ELSEVIER

Nuclear Physics B 426 (1994) 19–52

NUCLEAR PHYSICS B

Electric-magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang–Mills theory

N. Seiberg [a,b], E. Witten [b]

[a] *Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08855-0849, USA*
[b] *Institute for Advanced Study, Princeton, NJ 08540, USA*

Received 18 July 1994; accepted 26 July 1994

Abstract

We study the vacuum structure and dyon spectrum of $N = 2$ supersymmetric gauge theory in four dimensions, with gauge group SU(2). The theory turns out to have remarkably rich and physical properties which can nonetheless be described precisely; exact formulas can be obtained, for instance, for electron and dyon masses and the metric on the moduli space of vacua. The description involves a version of Olive–Montonen electric-magnetic duality. The "strongly coupled" vacuum turns out to be a weakly coupled theory of monopoles, and with a suitable perturbation confinement is described by monopole condensation.

1. Introduction

The dynamics of $N = 1$ supersymmetric gauge theories in four dimensions have been much explored, in part because of the possible phenomenological interest. Recently results have emerged about their strong coupling behavior [1–4]. Such results are largely consequences of the fact that the low energy effective superpotential is holomorphic both in the chiral superfields and in the parameters of the fundamental Lagrangian [1] (other aspects of such holomorphy were discussed in Refs. [5,6]). When the behavior of the superpotential near its singularities is combined with the symmetries and its holomorphy, the superpotential can often be determined exactly. The resulting superpotentials sometimes exhibit new physical phenomena [2–4] which might have applications also to non-supersymmetric gauge theories.

Our goal in the present paper is, among other things, to obtain similar strong coupling information about the corresponding $N = 2$ theories, at least in the basic case of SU(2)

SSDI 0550-3213(94)00330-0

gauge theory without matter multiplets. (Such multiplets will be incorporated in a subsequent paper [7].) In $N = 2$ theories, the Kähler potential and the masses of the stable particles are controlled by a holomorphic object, "the prepotential." Therefore, they can be studied in a way somewhat similar to the determination of the superpotential in $N = 1$ theories and ultimately determined.

Although our results are derived in the special case of $N = 2$ supersymmetry, they exhibit physical phenomena of general interest, including asymptotic freedom, chiral symmetry breaking, generation of a mass scale from strong coupling, confinement of electric charge via condensation of magnetic monopoles, and a version of electric-magnetic duality. These results are also likely to shed some light on the phenomena in $N = 4$ and perhaps in string theory, and are likely to help in understanding the topological field theories that can be obtained from $N = 2$ theories by twisting [8–11]. As electric-magnetic duality will play an important role in solving the model, we will review some of the background here. The existence of such duality in supersymmetric Yang–Mills theories was first suspected from properties of the dyons—particles carrying electric and magnetic charge—that exist at the classical level in the $N = 2$ and $N = 4$ gauge theories. Semiclassically one finds—by an argument that originated with the work of Prasad and Sommerfeld [10] and Bogomol'nyi [11]—that the mass of a dyon of magnetic and electric quantum numbers $(n_{\rm m}, n_{\rm e})$ is

$$M \geqslant \sqrt{2}|Z| \tag{1.1}$$

with

$$Z = v\left(n_{\rm e} + i\frac{1}{\alpha}n_{\rm m}\right). \tag{1.2}$$

Here v is the Higgs expectation value, and $\alpha = g^2/4\pi$ with g the gauge coupling constant.[1] States for which the inequality in (1.1) is actually an equality are said to be BPS-saturated. To see that a similar inequality must hold quantum mechanically, one interprets the inequality as a consequence of a central extension of the supersymmetry algebra [12]: we will see later that the parameters in the central extension (and thus the parameters appearing in the definition of Z) have an interesting renormalization for $N = 2$ but not for $N = 4$.

The mass formula (1.2) has a symmetry under $n_{\rm e} \leftrightarrow n_{\rm m}$, $\alpha \leftrightarrow 1/\alpha$, $v \leftrightarrow v/\alpha$. Olive and Montonen [13] pointed out this symmetry and conjectured that it was an exact property of a suitable quantum theory. According to this dramatic conjecture, the strong coupling limit of the theory is equivalent to the weak coupling limit with ordinary particles and solitons exchanged. The $N = 2$ theory appears not to possess Olive–Montonen duality because the electrons and monopoles have different Lorentz quantum numbers (electrons are in a supersymmetric multiplet with spins $\leqslant 1$, while the monopoles have spins $\leqslant 1/2$). On the other hand [14], for $N = 4$ the electrons and monopoles have the same quantum numbers, making Olive–Montonen duality more plausible. Actually, in this paper we will find that a version of Olive–Montonen duality holds for $N = 2$.

[1] We normalize the Higgs field such that its kinetic term is multiplied by $1/4\pi\alpha = 1/g^2$.

The version in question is necessarily quite different from the one that seems to hold for $N = 4$. The $N = 2$ theory is asymptotically free, so the coupling constant is equivalent to a choice of scale. And an anomaly in the $\mathrm{U}(1)_{\mathcal{R}}$ R symmetry prevents the existence of a physical microscopic theta angle. (There is still an effective theta angle at low energies.) For $N = 4$, the anomalies cancel in both conformal invariance and the $\mathrm{U}(1)$ symmetry, so the theory possesses both a natural dimensionless gauge coupling g and a theta angle θ; the Olive–Montonen duality should be extended [15,16] to an action of $\mathrm{SL}(2, \mathbb{Z})$ on $\tau = \theta/2\pi + i4\pi/g^2$, as was first recognized in lattice models [15,16] and in string theory [17]. (Dramatic new evidence for this has appeared recently [18–21] in considerations of multi-monopole bound states [18,19] and the partition function on various manifolds [20,21].) For $N = 2$, these natural dimensionless parameters are absent, and "duality" as we interpret it has to do with the behavior of the theory as a function of the expectation value of the Higgs field.

The organization of the paper is as follows. Section 2 is a review of some known facts about $N = 2$ supersymmetric gauge theories. The classical theory has a continuous manifold of inequivalent ground states—the "classical moduli space." Quantum corrections do not lift the vacuum degeneracy, so also the quantum theory has a manifold of inequivalent ground states—a "quantum moduli space." We construct a low energy effective Lagrangian for the light degrees of freedom. As described in Ref. [22], the metric on the quantum moduli space is a Kähler metric that is written locally in terms of a holomorphic function. In certain $N = 1$ models [2], the quantum corrections change the topology of the moduli space. For $N = 2$ we will argue that there is no change in topology but a marked change in geometry, singularity structure, and physical interpretation.

In Section 3, we analyze more fully the geometry of the low energy Lagrangian. The local structure is unique only up to transformations of a certain kind (flat space limits of the "special geometry" transformations [23–26] that appear in certain supergravity and string theories). Physically these correspond to duality transformations on the low energy fields. The low energy Lagrangian is mapped to another Lagrangian of the dual fields.

In Section 4 we use the coupling of the light fields to the massive dyons to find an expression for the dyon masses. It includes the quantum corrections to the classical result and is manifestly dual. We also (following a previous two-dimensional analysis [27]) describe conditions under which the spectrum of BPS-saturated states (that is, states obeying the BPS mass formula (1.2)) does *not* vary continuously. The possibility of this phenomenon turns out to be an essential difference between $N = 2$ and $N = 4$.

In Sections 5 and 6 we make our proposal for solving the model. We begin section 5 by explaining that to get a sensible Kähler metric on the quantum moduli space, at least two singularities are needed in strong coupling. We propose that these singularities arise when a magnetic monopole or dyon goes to zero mass. As a check we show that, under a further perturbation, condensation of monopoles occurs precisely when confinement of electric charge is expected. This for the first time gives a real relativistic field theory model in which confinement of charge is explained in this long-suspected fashion. We also show that the monodromies resulting from massless monopoles and dyons fit together in just the right way. Then in Section 6 we show that, with the assumption that the singularities come from massless monopoles and dyons, it is possible to get a unique metric on the moduli space (and unique formulas for particle masses) obeying all the necessary conditions.

2. Review of $N = 2$ SUSY

2.1. Representations

All $N = 2$ theories have a global $SU(2)_R$ symmetry which acts on the two supercharges of given chirality. Scale invariant $N = 2$ theories have also a $U(1)_{\mathcal{R}}$ symmetry under which the supercharges of positive chirality have charge minus one.

We will be studying two types of $N = 2$ multiplet:

(1) The first is the $N = 2$ chiral multiplet (sometimes called a vector multiplet), containing gauge fields A_μ, two Weyl fermions λ, ψ, and a scalar ϕ, all in the adjoint representation. We arrange the fields as

$$\begin{array}{ccc} & A_\mu & \\ \lambda & & \psi \\ & \phi & \end{array} \tag{2.1}$$

to exhibit the $SU(2)_R$ symmetry which acts on the rows; A_μ and ϕ are singlets and λ, ψ are a doublet. In terms of $N = 1$ supersymmetry, these fields can be organized into a vector multiplet W_α (containing (A_μ, λ)) and a chiral multiplet Φ (containing (ϕ, ψ)). In this formalism, only one generator of $SU(2)_R$, which we will call $U(1)_J$, is manifest. $U(1)_J$ and $U(1)_{\mathcal{R}}$ are both $N = 1$ R symmetries, acting as

$$\begin{aligned} U(1)_J: &\quad \Phi \to \Phi(e^{-i\alpha}\theta) \\ U(1)_{\mathcal{R}}: &\quad \Phi \to e^{2i\alpha}\Phi(e^{-i\alpha}\theta) \end{aligned} \tag{2.2}$$

(2) The second type of multiplet is the hypermultiplet (sometimes called the scalar multiplet), consisting of two Weyl fermions ψ_q and $\psi^\dagger_{\tilde{q}}$ and complex bosons q and $\tilde{q}^\dagger$; $SU(2)_R$ again acts on the rows of the diamond:

$$\begin{array}{ccc} & \psi_q & \\ q & & \tilde{q}^\dagger \\ & \psi^\dagger_{\tilde{q}} & \end{array} \tag{2.3}$$

In terms of $N = 1$ supersymmetry, these fields make up two chiral multiplets Q and $\tilde{Q}$. The symmetries in (2.2) act on them as

$$\begin{aligned} U(1)_J: &\quad Q \to e^{i\alpha} Q(e^{-i\alpha}\theta) \\ &\quad \tilde{Q} \to e^{i\alpha}\tilde{Q}(e^{-i\alpha}\theta) \\ U(1)_{\mathcal{R}}: &\quad Q \to Q(e^{-i\alpha}\theta) \end{aligned}$$

$$\widetilde{Q} \to \widetilde{Q}(e^{-i\alpha}\theta). \tag{2.4}$$

The gauge quantum numbers of $\widetilde{Q}$ are dual to those of Q.

2.2. Renormalizable Lagrangians and classical flat directions

We will consider an $N = 2$ gauge theory, described by chiral superfields in the adjoint representation of a gauge group G, possibly coupled to additional matter hypermultiplets. $N = 2$ supersymmetry relates the gauge couplings to certain interactions which in the $N = 1$ language are described by the superpotential. In fact the requisite term in the superpotential is (in the notation of Ref. [28])

$$W = \sqrt{2}\,\widetilde{Q}\Phi Q. \tag{2.5}$$

This makes sense since $\widetilde{Q}$ and Q transform in dual representations and Φ is in the adjoint representation.

Classically, the Lagrangian is invariant under the $SU(2)_R \times U(1)_{\mathcal{R}}$ symmetry (2.2), (2.4) as well as, possibly, some flavor symmetries depending on the gauge representations of the hypermultiplets. In the quantum theory, the $U(1)_{\mathcal{R}}$ symmetry is generally broken by an anomaly. For $SU(N_c)$ gauge symmetry with N_f massless hypermultiplets ("quarks") in the fundamental representation, $U(1)_{\mathcal{R}}$ is broken to $\mathbb{Z}_{4N_c-2N_f}$. In this paper, we will mainly consider the SU(2) gauge theory without quarks. In this case the global symmetry is $(SU(2)_R \times \mathbb{Z}_8)/\mathbb{Z}_2$ since $4N_c - 2N_f = 8$ (the division by $\mathbb{Z}_2$ arises because the center of $SU(2)_R$ is contained in $\mathbb{Z}_8$).

The classical potential of the pure $N = 2$ theory (without hypermultiplets) is

$$V(\phi) = \frac{1}{g^2}\mathrm{Tr}[\phi, \phi^\dagger]^2. \tag{2.6}$$

For this to vanish, it is not necessary that ϕ should vanish; it is enough that ϕ and $\phi^\dagger$ commute. The classical theory therefore has a family of vacuum states. For instance, if the gauge group is SU(2), then up to gauge transformation we can take $\phi = \frac{1}{2}a\sigma^3$, with $\sigma^3 = \mathrm{diag}(1, -1)$ and a a complex parameter labeling the vacua. The Weyl group of SU(2) acts by $a \leftrightarrow -a$, so the gauge invariant quantity parametrizing the space of vacua is $u = \frac{1}{2}a^2 = \mathrm{Tr}\,\phi^2$. For non-zero a the gauge symmetry is broken to U(1) and the global $\mathbb{Z}_8$ symmetry is broken to $\mathbb{Z}_4$. The residual $\mathbb{Z}_4$ acts trivially on the u plane since the $U(1)_{\mathcal{R}}$ charge of u is 4. The global symmetry group acts on the u plane as a spontaneously broken $\mathbb{Z}_2$, acting by $u \leftrightarrow -u$.

Classically, there is a singularity at $u = 0$, where the full SU(2) gauge symmetry is restored and more fields become massless.

2.3. Low energy effective action

We now study the low energy effective action of the light fields on the moduli space. For generic $\langle\phi\rangle$ the low energy effective Lagrangian contains a single $N = 2$ vector multiplet, $\mathcal{A}$. The terms with at most two derivatives and not more than four fermions are constrained by $N = 2$ supersymmetry. They are expressed in terms of a single holomorphic function $\mathcal{F}(\mathcal{A})$, as explained in Ref. [22]. In $N = 1$ superspace, the Lagrangian is

$$\frac{1}{4\pi}\mathrm{Im}\left[\int d^4\theta \frac{\partial\mathcal{F}(A)}{\partial A}\bar{A} + \int d^2\theta \frac{1}{2}\frac{\partial^2\mathcal{F}(A)}{\partial A^2}W_\alpha W^\alpha\right], \tag{2.7}$$

where A is the $N = 1$ chiral multiplet in the $N = 2$ vector multiplet $\mathcal{A}$ whose scalar component is a.

We would like to make a few comments:

(1) For large a, asymptotic freedom takes over and the theory is weakly coupled. Moreover, since it is impossible to add an $N = 2$ invariant superpotential to (2.7), the vacuum degeneracy cannot be removed quantum mechanically. Therefore, the quantum theory has a non-trivial moduli space which is in fact a one complex dimensional Kähler manifold. The Kähler potential can be written in terms of the effective low energy $\mathcal{F}$ function as

$$K = \mathrm{Im}\left(\frac{\partial\mathcal{F}(A)}{\partial A}\bar{A}\right). \tag{2.8}$$

The metric is thus concretely

$$(ds)^2 = \mathrm{Im}\,\frac{\partial^2\mathcal{F}(a)}{\partial a^2}\,da\,d\bar{a}. \tag{2.9}$$

In the classical theory, $\mathcal{F}$ can be read off from the tree level Lagrangian of the SU(2) gauge theory and is $\mathcal{F}(\mathcal{A}) = \frac{1}{2}\tau_{\mathrm{cl}}\mathcal{A}^2$ with $\tau_{\mathrm{cl}} = \theta/2\pi + i4\pi/g^2$. Asymptotic freedom means that this formula is valid for large a if g^2 is replaced by a suitable effective coupling. The small-a behavior will however turn out to be completely different. Classically, the θ parameter has no consequences. Quantum mechanically, the physics is θ dependent, but since there is an anomalous symmetry, it can be absorbed in a redefinition of the fields. Therefore, we will set $\theta = 0$.

(2) The formula for the Kähler potential does not look covariant—the Kähler potential can be written in this way only in a distinguished class of coordinate systems, which we will analyze later. In fact, A is related by $N = 2$ supersymmetry to the "photon" A_μ, which has a natural linear structure; this gives a natural coordinate system (or what will turn out to be a natural class of coordinate systems) for A.

(3) The low energy values of the gauge coupling constant and theta parameter can be read off from the Lagrangian. If we combine them in the form $\tau = \theta/2\pi + i4\pi/g^2$, and denote the effective couplings in the vacuum parametrized by a as $\tau(a)$, then

$$\tau(a) = \frac{\partial^2\mathcal{F}}{\partial a^2}. \tag{2.10}$$

(4) The generalization to an arbitrary compact gauge group G of rank r is as follows. The potential is always given by (2.6), so the classical vacua are labeled by a complex adjoint-valued matrix ϕ with $[\phi, \phi^\dagger] = 0$. The unbroken gauge symmetry at the generic point on the moduli space is the Cartan subalgebra and therefore the complex dimension

of the moduli space is r. The low energy theory is described in terms of r abelian chiral multiplets $\mathcal{A}^i$, and the generalization of (2.7) is [22]

$$\frac{1}{4\pi}\mathrm{Im}\left[\int d^4\theta \frac{\partial\mathcal{F}(A)}{\partial A^i}\overline{A}^i + \int d^2\theta \tfrac{1}{2}\frac{\partial^2\mathcal{F}(A)}{\partial A^i \partial A^j} W^i_\alpha W^{\alpha j}\right]. \tag{2.11}$$

Here i labels the generators in the Cartan subalgebra and locally $\mathcal{F}$ is an arbitrary holomorphic function of r complex variables.

(5) The SU(2) theory, studied on the flat direction with $u \neq 0$, has in addition to the massless chiral or vector multiplet $\mathcal{A}$, additional charged massive vector multiplets. One can easily write a gauge invariant effective action for the triplet of chiral multiplets $\mathcal{A}^a$, $a = 1, \ldots, 3$, which reduces at low energies to (2.7) for the massless fields and incorporates the massive ones. Using the same function $\mathcal{F}$ as above, we set $\mathcal{F}(\sqrt{\mathcal{A}\cdot\mathcal{A}}) = \mathcal{H}(\mathcal{A}\cdot\mathcal{A})$ and write

$$\frac{1}{2\pi}\mathrm{Im}\left[\int d^4\theta \mathcal{H}' A^a \left(e^V\right)_{ab} \overline{A}^b + \int d^2\theta \tfrac{1}{2}\left(\mathcal{H}'\delta^{ab} + 2\mathcal{H}'' A^a A^b\right) W^a_\alpha W^b_\alpha\right], \tag{2.12}$$

where we used the SU(2)-invariant metric δ^{ab} to raise and lower indices. Eq. (2.12) has $N = 2$ supersymmetry and manifest gauge invariance, and reduces at low energies to (2.7).

(6) The Lagrangian (2.7) is unchanged if we add to $\mathcal{F}$ terms linear in $\mathcal{A}$. This has the effect of shifting $\partial\mathcal{F}/\partial A$ by a constant. We will later assign physical meaning to

$$h(A) = \frac{\partial\mathcal{F}}{\partial A}. \tag{2.13}$$

The additive constant will always be fixed by comparing with the high energy theory as in (2.12).

As we have already mentioned, classically the $\mathcal{F}$ function is

$$\mathcal{F}_0 = \tfrac{1}{2}\tau_{\mathrm{cl}}\mathcal{A}^2. \tag{2.14}$$

The quantum corrections were analyzed in Ref. [22]. The tree level and one-loop contributions add up to

$$\mathcal{F}_{\text{one loop}} = i\frac{1}{2\pi}\mathcal{A}^2 \ln\frac{\mathcal{A}^2}{\Lambda^2}, \tag{2.15}$$

where Λ is the dynamically generated scale. This logarithm is related to the one-loop beta function and also ensures the anomalous transformation laws under $U(1)_{\mathcal{R}}$. Higher order perturbative corrections are absent. Instantons lead to new terms. The anomaly and the instanton action suggest that

$$\mathcal{F} = i\frac{1}{2\pi}\mathcal{A}^2 \ln\frac{\mathcal{A}^2}{\Lambda^2} + \sum_{k=1}^{\infty}\mathcal{F}_k\left(\frac{\Lambda}{\mathcal{A}}\right)^{4k}\mathcal{A}^2, \tag{2.16}$$

where the kth term arises as a contribution of k instantons. A detailed calculation of the $k = 1$ term [22] indicates that $\mathcal{F}_1 \neq 0$. We will soon see that infinitely many $\mathcal{F}_k$ are nonzero.

Corrections to the classical formula (2.14) are related to the beta function, and for $N = 4$ supersymmetric Yang–Mills theory, whose beta function vanishes, the formula (2.14) is exact.

3. Duality

We have noted above that locally, by virtue of $N = 2$ supersymmetry, the metric on the moduli space is of the form

$$(\mathrm{d}s)^2 = \mathrm{Im}\ \tau(a)\ \mathrm{d}a\,\mathrm{d}\bar{a}, \tag{3.1}$$

with $\tau(a)$ the holomorphic function $\tau = \partial^2\mathcal{F}/\partial a^2$. The one-loop formula (2.15) shows that for large $|a|$, $\tau(a) \approx i\left(\ln(a^2/\Lambda^2) + 3\right)/\pi$ is a multivalued function whose imaginary part is single-valued and positive. However, if $\mathrm{Im}\ \tau(a)$ is globally defined it cannot be positive definite as the harmonic function $\mathrm{Im}\ \tau$ cannot have a minimum. This indicates that the above description of the metric must be valid only locally.

To what extent is it possible to change variables from a to some other local parameter, while leaving the metric in the form (3.1)? The answer to this question is at the heart of the physics. We define $a_D = \partial\mathcal{F}/\partial a$. The metric can then be written

$$(ds)^2 = \mathrm{Im}\ \mathrm{d}a_D\,\mathrm{d}\bar{a} = -\frac{i}{2}\left(\mathrm{d}a_D\,\mathrm{d}\bar{a} - \mathrm{d}a\,\mathrm{d}\bar{a}_D\right). \tag{3.2}$$

This formula is completely symmetric in a and a_D, so if we use a_D as the local parameter, the metric will be in the same general form as (3.1), with a different harmonic function replacing $\mathrm{Im}\ \tau$. As we will see presently, this transformation corresponds to electric-magnetic duality. Before entering into that, let us identify the complete class of local parameters in which the metric can be written as in (3.2).

3.1. Mathematical description

To treat the formalism in a way that is completely symmetric between a and a_D, we introduce an arbitrary local holomorphic coordinate u, and treat a and a_D as functions of u. u is a local coordinate on a complex manifold $\mathcal{M}$—the moduli space of vacua of the theory. Eventually we will pick u to be the expectation value of $\mathrm{Tr}\,\phi^2$—a good physical parameter—but for now u is arbitrary.

Introduce a two dimensional complex space $X \cong \mathbb{C}^2$ with coordinates (a_D, a). Endow X with the symplectic form $\omega = \mathrm{Im}\ \mathrm{d}a_D \wedge \mathrm{d}\bar{a}$. The functions $(a_D(u), a(u))$ give a map f from $\mathcal{M}$ to X. The metric on $\mathcal{M}$ is

$$(\mathrm{d}s)^2 = \mathrm{Im}\ \frac{\mathrm{d}a_D}{\mathrm{d}u}\frac{\mathrm{d}\bar{a}}{\mathrm{d}\bar{u}}\mathrm{d}u\,\mathrm{d}\bar{u} = -\frac{i}{2}\left(\frac{\mathrm{d}a_D}{\mathrm{d}u}\frac{\mathrm{d}\bar{a}}{\mathrm{d}\bar{u}} - \frac{\mathrm{d}a}{\mathrm{d}u}\frac{\mathrm{d}\bar{a}_D}{\mathrm{d}\bar{u}}\right)\mathrm{d}u\,\mathrm{d}\bar{u}. \tag{3.3}$$

This formula is valid for an arbitrary local parameter u on $\mathcal{M}$. If one picks $u = a$, one gets back the original formula (2.9) for the metric. (The above formula can be described in a coordinate-free way by saying that the Kähler form of the induced metric on $\mathcal{M}$ is $f^*(\omega)$.) Notice that ω had no particular positivity property and thus, if $a(u)$

and $a_D(u)$ are completely arbitrary local holomorphic functions, the metric (3.3) is not positive. We will eventually construct $a(u)$ and $a_D(u)$ in a particular way that will ensure positivity.

It is easy to see what sort of transformations preserve the general structure of the metric. If we set $a^\alpha = (a_D, a)$, $\alpha = 1, 2$, and let $\epsilon_{\alpha\beta}$ be the antisymmetric tensor with $\epsilon_{12} = 1$, then

$$(\mathrm{d}s)^2 = -\frac{i}{2}\epsilon_{\alpha\beta}\frac{\mathrm{d}a^\alpha}{\mathrm{d}u}\frac{\mathrm{d}\overline{a}^\beta}{\mathrm{d}\overline{u}}\mathrm{d}u\,\mathrm{d}\overline{u}. \tag{3.4}$$

This is manifestly invariant under linear transformations that preserve ϵ and commute with complex conjugation (the latter condition ensures that a^α and $\overline{a}^\alpha$ transform the same way). These transformations make the group $\mathrm{SL}(2,\mathbb{R})$ (or equivalently $\mathrm{Sp}(2,\mathbb{R})$). Also, (3.4) is obviously invariant under adding a constant to a_D or a. So if we arrange (a_D, a) as a column vector v, the symmetries that preserve the general structure are

$$v \to Mv + c, \tag{3.5}$$

where M is a 2×2 matrix in $\mathrm{SL}(2,\mathbb{R})$, and c is a constant vector. Later we will find (from considerations involving the gauge fields and the electric and magnetic charges) that M must be in $\mathrm{SL}(2,\mathbb{Z})$ and that in the pure $N = 2$ gauge theory, c must vanish. (In coupling to matter, c will play an important role [7].) In general, the group of transformations (3.5) can be thought of as the group of 3×3 matrices of the form

$$\begin{pmatrix} 1 & 0 \\ c & M \end{pmatrix}, \tag{3.6}$$

acting on the three objects $(1, a_D, a)$.

Generalization to dimension greater than one.

Though it will not be exploited in the present paper, and can thus be omitted by the reader, let us briefly discuss the generalization to other gauge groups. If the gauge group G has rank r, then $\mathcal{M}$ has complex dimension r. Locally from (2.11) the metric is

$$(\mathrm{d}s)^2 = \mathrm{Im}\ \frac{\partial^2\mathcal{F}}{\partial a^i \partial a^j}\mathrm{d}a^i \mathrm{d}\overline{a}^j, \tag{3.7}$$

with distinguished local coordinates a^i and a holomorphic function $\mathcal{F}$. We again reformulate this by introducing

$$a_{D,j} = \frac{\partial\mathcal{F}}{\partial a^j}. \tag{3.8}$$

Then we can write

$$(\mathrm{d}s)^2 = \mathrm{Im}\ \sum_i \mathrm{d}a_{D,i}\,\mathrm{d}\overline{a}^i. \tag{3.9}$$

To formulate this invariantly, we introduce a complex space $X \cong \mathbf{C}^{2r}$ with coordinates $a^i, a_{D,j}$. We endow X with the symplectic form $\omega = \frac{1}{2}i\sum_i\left(\mathrm{d}a^i \wedge \mathrm{d}\overline{a}_{D,i} - \mathrm{d}a_{D,i} \wedge \mathrm{d}\overline{a}^i\right)$ of type $(1,1)$ and also with the holomorphic two-form $\omega_h = \sum_i \mathrm{d}a^i \wedge \mathrm{d}a_{D,i}$. Then we introduce arbitrary local coordinates u^s, $s = 1,\ldots,r$, on the moduli space $\mathcal{M}$, and describe a map $f : \mathcal{M} \to X$ by functions $a^i(u)$, $a_{D,j}(u)$. We require f to be such that $f^*(\omega_h) = 0$; this precisely ensures that locally, if we pick $u^i = a^i$, then $a_{D,j}$ must be of the form in (3.8) with some holomorphic function $\mathcal{F}$. Then we take the metric on $\mathcal{M}$ to be the one whose Kähler form is $f^*(\omega)$; in formulas the metric is

$$(\mathrm{d}s)^2 = \mathrm{Im} \sum_{s,t,i} \frac{\partial a_{D,i}}{\partial u^s} \frac{\partial \overline{a}^i}{\partial \overline{u}^t} \mathrm{d}u^s \mathrm{d}\overline{u}^t. \tag{3.10}$$

If again we arrange a, a_D as a $2r$-component column vector v, then the formalism is invariant under transformations $v \to Mv + c$ with M a matrix in $\mathrm{Sp}(2r,\mathbb{R})$ and c a constant vector. Again, considerations involving the charges will eventually require that M be in $\mathrm{Sp}(2r,\mathbb{Z})$ and impose restrictions on c.

3.2. Physical interpretation via duality

So far we have seen that the spin zero component of the $N = 2$ multiplet has a Kähler metric of a very special sort, constructed using a distinguished set of coordinate systems. This rigid structure is related by $N = 2$ supersymmetry to the natural linear structure of the gauge field. We have found that, for the spin zero component, the distinguished parametrization is not completely unique; there is a natural family of parametrizations related by $\mathrm{SL}(2,\mathbb{R})$. How does this $\mathrm{SL}(2,\mathbb{R})$ (which will actually be reduced to $\mathrm{SL}(2,\mathbb{Z})$) act on the gauge fields?

$\mathrm{SL}(2,\mathbb{R})$ is generated by the transformations

$$T_b = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, \text{ and } S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{3.11}$$

with real b. The former acts as $a_D \to a_D + ba$, $a \to a$; this acts trivially on the distinguished coordinate a, and can be taken to act trivially on the gauge field. By inspection of (2.12), the effect of $a_D \to a_D + ba$ on the gauge kinetic energy is just to shift the θ angle by $2\pi b$; in the abelian theory, this has no effect until magnetic monopoles (or at least non-trivial U(1) bundles) are considered. Once that is done, the allowed shifts in the θ angle are by integer multiples of 2π; that is why b must be integral and gives essentially our first derivation of the reduction to $\mathrm{SL}(2,\mathbb{Z})$.

The remaining challenge is to understand what S means in terms of the gauge fields. We will see that it corresponds to electric-magnetic duality. To see this, let us see how duality works in Lagrangians of the sort introduced above.

We work in Minkowski space and consider first the purely bosonic terms involving only the gauge fields. We use conventions such that $F_{\mu\nu}^2 = -({}^*F)_{\mu\nu}^2$ and ${}^*({}^*F) = -F$ where ${}^*F$ denotes the dual of F. The relevant terms are

$$\frac{1}{32\pi} \mathrm{Im} \int \tau(a) \cdot (F + i{}^*F)^2 = \frac{1}{16\pi} \mathrm{Im} \int \tau(a) \cdot (F^2 + i{}^*FF) \,. \tag{3.12}$$

Duality is carried out as follows. The constraint $\mathrm{d}F = 0$ (which in the original description follows from $F = \mathrm{d}A$) is implemented by adding a Lagrange multiplier vector field V_D.

Then F is treated as an independent field and integrated over. The normalization is set as follows. The $U(1) \subset SU(2)$ is normalized such that all SO(3) fields have integer charges (matter multiplets in the fundamental representation of SU(2) therefore have half integer charges). Then, a magnetic monopole corresponds to $\epsilon^{0\mu\nu\rho}\partial_\mu F_{\nu\rho} = 8\pi\delta^{(3)}(x)$. For V_D to couple to it with charge one, we add to (3.12)

$$\frac{1}{8\pi}\int V_{D\mu}\epsilon^{\mu\nu\rho\sigma}\partial_\nu F_{\rho\sigma} = \frac{1}{8\pi}\int {}^*F_D F = \frac{1}{16\pi}\mathrm{Re}\int ({}^*F_D - iF_D)(F + i^*F)\,, \tag{3.13}$$

where $F_{D\mu\nu} = \partial_\mu V_{D\nu} - \partial_\nu V_{D\mu}$ is the field strength of V_D. We can now perform the Gaussian functional integral over F and find an equivalent Lagrangian for V_D,

$$\frac{1}{32\pi}\mathrm{Im}\;\frac{-1}{\tau}(F_D + i^*F_D)^2 = \frac{1}{16\pi}\mathrm{Im}\;\frac{-1}{\tau}(F_D^2 + i^*F_D F_D)\,. \tag{3.14}$$

We now repeat these steps in $N = 1$ superspace. We treat W_α in $\frac{1}{8\pi}\mathrm{Im}\int d^2\theta\tau(A)W^2$ as an independent chiral field. The superspace version of the Bianchi identity $dF = 0$ is $\mathrm{Im}\,\mathcal{D}W = 0$ ($\mathcal{D}$ is the supercovariant derivative). It can be implemented by a real vector superfield V_D Lagrange multiplier. We add to the action

$$\begin{aligned}\frac{1}{4\pi}\mathrm{Im}\int d^4x\, d^4\theta\, V_D\mathcal{D}W &= \frac{1}{4\pi}\mathrm{Re}\int d^4x\, d^4\theta\, i\mathcal{D}V_D W\\ &= -\frac{1}{4\pi}\mathrm{Im}\int d^4x\, d^2\theta W_D W.\end{aligned} \tag{3.15}$$

Performing the Gaussian integral over W we find an equivalent Lagrangian

$$\frac{1}{8\pi}\mathrm{Im}\int d^2\theta\frac{-1}{\tau(A)}W_D^2. \tag{3.16}$$

To proceed further, we need to transform the $N = 1$ chiral multiplet A to A_D. The kinetic term

$$\mathrm{Im}\int d^4\theta h(A)\overline{A} \tag{3.17}$$

is transformed by

$$A_D = h(A) \tag{3.18}$$

to

$$\mathrm{Im}\int d^4\theta h_D(A_D)\overline{A}_D \tag{3.19}$$

where $h_D(h(A)) = -A$ is minus the inverse function. Then using $h'(A) = \tau(A)$ the coefficient of the gauge kinetic term (3.16) becomes

$$-\frac{1}{\tau(A)} = -\frac{1}{h'(A)} = h_D'(A_D) = \tau_D(A_D). \tag{3.20}$$

Note that a shift of h by a constant does not affect the Lagrangian. Therefore, the duality transformation has a freedom to shift A_D by a constant.

The relations $A_D = h(A)$ and $h_D = -A$ mean that the duality transformation precisely implements the missing $\mathrm{SL}(2,\mathbb{Z})$ generator S. The function $\tau = h'$ is mapped by

$$\tau_D(A_D) = -\frac{1}{\tau(A)} . \tag{3.21}$$

Remembering that $\tau(a) = \theta(a)/2\pi + i4\pi/g(a)^2$, we see that the duality transformation inverts τ rather than the low energy gauge coupling $g(a)$. (A similar phenomenon is known in $R \leftrightarrow 1/R$ duality in string theory [29], where $B_{ij} + iG_{ij}$, which is analogous to our τ, is inverted rather than G_{ij}.)

It is important to stress that unlike $\tau \to \tau + 1$, the duality transformation is not a symmetry of the theory. It maps one description of the theory to *another* description of the same theory.

For other gauge groups G the low energy Lagrangian has several abelian fields, A^i, in the Cartan subalgebra. Then

$$(A_D)_i = h_i(A^i) = \partial_i \mathcal{F}(A^i) \tag{3.22}$$

which leads to

$$h_D^i(h_j(A^k)) = -A^i \tag{3.23}$$

and the "metrics"

$$\begin{aligned} \tau_{ij}(A) &= \partial_i\partial_j \mathcal{F}(A) = \partial_j h_i(A) , \\ \tau_D^{ij}(A_D) &= \partial^i\partial^j \mathcal{F}_D(A_D) = \partial^j h_D^i(A_D) \end{aligned} \tag{3.24}$$

satisfy

$$f_{ij} f_D^{jk} = -\delta_i^k . \tag{3.25}$$

The above transformation together with the more obvious shifts $A_{Di} \to A_{D_i} + M_{ij}A^j$ generate $\mathrm{Sp}(2r,\mathbb{Z})$.

3.3. Coupling to gravity

Before concluding this section, we would like to compare the structure we have found to the "special geometry" that appears if the chiral multiplet is coupled to $N=2$ supergravity [23]. (This will not be used in the present paper.) In $N=2$ supergravity, the general Kähler metric for a system of r chiral superfields is described locally by a holomorphic function $\mathcal{G}_0(a^1,\dots,a^r)$ of r complex variables a^i. The Kähler potential is

$$K_{\mathrm{grav}} = -\ln\left(2i(\mathcal{G}_0 - \overline{\mathcal{G}}_0) + \frac{i}{2}\sum_i \left(\overline{a}^i\frac{\partial \mathcal{G}_0}{\partial a^i} - a^i\frac{\partial \overline{\mathcal{G}}_0}{\partial \overline{a}^i}\right)\right) . \tag{3.26}$$

In global supersymmetry we had a local holomorphic function $\mathcal{F}$ with

$$K = \frac{-i}{2}\sum_i \left(\overline{a}^i\frac{\partial \mathcal{F}}{\partial a^i} - a^i\frac{\partial \overline{\mathcal{F}}}{\partial \overline{a}^i}\right) . \tag{3.27}$$

One would expect that there is some limit in which gravitational effects are small and (3.26) would reduce to (3.27). How does this occur?

It suffices to set

$$\mathcal{G}_0 = -i\frac{M_{\rm Pl}{}^2}{4} + \mathcal{F}, \tag{3.28}$$

with $M_{\rm Pl}$ the Planck mass. Then if $M_{\rm Pl}$ is much larger than all relevant parameters, we get

$$K_{\rm grav} = -\ln M_{\rm Pl}{}^2 + \frac{K}{M_{\rm Pl}{}^2} + O(M_{\rm Pl}{}^{-4}). \tag{3.29}$$

The constant term $-\ln M_{\rm Pl}{}^2$ does not contribute to the Kähler metric, so up to a normalization factor of $1/M_{\rm Pl}^2$, the Kähler metric with supergravity reduces to that of global $N=2$ supersymmetry as $M_{\rm Pl} \to \infty$ keeping everything else fixed.

More fundamentally, we would like to compare the allowed monodromy groups. In supergravity, the global structure is exhibited as follows. One introduces an additional variable a^0 and sets $\mathcal{G} = (a^0)^2\mathcal{G}_0$. One also introduces $a_{D,j} = \partial\mathcal{G}/\partial a^j$ for $j = 0, \dots, r$. Then one finds that the special Kähler structure of (3.26) allows $\mathrm{Sp}(2r+2, \mathbb{R})$ transformations acting on (a_{Di}, a^j).[2] Now, in decoupling gravity, we consider $\mathcal{G}$ to be of the special form in (3.28). In that case,

$$a_{D,0} = -i\frac{M_{\rm Pl}}{2}. \tag{3.30}$$

The other a^i, $a_{D,j}$ are independent of $M_{\rm Pl}$. To preserve this situation in which $M_{\rm Pl}$ appears only in $a_{D,0}$, we must consider only those $\mathrm{Sp}(2r+2, \mathbb{R})$ transformations in which the transformations of all fields are independent of $a_{D,0}$. These transformations all leave a^0 invariant. There is no essential loss then in scaling the a's so that $a^0 = 1$. Arrange the $a_{D,i}, a^j$ with $i, j = 1, \dots, r$ as a column vector v. The $\mathrm{Sp}(2r+2, \mathbb{R})$ transformations that leave invariant $a^0 = 1$ act on v by $v \to Mv + c$ where $M \in \mathrm{Sp}(2r, \mathbb{R})$ and c is a constant. (The transformations with $c \neq 0$ do not leave $a_{D,0}$ invariant, but its variation is independent of $M_{\rm Pl}$ and so is negligible in the limit in which gravity is weak.) This is precisely the duality group that we found in the global $N=2$ theory.

4. Dyon masses

The SU(2) gauge theory under discussion has electrically and magnetically charged particles whose masses satisfy

$$M^2 = 2|Z|^2 \tag{4.1}$$

with

$$Z_{\rm cl} = a(n_{\rm e} + \tau_{\rm cl} n_{\rm m}) \tag{4.2}$$

[2] Once one considers the gauge fields, this is reduced to $\mathrm{Sp}(2r+2, \mathbb{Z})$. The symplectic form preserved by $\mathrm{Sp}(2r+2, \mathbb{R})$ is the usual one $\sum_i \mathrm{d}a^i \wedge \mathrm{d}a_{D,i}$.

($\tau_{\rm cl} = \theta/2\pi + i4\pi/g^2$) where $n_{\rm e}$ and $n_{\rm m}$ are the electric and magnetic charges; they are integers as long as all elementary fields are in representations of SO(3) (fields that are in SU(2) representations of half-integral spin have half-integral $n_{\rm e}$). The origin of this formula [12] is that Z arises as a central extension in the $N = 2$ supersymmetry algebra and Eq. (4.1) follows from the representation theory of $N = 2$ for "small" representations. One uses the algebra to show that for any state of given $(n_{\rm m}, n_{\rm e})$,

$$M \geqslant \sqrt{2}|Z| \tag{4.3}$$

with equality precisely for the "small" representations of $N = 2$ (four helicity states instead of sixteen). We will not review the argument here. States saturating the inequality are called BPS-saturated states.

As stressed in Ref. [12], the same interpretation should apply quantum mechanically, but the coefficients of $n_{\rm e}$ and $n_{\rm m}$ in Z might be modified. One way to find the modification is to calculate the central extension of the algebra from the low energy effective Lagrangian (2.7). This leaves an ambiguity in shifting A and A_D by a constant. As remarked above, such an ambiguity can be resolved by considering the full high energy theory in its effective form as in (2.12).

Alternatively, we can couple the theory based on (2.7) to a hypermultiplet—two $N = 1$ chiral multiplets M and $\widetilde{M}$—with electric charge $n_{\rm e}$ and a canonical kinetic term. As in (2.5), the coupling to the gauge field is $N = 2$ supersymmetric only with a superpotential

$$\sqrt{2}\, n_{\rm e} A M \widetilde{M}\,. \tag{4.4}$$

There is also a possibility of adding a mass term; this has the effect of shifting $n_{\rm e}A$ by a constant corresponding to the ambiguity mentioned above. By embedding the theory in a higher energy theory such as (2.12), this ambiguity is removed, and one learns in the pure gauge theory that the coefficient of $M\widetilde{M}$ in the superpotential is precisely $\sqrt{2}n_{\rm e}A$, with no additional additive constant. With additional matter multiplets included [7], such a term will arise and play an important role.

For non-zero a, the fields M and $\widetilde{M}$ are massive. As the corresponding states are in a "small" representation, their mass is determined by the central extension in the algebra [12]. Comparing (4.1) and (4.4), we conclude that $Z = an_{\rm e}$. Using the duality transformation, it is clear that for magnetic monopoles with magnetic charge $n_{\rm m}$, $Z = a_D n_{\rm m}$ and for dyons

$$Z = an_{\rm e} + a_D n_{\rm m}. \tag{4.5}$$

These formulas can be verified directly; for instance, to study the BPS bound for monopoles, we consider the full high energy theory in its effective form (2.12) and examine the bosonic terms in the Hamiltonian of a magnetic monopole for real ϕ^a:

$$\begin{aligned} E &= \frac{1}{4\pi}\mathrm{Im} \int \mathrm{d}^3x \left(\tau_{ab}(D_i\phi)^a(D_i\phi)^b + \tfrac{1}{2}\tau_{ab}B_i^a B_i^b\right) \\ &= \frac{1}{4\pi}\mathrm{Im} \int \mathrm{d}^3x \left(\left(\frac{-1}{\tau}\right)^{ab}(D_ih)_a(D_ih)_b + \tfrac{1}{2}\tau_{ab}B_i^a B_i^b\right) \end{aligned}$$

$$= \frac{1}{4\pi}\mathrm{Im}\left\{\int \mathrm{d}^3x \left(\frac{-1}{\tau}\right)^{ab}\left[(D_ih)_a \pm \frac{1}{\sqrt{2}}\tau_{ac}B_i^c\right]\left[(D_ih)_b \pm \frac{1}{\sqrt{2}}\tau_{bd}B_i^d\right]\right\}$$
$$\mp\sqrt{2}\,\partial_i(B_i^a h_a)$$
$$\geqslant \left|\frac{\sqrt{2}}{4\pi}\oint \mathrm{d}^2s\, B_i^a h_a\right| = \sqrt{2}\,|n_\mathrm{m} a_D|. \tag{4.6}$$

This inequality confirms that for monopoles $Z = n_\mathrm{m} a_D$.

We would like to make a few comments:

(1) The expression (4.5) has the correct semiclassical limit and manifest duality.

(2) Unlike (4.2), it is renormalization group invariant. Note that it does not differ from (4.2) merely by replacing $i4\pi/g^2$ by $\tau(a)$; i.e. by the running coupling. The distinction between them appears already at one loop.

(3) From this expression, it follows that the renormalization of the BPS formula vanishes for $N = 4$ supersymmetric Yang–Mills theory. For $N = 4$, the classical expression (2.14) is exact (note the comment at the end of Section 2) so there are no corrections to the classical formula $a_D = \tau_\mathrm{cl} a$.

(4) The generalization of (4.5) to an arbitrary gauge group is

$$Z = a^i n_{e,i} + h_i(a) n_\mathrm{m}{}^i = a \cdot n_\mathrm{e} + a_D \cdot n_\mathrm{m}\,, \tag{4.7}$$

where a^i are local coordinates on the quantum moduli space and $n_{e,i}$, $n_\mathrm{m}{}^i$ are the electric and magnetic charges.

Now let us discuss the restrictions that the dyon mass formula puts on the duality discussed in Section 3. First of all, it is clear that Z, since it determines particle masses (or appears in the supersymmetry algebra) must be invariant under the monodromies. In Section 3, we arranged (a_D, a) as a column vector v, and found that the analysis of the Kähler metric permitted monodromies $v \to Mv + c$. Since $n_\mathrm{e} a + n_\mathrm{m} a_D$ is not invariant under addition of a constant to a or a_D, and there is no way to compensate for this by any transformation of n_e or n_m, we must set $c = 0$. [3] Moreover, under $v \to Mv$, we need $w \to wM^{-1}$ where w is the row vector $w = (n_\mathrm{m}, n_\mathrm{e})$. But since $(n_\mathrm{m}, n_\mathrm{e})$ are integers, M^{-1} must be integer-valued. As the determinant of M is 1, it follows that also M is integer-valued; hence the monodromy group is at most $\mathrm{SL}(2,\mathbb{Z})$.

Stability of BPS-saturated states

Now we will discuss the stability of BPS-saturated states. Many of the following remarks are well known and none are new.

A BPS-saturated state of given $(n_\mathrm{m}, n_\mathrm{e})$ determines a vector $Z = n_\mathrm{m} a_D + n_\mathrm{e} a$ in the complex plane. Its mass is $\sqrt{2}$ times the length of that vector. According to (4.3), all other states with the same $(n_\mathrm{m}, n_\mathrm{e})$ are heavier. Assuming that the ratio a_D/a is not real, the complex numbers a and a_D generate a lattice in the complex plane, and Z is a point in that lattice.

Let us analyze a possible decay process of a BPS-saturated state S with $Z = an_\mathrm{e} + a_D n_\mathrm{m}$ and mass $M = \sqrt{2}|Z|$ to states S_i with $Z_i = n_{m,i} a_D + n_{e,i} a$ and masses $M_i \geqslant \sqrt{2}|Z_i|$.

[3] When matter is included, additional terms appear in Z and one no longer gets $c = 0$ [7].

Since the charges $(n_{\rm m}, n_{\rm e})$ must be conserved, $Z = \sum_i Z_i$. It is clear from the triangle inequality that

$$|Z| \leqslant \sum_i |Z_i| \tag{4.8}$$

and hence

$$M \leqslant \sum_i M_i. \tag{4.9}$$

Of course, if $M < \sum_i M_i$ the decay is impossible. Equality is achieved in (4.8) when and only when all the states are BPS-saturated and all the vectors Z and Z_i are aligned, that is $t_i = Z_i/Z$ is real and positive and $\sum_i t_i = 1$. Assuming that the ratio of a_D/a is not real, this is possible only if the charge vectors $(n_{\rm m}, n_{\rm e})$ of the initial particles and $(n_{m,i}, n_{e,i})$ of the final particles are proportional. This in turn is possible only if $n_{\rm m}$ and $n_{\rm e}$ are not relatively prime; i.e. $(n_{\rm m}, n_{\rm e}) = (qm, qn)$ with integers q, m and n.

Conversely, states with $(n_{\rm m}, n_{\rm e}) = (qm, qn)$ are in fact only neutrally stable against decay to q states with charges (m, n).

Now what happens to these stable particles (BPS saturated with $(n_{\rm m}, n_{\rm e})$ relatively prime) as we vary some of the parameters that determine the vacuum? Such small changes in the vacuum can be described by emission of zero momentum particles—the neutral u quanta. Possible emission of such particles does not affect the argument for stability of the stable BPS saturated states so those states must persist as the parameters are varied.

So far we have assumed that a_D/a is not real. As explained in Ref. [27], the above argument fails if, upon varying the parameters, a_D/a passes through the real axis.[4] When that happens, the two dimensional lattice generated by a_D and a collapses to a one dimensional configuration, and it becomes much easier for the triangle inequality to collapse to an equality. For instance, if a_D/a is real and irrational, there are infinitely many points $n_{m,1}a_D + n_{e,1}a$ on the segment between 0 and $Z = n_{\rm m}a_D + n_{\rm e}a$; letting Z_1 be one such point, and $Z_2 = Z - Z_1$, a BPS-saturated state of given Z is only neutrally stable against decaying to possible BPS-saturated particles of given Z_1 and Z_2.

Moreover—and this is the main point—it was shown in Ref. [27] that at least in two dimensions, a BPS-saturated particle S of given $(n_{\rm m}, n_{\rm e})$ can disappear (or appear) when one passes through the point in parameter space at which a_D/a is real. What happens is that, for, say, $\mathrm{Im}(a_D/a) > 0$, the S particle, if it exists, is stable against decay to, say, $S_1 + S_2$, but for $\mathrm{Im}(a_D/a) \to 0$, the mass of the S particle goes up to the $S_1 + S_2$ threshold. Varying the parameters still further, to $\mathrm{Im}(a_D/a) < 0$, the S particle no longer exists—it has decayed to $S_1 + S_2$. A BPS-saturated state of the given $(n_{\rm m}, n_{\rm e})$ would again be stable when $\mathrm{Im}(a_D/a) < 0$, but such a particle may not exist. In Ref. [27], the precise number of BPS-saturated states that appear or disappear in this way was computed in two dimensions. An analogous computation in four dimensions would be desirable.

[4] Their formulation is slightly more general; they do not assume that the allowed values of Z are integer linear combinations of two basic numbers a and a_D.

In the present paper, discontinuity of the BPS-saturated spectrum will be found (in Section 6) only for strong coupling where it is difficult to explicitly check what is going on. In a subsequent paper [7], we will see such jumping also for weak coupling.

Such jumping does not occur for the $N = 4$ theory since then one has the exact formula $a_D = \tau_{\rm cl} a$ ensuring that a_D/a is not real. Consequently, the spectrum of BPS-saturated states of given $(n_{\rm m}, n_{\rm e})$ is independent of the coupling and so can be computed semiclassically for weak coupling. This has, in fact, been assumed in tests of Olive–Montonen duality for $N = 4$.

5. Structure of the moduli space

In Section 2, we developed the general local framework for the low energy effective action of the $N = 2$ theory. At the outset of Section 3, we noted that this framework could not be satisfactory globally because the metric on the moduli space of vacua could not be positive definite. Instead, we found that the global structure could involve certain monodromies; as we have explained, the group generated by the monodromies is a subgroup of $\mathrm{SL}(2,\mathbb{Z})$.

5.1. The singularity at infinity

It is actually quite easy to see explicitly the appearance of non-trivial monodromies. In fact, asymptotic freedom implies a non-trivial monodromy at infinity. The renormalization group corrected classical formula $\mathcal{F}_{\text{one loop}} = i\mathcal{A}^2 \ln(\mathcal{A}^2/\Lambda^2)/2\pi$ gives for large a

$$a_D = \frac{\partial \mathcal{F}}{\partial a} \approx \frac{2ia}{\pi} \ln(a/\Lambda) + \frac{ia}{\pi}. \tag{5.1}$$

It follows that a_D is not a single-valued function of a for large a. If we recall that the physical parameter is really $u = \frac{1}{2}a^2$ (at least for large u and a), then the monodromy can be determined as follows. Under a circuit of the u plane at large u, one has $\ln u \to \ln u + 2\pi i$, and hence $\ln a \to \ln a + \pi i$. So the transformation are

$$\begin{aligned} a_D &\to -a_D + 2a \\ a &\to -a. \end{aligned} \tag{5.2}$$

Thus, there is a non-trivial monodromy at infinity in the u plane,

$$M_\infty = PT^{-2} = \begin{pmatrix} -1 & 2 \\ 0 & -1 \end{pmatrix}, \tag{5.3}$$

where P is the element -1 of $\mathrm{SL}(2,\mathbb{Z})$ and as usual

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{5.4}$$

The factor of P in the monodromy exists already at the classical level. As we said above, a and $-a$ are related by a gauge transformation (the Weyl subgroup of the $\mathrm{SU}(2)$

gauge group) and therefore we work on the u plane rather than its double cover, the a plane. In the anomaly free $\mathbb{Z}_8$ subgroup of the R symmetry group $U(1)_{\mathcal{R}}$, there is an operation that acts on a by $a \to -a$; when combined with a Weyl transformation, this is the unbroken symmetry that we call P. Up to a gauge transformation it acts on the bosons by $\phi \to -\phi$, so it reverses the sign of the low energy electromagnetic field which in terms of SU(2) variables is proportional to $\mathrm{Tr}\,(\phi F)$. Hence it reverses the signs of all electric and magnetic charges and acts as $-1 \in SL(2,\mathbb{Z})$. The P monodromy could be removed by (perhaps artificially) working on the a plane instead of the u plane.

The main new point here is the factor of T^{-2} which arises at the quantum level. This factor of T^{-2} has a simple physical explanation in terms of the electric charge of a magnetic monopole. As explained in Ref. [30], magnetic monopoles labeled by (n_m, n_e) have anomalous electric charge $n_e + (\theta_{\rm eff}/2\pi) n_m$. The appropriate effective theta parameter is the low energy one

$$\theta_{\rm eff} = 2\pi \,\mathrm{Re}\; \tau(a) = 2\pi \,\mathrm{Re}\; \frac{\mathrm{d}a_D}{\mathrm{d}a} = 2\pi \,\mathrm{Re}\; \frac{\mathrm{d}a_D/\mathrm{d}u}{\mathrm{d}a/\mathrm{d}u}. \tag{5.5}$$

For large $|a|$, we have $\theta_{\rm eff} \approx -4\arg(a)$ which can be understood from the anomaly in the $U(1)_{\mathcal{R}}$ symmetry. The monodromy at infinity transforms the row vector (n_m, n_e) to $(-n_m, -n_e - 2n_m)$, which implies that (a_D, a) transforms to $(-a_D + 2a, -a)$. The electric charge of the magnetic monopole can in fact be seen in the formula for Z, which if we take a_D from (5.1) and set $a = a_0 \mathrm{e}^{-i\theta_{\rm eff}/4}$ (with $a_0 > 0$) is

$$Z \approx a_0 \mathrm{e}^{-i\theta_{\rm eff}/4} \left\{ \left(n_e + \frac{\theta_{\rm eff} n_m}{2\pi} \right) + i n_m \left(\frac{2\ln a_0/\Lambda + 1}{\pi} \right) \right\}. \tag{5.6}$$

The monodromy under $\theta_{\rm eff} \to \theta_{\rm eff} + 4\pi$ is easily seen from this formula to transform (n_m, n_e) in the expected fashion. Of course, this simple formula depended on the semiclassical expression (5.1) for a_D; with the exact expressions we will presently propose, the results are much more complicated, in part because the effective theta angle is no longer simply the argument of a.

5.2. Singularities at strong coupling

The monodromy at infinity means that there must be an additional singularity (or topological complication) somewhere in the u plane. If $\mathcal{M}'$ is the moduli space of vacua with all singularities deleted, then the monodromies must give a representation of the fundamental group of $\mathcal{M}'$ in $SL(2,\mathbb{Z})$. Can this representation be abelian? If the monodromies all commute with PT^{-2}, then a^2 is a good global complex coordinate, and the metric is globally of the form (3.1) with a global harmonic function $\mathrm{Im}\;\tau(a)$. As we have already noted, such a metric could not be positive.

The alternative is to assume a nonabelian representation of the fundamental group. This requires at least two more punctures of the u plane (in addition to infinity). Since there is a symmetry $u \leftrightarrow -u$ acting on the u plane, the minimal assumption is that there are precisely two more punctures exchanged by this symmetry. In this paper, we will find that this assumption leads to a unique and elegant solution that passes many tests.

(In a following paper [7], we will in some sense derive this assumption from more general properties of $N = 2$ systems with matter.)

The most natural physical interpretation of singularities in the u plane is that some additional massless particles are appearing at a particular value of u. Such a phenomenon of singularities in moduli space associated with the occurrence of extra massless particles has already been observed in $N = 1$ theories [2] and we will argue that it also happens in our problem.

For instance, in the classical theory, at $u = a = 0$, the SU(2) gauge symmetry is restored; all the gluons become massless. In fact classically $a_D = 4\pi i a/g^2$ also vanishes at this point, and the monopoles and dyons become massless as well. One might be tempted to believe that the missing singularity comes from an analogous point in the quantum theory at which the gauge boson masses vanish. Though this behavior might seem unusual in asymptotically free theories in general, there are good indications that some $N = 1$ theories have an infrared fixed point with massless nonabelian gluons [2,31].

However, there are good reasons to doubt that the $N = 2$ theory has this behavior. First of all, to make sense of the monodromies, one needs (as we saw above) not a single singularity but (at least) a pair of singularities at non-zero $\langle \mathrm{Tr}\,\phi^2 \rangle$. One might be willing to believe that $\langle \mathrm{Tr}\,\phi^2 \rangle \neq 0$ in the theory with massless nonabelian gauge bosons because of spontaneous breaking of the discrete chiral symmetry. This assumption, however, clashes with the asymptotic conformal invariance that one would expect in the infrared if the gauge bosons are massless. In fact, a non-zero expectation value of $\mathrm{Tr}\,\phi^2$ contradicts conformal invariance unless $\mathrm{Tr}\,\phi^2$ is of dimension zero. In a unitary quantum field theory, the only operator of zero dimension is the identity; $\mathrm{Tr}\,\phi^2$ cannot mix with the identity under renormalization because it is odd under a global symmetry.

Moreover, a conformally invariant point for the $N = 2$ theory is far-fetched because conformal invariance together with $N = 2$ supersymmetry implies invariance under the full $N = 2$ superconformal algebra including the $U(1)_{\mathcal{R}}$ symmetry. Thus, the instanton anomaly in the $U(1)_{\mathcal{R}}$ symmetry would have to somehow disappear. Moreover, for a field such as $\mathcal{O} = \mathrm{Tr}\,\phi^2$ which is in a chiral multiplet, superconformal invariance implies that the dimension $D(\mathcal{O})$ and $U(1)_{\mathcal{R}}$ charge $\mathcal{R}(\mathcal{O})$ are related by $D(\mathcal{O}) = \mathcal{R}(\mathcal{O})/2$. Thus at a conformal point the dimension of $\mathrm{Tr}\,\phi^2$ would have the canonical value 2, not the value 0 that it should have in order to have an expectation value.

5.3. *Interpretation of the singularities*

Since the above discussion of massless gauge bosons does not appear promising, we will assume that the singularities come from massive particles of spin $\leqslant 1/2$ that become massless at particular points in the moduli space. Since there are no such elementary multiplets, these must be bound states or collective excitations. It might sound counter intuitive that such objects can become massless. However, a similar phenomenon of massless bound states (which are not Goldstone bosons) has been observed in $N = 1$ theories [2] and we will argue that it also happens in $N = 2$.

The possibilities are severely restricted by the structure of $N = 2$ supersymmetry: a massive multiplet of particles of spins $\leqslant 1/2$ must be a hypermultiplet that saturates the

BPS bound.

In the semiclassical approximation the only such hypermultiplets in the $N = 2$ gauge theory are the monopoles and dyons whose mass renormalization was the subject of Section 4. We will interpret the needed singularities as arising when these particles become massless. For instance, from the discussion of masses in Section 4, the monopole becomes massless, while the gluons remain massive, at a point where $a_D = 0$ while $a \neq 0$. Similarly a $(1, 1)$ dyon becomes massless if $a + a_D = 0$ while $a, a_D \neq 0$.

Proposing that magnetic monopoles become massless and dominate the low energy landscape (at certain points in the moduli space of vacua) may seem bold. In the rest of this paper we will give evidence for this hypothesis as follows:

(1) We will use the renormalization group to compute the $\mathrm{SL}(2, \mathbb{Z})$ monodromy that arises near a point at which a hypermultiplet becomes massless. We will then show that if the hypermultiplets that are relevant are the monopoles and dyons that are visible semiclassically, then the monodromies work out consistently.

(2) The underlying $N = 2$ theory can be perturbed to an $N = 1$ theory by adding $\mathrm{Tr}\,\Phi^2$ as a superpotential. It is believed that this causes confinement of quarks. We will see that the same perturbation, added near the point at which monopoles are becoming massless, causes the monopoles to condense (develop a vacuum expectation value). This naturally leads to confinement of charges, giving—for the first time—an example in which confinement in a nonabelian theory is naturally understood in terms of monopole condensation.

(3) Finally, we will show that if it is assumed that there is a minimal set of singularities coming from massless monopoles and dyons, then one can determine uniquely and exactly the full structure of the low energy theory, including the Kähler metric on the moduli space of vacua and the particle masses as a function of the parameters. In particular the puzzle with positivity of the metric mentioned at the outset of section 3 is naturally overcome.

Explaining these points will occupy the rest of this paper.

5.4. *Effects of a massless monopole*

Our first task is to analyze the behavior of the effective Lagrangian near a point u_0 on the moduli space where magnetic monopoles become massless, that is, where

$$a_D(u_0) = 0. \tag{5.7}$$

Since monopoles couple in a non-local way to the original photon, we cannot use that photon in our effective Lagrangian. Instead, we should perform a duality transformation and write the effective Lagrangian in terms of the dual vector multiplet $\mathcal{A}_D$. The low energy theory is therefore an abelian gauge theory with matter (an $N = 2$ version of QED). The unusual fact that the light matter fields are magnetically charged rather than electrically charged does not make any difference to the low energy physics. The only reason we call these particles monopoles rather than electrons is that this language is appropriate at large $|u|$ where the theory is semiclassical.

The dominant effect on the low energy gauge coupling constant is due to loops of light fields. In our case, these are the light monopoles. The low energy theory is not asymptotically free and therefore its gauge coupling constant becomes smaller as the mass of the monopoles becomes smaller. Since the mass is proportional to a_D, the low energy coupling goes to zero as $u \to u_0$. The electric coupling constant which is the inverse of the magnetic one diverges at that point.

More quantitatively, using the one-loop beta function, near the point where $a_D = 0$, the magnetic coupling is

$$\tau_D \approx -\frac{i}{\pi} \ln a_D. \tag{5.8}$$

Since a_D is a good coordinate near that point,

$$a_D \approx c_0(u - u_0) \tag{5.9}$$

with some constant c_0. Using $\tau_D = \mathrm{d}h_D/\mathrm{d}a_D$, we learn that

$$a(u) = -h_D(u) \approx a_0 + \frac{i}{\pi} a_D \ln a_D \approx a_0 + \frac{i}{\pi} c_0(u - u_0) \ln(u - u_0) \tag{5.10}$$

for some constant $a_0 = a(u = u_0)$. This constant a_0 cannot be zero because if it had been zero, all the electrically charged particles would have been massless at $u = u_0$ and the computation using light monopoles only would not be valid. (In fact, there would be no local effective field theory incorporating both the light electrons and light monopoles.)

Now we can read off the monodromy. When u circles around u_0, so $\ln(u - u_0) \to \ln(u - u_0) + 2\pi i$, one has

$$\begin{aligned} a_D &\to a_D\,, \\ a &\to a - 2a_D. \end{aligned} \tag{5.11}$$

This effect is a sort of dual of the monodromy at infinity. Near infinity, the monopole gains electric charge, and near $u = u_0$, the electron gains magnetic charge. (It does not come back as a dyon but as a pair of particles for reasons explained at the end of section 6.) (5.11) can be represented by the 2×2 monodromy matrix

$$M_1 = ST^2S^{-1} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}. \tag{5.12}$$

5.5. The third singularity

With our assumption that there are only three singularities (counting $u = \infty$) and with two of the three monodromies determined in (5.3) and (5.12), we can now determine the third monodromy, which we will call M_{-1}. (The motivation for the notation is that we will eventually introduce parameters in which the singularities at finite u are at $u = 1$ and $u = -1$.) With all of the monodromies taken in the counter clockwise direction as in Fig. 1, the monodromies must obey $M_1 M_{-1} = M_\infty$, and from this we get

$$M_{-1} = (TS)T^2(TS)^{-1} = \begin{pmatrix} -1 & 2 \\ -2 & 3 \end{pmatrix}. \tag{5.13}$$

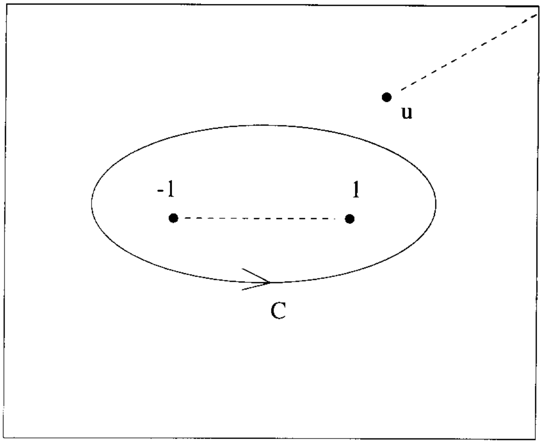

Fig. 1. The u plane with monodromies around 1, -1, and ∞. Note the choice of base point in the definition of the monodromies.

The matrix M_{-1} is conjugate to M_1. In fact, if

$$A = TM_1 = \begin{pmatrix} -1 & 1 \\ -2 & 1 \end{pmatrix}, \tag{5.14}$$

then

$$M_{-1} = AM_1A^{-1}. \tag{5.15}$$

Hence, M_{-1} can arise from a massless particle, just like M_1. Eq. (5.15) would also hold if A is replaced by $AM_1{}^r$ for any integer r.

What kind of particle should become massless to generate this singularity? If one arranges the charges as a row vector $q = (n_{\rm m}, n_{\rm e})$, then the massless particle that produces a monodromy M has $qM = q$. For instance, monodromy M_1 arises from a massless monopole of charge vector $q_1 = (1, 0)$, and using the known form of M_1, one has $q_1M_1 = q_1$. Duality symmetry implies that this must be so not just for the particular monodromy M_1 but for any monodromy coming from a massless particle. Upon setting $q_{-1} = (1, -1)$, we get $q_{-1}M_{-1} = q_{-1}$, and hence the monodromy M_{-1} arises from vanishing mass of a dyon of charges $(1, -1)$.

It seems that we are seeing massless particles of charges $(1, 0)$ or $(1, -1)$. However, there is in fact a complete democracy among dyons. The BPS-saturated dyons that exist semiclassically have charges $(1, n)$ (or $(-1, -n)$) for arbitrary integer n. The monodromy at infinity brings about a shift $(1, n) \to (1, n - 2)$. If one carries out this shift n times before proceeding to the singularity at $u = 1$ or $u = -1$, the massless particles producing those singularities would have charges $(1, -2n)$ and $(1, -1 - 2n)$, respectively. This amounts to conjugating the representation of the fundamental group by $M_\infty{}^n$.

The particular matrix A in (5.14) obeys $A^2 = -1$, which is equivalent to the identity as an automorphism of $\mathrm{SL}(2, \mathbb{Z})$. Conjugation by A implements the underlying $\mathbb{Z}_2$ symmetry of the quantum moduli space which according to our assumptions exchanges the two singularities. The $\mathbb{Z}_2$ maps $M_1 \to M'_1 = M_{-1}$, $M_{-1} \to M'_{-1} = M_1$ and $M_\infty \to M'_\infty = M'_1 M'_{-1} = M_{-1}M_1$. Note that M'_∞ is not just obtained from M_∞ by conjugation, but the relation $M_\infty = M_1 M_{-1}$ is preserved. The reason for that is that (as in any situation in which one is considering a representation of the fundamental group of a manifold in a nonabelian group), the definition of the monodromies requires a choice of base point, as shown in Fig. 1. The operation $u \to -u$ acts on the base point, and this has to be taken into account in determining how M_∞ transforms under $\mathbb{Z}_2$.

One can go farther and show that if one assumes the existence of a $\mathbb{Z}_2$ symmetry between M_1 and M_{-1}, then they must be conjugate to T^2, and not some other power of T. In our derivation of the monodromy (5.12), the 2 came from something entirely independent of the assumption of a $\mathbb{Z}_2$ symmetry, namely, from the charges and multiplicities of the monopoles that exist semiclassically.

5.6. Monopole condensation and confinement

We will now explain a quite satisfying physical phenomenon which was in fact at the heart of how some of these things were originally discovered.

We recall that the underlying $N = 2$ chiral multiplet $\mathcal{A}$ decomposes under $N = 1$ supersymmetry as a vector multiplet W_α and a chiral multiplet Φ. Breaking $N = 2$ down to $N = 1$, one can add a superpotential $W = m\mathrm{Tr}\,\Phi^2$ for the chiral multiplet. This gives a bare mass to Φ, reducing the theory at low energies to a pure $N = 1$ gauge theory. The low energy theory has a $\mathbb{Z}_4$ chiral symmetry. This theory is strongly believed to generate a mass gap, with confinement of charge and spontaneous breaking of $\mathbb{Z}_4$ to $\mathbb{Z}_2$. Furthermore, there is no vacuum degeneracy except what is produced by this symmetry breaking, so that there are precisely two vacuum states [32].

How can this be mimicked in the low energy effective $N = 2$ theory? That theory has a moduli space $\mathcal{M}$ of quantum vacua. The massless spectrum at least semiclassically consists solely of the abelian chiral multiplet $\mathcal{A}$ of the unbroken U(1) subgroup of SU(2). If those are indeed the only massless particles, the effect in the low energy theory of turning on m can be analyzed as follows. The operator $\mathrm{Tr}\,\Phi^2$ is represented in the low energy theory by a chiral superfield U. Its first component is the scalar field u whose expectation value is

$$\langle u\rangle = \langle \mathrm{Tr}\,\phi^2\rangle \tag{5.16}$$

(ϕ is the $\theta = 0$ component of the superfield Φ). This is a holomorphic function on the moduli space. At least for small m we should add to our low energy Lagrangian an effective superpotential $W_{\text{eff}} = mU$ (soon we will show that this is the exact expression also for large m).

Turning on the superpotential mU would perhaps eliminate almost all of the vacua and in the surviving vacua give a mass to the scalar components of $\mathcal{A}$. But if there are no extra degrees of freedom in the discussion, the gauge field in $\mathcal{A}$ would remain massless. To get a mass for the gauge field, as is needed since the microscopic theory has a mass gap for $m \neq 0$, one needs either (i) extra light gauge fields, giving a non-abelian gauge theory and possible strong coupling effects, or (ii) light charged fields, making possible a Higgs mechanism.

Thus we learn, as we did in discussing the monodromies, that somewhere on $\mathcal{M}$ extra massless states must appear. The option (i) does not seem attractive, for reasons that we have already discussed. Instead we will consider option (ii), with the further proviso, from our earlier discussion, that the light charged fields in question are monopoles and dyons.

Near the point at which there are massless monopoles, the monopoles can be represented in an $N = 1$ language by ordinary (local) chiral superfields M and $\widetilde{M}$, as long as we describe the gauge field by the dual to the original photon, $\mathcal{A}_D$. The superpotential is

$$\widehat{W} = \sqrt{2}A_D M\widetilde{M} + mU(A_D), \tag{5.17}$$

where the first term is required by $N = 2$ invariance of the $m = 0$ theory, and the second term is the effective contribution to the superpotential induced by the microscopic perturbation $m\,\mathrm{Tr}\,\Phi^2$.

The fact that the superpotential (5.17) is exact can be established by using the non-renormalization theorem of [1] as follows. For $m = 0$ the theory is invariant under $\mathrm{SU}(2)_R$. It will suffice to consider its $\mathrm{U}(1)_J$ subgroup (2.2). This is an $N = 1$ R symmetry under which Φ has charge zero. The two $N = 1$ chiral fields M and $\widetilde{M}$ are in an $N = 2$ hypermultiplet. Therefore, according to (2.4), they both have charge one. The presence of a term $m\mathrm{Tr}\,\Phi^2$ in the microscopic superpotential shows that the parameter m carries charge two. The low energy superpotential is holomorphic in its variables $\widehat{W}(m, M\widetilde{M}, A_D)$ and should have charge two under $\mathrm{U}(1)_J$. Imposing that it is regular at $m = M\widetilde{M} = 0$, we find that it is of the form $\widehat{W} = mf_1(A_D) + M\widetilde{M}f_2(A_D)$. The functions f_1 and f_2 are independent of m and can be determined by examining the limit of small m, leading to (5.17).

The low energy vacuum structure is easy to analyze. Vacuum states correspond to solutions, up to gauge transformation, of

$$\mathrm{d}\widehat{W} = 0 \tag{5.18}$$

that obey the additional condition

$$|M| = |\widetilde{M}| \tag{5.19}$$

(we denote by M and $\widetilde{M}$ both the superfields and their first components). The latter condition comes from vanishing of the D terms. Implementing these conditions, one finds if $m = 0$ that vacuum states correspond to $M = \widetilde{M} = 0$ with arbitrary a_D; this is simply the familiar moduli space $\mathcal{M}$. If $m \neq 0$ the result is quite different. We get

$$\begin{aligned} \sqrt{2}M\widetilde{M} + m\frac{\mathrm{d}u}{\mathrm{d}a_D} &= 0\,, \\ a_D M = a_D\widetilde{M} &= 0. \end{aligned} \tag{5.20}$$

Assuming that $\mathrm{d}u \neq 0$, the first equation requires $M, \widetilde{M} \neq 0$, whence the second equation requires $a_D = 0$. Imposing also (5.19), we get a unique solution up to gauge transformation, with

$$M = \widetilde{M} = \left(-mu'(0)/\sqrt{2}\right)^{1/2}. \tag{5.21}$$

Expanding around this vacuum, it is easy to see that there is a mass gap. For instance, the gauge field gets a mass by the Higgs mechanism, since $M, \widetilde{M} \neq 0$. The Higgs mechanism in question is a magnetic Higgs mechanism, since the fields with expectation values are monopoles! Condensation of monopoles will induce confinement of electric charge. Thus, we get an explanation in terms of the low energy effective action of why the microscopic theory becomes confining when the $m\,\mathrm{Tr}\,\Phi^2$ superpotential is added.

We have also noted that in the presence of the perturbing superpotential, the microscopic theory has a $\mathbb{Z}_4$ global symmetry, spontaneously broken down to $\mathbb{Z}_2$. This symmetry breaking is manifest in the effective theory, since the broken symmetry exchanges the point where $a_D = 0$ and there is a massless monopole with a point where $a - a_D = 0$ and there is a massless dyon. Thus, the effective theory has two vacuum states from the two points of extra massless particles (corresponding to the two singularities in the u plane that were discussed above), related by a broken symmetry, in parallel with what is expected microscopically.

In the microscopic description, one attributes the properties of the two vacuum states (for $m \neq 0$) to a difficult-to-understand strong gauge coupling. In the low energy theory, we have found a perfectly peaceful description involving a *weakly coupled* theory of monopoles and photons; the coupling constant flows to zero in the infrared if $m = 0$, and in general flows to a value of order $-1/\ln m$, since this is the behavior of weakly coupled QED. The original, electric, gauge coupling is the inverse of the magnetic coupling, so flows in the infrared to a value of order $-\ln m$; in this sense the two vacua that survive when m is small and non-zero are strongly coupled.

The effective superpotential (5.17) gives a good description of the low energy physics near the point u_0 where the monopoles are light. What is its meaning far from u_0? Then, except for the expectation values of the monopoles and their masses, few of their properties could be correctly inferred from the effective Lagrangian. The effective Lagrangian has even greater difficulties if one tries to continue it to $u = -u_0$ where massless dyons should appear. There cannot be an effective field theory containing both the monopoles and dyons as elementary fields, as they are not relatively local. The theory we are discussing is an interesting example of a theory in which, while in the neighborhood of any one vacuum there is a good description by a low energy effective theory, there is no low energy effective theory that is reasonable everywhere.

5.7. The effective parameter

Finally, we will tie up some remaining loose ends and at the same time prepare for the solution of the model in the next section.

In the above, we found two vacuum states for $m \neq 0$, while assuming that $\mathrm{d}u \neq 0$. We would get additional states, with $M = \widetilde{M} = 0$, for any point at which $\mathrm{d}u = 0$. We do not want any additional vacuum states, since two is the correct number of vacua in the microscopic theory, so we assume that $du \neq 0$ everywhere.

The fact that $du \neq 0$ means that u is everywhere a good local coordinate on the space of vacua. Is it also a good coordinate globally? The question is whether there is precisely one vacuum for given u. According to (5.16), u is simply the expectation value of $\mathrm{Tr}\,\phi^2$, regarded as a function on the space of vacua. At least when $u = \mathrm{Tr}\phi^2$ is large, perturbation theory is reliable and there is precisely one vacuum for given u. This combination of facts strongly suggests that u is a good global coordinate on the space $\mathcal{M}$ of vacua.

In proceeding in the next section to propose a solution of the model, we will assume that $\mathcal{M}$ is just the u plane (with u equal to the expectation value of $\mathrm{Tr}\,\phi^2$) with precisely the two singularities that we proposed above. We will pick a renormalization convention such that the singularities in the u plane are at 1 and -1. Using the solution of the model that we will propose in the next section, it would be possible to compare to perturbation theory and determine how this convention compares to other conventions such as $\overline{\mathrm{MS}}$.

6. The solution of the model

In this section, we will, finally, put the pieces together, with a couple of new ingredients, and make our proposal for the solution of the model.

The moduli space $\mathcal{M}$ of quantum vacua is to be the u plane with singularities at $1, -1$, and ∞ and a $\mathbb{Z}_2$ symmetry acting by $u \leftrightarrow -u$. Duality means that over the punctured u plane there is a flat $\mathrm{SL}(2,\mathbb{Z})$ bundle V with the following monodromies around $\infty, 1$, and -1:

$$
\begin{aligned}
M_\infty &= \begin{pmatrix} -1 & 2 \\ 0 & -1 \end{pmatrix} , \\
M_1 &= \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} , \\
M_{-1} &= \begin{pmatrix} -1 & 2 \\ -2 & 3 \end{pmatrix} .
\end{aligned}
\tag{6.1}
$$

The quantities $(a_D(u), a(u))$ are a holomorphic section of the bundle V (or of $V \otimes \mathbb{C}$ to be precise). This section is to be determined by its asymptotic behavior. Near $u = \infty$,

$$
\begin{aligned}
a &\approx \sqrt{2u} , \\
a_D &\approx i\frac{\sqrt{2u}}{\pi} \ln u .
\end{aligned}
\tag{6.2}
$$

Near $u = 1$,

$$
\begin{aligned}
a_D &\approx c_0(u-1) , \\
a &\approx a_0 + \frac{i}{\pi} a_D \ln a_D ,
\end{aligned}
\tag{6.3}
$$

with constants a_0, c_0. The behavior near $u = -1$ is similar, with $a - a_D$ replacing a_D.

There is also one more important constraint. The metric on the u plane can be written

$$(\mathrm{d}s)^2 = \mathrm{Im}(\tau)\cdot|\mathrm{d}a|^2 \tag{6.4}$$

with

$$\tau = \frac{\mathrm{d}a_D/\mathrm{d}u}{\mathrm{d}a/\mathrm{d}u.} \tag{6.5}$$

To ensure positivity of the metric, $\mathrm{Im}(\tau)$ must be positive definite. We need not specify the asymptotic behavior of τ near $-1, 1$, and ∞, since this is determined by the asymptotic behavior of a and a_D.

The key point in making it practical to solve these conditions is that the flat $\mathrm{SL}(2,\mathbb{Z})$ bundle has a nice interpretation. To begin with, note that the monodromy matrices in equation (6.1) are all congruent to 1 modulo 2. So these matrices do not generate $\mathrm{SL}(2,\mathbb{Z})$—at most they could generate the subgroup of $\mathrm{SL}(2,\mathbb{Z})$ consisting of matrices congruent to 1 modulo 2. This subgroup is usually called $\Gamma(2)$. In fact, M_∞ and M_1 do generate $\Gamma(2)$.[5] Moreover, the u-plane punctured at 1, -1, and ∞ has a very special interpretation. It can be thought of as the quotient of the upper half plane H by $\Gamma(2)$. Indeed, as $\Gamma(2)$ is of index six in $\mathrm{SL}(2,\mathbb{Z})$, the quotient $H/\Gamma(2)$ is a six-fold cover of the usual modular domain. This quotient has three cusps, which we can take to be at $1, -1$, and ∞, with precisely the monodromies in (6.1).

The family of curves parametrized by $H/\Gamma(2)$ can be described very explicitly, by the equation

$$y^2 = (x-1)(x+1)(x-u). \tag{6.6}$$

The idea here is that for every u, there is a genus one Riemann surface E_u, determined by Eq. (6.6). This equation describes a double cover of the x plane branched over $-1, 1, \infty$, and u. The curve E_u becomes singular when (and only when) two branch points coincide. This occurs precisely for $u = 1, -1$, or ∞.

Note among other things that (6.6) has a manifest symmetry w that maps $u \to -u$, $x \to -x$, $y \to \pm iy$. This generates a $\mathbb{Z}_4$ symmetry, but only a $\mathbb{Z}_2$ quotient acts on the u plane. Indeed, $P = w^2$ acts trivially on u and x while mapping $y \to -y$. It will turn out that a and a_D are odd under $y \to -y$ and so odd under P. The symmetry structure just described is precisely that of the field theory that we are aiming to solve.

The Riemann surface E_u has a two-dimensional first homology group $V_u = H^1(E_u, \mathbb{C})$. The V_u are fibers of a flat bundle V over the punctured u plane, and this is the bundle of which the pair (a_D, a) is a section. This is a convenient description of the bundle, as we will see. The bundle V can be trivialized locally by picking a pair of independent and continuously varying one-cycles γ_1, γ_2 on E_u; these can be normalized so that the intersection number is

$$\gamma_1 \cdot \gamma_2 = 1. \tag{6.7}$$

The space $H^1(E_u, \mathbb{C})$ can be thought of as the space of meromorphic $(1,0)$-forms on E_u of vanishing residues, modulo exact forms (or total derivatives). The heuristic

[5] For a quick proof of this and other assertions made presently, see pp. 92-3 of Ref. [33].

idea is that if λ is such a one-form, then it represents an element of $H^1(E_u,\mathbb{C})$ because it can be paired with one-cycles by

$$\gamma \to \oint_\gamma \lambda. \tag{6.8}$$

The condition that the residues of λ vanish ensures that this pairing is invariant under deformation of γ even across a pole of λ. One identifies $\lambda \sim \lambda + \mathrm{d}w$ (with w a meromorphic function) since the exact differential dw would not contribute to the contour integral (6.8).

As $H^1(E_u,\mathbb{C})$ is two-dimensional, a basis is provided by any two linearly independent elements, for instance

$$\lambda_1 = \frac{\mathrm{d}x}{y} \text{ and } \lambda_2 = \frac{x\,\mathrm{d}x}{y}. \tag{6.9}$$

Here λ_1 is actually a holomorphic differential, having no poles even at infinity; it is up to a scalar multiple the unique holomorphic differential on E_u. (In terms of the Hodge decomposition of $H^1(E_u,\mathbb{C})$, it represents an element of $H^{1,0}$.) If one picks on E_u a basis of one-cycles normalized as in (6.7), then the periods

$$b_i = \oint_{\gamma_i} \lambda_1, \quad i = 1,2 \tag{6.10}$$

obey

$$\frac{b_1}{b_2} = \tau_u, \tag{6.11}$$

with τ_u the usual τ parameter of the elliptic curve E_u, which has the fundamental property

$$\mathrm{Im}(\tau_u) > 0. \tag{6.12}$$

Under a change in the basis of γ's, τ_u would be transformed by the standard action of $SL(2,\mathbb{Z})$ on the upper half plane. As for λ_2, its only pole is a double pole at infinity (where the residue vanishes, since in any case the sum of the residues of a meromorphic differential is zero).

As (6.9) gives a basis of $H^1(E_u,\mathbb{C})$, an arbitrary section of the flat bundle V_u can be represented in the form

$$\lambda = a_1(u)\lambda_1 + a_2(u)\lambda_2. \tag{6.13}$$

This representation is not necessarily convenient, as it may be convenient to use the freedom of adding dw for some function w. In any event, once a section λ is chosen, to extract the components a_D and a of this section, one picks the basis γ_1, γ_2 of one-cycles and defines

$$a_D = \oint_{\gamma_1} \lambda$$

$$a = \oint_{\gamma_2} \lambda. \tag{6.14}$$

If a different choice is made for the γ_i, the pair (a_D, a) will be transformed by an element of $\mathrm{SL}(2,\mathbb{Z})$.

Now we would like to impose the condition that τ as defined in (6.5) has $\mathrm{Im}(\tau) > 0$. There is an obvious way to satisfy this condition. One has

$$\frac{\mathrm{d}a_D}{\mathrm{d}u} = \oint_{\gamma_1} \frac{\mathrm{d}\lambda}{\mathrm{d}u},$$
$$\frac{\mathrm{d}a}{\mathrm{d}u} = \oint_{\gamma_2} \frac{\mathrm{d}\lambda}{\mathrm{d}u}. \tag{6.15}$$

Suppose that

$$\frac{\mathrm{d}\lambda}{\mathrm{d}u} = f(u)\lambda_1 = f(u)\frac{\mathrm{d}x}{y}, \tag{6.16}$$

with $f(u)$ a function of u only. Then we get

$$\frac{\mathrm{d}a_D}{\mathrm{d}u} = f(u) b_1,$$
$$\frac{\mathrm{d}a}{\mathrm{d}u} = f(u) b_2, \tag{6.17}$$

with b_i defined in (6.10). Consequently,

$$\tau = \frac{\mathrm{d}a_D/\mathrm{d}u}{\mathrm{d}a/\mathrm{d}u} = \frac{b_1}{b_2} = \tau_u, \tag{6.18}$$

and therefore $\mathrm{Im}(\tau) > 0$. Conversely, if $\mathrm{Im}(\tau) > 0$ everywhere, then $\tau = \tau_u$ and λ is as above. To see this, note first if $\mathrm{Im}(\tau) > 0$ everywhere, then for every u, $\tau(u)$ is the τ-parameter of an elliptic curve. Also, the pair $(\mathrm{d}a_D/\mathrm{d}u, \mathrm{d}a/\mathrm{d}u)$ transforms under $\mathrm{SL}(2,\mathbb{Z})$ the same way as (a_D, a), and so gives a section of the same flat bundle. The family of curves determined by $u \to \tau(u)$ consequently has the same monodromies (and singularities) as the family E_u determined by $u \to \tau_u$. It follows on general grounds that these families coincide.

It remains to select f and verify the desired asymptotic behavior of a_D and a near $u = 1, -1, \infty$. We will in fact show that everything works out correctly with $f = -\sqrt{2}/4\pi$. It will then be clear that this is the unique choice that works; in particular, a non-constant f would somewhere introduce unwanted poles or zeroes.

First of all, given (6.16), there is no problem in finding λ. We can take

$$\lambda = \frac{\sqrt{2}}{2\pi} \frac{\mathrm{d}x\,\sqrt{x-u}}{\sqrt{x^2-1}} = \frac{\sqrt{2}}{2\pi} \frac{\mathrm{d}x\, y}{x^2-1}. \tag{6.19}$$

Obviously, the u derivative of this expression reproduces (6.16) with $f = -\sqrt{2}/4\pi$. The residues of λ vanish since

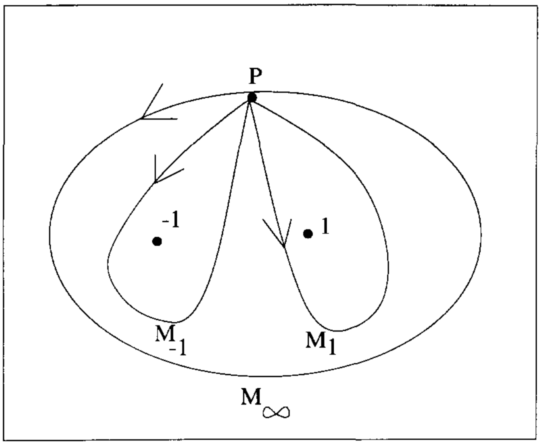

Fig. 2. A non-trivial one-cycle on E_u comes from a contour in the x plane that loops around two of the branch points. The cuts in the x plane are indicated with dotted lines.

$$\lambda = \frac{\sqrt{2}(\lambda_2 - u\lambda_1)}{2\pi} \tag{6.20}$$

with λ_i as above. λ is odd under $P = w^2$, as the second formula in (6.19) makes clear, and that is why P reverses the sign of a and a_D as defined presently.

Now we need an explicit basis of one-cycles on E_u. These can be constructed as follows. A circle C in the x plane that loops around two of the branch points lifts to a non-trivial cycle γ on E_u. (Apart from any topological arguments, the non-triviality of this cycle follows from the non-vanishing of the integral that we are about to write.) Suppose, for instance, that as in Fig. 2 we take C to loop once around the branch points at $x = 1, -1$. The differential form λ that we wish to integrate, viewed by the first formula in (6.19) as a form on the x plane, has a factor $\sqrt{(x-u)/(x^2-1)}$ that requires branch cuts. We can take the cuts to run from -1 to 1 and from u to ∞, as shown in Fig. 2. Then we can deform C so it runs just around the cut between -1 and 1. So

$$\oint_\gamma \lambda = \frac{\sqrt{2}}{2\pi} \oint_C \frac{\mathrm{d}x\,\sqrt{x-u}}{\sqrt{x^2-1}} = \frac{\sqrt{2}}{\pi} \int_{-1}^{1} \frac{\mathrm{d}x\,\sqrt{x-u}}{\sqrt{x^2-1}}. \tag{6.21}$$

A factor of two in the last step comes because the integral over C contains an integral from -1 to 1 then from 1 back to -1 on the other side of the cut; the two segments make equal contributions.

Now we can implement the definition of a and a_D in (6.14). Choosing the contour just described to be γ_2, we have

$$a = \frac{\sqrt{2}}{\pi} \int_{-1}^{1} \frac{\mathrm{d}x\,\sqrt{x-u}}{\sqrt{x^2-1}}. \tag{6.22}$$

Similarly, defining another cycle γ_1 by using a circle that loops around the branch points at 1 and u, we can take

$$a_D = \frac{\sqrt{2}}{\pi} \int_1^u \frac{\mathrm{d}x\,\sqrt{x-u}}{\sqrt{x^2-1}}. \tag{6.23}$$

Of course, this particular basis of one-cycles was picked with a view to getting the desired behavior near the singularities in the u plane. The occurrence of a square root means that the overall signs of a and a_D are ill-defined. That is in keeping with the fact that, from the outset, the classical relation $a^2 = 2u$ means that a can be recovered from the gauge invariant quantity u only up to sign.

It will suffice to study the behavior of a and a_D near $u = \infty$ and $u = 1$, since the behavior near $u = -1$ is determined by the $\mathbb{Z}_2$ symmetry of the u plane that was described earlier.

Near $u = \infty$, we get

$$a \approx \frac{\sqrt{2u}}{\pi} \int_{-1}^1 \frac{\mathrm{d}x}{\sqrt{1-x^2}} = \sqrt{2u} \tag{6.24}$$

and (after a change of variables $x = uz$)

$$a_D = \frac{\sqrt{2u}}{\pi} \int_{1/u}^1 \frac{\mathrm{d}z\,\sqrt{z-1}}{\sqrt{z^2-u^{-2}}}. \tag{6.25}$$

For $u \to \infty$ the integral develops a logarithmic divergence near $z = 0$; extracting the divergent term, we get

$$a_D \approx i\frac{\sqrt{2u}\ln u}{\pi}. \tag{6.26}$$

Of course, this logarithm is the signal of asymptotic freedom in the semiclassical region of large u.

The behavior near $u = 1$ is equally easy to determine. From (6.25), we get

$$a_D = \frac{\sqrt{2u}}{\pi} \int_{1/u}^1 \frac{\mathrm{d}z\sqrt{z-1}}{\sqrt{z+u^{-1}}\sqrt{z-u^{-1}}} \approx \frac{1}{\pi} \int_{1/u}^1 \frac{\mathrm{d}z\,\sqrt{z-1}}{\sqrt{z-u^{-1}}}$$
$$= \frac{i}{2}\left(1-\frac{1}{u}\right) \approx \frac{i(u-1)}{2}. \tag{6.27}$$

What about a? At $u = 1$ the integral for a is

$$a(u=1) = \frac{\sqrt{2}}{\pi} \int_{-1}^1 \frac{\mathrm{d}x}{\sqrt{x+1}} = \frac{4}{\pi}. \tag{6.28}$$

However, the derivative of a with respect to u is given by an integral

$$\frac{\mathrm{d}a}{\mathrm{d}u} = -\frac{\sqrt{2}}{2\pi}\int_{-1}^{1} \frac{\mathrm{d}x}{\sqrt{(x+1)(x-1)(x-u)}} \tag{6.29}$$

that—for $u \to 1$—becomes logarithmically divergent near $x = 1$. Extracting the coefficient of the logarithm, we find that the expansion of a is

$$a = \frac{4}{\pi} - \frac{(u-1)\ln(u-1)}{2\pi} + \dots . \tag{6.30}$$

Comparing (6.30) and (6.27), we get the desired monodromy $a \to a - 2a_D$ near $u = 1$.

This completes our verification of the expected properties. But there is still one point to discuss.

The spectrum.

The remaining point concerns, in a sense, the physical meaning of the duality that we have used to solve the theory.

In looping around $u = 1$ or $u = -1$, the pair (a_D, a) are transformed by monodromies M_1 and M_{-1}. The charges $(n_{\mathrm{m}}, n_{\mathrm{e}})$ are transformed similarly. Naively, one would think that the spectrum of BPS-saturated states would be transformed by the monodromy matrices. In that case, since the monodromies generate $\Gamma(2)$, the spectrum of BPS-saturated states would be $\Gamma(2)$ invariant.

In fact, that is not true. In the semiclassical region of large u, the BPS-saturated states are the electrons and W bosons of $(n_{\mathrm{m}}, n_{\mathrm{e}}) = (0, \pm 1)$ and the dyons of $(n_{\mathrm{m}}, n_{\mathrm{e}}) = (\pm 1, n)$; moreover the W bosons are chiral multiplets (spin $\leqslant 1$) and the dyons are hypermultiplets (spin $\leqslant 1/2$). The fact that this spectrum is not duality invariant is precisely the reason that it was concluded many years ago that Olive–Montonen duality did not hold for $N = 2$ super Yang–Mills theory.

From the discussion at the end of Section 4, there is a possible mechanism for curing the problem. The spectrum of BPS-saturated states can jump on crossing a curve in the u plane on which the ratio a_D/a is real. If a curve on which a_D/a is real passes through $u = 1$, then in looping around $u = 1$, one would have to cross that curve and the resulting jumping of the spectrum would invalidate conclusions based on the monodromies.

It is easy to see that this situation does arise. Near $u = 1$, a is nearly real - in fact $a(u = 1) = 4/\pi$. But a_D can have an arbitrary phase near $u = 1$ since $a_D \approx i(u-1)/2$. Thus, jumping can occur on a curve that near $u = 1$ looks like $u = 1 + it$, t real, where a_D/a is real. A similar curve on which a_D/a is real passes through $u = -1$.

It is consistent with everything we know, and will resolve all the puzzles about lack of duality in the spectrum, if the curve on which a_D/a is real looks something like $|u| = 1$. Then one could avoid the jumping phenomenon only if one stays in the region $u > 1$; the only monodromy that can be seen in that region is M_∞, under which the spectrum of BPS-saturated states is indeed invariant. However, we do not know a practical way to determine the curve on which a_D/a is real.

There is, however, one important region in which one can easily prove that a_D/a is *not* real. This is on the real u axis for $|u| > 1$, as one can easily see from (6.22) and (6.23). For u real and $u > 1$, the quantity $(x-u)/(x^2-1)$ is positive in the integration region for a and negative in the integration region for a_D, so a is real and a_D is imaginary. Using the $\mathbb{Z}_2$ symmetry $u \to -u$, it is clear that a_D/a is also not real for u real and $u < -1$. Therefore, by sticking to the real u axis, one can come in from the semiclassical region of large u to the singularities at $u = \pm 1$ without crossing any jumping curves. Therefore, whatever particles become massless at $u = \pm 1$ must evolve continuously from the BPS-saturated states that can be seen semiclassically near infinity. So for our picture of the strong coupling region to make sense, the monopoles and dyons that we need for the singularities must exist in the semiclassical region. Happily, they do.

Acknowledgement

We would like to thank T. Banks for a useful discussion on confinement and condensation of magnetic monopoles, P. Deligne for some conceptual assistance, G. Moore for suggesting a possible role of hypergeometric functions, and C. Vafa for explaining certain two-dimensional results. We have also benefited from discussions with K. Intriligator, R. Leigh and S. Shenker. This work was supported in part by DOE grant #DE-FG05-90ER40559 and in part by NSF grant PHY92-45317.

References

[1] N. Seiberg, Phys. Lett. B 318 (1993) 469.
[2] N. Seiberg, hep-th/9402044, Phys. Rev. D 49 (1994) 6857.
[3] K. Intriligator, R. Leigh and N. Seiberg, hep-th/9403198, RU-94-26, Phys. Rev. in press.
[4] V. Kaplunovsky and J. Louis, hep-th/9402005.
[5] M.A. Shifman and A.I. Vainshtein, Nucl. Phys. B 277 (1986) 456; B 359 (1991) 571.
[6] D. Amati, K. Konishi, Y. Meurice, G.C. Rossi and G. Veneziano, Phys. Rep. 162 (1988) 169 and references therein.
[7] N. Seiberg and E. Witten, to appear.
[8] E. Witten, Commun. Math. Phys. 117 (1988) 353.
[9] E. Witten, IASSNS-HEP-94-5, hep-th/9403195.
[10] M.K. Prasad and C.M. Sommerfield, Phys. Rev. Lett. 35 (1975) 760.
[11] E.B. Bogomol'nyi, Sov. J. Nucl. Phys. 24 (1976) 449.
[12] E. Witten and D. Olive, Phys. Lett. B 78 (1978) 97.
[13] C. Montonen and D. Olive, Phys. Lett. B 72 (1977) 117;
P. Goddard, J. Nuyts and D. Olive, Nucl. Phys. B 125 (1977) 1.
[14] H. Osborn, Phys. Lett. B 83 (1979) 321.,
[15] J. Cardy and E. Rabinovici, Nucl. Phys. B 205 (1982) 1;
J. Cardy, Nucl. Phys. B205 (1982) 17.
[16] A. Shapere and F. Wilczek, Nucl. Phys. B 320 (1989) 669.
[17] A. Font, L. Ibanez, D. Lüst and F. Quevedo, Phys. Lett. B 249 (1990) 35.
[18] A. Sen, TIFR-TH-94-08, hep-th/9402002.
[19] G. Segal, to appear.
[20] L. Girardello, A. Giveon, M. Porrati and A. Zaffaroni, NYU-TH-94/06/02, IFUM/472/FT, SISSA 80/94/EP, RI-6-94, hep-th/9406128.
[21] C. Vafa and E. Witten, to appear.
[22] N. Seiberg, Phys. Lett. B 206 (1988) 75.,

[23] B. De Wit and A. Van Proeyen, Nucl. Phys. B 245 (1984) 89.
[24] S. Ferrara, Mod. Phys. Lett. A6 (1991) 2175.
[25] A. Strominger, Commun. Math. Phys. 133 (1990) 163.
[26] P. Candelas and X. de la Ossa, Nucl. Phys. B 355 (1991) 455.
[27] S. Cecotti, P. Fendley, K. Intriligator and C. Vafa, Nucl. Phys. B 386 (1992) 405;
S. Cecotti and C. Vafa, Commun. Math. Phys. 158 (1993) 569.
[28] J. Wess and J. Bagger, Supersymmetry and supergravity (Princeton Univ. Press, Princeton, 1982).
[29] For a review, see A. Giveon, M. Porrati and E. Rabinovici, RI-1-94, hep-th/9401139.,
[30] E. Witten, Phys. Lett. B 86 (1979) 283.
[31] N. Seiberg, to appear.
[32] E. Witten, Nucl. Phys. B 202 (1982) 253.
[33] C.H. Clemens, A scrapbook of complex curve theory (Plenum, New York, 1980).