; in that case one can see that S -duality induces a Fourier transform on the boundary conditions, a result similar to the large k limit of the Λ -dependent factor in (4.45).

Note also that (4.44) tells us that the modular weight is determined by $w = -(N-1)$, which is in some agreement with our earlier formulas as $N-1$ is the Euler characteristic of the L^2 cohomology of X_Γ . It is interesting that this value is independent of the gauge

group. The value of c , however, is fractional now, and deviates from the formula $c = (k - 1)\chi$ that we found for compact four-manifolds without boundary.

5. Computation by physical methods

5.1. Reduction to $N = 1$

In this section, by imitating an analogous computation for Donaldson theory [35], we will analyze the partition function of the twisted $N = 4$ theory (with gauge group $SU(2)$ or $SO(3)$) on an arbitrary Kähler manifold X with $H^{2,0}(X) \neq 0$. An answer emerges that satisfies an impressive set of constraints and gives support therefore to the overall picture of S -duality.

The key input, mentioned at the end of the introduction to Section 2, is that on a Kähler manifold the theory actually has four topological symmetries, one from each of the four underlying supersymmetries. Therefore, a perturbation that breaks the $N = 4$ supersymmetry down to $N = 1$ would still leave us with one topological symmetry, enough to control the theory by reducing computations to classical configurations and quantum vacuum states. We can attempt to find an $N = 1$ -invariant perturbation whose addition to the theory brings about some simplification.

The $N = 4$ theory can be viewed as an $N = 1$ theory with three chiral superfields, say T , U , and V , in the adjoint representation of the gauge group. The superpotential is

$$W = \text{Tr } T[U, V]. \quad (5.1)$$

For the time being the gauge group is an arbitrary compact simple Lie group G . While preserving $N = 1$ supersymmetry, we could add a perturbation to the superpotential, say the mass term $\Delta W = -\frac{1}{2}m\text{Tr}(T^2 + U^2 + V^2)$ (or any other nondegenerate quadratic expression, which would have essentially the same effect). As in Ref. [35], the resulting perturbation of the Lagrangian is the sum of a term of the form $\{Q, \dots\}$ plus a BRST-invariant operator \mathcal{O} of positive ghost number. In Donaldson theory, \mathcal{O} has non-vanishing matrix elements and affects the theory in an important, but calculable, way. A simplification occurs for $N = 4$: the “vacuum” has zero ghost number (as there is no anomaly), so all matrix elements of \mathcal{O} vanish; hence the partition function is invariant under the perturbation.

Now let us find the vacuum states of the theory. First we do this classically. The superpotential including the perturbation is

$$\widehat{W} = -\frac{1}{2}m\text{Tr}(T^2 + U^2 + V^2) + \text{Tr } T[U, V]. \quad (5.2)$$

The equations for a critical point of W are

$$\begin{aligned} [U, V] &= mT, \\ [V, T] &= mU, \\ [T, U] &= mV. \end{aligned} \quad (5.3)$$

These equations have the following meaning: apart from a factor of m , which can be scaled out, they are the standard commutation relations of the Lie algebra of $SU(2)$. To complete the determination of the classical vacuum states, one must divide the space of solutions of (5.3) by the gauge group G , and set the D terms to zero. Those two steps combined are equivalent to dividing by the complexification $G_{\mathbb{C}}$ of G .

The conclusion then is that the classical vacua are in one to one correspondence with complex conjugacy classes of homomorphisms of the $SU(2)$ Lie algebra to that of G . For instance, for $G = SU(m)$, there is a single irreducible representation – since $SU(2)$ has up to conjugacy a single irreducible m dimensional representation – and various reducible representations. The irreducible embedding breaks the gauge symmetry completely, while the reducible embeddings leave various subgroups of G unbroken. At the opposite extreme from the irreducible solution of (5.3) is the trivial embedding, with $T = U = V = 0$. Here the unbroken symmetry group is G itself.

In what follows, we will mainly consider G to be $SU(2)$ or $SO(3)$, in which case the only embeddings are the irreducible one and the trivial one. For other groups, there are other, intermediate cases. For $G = SU(N)$ with prime N , a simplification arises: the two extreme cases are the only ones in which the unbroken symmetry group is semi-simple; the vacua coming from reducible but non-trivial embeddings all have $U(1)$ factors in the unbroken symmetry group. The partition function of the twisted $N = 4$ theory vanishes for gauge group $U(1)$ on any four-manifold, because of fermion zero modes (which cannot be lifted as the $U(1)$ theory is free). This presumably means that in the generalization of the computation that follows to $SU(N)$ with prime N , the intermediate vacua do not contribute. That would not be so for other groups. For instance, if the gauge group is $G = SU(nm)$ with $n, m > 1$, one can consider an embedding of $SU(2)$ in G such that the nm -dimensional representation of G decomposes as n copies of the m -dimensional representation of $SU(2)$; for this embedding, the unbroken gauge group is $SU(n)$, and hence one would expect the corresponding vacuum to contribute to the $SU(nm)$ theory.

Quantum vacua

Now we consider the vacuum structure at the quantum level. The trivial embedding gives at low energies the pure $N = 1$ supersymmetric gauge theory; the supermultiplets T, U , and V are all massive. The pure $N = 1$ super Yang–Mills theory is asymptotically free and is believed to generate a mass gap (and to undergo confinement) at low energies. For gauge group $SU(m)$, the model has a \mathbb{Z}_{2m} global chiral symmetry,³⁴ which is believed to be spontaneously broken down to the \mathbb{Z}_2 subgroup that permits fermion masses. This gives an m -fold vacuum degeneracy, which is believed to account for the full vacuum degeneracy of the theory.

The chiral symmetry and chiral symmetry breaking have the following interpretation. The gluino field λ of the $N = 1$ theory has $2m$ zero modes in an instanton field; the chiral symmetry is therefore $\lambda \rightarrow e^{\pi i/m} \lambda$. In an instanton field, one finds an expectation value for the operator $(\lambda\lambda)^m$:

³⁴ This is explicitly broken by the coupling to the massive superfields T, U , and V , but the breaking is irrelevant at low energies and does not affect the following remarks.

$$\langle (\lambda\lambda)^m \rangle = A^{3m} e^{i\theta}. \quad (5.4)$$

Here A is the mass scale of the theory and θ is the θ angle. The θ dependence of (5.4), which is seen most naively from the fact that the left-hand side is non-zero at the one instanton level, follows in a standard fashion from the quantum numbers of λ under an anomalous $U(1)$ symmetry of the low energy $N = 1$ theory. Quantum mechanically, $\lambda\lambda$ has a vacuum expectation value; by cluster decomposition this must be an m th root of (5.4):

$$\langle \lambda\lambda \rangle_r = A^3 e^{i\theta/m} e^{2\pi i r/m}. \quad (5.5)$$

Here $r = 1, \dots, m$ labels the choice of vacuum state and $\langle \rangle_r$ is the expectation value in the r th vacuum. Thus, (5.5) shows that the correlation functions in a particular vacuum are not invariant under $\theta \rightarrow \theta + 2\pi$; rather, that transformation cyclically permutes the m vacuum states. In a particular vacuum, the correlators are invariant only under $\theta \rightarrow \theta + 2\pi m$. In the context of strong-weak duality, with modular functions of $\tau = (\theta/2\pi) + (4\pi i/g^2)$, this means that the contributions of individual vacua are invariant only under

$$\tau \rightarrow \tau + m. \quad (5.6)$$

Under $\tau \rightarrow \tau + 1$, the m vacua are cyclically permuted.

That completes our discussion of the trivial embedding. The irreducible vacuum is much easier to analyze since the gauge symmetry group is completely broken at the classical level, so there are no strong quantum effects of any kind (if the gauge coupling is small) and the trivial embedding leads to a unique quantum vacuum. The intermediate vacua can be analyzed as a combination of the above cases, with the extra phenomenon that because of the $U(1)$ factors they have no mass gap.

Thus, for instance, for $G = SU(2)$ there are three vacuum states, two from the trivial embedding and one from the irreducible embedding. They all have mass gaps. The first two are related by a broken symmetry. If indeed w is the generator of the \mathbb{Z}_4 chiral symmetry, then w acts by

$$w : \tau \rightarrow \tau + 1 \quad (5.7)$$

and exchanges the two vacua that come from the trivial embedding. They are each invariant under w^2 , which acts by $\tau \rightarrow \tau + 2$ and generates the unbroken \mathbb{Z}_2 .

5.2. Partition functions of some simple theories

Now we need to practice with certain general points about quantum field theory.

If the gauge group is $G = SU(2)$, the irreducible vacuum does not quite have completely broken gauge symmetry; $SU(2)$ is broken down to its center \mathbb{Z}_2 . On a four-manifold X , the theory in the irreducible vacuum reduces at low energies to \mathbb{Z}_2 gauge theory. By \mathbb{Z}_2 gauge theory, we mean a theory in which one sums over \mathbb{Z}_2 -valued flat

connections – there are 2^{b_1} of them³⁵ – and in which dividing by the volume of the gauge group means dividing by the number of elements in \mathbb{Z}_2 , which is 2. The partition function of \mathbb{Z}_2 gauge theory on X is therefore

$$Z = 2^{b_1-1}. \quad (5.8)$$

(Of course, this formula has a similar origin to the prefactor in (3.17).) For a general finite abelian gauge group Γ , the result would be $(\#\Gamma)^{b_1-1}$, with $\#\Gamma$ the number of elements in Γ .

Now let us consider another simple case, the partition function of a theory that has a mass gap and in which all degrees of freedom can be measured locally (there are no unbroken gauge invariances). The mass gap ensures that a simple result will emerge if one scales up the metric by $g \rightarrow tg$ and takes t large. The contribution to the partition function will behave for large t like $\exp(-L_{\text{eff}})$, where – because of the mass gap – L_{eff} has an expansion in local operators:

$$L_{\text{eff}} = \int_X d^4x \sqrt{g} (u + vR + wR^2 + \dots). \quad (5.9)$$

Now suppose in addition one knows for some reason that the partition function $Z = \exp(-L_{\text{eff}})$ is a topological invariant. Topological invariance means that the operators that arise in (5.9) integrate to give topological invariants. The only topological invariants of a four-manifold that can be written as such integrals of a local operator are the Euler characteristic χ and the signature σ . The contribution to the partition function of a vacuum with a mass gap and in which all degrees of freedom can be measured locally is thus of the form

$$e^{a\chi+b\sigma}, \quad (5.10)$$

where a and b are independent of the particular four-manifold.

For a vacuum with a mass gap that also has an abelian group Γ of local gauge invariances (Γ is a finite group or there would not be a mass gap), the partition function is the product of (5.10) with the partition function of the finite gauge theory:

$$Z = (\#\Gamma)^{b_1-1} e^{a\chi+b\sigma}. \quad (5.11)$$

5.3. Partition function on a hyper-Kähler manifold

Now we will begin an analysis that will lead to a proposal for the partition function of the $SU(2)$ or $SO(3)$ theory on a compact Kähler manifold X with $H^{2,0}(X) \neq 0$. As in [35], we first neglect the difference between the physical and topological theories. This means that we will obtain formulas that are really valid only for hyper-Kähler manifolds (of which there are very few examples) since in the hyper-Kähler case the physical and

³⁵ b_1 is the first Betti number of X . We assume for simplicity that there is no torsion (or at least no 2-torsion) in the cohomology of X , so that the ordinary Betti numbers coincide with the \mathbb{Z}_2 -valued ones.

topological theories coincide. Then we make a correction involving the twisting and the canonical divisor of X to understand the general case.

Because the various vacua of the twisted theory all have mass gaps, we can use (5.11) with Γ trivial for the trivial embeddings and $\Gamma = \mathbb{Z}_2$ for the irreducible embedding. For hyper-Kähler manifolds, the parameters a and b in (5.11) cannot both be detected. The reason is that on a four-dimensional Kähler manifold, the canonical divisor K obeys

$$K \cdot K = 2\chi + 3\sigma. \quad (5.12)$$

(Here $K \cdot K$ is the intersection pairing.) In particular $2\chi + 3\sigma$ vanishes for hyper-Kähler manifolds, which have $K = 0$. Therefore, as long as we are on hyper-Kähler manifolds, we can only see one linear combination of χ and σ (the other will appear when we make corrections involving the canonical divisor). It turns out that particularly nice answers emerge if we choose the combination

$$\nu = \frac{\chi + \sigma}{4}, \quad (5.13)$$

which is always an integer for Kähler manifolds. As long as the twisting can be ignored, the contribution to the partition function from a vacuum state with a mass gap should be $e^{a\nu}$ with a universal a (times a factor involving the discrete gauge symmetry). In [35], as there was no dimensionless coupling constant, a was simply a constant. In our present problem, there is a dimensionless coupling constant τ and a is a function of τ .

For the $SU(2)$ theory, there are three vacuum states, all with mass gaps, so the partition function ignoring the twisting would therefore be

$$Z = 2^{b_1-1} e^{a(\tau)\nu} + e^{b(\tau)\nu} + e^{c(\tau)\nu}, \quad (5.14)$$

where the first term is the contribution from the irreducible embedding (with its unbroken \mathbb{Z}_2 gauge invariance) and the last two terms are the contributions from the two vacua that come from reducible embeddings. Further, the contribution of the irreducible vacuum is periodic in θ , while the other two are related by a broken symmetry, so

$$\begin{aligned} e^{a(\tau+1)} &= e^{a(\tau)}, \\ e^{b(\tau+1)} &= i^{-1} e^{c(\tau)}, \\ e^{c(\tau+1)} &= i^{-1} e^{b(\tau)}. \end{aligned} \quad (5.15)$$

Moreover a and b are such that the sum in (5.14) transforms as a modular form for an appropriate subgroup of $SL(2, \mathbb{Z})$ as explained in Section 3. Only the factors of i^{-1} on the right-hand side of (5.15) may need special explanation here. The reason for these factors is that, although the bulk theory is invariant under the \mathbb{Z}_4 symmetry generated by w , on a four-manifold X the path integral measure transforms with a factor of i^ν ; this can be seen by counting λ zero modes and is explained in the derivation of equation (2.46) of [35] (in comparing to that equation note that $\Delta \cong \nu$ modulo 4).

Comparing to the answer found for $K3$ for $SU(2)$ gauge group in Section 4.1, we see that it is naturally written in precisely this form, with

$$e^{a(\tau)} = \frac{1}{2} G(q^2)^{1/2} = \frac{1}{2q \prod_{n=1}^{\infty} (1 - q^{2n})^{12}},$$

$$\begin{aligned}
e^{b(\tau)} &= \frac{1}{2} G(q^{1/2})^{1/2} = \frac{1}{2 q^{1/4} \prod_{n=1}^{\infty} (1 - q^{n/2})^{12}}, \\
e^{c(\tau)} &= \frac{1}{2} G(-q^{1/2})^{1/2} = \frac{1}{2 q^{1/4} \prod_{n=1}^{\infty} (1 - (-1)^n q^{n/2})^{12}}.
\end{aligned} \tag{5.16}$$

Since $\nu = 2$ for $K3$, the signs on the right-hand side of (5.16) are undetermined, and have been selected for later convenience.³⁶ Of course, this identification of a , b , and c is not really rigorous, but it is so natural that we will accept it. Notice that – since e^a , e^b , and e^c cannot have zeroes – the fact that the η function has no zeroes in the upper half plane is essential for the formula to make sense. Also, the factors of i on the right hand side of (5.15) come from the leading powers $q^{-1/4}$ in e^b and e^c (e^a is instead strictly invariant under $\tau \rightarrow \tau + 1$ as the leading power of q is integral).

For $SU(m)$ the analog of (5.14) would be

$$Z = m^{b_1-1} e^{a(\tau)\nu} + e^{b_1(\tau)\nu} + e^{b_2(\tau)\nu} + \dots + e^{b_m(\tau)\nu}, \tag{5.17}$$

where the first term comes from the irreducible embedding and the other m terms from the trivial embedding. In addition to modular invariance of the whole sum, the individual terms should obey

$$\begin{aligned}
e^{a(\tau+1)} &= e^{a(\tau)}, \\
e^{b_r(\tau+1)} &= e^{-i\pi/m} e^{b_{r+1}(\tau)}.
\end{aligned} \tag{5.18}$$

The proposal for the $SU(m)$ partition function on $K3$ that was written at the end of Section 4.1 is naturally written in this form with obvious choices of a and the b_r . Eq. (5.17) should be valid for m prime; otherwise, intermediate vacua contribute, as explained above.

5.4. Effects of twisting

Now we want to consider what happens on more general Kähler manifolds with a non-trivial canonical divisor. The physical and twisted models no longer coincide, and the twisting has the following effect. Of the three superfields T, U , and V , two, say T and U , are scalars in the twisted model and the third, V , is a $(2,0)$ -form. (This is already clear from the description of the quantum numbers in Section 2.4.) Moreover, the superpotential W now transforms as a $(2,0)$ form on X (as in Eq. (2.42) of Ref. [35]).

Given the quantum numbers, a mass term combining V and one of the other superfields, say U , can be written without any difficulty: $W_1 = -m \text{Tr } UV$. However, to give a mass term to all three superfields, one needs as in Ref. [35] to pick a c -number holomorphic two-form ω on X . (So, in particular, $H^{2,0}(X)$ must be non-zero.) For instance, once such an ω is picked, one can give T a spatially dependent mass term

$$W_2 = -\omega \text{Tr } T^2. \tag{5.19}$$

³⁶ It would be possible to multiply this formula and all subsequent ones that involve a , b , and c by a factor of $(-1)^\nu$, which we are unable to determine since the available computations are for manifolds with even ν .

We now consider the effects of this perturbation. Away from the zeroes of ω , the superpotential $W_1 + W_2$ can be analyzed just as we have done above, leading to the familiar classification of vacuum states in terms of $SU(2)$ embeddings. Near zeroes of ω , the picture changes. In analyzing this, we will consider only gauge group $SU(2)$ or $SO(3)$, so in the bulk theory we have only the trivial and irreducible embeddings to consider. Both exhibit new behavior near zeroes of ω , but in very different ways. In working out the details, we assume, for simplicity (as in [35]), that the zeroes of ω are a union of disjoint, smooth complex curves C_i , $i = 1, \dots, n$ of genus g_i , on which ω has a simple zero.

The trivial embedding

First we consider the strongly coupled vacua that come from the trivial embedding. The C_i behave like the world-sheets of “superconducting cosmic strings” that can trap charges (see Section 3.3 of [35]). Even after understanding in bulk the dynamics of the strongly coupled four-dimensional theory, one must analyze the behavior of the effective two-dimensional theory near C_i .

The main points that one can make come from the symmetry structure (see Section 2.7 of Ref. [35]). In bulk, the theory has a mass gap with \mathbb{Z}_4 spontaneously broken down to \mathbb{Z}_2 . This \mathbb{Z}_2 is realized near the C_i as a chiral symmetry that prevents fermion masses in the effective two-dimensional theory, so it is natural that it should be spontaneously broken, giving a two-fold vacuum degeneracy. It is natural to assume that this strongly coupled vacuum state has a mass gap and no vacuum degeneracy except what follows from the symmetry breaking.

The vacuum structure (for states coming from the trivial embedding of $SU(2)$) is thus as follows. In bulk there are two vacuum states, say $|+\rangle$ and $|-\rangle$, related by the $\mathbb{Z}_4 \rightarrow \mathbb{Z}_2$ symmetry breaking. Near any of the C_i there is a further two-fold bifurcation of the vacuum; $|+\rangle$, for instance, splits into $|++\rangle$ and $|+-\rangle$.

According to (5.7), the generator w of the underlying \mathbb{Z}_4 symmetry acts on the contributions of these vacua to the partition function by $\tau \rightarrow \tau + 1$. The \mathbb{Z}_2 that is unbroken in bulk is generated by w^2 , which acts by $\tau \rightarrow \tau + 2$. Hence the $|++\rangle$ and $|+-\rangle$ vacua are exchanged by $\tau \rightarrow \tau + 2$. This also exchanges $|--\rangle$ with $|+-\rangle$.

Now, consider a particular vacuum, say $|++\rangle$, along a cosmic string component C_i . Its contribution to the partition function is again of the form $\exp(-L_{\text{eff}})$, where L_{eff} has an expansion in local operators as in (5.9). The only difference is that now they are two-dimensional local operators since we are determining the contribution localized near C_i where the bulk description fails. The only topological invariant of C_i that can be written as the integral of a local operator (indeed, its only topological invariant) is its Euler characteristic $\chi(C_i) = 2 - 2g_i$. Hence the contribution of $|++\rangle$ near C_i is a factor of $e^{(1-g_i)u(\tau)}$, where the function $u(\tau)$ is independent of the details of X and C_i . Similarly, the $|+-\rangle$ vacuum contributes a factor $e^{(1-g_i)v(\tau)}$, where u and v are exchanged by $\tau \rightarrow \tau + 2$:

$$\begin{aligned} u(\tau + 2) &= v(\tau) , \\ u(\tau + 4) &= u(\tau) . \end{aligned} \tag{5.20}$$

With all its possible bifurcations, which must be chosen independently on each C_i , the $|+\rangle$ vacuum contributes a factor of

$$e^{b(\tau)\nu} \prod_{i=1}^s (e^{u(\tau)(1-g_i)} + t_i e^{\nu(\tau)(1-g_i)}). \quad (5.21)$$

The factors of $t_i = \pm 1$ have the same origin as the factors of i^{-1} on the right-hand side of (5.15) – they incorporate a global anomaly (in the coupling to gravity) in the \mathbb{Z}_2 symmetry that permutes $|++\rangle$ and $|+-\rangle$.³⁷ Similarly the possible bifurcations of the $|-\rangle$ vacuum contribute the transform of this under $\tau \rightarrow \tau + 1$, or

$$e^{b(\tau+1)\nu} \prod_{i=1}^s (e^{u(\tau+1)(1-g_i)} + t_i e^{\nu(\tau+1)(1-g_i)}). \quad (5.22)$$

Now, (5.12) is equivalent to the formula

$$2\chi + 3\sigma = \sum_i (g_i - 1). \quad (5.23)$$

This formula is the real reason that we need not in the bulk theory consider an extra factor of the form $\exp(\gamma(\tau)(2\chi + 3\sigma))$. It could be absorbed in redefining the function u .

The non-trivial embeddings

Now we analyze what becomes of the non-trivial embeddings when the mass terms involve a choice of holomorphic two-form. What this means is that we consider the theory with the combined superpotential

$$W_{\text{tot}} = -\omega \text{Tr} T^2 - m \text{Tr} UV + \text{Tr} T[U, V], \quad (5.24)$$

and look for a classical solution with zero action in the twisted theory. The analysis is feasible because by arguments similar to those in Section 2.4, the $(2, 0)$ part of the curvature can be assumed to vanish, and the structure is therefore holomorphic.

The condition for a critical point of W_{tot} is

$$\begin{aligned} [T, U] &= mU, \\ [T, V] &= -mV, \\ [U, V] &= 2\omega T. \end{aligned} \quad (5.25)$$

These equations have the following immediate consequence. Regardless of what ω does, the matrix T is at each point in X conjugate to

$$\frac{m}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (5.26)$$

and in particular never vanishes. Thus T breaks the gauge group down to $U(1)$.

³⁷ This factor appears in Eq. (2.66) of Ref. [35] and is explained in the derivation of that equation.

For the kinetic energy of T , U , and V to vanish in the twisted theory does not require that they should be covariantly constant, but only that they should be holomorphic. The existence of a holomorphic T that is everywhere in the conjugacy class indicated in (5.26) means that the $SU(2)$ gauge bundle E splits as a sum $E \cong L \oplus L^{-1}$ with some holomorphic line bundle L . With such a splitting, T then *globally* takes the form of (5.26). Eq. (5.25) then implies that U and V are of the form

$$\begin{aligned} U &= \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}, \\ V &= \begin{pmatrix} 0 & 0 \\ c & 0 \end{pmatrix}, \end{aligned} \quad (5.27)$$

with

$$bc = m\omega. \quad (5.28)$$

Here b is a holomorphic section of the line bundle L^2 , and c is a holomorphic section of the line bundle $K \otimes L^{-2}$. (K enters because V is a $(2,0)$ -form.)

In order to make the exposition in the rest of this section as simple as possible, we will assume first that the zeroes of ω consist of a *single* smooth connected curve C , on which ω has a simple zero. There is then only a single t_i in (5.21) and (5.22), and according to Ref. [35], Eq. (2.61), it is

$$t = (-1)^\nu. \quad (5.29)$$

At the very end of the section we will return to the case that the canonical divisor is a union of disjoint smooth curves C_i .

With the given assumption about ω , (5.28) is easy to analyze. The product bc vanishes with multiplicity one on the irreducible smooth curve C , so either c vanishes on C and b has no zeroes, or vice-versa. If b has no zeroes, the line bundle L is trivial, and up to a complex gauge transformation, $b = m$ and $c = \omega$. If c has no zeroes, then $K \otimes L^{-2}$ is trivial, so $L \cong K^{1/2}$ (that is, L is a line bundle with $L^{\otimes 2} \cong K$) and up to a complex gauge transformation, $b = \omega$ and $c = m$.

Once the holomorphic data are known the metric on the gauge bundle is determined (presumably uniquely, by a convexity argument, but certainly not explicitly) by the condition (analogous to Eq. (2.67) in Section 2.4) for the $(1,1)$ part of the curvature.

Thus, we have determined the possible vacuum solutions: there is one on the trivial bundle E_0 , which has instanton number zero, and one on the bundle $E_1 = K^{1/2} \oplus K^{-1/2}$, whose instanton number is

$$-\frac{K \cdot K}{4} = -\frac{g-1}{4}, \quad (5.30)$$

with g the genus of C . Note that this number is typically negative. Actually, we saw in Section 2.4 that bundles of negative instanton number can contribute when the vanishing theorem fails, and we saw (under a hypothesis that was equivalent to having the canonical divisor connected and $K \cdot K > 0$) that the most negative possible value of a bundle that would contribute was $-K \cdot K/4$.

Of course, it might happen that X is not a spin manifold. Then $K^{1/2}$ does not exist. In that case, only E_0 will contribute to the $SU(2)$ theory. However, E_1 will always contribute in the $SO(3)$ theory. Indeed, the $SO(3)$ bundle $\text{ad}(E_1)$ derived from E_1 , which is $K \oplus K^{-1} \oplus \mathcal{O}$ (with \mathcal{O} a trivial bundle) always exists. Its second Stieffel–Whitney class is $w_2(\text{ad}(E_1)) = w_2(X)$, where $w_2(X)$ is the second Stieffel–Whitney class of the tangent bundle of X , which is the same as the reduction modulo two of $c_1(K)$. Indeed, $\text{ad}(E_1)$ is the bundle of self-dual two-forms, while E_1 , if it exists, is one of the two chiral spin bundles of X ; the obstruction to constructing the latter from the former is $w_2(X)$.

The general structure

The general structure of the partition function is then as follows, for the $SU(2)$ theory:

$$\begin{aligned} Z_{SU(2)} = & 2^{b_1-1} e^{a(\tau)\nu} (F + G \delta_{w_2(X)=0}) \\ & + e^{b(\tau)\nu} (e^{(1-g)u(\tau)} + (-1)^\nu e^{(1-g)u(\tau+2)}) \\ & + e^{c(\tau)\nu} (e^{(1-g)u(\tau+1)} + (-1)^\nu e^{(1-g)u(\tau+3)}). \end{aligned} \quad (5.31)$$

(The factors of $(-1)^\nu$ come from the t_i in (5.21) and (5.22) via (5.29).) F and G are corrections near the canonical divisor to contributions of the vacuum bundles E_0 and E_1 ; of course E_1 only contributes to the $SU(2)$ theory if $w_2(X) = 0$. In what follows we abbreviate $x_0 = w_2(X)$. The functions a, b, c , and u are universal (independent of X) because of the mass gaps. It might be possible to argue *a priori* for a similar universality of F and G ,³⁸ but in any case S -duality implies such universality.

Now given (2.27), we want to determine the analogous formula for the $SO(3)$ theory. In the $SO(3)$ theory, one must sum over all values of $w_2(E)$. One can nearly interpret (5.31) as the contribution of bundles with $w_2(E) = 0$ to the $SO(3)$ theory, but it is necessary to make a correction that involves comparing the volumes of the $SU(2)$ and $SO(3)$ gauge groups, as discussed in Section 3. So to get from (5.31) the contribution to the $SO(3)$ theory from bundles with $w_2(E) = 0$, one simply divides by 2^{b_1-1} .

Once this is known, there is no problem getting the contribution in the $SO(3)$ theory from bundles with an arbitrary value of $x = w_2(E)$. The F and G functions anyway only contribute for a particular value of x , and the others because of the mass gap only see the global object x through its influence on anomalies which determine the relative phases between contributions of different vacua. The x -dependence of these anomalies is explained in Ref. [35] in the derivation of Eq. (2.79). Putting together the anomalies and the factor of 2^{-b_1+1} , the contribution in the $SO(3)$ theory of bundles with $w_2(E) = x$ is

$$\begin{aligned} Z_x = & e^{a(\tau)\nu} (F \delta_{x=0} + G \delta_{x=x_0}) \\ & + 2^{1-b_1} e^{b(\tau)\nu} (e^{(1-g)u(\tau)} + (-1)^{\nu+x \cdot x_0} e^{(1-g)u(\tau+2)}) \\ & + i^{-x^2} 2^{1-b_1} e^{c(\tau)\nu} (e^{(1-g)u(\tau+1)} + (-1)^{\nu+x \cdot x_0} e^{(1-g)u(\tau+3)}). \end{aligned} \quad (5.32)$$

³⁸ It is not clear whether for the irreducible vacua, the weakly coupled theories that arise near the canonical divisor have mass gaps.

Of course, the F function, associated with the vacuum bundle E_0 , only contributes for $x = 0$, and the G function, associated with the vacuum bundle E_1 , only contributes if $x = x_0 = w_2(X)$.

The unknown functions can be determined as follows. According to the blowing-up formula of Yoshioka, blowing up a point in a Kähler manifold X_0 multiplies the partition function by a factor of θ_0/η^2 with η the Dedekind eta function and

$$\theta_0 = \sum_{n \in \mathbb{Z}} q^{n^2}. \quad (5.33)$$

(In this assertion, x is taken to be a pullback from X_0 .) If we consider the special case that X_0 is $K3$ (so the partition function on its one-point blow-up X is governed by (5.32) with $g = 0, \nu = 2, x \cdot x_0 = 0$),³⁹ this implies that

$$F(\tau) = e^{u(\tau)} + e^{u(\tau+2)} = e^{u(\tau+1)} + e^{u(\tau+3)} = \frac{\theta_0}{\eta^2}. \quad (5.34)$$

The function θ_0/η^2 transforms in a two-dimensional representation of $SL(2, \mathbb{Z})$, the other function that enters being θ_1/η^2 with

$$\theta_1 = \sum_{n \in \mathbb{Z} + \frac{1}{2}} q^{n^2}. \quad (5.35)$$

Under $\tau \rightarrow -1/\tau$, F and G are mapped to $e^{u(\tau)}$ and $e^{u(\tau+2)}$, respectively, as will be clear presently when we check the modular transformation laws in detail. Since F is known (at least in this special case) $e^{u(\tau)}$ is thereby determined – once and for all, since u is a universal function. By applying $\tau \rightarrow -1/\tau$ one now determines F and G in general. In particular, one finds that $G \sim (\theta_1/\eta^2)^{1-g} \sim q^{(1-g)/4} (1 + \dots) \eta^2$ where the leading exponent $q^{(1-g)/4}$ is in happy agreement with the instanton number of the classical solution that is responsible for the presence in the formula of the G function.

In the following subsection, we will write down a precise formula that was found by the reasoning just sketched and verify that it has all of the right properties. The reader who works through the verification should be able to see that given the general structure that we have proposed, the formula we write down is the only one that works.

5.5. The formula

We continue to assume temporarily that the canonical divisor of X is connected; its genus is

$$g - 1 = 2\chi + 3\sigma. \quad (5.36)$$

b_2^+ and b_2^- will denote the dimensions of the spaces $H^{2,+}$ and $H^{2,-}$ of self-dual and anti-self-dual harmonic two-forms on X ; then we have

³⁹ Recall from the end of Section 2.4 that the vanishing theorem holds for the blowup of $K3$ – so we can apply instanton results such as those of Yoshioka. Similarly, we can consider $K3$ with an arbitrary number of points blown up.

$$\begin{aligned}
b_2 &= b_2^+ + b_2^-, \\
\sigma &= b_2^+ - b_2^-, \\
\chi &= 2 - 2b_1 + b_2^+ + b_2^-,
\end{aligned} \tag{5.37}$$

with b_2 the second Betti number of X .

The formula we propose is then

$$\begin{aligned}
Z_x &= \left(\frac{1}{4} G(q^2) \right)^{\nu/2} \left(\delta_{x,0} (-1)^\nu \left(\frac{\theta_0}{\eta^2} \right)^{1-g} + \delta_{x,x_0} \left(\frac{\theta_1}{\eta^2} \right)^{1-g} \right) \\
&+ 2^{1-b_1} \left(\frac{1}{4} G(q^{1/2}) \right)^{\nu/2} \left(\left(\frac{\theta_0 + \theta_1}{2\eta^2} \right)^{1-g} + (-1)^{\nu+x \cdot x_0} \left(\frac{\theta_0 - \theta_1}{2\eta^2} \right)^{1-g} \right) \\
&+ 2^{1-b_1} i^{-x^2} \left(\frac{1}{4} G(-q^{1/2}) \right)^{\nu/2} \left(\left(\frac{\theta_0 - i\theta_1}{2\eta^2} \right)^{1-g} + (-1)^{\nu+x \cdot x_0} \left(\frac{\theta_0 + i\theta_1}{2\eta^2} \right)^{1-g} \right)
\end{aligned} \tag{5.38}$$

This formula manifestly transforms correctly under $\tau \rightarrow \tau + 1$. According to the precise version of the strong-weak duality conjecture proposed in Section 3.3, the transformation law under $\tau \rightarrow -1/\tau$ should be

$$Z_y(-1/\tau) = 2^{-b_2/2} (-1)^\nu \left(\frac{\tau}{i} \right)^{-\chi/2} \sum_x (-1)^{x \cdot y} Z_x(\tau). \tag{5.39}$$

To evaluate the sum over x on the right-hand side, we need various formulas:

$$\begin{aligned}
\sum_x (-1)^{x \cdot y} \delta_{x,0} &= 1, \\
\sum_x (-1)^{x \cdot y} \delta_{x,x_0} &= (-1)^{y \cdot x_0}, \\
\sum_x (-1)^{x \cdot y} &= 2^{b_2} \delta_{y,0}, \\
\sum_x (-1)^{x \cdot y + x \cdot x_0} &= 2^{b_2} \delta_{y,x_0}, \\
\sum_x (-1)^{x \cdot y} i^{-x^2} &= 2^{b_2/2} i^{-\sigma/2 + y^2}, \\
\sum_x (-1)^{x \cdot y} i^{-x^2} (-1)^{x \cdot x_0} &= 2^{b_2/2} i^{\sigma/2 - y^2}.
\end{aligned} \tag{5.40}$$

The only formulas here that require comment are the last two. The next-to-last sum in (5.40), on completing the square, is $\sum_x i^{-(x+y)^2 + y^2} = i^{y^2} \sum_x i^{-x^2}$, and so is equivalent to the sum $\sum_x i^{-x^2}$, which is a special case of a sum that we met in Section 3.3. The last sum in (5.40) can be treated the same way; alternatively, using the Wu formula,

$$(-1)^{x \cdot x_0} = (-1)^{x^2}, \tag{5.41}$$

the last sum is the complex conjugate of the one before.

Using these sums, we find that the right-hand side of (5.39) is (dropping the modular weight)

$$\begin{aligned} \text{R.H.S.} = & 2^{-b_2/2} \left(\frac{G(q^2)}{4} \right)^{\nu/2} \left(\left(\frac{\theta_0}{\eta^2} \right)^{1-g} + (-1)^{\nu+x \cdot x_0} \left(\frac{\theta_1}{\eta^2} \right)^{1-g} \right) \\ & + 2^{b_2/2+1-b_1} \left(\frac{G(q^{1/2})}{4} \right)^{\nu/2} \left(\delta_{x,0} (-1)^\nu \left(\frac{\theta_0 + \theta_1}{2\eta^2} \right)^{1-g} + \delta_{x,x_0} \left(\frac{\theta_0 - \theta_1}{2\eta^2} \right)^{1-g} \right) \\ & + 2^{1-b_1} \left(\frac{G(-q^{1/2})}{4} \right)^{\nu/2} \left((-1)^\nu i^{x^2-\sigma/2} \left(\frac{\theta_0 - i\theta_1}{2\eta^2} \right)^{1-g} + i^{-x^2+\sigma/2} \left(\frac{\theta_0 + i\theta_1}{2\eta^2} \right)^{1-g} \right). \end{aligned} \quad (5.42)$$

To verify (5.39), we need the modular transformations of the various functions. Under $\tau \rightarrow -1/\tau$, we have (apart from modular weight)

$$\begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} \rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}. \quad (5.43)$$

This means, in particular, that θ_0 is exchanged with $(\theta_0 + \theta_1)/\sqrt{2}$ and θ_1 with $(\theta_0 - \theta_1)/\sqrt{2}$, while $(\theta_0 \pm i\theta_1)$ is mapped to $i^{\pm 1/2}(\theta_0 \mp \theta_1)$. These facts are convenient, since the functions just mentioned all appear raised to the $(1-g)$ power in (5.38). We also have (apart from the modular weight)

$$\begin{aligned} G(q^2)^{\nu/2} & \rightarrow 2^{\frac{3}{2}(\chi+\sigma)} G(q^{1/2})^{\nu/2}, \\ G(q^{1/2})^{\nu/2} & \rightarrow 2^{-\frac{3}{2}(\chi+\sigma)} G(q^2)^{\nu/2}, \\ G(-q^{1/2})^{\nu/2} & \rightarrow G(-q^{1/2})^{\nu/2}. \end{aligned} \quad (5.44)$$

With the aid of these transformation laws and identities (5.37) and (5.41), it is straightforward to verify (5.39).

Disconnected canonical divisor

It remains only to generalize the above formulas for the case that ω vanishes (with multiplicity one) on a union of disjoint smooth components C_i , $i = 1, \dots, n$, of genus g_i .

This has two effects: the bulk vacua $|+\rangle$ and $|-\rangle$ that come from the trivial embedding of $SU(2)$ bifurcate along the C_i into 2^n separate ground states – in a fashion that we have already described in (5.22) and (5.21), before we specialized to $n = 1$. Also, the contribution of the irreducible embedding is more complicated.

It is straightforward to write down the contribution of the trivial embedding since we now have determined all of the functions that appear in (5.22) and (5.21). The contribution is

$$2^{1-b_1} \left(\frac{G(q^{1/2})}{4} \right)^{\nu/2} \prod_{i=1}^n \left(\left(\frac{\theta_0 + \theta_1}{2\eta^2} \right)^{1-g_i} + t_i (-1)^{x \cdot C_i} \left(\frac{\theta_0 - \theta_1}{2\eta^2} \right)^{1-g_i} \right)$$

$$+2^{1-b_1} i^{-x^2} \left(\frac{G(-q^{1/2})}{4} \right)^{\frac{x}{2}} \prod_{i=1}^n \left(\left(\frac{\theta_0 - i\theta_1}{2\eta^2} \right)^{1-g_i} + t_i (-1)^{x \cdot C_i} \left(\frac{\theta_0 + i\theta_1}{2\eta^2} \right)^{1-g_i} \right). \quad (5.45)$$

The t_i obey

$$\prod_i t_i = (-1)^\nu, \quad (5.46)$$

as shown in Eq. (3.57) of Ref. [35]). Also, $\prod_i (-1)^{x \cdot C_i} = (-1)^{x \cdot x_0}$ as the union of the C_i is the canonical divisor, whose Poincaré dual reduces modulo two to x_0 .

To find the contributions of the irreducible embedding, we must reexamine the factorization $bc = m\omega$ of equation (5.28). The possible solutions are as follows. Given that ω vanishes with multiplicity one on the union of the C_i , b can vanish with multiplicity one on any subset of the C_i , with c vanishing on the others. If then $L_i = \mathcal{O}(C_i)$ is the line bundle whose sections are functions with a simple zero on C_i , the $SU(2)$ bundle E is $E_{\vec{\epsilon}} = L_{\vec{\epsilon}}^{-1/2} \oplus L_{\vec{\epsilon}}^{-1/2}$, where

$$L_{\vec{\epsilon}} = \otimes_{i=1}^n L_i^{\epsilon_i}, \quad (5.47)$$

where $\epsilon_i = 0$ or 1 are chosen independently. There are thus 2^n solutions in all (a fact which under $SL(2, \mathbb{Z})$ transforms into the fact that the $|+\rangle$ or $|-\rangle$ vacuum in bulk bifurcates into 2^n choices along the cosmic strings). Of course, as $L_{\vec{\epsilon}}$ may not have a square root, the bundles $E_{\vec{\epsilon}}$ may not really exist as $SU(2)$ bundles, but the corresponding $SO(3)$ bundles $\text{ad}(E_{\vec{\epsilon}})$ always exist – with second Stiefel–Whitney class

$$w_2(\vec{\epsilon}) = \sum_i \epsilon_i [C_i]. \quad (5.48)$$

Here $[C_i]$ is the reduction modulo two of the Poincaré dual of C_i ; one has $w_2(X) = \sum_i [C_i]$. The instanton number of the bundle $E_{\vec{\epsilon}}$ is

$$-\sum_i \epsilon_i \frac{g_i - 1}{4}. \quad (5.49)$$

The contribution of all 2^n solutions of $bc = m\omega$ is

$$\left(\frac{G(q^2)}{4} \right)^{\nu/2} \sum_{\vec{\epsilon}} \delta_{x, w_2(\vec{\epsilon})} \prod_{i=1}^n \left\{ t_i^{\epsilon_i} \left(\frac{\theta_0}{\eta^2} \right)^{(1-\epsilon_i)(1-g_i)} \left(\frac{\theta_1}{\eta^2} \right)^{\epsilon_i(1-g_i)} \right\}. \quad (5.50)$$

This formula was found by requiring that it is zero except if x is equal to $w_2(\vec{\epsilon})$ for some $\vec{\epsilon}$ and that the sum of (5.50) and (5.45) transforms correctly under $\tau \rightarrow -1/\tau$. That last assertion can indeed be verified using the identities that have already been exploited above. Note also that the smallest power of q in the contribution of a given solution agrees with the instanton number in (5.49).

6. Connections with RCFTs and strings

In this paper, especially in Section 4, formulas familiar in the context of two-dimensional rational conformal field theories (RCFTs) and strings made unexpected appearances. In this section we discuss some aspects of this mysterious phenomenon. We will first consider relations to RCFTs and then speculate about the potential applications to the question of S -duality in string theory.

We have seen in Section 4 that blowing up of a point has the effect of multiplying the $N = 4$ partition function with $SU(2)$ gauge group by essentially the level one characters of the $SU(2)$ WZW model. The magnetic flux vector labels the conformal blocks, and the instanton number mod 1 in the gauge theory corresponds to the conformal dimension mod 1 in the RCFT. Moreover we saw that the level k characters of $SU(n)$ arise for $N = 4$ super Yang–Mills theory with gauge group $U(k)$ on ALE spaces A_{n-1} . One might think that the appearance of conformal field theory characters simply reflects the fact that those objects are modular and that there are not too many modular objects of low weight and level. However, the connections go beyond the partition function. For the ALE spaces, for instance, Nakajima does not just get a WZW character but finds an action of the affine Kac–Moody algebra on the Hilbert space consisting of the cohomology of instanton moduli spaces.

Generalizing the structure found by Nakajima, one might wonder, for instance, whether for a more general four-manifold there is a natural action of the Virasoro algebra on the cohomology of instanton moduli spaces if one consider all instanton numbers at once. And what is the analog of the two-dimensional operator product expansion? Can one find for each rational conformal theory a (possibly non-compact) four-manifold whose $N = 4$ twisted partition function gives the characters of that RCFT?

In two-dimensional RCFT's, the conformal blocks χ_i form a representation of $SL(2, \mathbb{Z})$, much in the same way as do our partition functions \mathbb{Z}_v for $v \in H^2(M, \mathbb{Z}_n)$. RCFT's have additional structure: From the underlying operator-product relations one deduces a commutative, associative multiplication law called the Verlinde algebra on the space of blocks [58]. Does it have an analog in our problem? Since the Verlinde algebra is determined by the S matrix, and since the four-dimensional problem has an S matrix given in (3.15), we can deduce that the analog of the Verlinde-algebra should be the multiplication law

$$\mathbb{Z}_{v_1} \cdot \mathbb{Z}_{v_2} = \mathbb{Z}_{v_1+v_2}.$$

This operation is indeed commutative and associative (it is associated with the ordinary addition law of the \mathbb{Z}_n -valued flux) further enhancing the analogy between four dimensional gauge theories and rational conformal field theory.

Stringy spectrum?

Before going on, let us recall some aspects of the tests of S -duality in string theory [9]. One basic test is to identify all the BPS-saturated states with given “electric” charges and find what “magnetic states” they lead to under $SL(2, \mathbb{Z})$ transformation. One can then try to see if these states exist. This is usually done for the field theoretic modes. However, in string theory there is actually a full stringy spectrum of massive

BPS-saturated states. For heterotic strings the number of such states is given by the (left-moving) bosonic string oscillator partition function [59]. To see this one notes that if one takes the supersymmetric oscillators to be in the ground state, then regardless of the lattice momenta, one gets a “small” supersymmetry representation corresponding to BPS-saturated states [2]. Thus all the states which come from $N_R = 0$ and arbitrary N_L saturate the BPS bound and, if S -duality is to hold, there must be corresponding magnetic states. For $N_L = 0$, these states are the BPS monopoles that can be seen in field theory. For $N_L = 1$ it was suggested in [9] that they have the right quantum numbers to be identifiable with so-called H -monopoles [60,61]. Based on the $N = 4$ structure, and the fact that H -monopoles saturate the BPS bound, one expects the moduli space of H -monopoles to have a hyper-Kähler structure. Moreover the “magnetic” states would correspond to the cohomology of this moduli space. Unfortunately this moduli space is little understood. However, $K3$ was proposed as a candidate in Ref. [36], in part because it is four dimensional and hyper-Kähler and has 24 dimensional cohomology. The fact that $K3$ has 24 dimensional cohomology matches the fact that the bosonic string at $N_L = 1$ has 24 physical states (namely $\alpha_{-1}^i |0\rangle$, where i runs over the transverse directions in the light-cone gauge). Moreover, the Lorentz spin of these 24 states agrees with that of the cohomology of $K3$ if we identify the helicity operator in space-time with $(F_L + F_R)/2$, where we define an element of the cohomology group $H^{p,q}$ to have $F_L = p - 1$ and $F_R = q - 1$. Indeed, there are 22 states with helicity 0, one with helicity +1 and one with helicity -1.⁴⁰

Is there an analog of this for $N_L > 1$? The discussion in Section 4 makes clear a possible analog: the string states at arbitrary N_L can be associated with the cohomology of the symmetric product of N_L copies of $K3$! The helicities work out correctly, just as they do for $K3$. Note that the main thing needed for $K3$ to give the bosonic string partition function is its Euler characteristic (and the hyper-Kähler structure which enables one to use the orbifold formula to compute the cohomology of the symmetric product). So it is conceivable that $K3$ could be replaced here by another, perhaps non-compact, space.

If one compactifies a Type II superstring on a six-torus, there is again a stringy spectrum of BPS-saturated states, this time in 1-1 parallel with fermionic oscillator states. Is there any way to generate the oscillator states of a fermionic string by taking the cohomology of a symmetric product? The $N = 4$ supersymmetry again suggests that the relevant moduli space would be a hyper-Kähler manifold. In dimension four, other than $K3$ there is a unique compact example, namely the four-torus T^4 . The cohomology of T^4 is 16 dimensional, with 8 “bosonic” states (of even degree 0, 2, or 4) and 8 “fermionic” ones (of odd degree 1 or 3). This agrees with the fact that the fermionic string has 8 transverse Bose oscillators and 8 Fermi ones. If we take the helicity operator to be $(F_L + F_R)/2$, then the bosonic states in the cohomology of T^4 include six states of helicity zero and two of helicity ± 1 while the Fermi states include four of helicity $1/2$ and four of helicity $-1/2$. This agrees with the quantum numbers of the transverse

⁴⁰ If we compactify to six dimensions, there are two light cone helicity operators; identifying them with $(F_L + F_R)/2$ and $(F_L - F_R)/2$, the massless states of the bosonic string compactified to six dimensions agree with the cohomology of $K3$. This comment also applies to the type II superstring considered above.

oscillators of the string! The cohomology of a symmetric product of T^4 's can be analyzed by the same orbifold techniques that we used for $K3$. It is a Fock space derived from “one-particle states” which are the cohomology of T^4 , and so agrees with the fermionic string spectrum.

While we have little insight to offer at the moment, this relation of the oscillator spectrum of bosonic and fermionic strings to the cohomology of the two compact hyper-Kähler manifolds in four dimensions is certainly provocative.

Acknowledgements

We would like to thank P. Kronheimer for explaining the background about $K3$ surfaces, H. Nakajima and K. Yoshioka for patient explanations of their work and helpful suggestions, D. Gross, D. Morrison and D. Zagier for explanations of some points relevant to Section 2–4, C. Taubes for bringing to our attention various relevant work in the literature as well as discussions on instanton moduli spaces, T. Miyake for advice about some technical points, and G. Shimura and E. Verlinde for valuable comments. In addition C.V. would like to thank the Institute for Advanced Study for its hospitality and for providing a stimulating research environment where this work was done.

The research of C.V. was supported in part by the Packard fellowship and NSF grants PHY-92-18167 and PHY-89-57162. The work of E.W. was supported in part by NSF Grant PHY92-45317.

References

- [1] C. Montonen and D. Olive, Phys. Lett. B 72 (1977) 117;
P. Goddard, J. Nyuts and D. Olive, Nucl. Phys. B 125 (1977) 1.
- [2] E. Witten and D. Olive, Phys. Lett. B 78 (1978) 97.
- [3] H. Osborn, Phys. Lett. B 83 (1979) 321.
- [4] N. Seiberg and E. Witten, Monopoles, duality, and chiral symmetry breaking in $N = 2$ supersymmetric QCD, to appear.
- [5] J. Cardy and E. Rabinovici, Nucl. Phys. B 205 (1982) 1;
J. Cardy, Nucl. Phys. B 205 (1982) 17.
- [6] A. Shapere and F. Wilczek, Nucl. Phys. B320 (1989) 669.
- [7] A. Font, L. Ibanez, D. Lust, and F. Quevedo, Phys. Lett. B 249 (1990) 35.
- [8] S.J. Rey, Phys. Rev. D 43 (1991) 526.
- [9] A. Sen, preprint TIFR-TH-94-03 (hep-th/9402002).
- [10] A. Sen, preprint TIFR-TH-94-08 (hep-th/9402032).
- [11] G. Segal, to appear.
- [12] J. H. Schwarz and A. Sen, Phys. Lett. B 312 (1993) 105.
- [13] L. Dixon, Lectures given at the 1987 ICTP Summer Workshop in High Energy Physics and Cosmology, Trieste, Italy, 1987.
- [14] W. Lerche, C. Vafa and N. Warner, Nucl. Phys. B 324 (1989) 427.
- [15] B.R. Greene and M.R. Plesser, Nucl. Phys. B 338 (1990) 15.
- [16] P. Candelas, X.C. de la Ossa, P.S. Green and L. Parkes, Nucl. Phys. B 359 (1991) 21.
- [17] Essays on mirror manifolds, ed. S.-T. Yau (International Press, Hong Kong, 1992).

- [18] M. Duff, *Class. Quant. Grav.* 5 (1988) 189;
M. Duff and J. Lu, *Nucl. Phys. B* 354 (1991) 141; *Phys. Rev. Lett.* 66 (1991) 1402; *Class. Quant. Grav.* 9 (1991) 1;
M. Duff, R. Khuri and J. Lu, *Nucl. Phys. B* 377 (1992) 281;
J. Dixon, M. Duff and J. Plefka, *Phys. Rev. Lett.* 69 (1992) 3009.
- [19] A. Strominger, *Nucl. Phys. B* 343 (1990) 167.
- [20] A. Sen, to appear.
- [21] E. Witten, *Commun. Math. Phys.* 117 (1988) 353.
- [22] J. Yamron, *Phys. Lett. B* 213 (1988) 325.
- [23] S. Mukai, *Inv. Math.* 77 (1984) 101.
- [24] T. Nakashima, *Archiv. der Mathematik* 61 (1993) 100.
- [25] Z. B. Qin, *Manuscripta Mathematica* 79 (1993) 253.
- [26] K. Yoshioka, The Betti numbers of the moduli space of stable sheaves of rank 2 on P^2 , preprint, Kyoto University.
- [27] K. Yoshioka, The Betti numbers of the moduli space of stable sheaves of rank 2 on a ruled surface, preprint, Kyoto University.
- [28] A. A. Klyachko, *Funct. Anal. and Appl.* 25 (1991) 67.
- [29] H. Nakajima, Instantons on ALE spaces, quiver varieties, and Kac-Moody algebras, Tohoku University preprint 1993; *Int. Math. Res. Notices* (1994) 61.
- [30] G. 't Hooft, *Nucl. Phys. B* 138 (1978) 1; B153 (1979) 141.
- [31] L. Girardello, A. Giveon, M. Porrati and A. Zaffaroni, *Phys. Lett. B* 334 (1994) 331.
- [32] D. Zagier, *C.R. Acad. Sci. Paris*, 281 A (1975) 879.
- [33] F. Hirzebruch and D. Zagier, *Invent. Math.* vol. 36 (1976) 92.
- [34] M. Bershadsky, S. Cecotti, H. Ooguri and C. Vafa, *Nucl. Phys. B* 405 (1993) 279; Kodaira-Spencer theory of gravity and exact results for quantum string amplitudes, preprint HUTP-93-A025.
- [35] E. Witten, Supersymmetric Yang-Mills theory on a four-manifold, IASSNS-HEP-94/5.
- [36] J.P. Gauntlett and J.A. Harvey, unpublished; see also hep-th/9407111.
- [37] V. Mathai and D. Quillen, "Superconnections, Thom classes, and equivariant differential forms, *Topology* 25 (1986) 85.
- [38] M. F. Atiyah and L. Jeffrey, *J. Geom. Phys.* 7 (1990) 119.
- [39] L. Baulieu and I. M. Singer, *Commun. Math. Phys.* 117 (1988) 253.
- [40] E. Witten, *Int. J. Mod. Phys. A* 6 (1991) 2775.
- [41] E. Witten, *Nucl. Phys. B* 371 (1992) 191, Section 3.3.
- [42] J. Distler, Notes on $N = 2$ sigma models, hep-th-9212062.
- [43] S. Cordes, G. Moore, and S. Ramgoolam, Large N 2D Yang-Mills theory & topological string theory, hep-th-9402107, Section 6.7.
- [44] M. Blau and G. Thompson, *Commun. Math. Phys.* 152 (1993) 41; *Int. J. Mod. Phys. A* 8 (1993) 573.
- [45] M.F. Atiyah, N. Hitchin and I.M. Singer, *Proc. R. Soc. Lond. P A* 362 (1978) 425.
- [46] N. Hitchin, *Proc. London Math. Soc.* 3(1987) 55, 59.
- [47] N. Koblitz, *Introduction to elliptic curves and modular forms* (Springer-Verlag, Berlin, 1984).
- [48] S. Donaldson and P. Kronheimer, *The geometry of four manifolds* (Oxford Univ. Press, Oxford, 1990) p. 6.
- [49] J.-P. Serre, *A course in arithmetic* (Springer-Verlag, New York, 1973).
- [50] D. Husemoller and J. Milnor, *Symmetric bilinear forms* (Springer-Verlag, New York, 1973).
- [51] L. Dixon, J. Harvey, C. Vafa and E. Witten, *Nucl. Phys. B* 261 (1985) 678; B274 (1986) 285.
- [52] L. Götsche, *Math. Ann.* 286 (1990) 193.
- [53] F. Hirzebruch and T. Höfer, *Math. Ann.* 286 (1990) 255.
- [54] P. Kronheimer, private communication.
- [55] T. Miyake, *Modular forms* (Springer-Verlag, New York, 1989).
- [56] S. Donaldson, *J. Diff. Geom.* 26 (1987) 141.
- [57] V.G. Kac, *Infinite dimensional Lie algebras* (Birkhauser, Boston, 1983).
- [58] E. Verlinde, *Nucl. Phys. B* 300 (1999) 360.
- [59] A. Dabholkar and J.A. Harvey, *Phys. Rev. Lett.* 63 (1989) 478;
A. Dabholkar, G. Gibbons, J.A. Harvey and F. Ruiz, *Nucl. Phys. B* 340 (1990) 33.

- [60] R. Khuri, Phys. Lett. B 259 (1991) 261; B294 (1992) 325; Nucl. Phys. B387 (1992) 315.
- [61] J. Gauntlett, J. Harvey and J. Liu, Nucl. Phys. B409 (1993) 363.