Non-Perturbative Dynamics Of Four-Dimensional Supersymmetric Field Theories

Philip C. Argyres

Newman Laboratory, Cornell University, Ithaca NY 14853
argyres@mail.lns.cornell.edu

Abstract

An introduction to the construction and interpretation of supersymmetric low energy effective actions in four space-time dimensions is given. These effective actions are used to extract exact strong-coupling information about $N = 4$ and $N = 2$ supersymmetric gauge theories. The M-theory 5-brane construction which derives the effective action of certain $N = 2$ theories is described.

1 Introduction

The aim of these lectures is to introduce some of the arguments that have been used successfully in the last five years to obtain exact information about strongly coupled field theories. I will focus on four-dimensional field theories without gravity, although the techniques described here have been applied to theories in other dimensions and to string/M theory as well. I will also focus on theories with at least $N = 2$ supersymmetry (8 conserved supercharges) since these are physically rich theories with many open problems, but are still highly constrained by the symmetry.

The basic notion is that of a *low energy* (or *Wilsonian*) *effective action*. This is simply a local action describing a theory's degrees of freedom at energies below a given scale E. An example is the low energy effective action for QCD, chiral perturbation theory describing the interactions of pions at energies $E < \Lambda_{QCD}$. In such a theory particles heavier than Λ_{QCD} are included in the pion theory as classical sources. Other examples are the various ten and eleven-dimensional supergravity theories, which appear as effective actions for string/M theory at energies below their Planck scales. The effective action is obtained by averaging over (integrating out) the short distance fluctuations of the theory. If there is a sufficiently small ratio E/Λ between the cutoff energy scale E and the energy scale Λ characteristic of the dynamics of the degrees of freedom being averaged over, renormalization group arguments imply that the effective action can be systematically expanded as a power series in E/Λ—essentially an expansion in the number of derivatives of the fields.

We will use low energy effective actions to analyze four dimensional field theories by taking the limit as the cutoff energy scale E goes to zero, or equivalently, by just keeping the leading terms (up to two derivatives) in the low energy fields. I will call such $E \to 0$ low energy effective actions *infrared effective actions* (IREAs). The idea is to guess an IR effective field content for the microscopic (UV) theory in question and write down all possible IREAs built from these fields consistent with the global symmetries of the UV theory. For a "generic" UV theory this is no better than doing chiral perturbation theory for QCD, and would seem to give little advantage for obtaining exact results. However, if the theory has a continuous set of inequivalent vacua, it turns out that selection rules from global symmetries of the UV theory can sometimes constrain the IREA sufficiently to deduce exact results. There are a number of reviews deriving these exact results [1] assuming the constraints from supersymmetry. In particular, these lectures are a continuation of [2], where the construction of four-dimensional IREAs is explained in a relatively non-technical way.

We start in Sec. 2 with a brief review of IREAs with various amounts of supersymmetry. The constraints on the IREAs become progressively more restrictive as the number of supersymmetries is increased. In the $N=2$ case they are strong enough to allow quite general and restrictive properties of the moduli space of vacua of gauge theories to be deduced. The remainder of these lectures is devoted to using these IREAs to extract exact strong-coupling information about supersymmetric gauge theories. In particular, Sec. 3 discusses the exact IREAs of $N=4$ theories, while Sec. 4 discusses some general things that can be said about those of $N=2$ theories (Seiberg-Witten theory [3]). Sec. 5 will elaborate on the mathematical formulation of $N=2$ IREAs in preparation for Sec. 6 which presents an account of an M-theory 5-brane construction [4] which allows one to derive the IREA of certain $N=2$ theories.

Important topics omitted include the properties of interacting IREAs—the representation theory of superconformal algebras [5] and their use in analyzing IREAs [6]; instead these lectures concentrate on IR free effective actions. Also missing are details of supersymmetry algebras and the construction of their representations—many good texts and review articles cover this material [7]—or the application of the ideas presented here to theories in other dimensions [8].

2 IREAs

Since an IREA describes physics only for arbitrarily low energies, it is, by definition, scale invariant: we simply take the cutoff scale E below any finite scale in the theory. Scale invariant theories and therefore IREAs fall into one of the following categories:

Trivial theories in which all fields are massive, so there are no propagating degrees of freedom in the far IR.

Free theories in which all massless fields are non-interacting in the far IR. (They can still couple to massive sources, but these sources should not be treated dynamically in the IREA.)

Interacting theories of massless degrees of freedom which are usually assumed to be conformal field theories [9].

We generally have no effective description of interacting conformal field theories in four dimensions [10] so we must limit ourselves to free or trivial theories in the IR. A large class of these is given by the Coleman-Gross theorem [11] which states that for small enough couplings any theory of scalars, spinors, and $U(1)$ vectors in four dimensions flows in the IR to a free theory. We thus take the field content of our IREA to be a collection of real scalars ϕ^i, Weyl spinors ψ^a_α, and $U(1)$ vector fields A^I_μ. Here α and μ are the space-time spinor and vector indices, while i, a, and I label the different field species.

Since this theory is free in the IR, no interesting dynamics involving the spinor fields can occur, so the vacuum structure of this theory is governed by the scalar potential. Dropping the other fields we write the general Lagrangian with up to two derivatives for a set of real scalars

$$\mathcal{L} = -V(\phi) + \tfrac{1}{2} g_{ij}(\phi)\partial_\mu\phi^i\partial^\mu\phi^j. \tag{1}$$

Here the potential V is an arbitrary real function of the ϕ^i which is bounded below (for stability), while the coefficient g_{ij} of the generalized kinetic term is a real, symmetric and positive definite tensor (for unitarity). We assume V attains its minimum value, which without loss of generality we take to be $V = 0$.

Minimizing the generalized kinetic energy term implies that in the vacuum the scalars should all be constant. Denoting these constant values by the same symbols as for the fields, the set of all possible vacua is then seen to naturally have the structure of a Riemannian manifold $\mathcal{M}_0 = \{\phi^i\}$ with metric g_{ij} since an arbitrary non-singular field redefinition $\phi^i \to \widetilde{\phi}^i(\phi)$ transforms g_{ij} in the same way as a metric transforms under a change of coordinates.

If $V = 0$ identically, then $\mathcal{M}_0$ would describe a manifold of vacua of this theory. We call such a manifold of vacua the *moduli space* of the theory. Without any extra symmetries to constrain it, generically $V \neq 0$, so $\mathcal{M}_0$ is not the moduli space, but instead $\mathcal{M}_V = \mathcal{M}_0/\{V = 0\}$ is. At least locally $\mathcal{M}_V$ has the structure of a submanifold of $\mathcal{M}_0$.

Now let us incorporate the $U(1)$ gauge fields into our discussion of the moduli space. Some of the scalar fields may be charged under the $U(1)^n$ gauge group of the IREA. The infinitesimal $U(1)^n$ action of the gauge group on the scalars then generates a diffeomorphism of $\mathcal{M}_0$. For this to be a symmetry of the IREA it must both leave V invariant and be an isometry of the metric g_{ij}. In that case the IREA can be written (excluding the spinors) as

$$\mathcal{L} = -V(\phi) + \tfrac{1}{2} g_{ij}(\phi)D_\mu\phi^i D^\mu\phi^j - \frac{1}{16\pi}\mathrm{Im}\left[\tau_{IJ}(\phi)\mathcal{F}^I \wedge *\mathcal{F}^J\right], \tag{2}$$

where $D_\mu = \partial_\mu + A^I_\mu \xi_I$, treating the ξ^i_I as Killing vectors generating the isometry. The last term in Eq. 2 is a generalized Maxwell term for the $U(1)$ field strengths $F^I_{\mu\nu} = \partial_\mu A^I_\nu - \partial_\nu A^I_\mu$, where we have defined

$$\mathcal{F}^I = F^I - i{*F^I} \tag{3}$$

in terms of 2-form field strengths and the Hodge star operator $*F_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}$. τ_{IJ} is a complex (gauge invariant) function of the ϕ^i symmetric in I and J and whose imaginary part is positive definite (for unitarity). Defining the real and imaginary parts of the couplings by

$$\tau_{IJ} = \frac{\theta_{IJ}}{2\pi} + i\frac{4\pi}{(e^2)_{IJ}}, \tag{4}$$

the generalized Maxwell term can be expanded as

$$\mathcal{L}_{U(1)} = -\frac{1}{2(e^2)_{IJ}}F^I \wedge *F^J + \frac{\theta_{IJ}}{64\pi^2}F^I \wedge F^J, \tag{5}$$

showing that the imaginary part of τ_{IJ} is a matrix of couplings and the real part are theta angles.

The addition of the $U(1)$ gauge fields affects the moduli space because two points of $\mathcal{M}_0$ which are related by a gauge transformation must be identified. Thus $\mathcal{M}_0$ or $\mathcal{M}_V$ (since V is gauge invariant) is replaced by $\mathcal{M}$, formed by dividing by the action of the gauged isometry group $U(1)^n$: $\mathcal{M} = \mathcal{M}_V/U(1)^n$.

Note that the vacuum expectation values (vevs) of charged scalars can not parameterize the moduli space, because when a charged scalar gets a nonzero vev it Higgses the $U(1)$ it is charged under and thereby gets a mass. It is therefore not a flat direction—*i.e.* changing its vev takes us off the moduli space $\mathcal{M}$. Since we are interested only in the extreme IR limit, we only need to keep the *neutral* scalars which parameterize $\mathcal{M}$. In this case the IREA (2) simplifies since $V = 0$ on $\mathcal{M}$ by definition and $D_\mu = \partial_\mu$ on neutral scalars. Thus only the metric $g_{ij}(\phi)$ and couplings $\tau_{IJ}(\phi)$ need to be specified. (If we included the fermions, there would also be the coefficient functions of their kinetic terms as well.)

The IR free low energy $U(1)^n$ dynamics is form-invariant under *electric-magnetic duality* transformations. These are simply relabellings of the fields, interchanging electric and magnetic fields and charges, and, because of the Dirac quantization condition relating electric and magnetic charges [12], also inverting the couplings $\tau_{IJ} \to -\tau^{IJ}$, where τ^{IJ} is the matrix inverse of τ_{IJ}: $\tau^{IJ}\tau_{JK} = \delta^I_K$. This electric-magnetic duality transformation together with the invariance of the physics under 2π shifts of the theta angles (integer shifts of $\mathrm{Re}\tau_{IJ}$) $\tau_{IJ} \to \tau_{IJ} + \delta^K_I\delta^L_J + \delta^L_I\delta^K_J$, generate a discrete group of duality transformations:

$$\tau_{IJ} \to ({A_I}^L\tau_{LM} + B_{IM})(C^{JN}\tau_{NM} + {D^J}_M)^{-1}, \tag{6}$$

where

$$M \equiv \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in Sp(2n, \mathbf{Z}). \tag{7}$$

The conditions on the $n \times n$ integer matrices A, B, C, and D for M to be in $Sp(2n, \mathbf{Z})$ are

$$\begin{aligned} AB^T = B^T A, \qquad & B^T D = D^T B, \\ A^T C = C^T A, \qquad & D^T C = C D^T, \\ A^T D - C^T B \quad = & \quad AD^T - BC^T = 1, \end{aligned} \tag{8}$$

and imply that

$$M^{-1} = \begin{pmatrix} D^T & -B^T \\ -C^T & A^T \end{pmatrix}. \tag{9}$$

The action of an electric-magnetic duality transformation on the $2n$-component row vector of magetic and electric charges $(n_m^I, n_{e,J})$ of massive states is

$$(n_m \; n_e) \rightarrow (n_m \; n_e) \cdot M^{-1}. \tag{10}$$

Electric-magnetic duality transformations are not symmetry transformations since they acts on the couplings. Instead, electric-magnetic duality simply expresses the equivalence of free $U(1)$ field theories coupled to classical (massive) sources under $Sp(2n, \mathbf{Z})$ redefinitions of electric and magnetic charges. The importance of this redundancy in the Lagrangian description of IREAs becomes apparent when there is a moduli space $\mathcal{M}$ of inequivalent vacua. In that case, upon traversing a closed loop in $\mathcal{M}$ the physics must, by definition, be the same at the beginning and end of the loop, but the Lagrangian description need not—it may have suffered an electric-magnetic duality transformation. This possibility is often expressed by saying that the coupling matrix τ_{IJ}, in addition to being symmetric and having positive definite imaginary part, is also a section of a (flat) $Sp(2n, \mathbf{Z})$ bundle with action given by (6).

2.1 N=2 Supersymmetric IREAs

The basic (no central charges) $N = 2$ superalgebra is, in an indexless notation,

$$\{Q_m, \overline{Q}_n\} = \delta_{mn} P, \qquad \{Q_m, Q_n\} = 0, \qquad m, n = 1, 2, \tag{11}$$

where Q_m are two Weyl spinor supercharges, and P is the energy-momentum vector. Note that the $N = 2$ algebra has an $SU(2)_R$ group of automorphisms under which Q_m transforms as a doublet. (Global symmetries under which the supercharges transform are called R *symmetries*.)

On shell irreducible representations of (11) are easy to construct. There are two solutions with no spins greater than one: the *hypermultiplet*, containing two propagating complex scalars, ϕ and $\widetilde{\phi}$, as well as two Weyl fermions, ψ and $\widetilde{\psi}$; and the *vector multiplet*, made from one complex scalar a, two Weyl spinors λ and $\widetilde{\lambda}$, and a vector field A_μ. An important distinguishing feature of the hypermultiplet is that its scalars form a complex $SU(2)_R$ doublet. The bosonic degrees of freedom in a vector multiplet, by contrast, are a single complex scalar and a vector field, both transforming in the adjoint of the gauge group, and both singlets under $SU(2)_R$. In particular, in the case of $U(1)^n$ gauge group, which we are interested in for describing IREAs, the vector multiplet scalars are necessarily neutral.

An $N=2$ IREA with Abelian gauge group and neutral hypermultiplets, *a priori* has an action of the form (2) where the ϕ^i fields run over all the bosons (in both the vector and hypermultiplets), and $\mathcal{F}^I$ run over the $U(1)$ gauge fields. Compatibility with the $N=2$ global supersymmetry tightly constrains this action; see, for example, [2]. The result is that the general $N=2$ IREA gauge group $U(1)^n$ (labelled by indices $I,J=1,\ldots,n$) and n_f neutral hypermultiplets (labelled by indices $i,j=1,\ldots,n_f$) has the form

$$\mathcal{L} = g_{i\bar{\jmath}}(\phi,\widetilde{\phi})\left(\partial\phi^i\cdot\partial\overline{\phi}^{\bar{\jmath}} + \partial\widetilde{\phi}^i\cdot\partial\overline{\widetilde{\phi}}^{\bar{\jmath}}\right) + \mathrm{Im}\tau_{IJ}(a)\left(\partial a^I\cdot\partial\overline{a}^J + \mathcal{F}^I\cdot\mathcal{F}^J\right), \tag{12}$$

with

$$\partial_{[I}\tau_{J]K} = 0, \tag{13}$$

and τ_{IJ} a holomorphic function (really $Sp(2n,\mathbf{Z})$ section) of the vector multiplet scalars a^I.

This form of the IREA of $N=2$ supersymmetric theories has many important consequences. The first is the absence of any potential terms for the scalars which implies that in $N=2$ theories there will be a moduli space of vacua as long as $U(1)$ vector multiplets or neutral hypermultiplets can be shown to occur in the IREA.

The next $N=2$ selection rule follows from the fact that there are no kinetic cross terms between the vector and hypermultiplets, implying that the moduli space has a natural (local) product structure $\mathcal{M}=\mathcal{M}_H\times\mathcal{M}_V$, where $\mathcal{M}_H$ is the subspace of $\mathcal{M}$ along which only the hypermultiplet vevs vary while the vector multiplet vevs remain fixed, and *vice versa* for $\mathcal{M}_V$. In cases where $\mathcal{M}_V$ is trivial (a point), $\mathcal{M}=\mathcal{M}_H$ is called a *Higgs branch* of the moduli space; when $\mathcal{M}_H$ is trivial $\mathcal{M}_V$ is called the *Coulomb branch* (since there are always the massless $U(1)$ vector bosons from the vector multiplets). Cases where both $\mathcal{M}_H$ and $\mathcal{M}_V$ are non-trivial are called *mixed branches.*

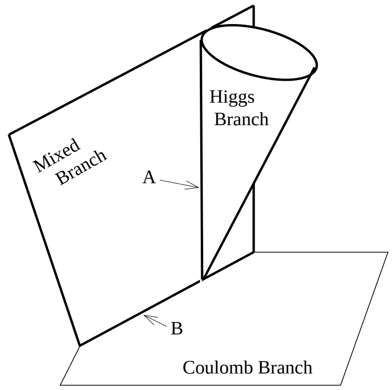

Figure 1: Cartoon of a classical $N = 2$ moduli space. The Higgs and mixed branches intersect along a Higgs submanifold A, while the mixed branch intersects the Coulomb branch along a Coulomb submanifold B.

In general the total moduli space of a given theory need not be a smooth manifold—it may have "jumps" where submanifolds of different dimensions meet. Classically this occurs as a result of the Higgs mechanism: a charged scalar vev Higgses some vector multiplets, typically lifting them (making them massive). But at the special point where the charged vev is zero, the vector multiplets become massless, leading to extra flat directions and a jump in the dimensionality of the moduli space. Hence, at least classically, the general picture of an $N = 2$ moduli space is a collection of intersecting manifolds, which can be Higgs, Coulomb, or mixed branches [16, 17], see Fig. 1.

This classical picture is, of course, modified quantum mechanically. A microscopic (UV) theory is characterized by some parameters (*e.g.* masses, strong coupling scales, theta angles, dimensionless couplings); we can always take ratios of these parameters to describe them by at most one scale Λ and a set of dimensionless parameters λ_k. The coefficient functions g_{ij} and τ_{IJ} of the IREA will, in general, depend on Λ and the λ_k. Determining this dependence of these IR quantities on UV parameters is the ultimate goal of the techniques reviewed in these lectures.

For asymptotically free gauge theories, the important UV parameter is the (complex) strong coupling scale of the theory, Λ (whose definition we'll recall in Section 4, below). The important property of asymptotically free theories is that they are nearly free at energy scales above Λ, so the classical theory is obtained in the limit $\Lambda \to 0$. Since Λ appears in τ_{IJ} (at, say, one loop), it appears in the Lagrangian in the same way a scalar vev a^I of an $N = 2$ vector multiplet would. Therefore, we can think of Λ as a background $U(1)$ vector superfield—in other words it is consistent to *assign* Λ

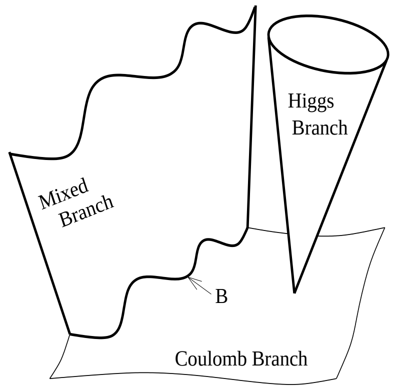

Figure 2: Cartoon of a quantum $N = 2$ moduli space. The Higgs branch and the Higgs (hypermultiplet vev) directions of the mixed branch remain unmodified from their classical geometries, though they may be deformed in the Coulomb (vector multiplet vev) directions. The Coulomb branch is generally different from its classical geometry.

supersymmetry transformation properties as if it were the lowest component of a chiral superfield. This implies that whatever strong dynamics takes place upon flowing to the IR, Λ will only enter the IREA in the way chiral multiplet scalars do. In particular, τ_{IJ} will be a holomorphic function of Λ [18]. Also, since the metric on the Higgs branch is independent of vector superfields, it is independent of Λ. Finally, we can use the fact that the classical theory is obtained in the limit $\Lambda \to 0$ to conclude that *the Higgs branch metric is given exactly by the classical answer* [16]. We thus learn that only the Coulomb branch can receive quantum corrections, and that any mixed branches will retain their classical product structure of a hypermultiplet manifold times the vector multiplet manifold corresponding to the subspace of the Coulomb branch along which the mixed and Coulomb branches intersect; see Fig. 2. Since the hypermultiplet manifolds can be determined classically in $N = 2$ supersymmetric gauge theories, we will not consider them further.

(It is worth examining more closely the logic of this argument. We are *assuming* that the IREA will be described by a nonlinear sigma model of some set of light chiral fields which are not necessarily simply a subset of those of the UV theory. We have no derivation of this hypothesis—we can only test it to see if it gives consistent answers. The couplings of the effective theory will be some functions of the couplings of the microscopic theory, which we would like to solve for. The next step of thinking of the couplings in the superpotential as background chiral superfields is just a trick—we are certainly allowed to do so if we like since the couplings enter in the microscopic theory in the same way a background chiral superfield would. The point of this trick is that it makes the restrictions on possible quantum corrections allowed by supersymmetry apparent. These restrictions are just a supersymmetric version of the familiar "selection rules" of quantum mechanics.)

Finally, a key fact about the Coulomb branch is that though it can be corrected quantum mechanically, it is never wholly lifted in asymptotically free $N = 2$ gauge theories. We will see this when we describe $N = 2$ non-Abelian gauge theories in Section 4, below. This means that $N = 2$ supersymmetric theories *generically* have a moduli space of vacua. The challenge of solving for the vacuum structure of $N = 2$ gauge theories is thus that of determining the geometry on the Coulomb branch. This geometry is encoded in the IREA (12), the integrability condition (13), and the $Sp(2n, \mathbf{Z})$ transformation properties (6), and is known as *rigid special Kähler* geometry. We will develop the mathematics of these manifolds further in Section 5 below. But for the moment, let us move on to $N = 4$ supersymmetry.

2.2 N=4 IREAs

The $N = 4$ superalgebra is

$$\{Q_m, \overline{Q}_n\} = \delta_{mn} P, \qquad \{Q_m, Q_n\} = 0, \qquad m, n = 1, \ldots, 4. \tag{14}$$

This algebra has an $SU(4)_R$ group of automorphisms under which Q_m transforms as a $\mathbf{4}$. There is one on-shell irreducible representation with no spins greater than one, which decomposes under an $N = 2$ subalgebra as a vector multiplet plus a hypermultiplet. Its field content can be organized into six real scalars a_i transforming as a $\mathbf{6}$ of $SU(4)_R$, four Weyl fermions ψ_n in the $\mathbf{4}$ of $SU(4)_R$, and a vector field A_μ. All these fields must transform in the adjoint of the gauge group since they are in the same multiplet as a vector boson; in the case of $U(1)^n$ gauge group the scalars are necessarily neutral.

The $N = 4$ IREA with Abelian gauge group has the same form as the $N = 2$ IREA (12). But now since the $N = 2$ hypermultiplet and vector multiplet scalars are related by the $SU(4)_R$ global symmetry, they must have the same metric, τ^{IJ}:

$$\mathcal{L} = \mathrm{Im}\tau_{IJ} \left(\partial a_i^I \cdot \partial \overline{a}_i^J + \mathcal{F}^I \cdot \mathcal{F}^J \right). \tag{15}$$

Furthermore, since by the $N = 2$ selection rules, the vector multiplet metric and the hypermultiplet metric cannot depend on the same fields, we must have

$$\tau_{IJ} = \text{constant}. \tag{16}$$

This has the immediate consequence that the moduli space of the $N = 4$ theory must locally be flat:

$$\mathcal{M} = \mathbf{R}^{6n} = \{a_i^I\}. \tag{17}$$

3 N=4 Exact Results

We will now turn to our main task of using the supersymmetric IREAs found in the last section to deduce exact non-perturbative information about supersymmetric gauge theories. We start with the $N = 4$ case since it is the most constrained, and so gives the simplest illustration of the basic idea.

Consider an $N = 4$ super Yang-Mills theory. Its (UV) bosonic action is given by

$$\mathcal{L} = \mathrm{Im}\left\{\tau \mathrm{tr}\left(D_\mu \Phi_i D^\mu \Phi_i + \mathcal{F} \cdot \mathcal{F} + \sum_{i>j} [\Phi_i, \Phi_j]^2\right)\right\}, \tag{18}$$

where $i, j = 1, \ldots, 6$ for the six adjoint scalars in the $N = 4$ vector multiplet. For definiteness, let us take the gauge group to be $SU(n + 1)$.

Classically, the vacua of this theory occur for Φ_i vevs of the form (up to gauge transformations)

$$\langle \Phi_i \rangle = \begin{pmatrix} \alpha_i^1 & & \\ & \ddots & \\ & & \alpha_i^{n+1} \end{pmatrix}, \tag{19}$$

with

$$\sum_{K=1}^{n+1} \alpha_i^K = 0. \tag{20}$$

This tracelessness condition is required for an $(n + 1) \times (n + 1)$ matrix representation of the adjoint representation of $SU(n + 1)$. By the usual Higgs mechanism, a generic such vev spontaneously breaks $SU(n + 1) \to U(1)^n$. There are special vacua where two or more of the α_i^K are equal where $SU(n + 1)$ is not completely broken down to $U(1)$'s, but has some $SU(m)$ subgroups left unbroken.

The classical moduli space is thus the flat $6n$-dimensional manifold $\mathcal{M} = \{\alpha_i^I, I = 1, \ldots, n\}$. Actually, choosing the vevs of the form (19) does not completely fix the gauge invariance: the Weyl subgroup of $SU(n + 1)$ acts on the α_i^K by permutations on the K index. Thus the moduli space must be divided out by this S_{n+1} group of permutations, so

$$\mathcal{M} = \mathbf{R}^{6n}/S_{n+1}. \tag{21}$$

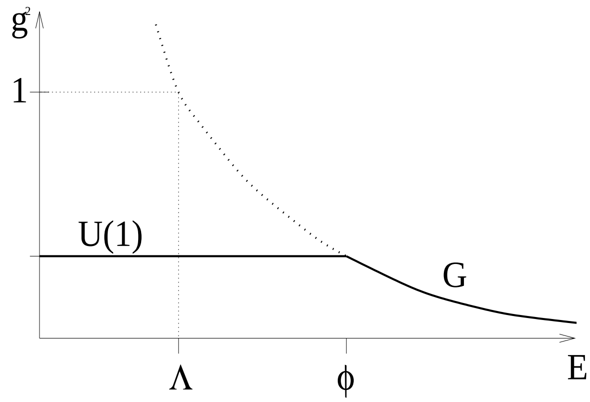

Figure 3: Running of the coupling of an asymptotically free gauge theory with gauge group G Higgsed to $U(1)$'s at a scale $\langle\Phi\rangle \gg \Lambda$. The $U(1)$ couplings do not run below $\langle\Phi\rangle$ because there are no charged fields lighter than ϕ.

The orbifold submanifolds at the fixed points of the S_{n+1} action occur at precisely the places where the low energy $U(1)^n$ gauge group is enhanced.

Now we turn to the quantum mechanical theory. The first question is whether the UV coupling τ (which is classically dimensionless) suffers some renormalization group running, thus generating some strong coupling scale Λ? Our IREA selection rules can immediately rule this out, however. For we have seen that the IR effective $U(1)^n$ coupling τ_{IJ} must be a constant, independent of the values of any of the vevs α_i^K. But if the UV coupling ran at high energies, we would detect this in a vev-dependence of the τ_{IJ}, for at weak enough UV coupling (large vevs in an asymptotically free theory) the classical Higgs mechanism picture of the classical picture can be made arbitrarily precise, implying that the τ_{IJ} will be equal to the value of the microscopic τ at the scale $\langle\Phi\rangle$; see Fig. 3. Since, in fact, the τ_{IJ} are independent of $\langle\Phi\rangle$, so τ must be exactly (even non-perturbatively) independent of scale.

We have thus learned that the $N=4$ super Yang-Mills theory is a scale invariant, or conformal, field theory. Indeed, it is easy to check at one loop that the beta-function for the running of the gauge coupling vanishes, and can also be verified to all orders in perturbation theory. The form of the $N=4$ IREA also shows it to be true non-perturbatively.

To make further progress on the quantum vacuum structure of these $N=4$ theories, consider such theories at weak coupling,

$$\tau \to +i\infty. \tag{22}$$

Then the classical description of the $SU(n+1) \to U(1)^n$ Higgs mechanism is good, giving an IREA

$$\mathcal{L} = \mathrm{Im}\tau_{IJ}\left(\partial a_i^I \cdot \partial \overline{a}_i^J + \mathcal{F}^I \cdot \mathcal{F}^J\right), \tag{23}$$

for $I, J = 1, \ldots, n$ where we have defined

$$a_i^I \equiv \sum_{K=1}^{I} \alpha_i^K, \tag{24}$$

and

$$\tau_{IJ} = \tau \begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & \ddots \\ & & \ddots & \ddots \end{pmatrix}, \tag{25}$$

which is proportional to the Cartan matrix of $SU(n+1)$. Note that the change of basis (24) is integer-valued, *i.e.* an element of $GL(n, \mathbf{Z})$; this was a necessary restriction in order to preserve the integrality of the magnetic and electric charges (n_m, n_e) of any massive states in the theory.

The moduli space of this theory is just the classical one (21), including dividing by the S_{n+1} Weyl group action on the a_i^I. For example, the simplest case is $SU(2)$ gauge group, where $n = 1$. Then the moduli space is $\mathcal{M} = \mathbf{R}^6/S_2$ where the $S_2 \simeq \mathbf{Z}_2$ acts on the six coordinates a_i as $S_2 :\ a_i \to -a_i$. This has a single fixed point at the origin. Thus the vacuum structure is simple: at the origin of moduli space there is a scale invariant vacuum with an unbroken $SU(2)$ gauge invariance, and there is a six dimensional space of flat directions leading away from it where the scale invariance is spontaneously broken by the non-zero a_i vevs and the low energy theory is the $N = 4$ $U(1)^n$ theory (23).

At points in the moduli space where new degrees of freedom (not included in the fields of the IREA) become massless, the IREA description of the physics breaks down. Generally this break down is signalled by a singularity in the metric on the moduli space. In the above example the origin was singular in this way: the $W^{\pm}$ bosons (and their superpartners) filling out the $SU(2)$ adjoint multiplet became massless there, and the metric was singular there (it is a $\mathbf{Z}_2$ orbifold point).

Finally, we can deduce what happens to the vacuum structure for couplings not near weak coupling. Since the low energy τ_{IJ} cannot depend on the vevs, it can only be a function of the UV coupling τ. Treating the IREA (23) as an $N = 2$ supersymmetric action, τ must enter only holomorphically in τ_{IJ}. Furthermore, by the angularity of the theta angle, *i.e.* the invariance of the physics under $\tau \to \tau + 1$, τ_{IJ} can depend on

τ only as

$$\tau_{IJ} = \tau C_{IJ} + \sum_{\ell=0}^{\infty} C_{IJ}^{(\ell)} e^{2\pi i \ell \tau}, \tag{26}$$

where C_{IJ} is the Cartan matrix appearing in (25), and $C_{IJ}^{(\ell)}$ are arbitrary independent complex matrices. Note that the first term is not invariant under $\tau \to \tau + 1$, but shifts by the interger matrix C_{IJ}. This has no effect on the physics since it is just a low-energy electric-magnetic duality transformation in the $U(1)^n$ theory.

I do not know of a first principles argument to determine the $C_{IJ}^{(\ell)}$'s, but various indirect arguments from string and M theory (mentioned below) imply that they are all proportional to the Cartan matrix C_{IJ}. In that case we have $\tau_{IJ} = f(\tau) C_{IJ}$, with $f(\tau) \to \tau$ as $\tau \to +i\infty$. In particular, $f(\tau)$ differs from τ only by nonperturbative terms. Since we have no alternative non-perturbative definition of the UV coupling τ, *we are free to define $f(\tau)$ itself to be the UV coupling*: $f(\tau) = \tau$. So, finally, the IR $U(1)^n$ couplings are

$$\tau_{IJ} = \tau C_{IJ}. \tag{27}$$

The $Sp(2n, \mathbf{Z})$ electric-magnetic duality transformations (6) include transformations taking $\tau_{IJ} \to \tau'_{IJ}$ such that $\tau \to \tau + 1$ and $\tau \to -1/\tau$. These generate an $SL(2, \mathbf{Z})$ group of transformations on the microscopic coupling which leave the IR physics invariant. This is evidence for the existence of an *S-duality* of the $N = 4$ theories [19, 20] which is simply the statement that the theories with UV couplings τ related by $SL(2, \mathbf{Z})$ transformations are physically equivalent. S-dualities, also known as strong-weak coupling dualities, or Montonen-Olive dualities, and are conceptually distinct from electric-magnetic dualities. It is worth emphasizing that electric-magnetic dualities are equivalences of the free IR effective $U(1)^n$ theories, whereas S-dualities identify interacting theories with *a priori* distinct couplings.

Further evidence for $N = 4$ S-duality comes from the spectrum of BPS states in the theories. BPS states are states preserving some of the supersymmetries. The masses of states preserving half the supersymmetries are known exactly in terms of the *central charges* of the supersymmetry algebra [20]. For example, for $SU(2)$, the mass of a $\frac{1}{2}$-BPS state with magnetic and electric charges n_m and n_e is

$$M^2 = \frac{1}{\mathrm{Im}\tau} |n_e + \tau n_m|^2 (a_i^I C_{IJ} a_i^J). \tag{28}$$

It is easy to check that this formula is invariant under electric-magnetic duality transformations. Furthermore, under S-duality transformations it takes states with given

(n_m, n_e) into ones with different charges. In particular, given that the massive W-bosons of the $N = 4$ theory are $\frac{1}{2}$-BPS states with $(n_m, n_e) = (0, \pm 1)$, if S-duality is correct it follows that there must be also be a $\frac{1}{2}$-BPS state for all relatively prime choices of electric and magnetic charges. Such states have been constructed [21], adding to the evidence for the S-duality.

The best evidence for $N = 4$ S-duality comes from string and M theory. There are a set of similar dualities in these theories which fit together in an intricate way, and also imply the $N = 4$ S-duality. The self-consistency of this "web of dualities" thus lends strong support to the existence of $N = 4$ S-duality. Scale-invariant $N = 2$ theories also have S-dualities [22, 23, 24]. It is striking that, unlike their $N = 4$ counterparts, many of the $N = 2$ dualities can be proved using purely field theoretic arguments [25, 26].

4 Seiberg-Witten Theory

We now turn to deriving the vacuum properties of $N = 2$ supersymmetric gauge theories. Since there are now both hypermultiplets and vector multiplets at our disposal, we can construct a much richer set of $N = 2$ theories than $N = 4$ theories. For simplicity we will focus only on the $N = 2$ Yang-Mills theories, that is, those with only vector multiplets appearing in the microscopic action. The treatment of theories with hypermultiplets does not differ much from the pure Yang-Mills theories, especially as we are primarily interested in the Coulomb branch of the moduli space.

Taking $SU(n+1)$ as our example again, denote the complex adjoint scalar field of the vector multiplet by Φ, an $(n+1) \times (n+1)$ complex traceless matrix. Then the $N = 2$ Yang-Mills action looks much like the $N = 4$ action (18),

$$\mathcal{L} = \mathrm{Im}\left\{\tau \,\mathrm{tr}\left(D_\mu \Phi D^\mu \Phi + [\Phi, \Phi]^2 + \mathcal{F} \cdot \mathcal{F}\right)\right\}. \tag{29}$$

Classically, the vacua of this theory occur for Φ vevs of the form (up to gauge transformations)

$$\langle \Phi \rangle = \begin{pmatrix} a^1 & & \\ & \ddots & \\ & & a^{n+1} \end{pmatrix}, \tag{30}$$

with

$$\sum_{K=1}^{n+1} a^K = 0, \tag{31}$$

and the a^K complex. Such a vev spontaneously breaks $SU(n+1) \to U(1)^n$ except when two or more of the a^K are equal so that $SU(n+1)$ is not completely broken down to $U(1)$'s, but has some $SU(m)$ subgroups left unbroken. Choosing the vevs of the form (30) does not completely fix the gauge invariance since the Weyl subgroup $S_{n+1} \subset SU(n+1)$ acts on the a^K by permutations on the K index. The classical moduli space is thus a flat n-complex-dimensional manifold with orbifold singularities

$$\mathcal{M} = \mathbf{C}^n / S_{n+1}. \tag{32}$$

Gauge-invariant coordinates on this space can be taken to be the n independent complex symmetric polynomials in the a^K:

$$\begin{aligned} s_2 &= \sum_{J<K} a^J a^K, \\ s_3 &= \sum_{J<K<L} a^J a^K a^L, \\ &\vdots \\ s_{n+1} &= a^1 a^2 \cdots a^{n+1}. \end{aligned} \tag{33}$$

(Note that there is no s_1 since the sum of the a^K's vanishes by the tracelessness condition.)

Now we turn to the quantum mechanical theory. Unlike the $N=4$ super Yang-Mills theories, the $N=2$ theories are asymptotically free and their UV coupling τ (which is classically dimensionless) runs with scale, generating a strong coupling scale Λ. Let us recall how this scale is defined. Consider an asymptotically free gauge theory with kinetic term $-(1/4g_0^2)\mathrm{tr}F^2$ in an effective action at a scale μ_0, with g_0 the coupling at that scale. For g_0 small enough we can calculate with arbitrary accuracy the renormalization group running of the coupling from the one loop result $8\pi^2 g^{-2}(\mu) \simeq -b_0 \log(|\Lambda|/\mu)$, where we have defined $|\Lambda| \equiv \mu_0 e^{-8\pi^2/b_0 g_0^2}$, the strong coupling scale of the gauge group. It is then convenient to introduce a complex "scale" $\Lambda \equiv |\Lambda| e^{i\theta/b_0}$ so that the complex coupling $\tau \equiv (\theta/2\pi) + i(4\pi/g^2) = (b_0/2\pi i)\log(\Lambda/\mu)$ at one loop.

(The coefficient of the one-loop beta function is given by

$$b_0 = \frac{11}{6}T(adj) - \frac{1}{3}\sum_a T(R_a) - \frac{1}{12}\sum_i T(R_i) \tag{34}$$

where the indices a run over Weyl fermions in representations R_a of the gauge group, and i runs over real scalars in the representations R_i. $T(R)$ is the index of the representation R; for $SU(n+1)$, for example, the index of the fundamental representation is 1, and of the adjoint representation is $2(n+1)$. For the $N=2$ Yang-Mills theory, all fields are in the adjoint representation, and we have one complex scalar and two Weyl fermions, thus giving $b_0 = T(adj) = 2(n+1)$.)

Far out on the on the Coulomb branch, where $a^K \gg \Lambda$, the $SU(n+1)$ theory is Higgsed to the $U(1)^n$ gauge group at a scale where the microscopic theory is very weakly coupled; see Fig. 3. Thus, the low-energy effective $U(1)^n$ couplings τ_{IJ} will be proportional to the running microscopic coupling at the scale of the $a^K \sim \langle a \rangle$ vevs:

$$\tau_{IJ} \sim \frac{b_0 C_{IJ}}{2\pi i} \log\left(\frac{\Lambda}{\langle a \rangle}\right) \tag{35}$$

for C_{IJ} some constant matrix that can be computed in perturbation theory. We see here the parameter Λ enters the IREA along with the vector multiplet scalar vevs, so we can treat Λ as if it were such a vev. In particular, Λ can only enter τ_{IJ} holomorphically. Furthermore, due to the angular nature of the theta angle, as $\tau \to \tau + 1$, or $\Lambda^{b_0} \to e^{2\pi i}\Lambda^{b_0}$, the physics must remain invariant. Thus $\tau_{IJ} = \tau_{IJ}(s_k, \Lambda^{b_0})$ is a holomorphic function of the s_k and Λ^{b_0} which matches on to (35) as $\Lambda \to 0$.

We can now derive a key fact about the Coulomb branch: though it can be corrected quantum mechanically, it is never lifted in asymptotically free $N = 2$ gauge theories. This is because there is a Coulomb branch for large adjoint scalar vevs where the asymptotically free gauge theory is Higgsed to $U(1)^n$ at arbitrarily weak coupling. Quantum corrections in the resulting $N = 2$ IREA cannot lift these flat directions since the only way (at weak coupling) to give mass to the $U(1)$ photons in the vector multiplets is by the Higgs mechanism; but there are no charged scalars in the vector multiplet. Thus it is not lifted for large enough s_k, and so by analytic continuation it cannot be lifted even for $s_k \sim \Lambda$ where perturbation theory is no longer valid. In general, complex manifolds like the Coulomb branch of $N = 2$ theories can become singular only on complex submanifolds, that is to say submanifolds at least 2 real dimensions smaller than the moduli space. Thus these singularities cannot be barriers preventing analytic continuation into a region of strong coupling.

The simplest example is the Coulomb branch of the $SU(2)$ Yang-Mills theory. The microscopic potential terms imply the equation $[\overline{\Phi}, \Phi] = 0$ for the complex adjoint scalar field, implying that Φ can be diagonalized by color rotations:

$$\Phi = \begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix}, \tag{36}$$

and there is a discrete gauge identification $a \simeq -a$. The gauge-invariant variable is

$$s_2 = -a^2 \equiv -\frac{1}{2}U, \tag{37}$$

where we have introduced the traditional name "U" for this Coulomb branch coordinate. It is easy to see that (36) leaves the diagonal $U(1) \subset SU(2)$ unbroken, and the light field U is neutral under this $U(1)$. We can thus think of the light degrees of freedom appearing in the IREA as those of an $N = 2$ $U(1)$ vector multiplet with complex scalar field U and a vector boson A_μ, as well as two Weyl fermions. The IREA can thus be written as

$$\mathcal{L} = \mathrm{Im}\, \tau(U, \Lambda) \left(\partial a(U) \cdot \partial \overline{a(U)} + \mathcal{F} \cdot \mathcal{F} \right), \tag{38}$$

where, by the arguments of the preceeding paragraphs, $a(U)$ is some holomorphic function of U and Λ^4, and the effective $U(1)$ gauge coupling will have the form

$$\tau(U) = \frac{1}{2\pi i} \log \left(\frac{\Lambda^4}{U^2} \right) + \sum_{n=0}^{\infty} c_n \left(\frac{\Lambda^4}{U^2} \right)^n . \tag{39}$$

The fact that only U^2 enters this formula follows from matching dimension with Λ^4, whose power follows from the coefficient of the one-loop beta function; it reflects a global $\mathbf{Z}_2$ symmetry acting on the Coulomb branch under $U \to -U$.

Solving for the vacuum structure of the $SU(2)$ theory is thus reduced to determining this function $\tau(U)$. It is worth examining the formula (39) in some detail. The first, logarithm, term came from matching to the one-loop running of the microscopic coupling for $U \gg \Lambda^2$. Because under theta-angle rotations, corresponding to 2π phase rotations of Λ^4, the physics must remain invariant, the low energy $\tau(U)$ can at most suffer an $Sp(2, \mathbf{Z}) \simeq SL(2, \mathbf{Z})$ electric-magnetic duality transformation. The terms included in (39) imply that $\tau(U) \to \tau(U) + 1$ under such a rotation, which is indeed in $SL(2, \mathbf{Z})$. Any other terms containing multiple logarithms, or any non-constant coefficient of the single logarithm term are not allowed, since they would necessarily imply $\tau(U)$ transformations under theta-angle rotations which are U-dependent, and therefore not in $SL(2, \mathbf{Z})$ since $SL(2, \mathbf{Z})$ is a discrete group of transformations. The absence of these higer logarithm terms is equivalent to the absence of all higher-loop corrections to the running of the microscopic coupling.

The terms proportional to Λ^{4n} correspond to a non-perturbative n-instanton contribution. Since the model is Higgsed for large U, the instantons have an effective IR cutoff at the scale U, so these instanton effects are calculable; the first two coefficients have been calculated [27]. In principle one could compute $\tau(U)$ by calculating all the n-instanton contributions, and then analytically continuing (39) to the whole U-plane; in practice this is too hard. Instead, we follow N. Seiberg and E. Witten's more physical approach to determining $\tau(U)$ [3].

There are two puzzles which indicate that we are missing some basic physics:

(1.) The effective coupling $\tau(U)$ is holomorphic, implying that $\mathrm{Re}\tau$ and $\mathrm{Im}\tau$ are harmonic functions on the U-plane. Since they are not constant functions, they therefore must be unbounded both above and below. In particular this implies that $\mathrm{Im}\tau = \frac{1}{g^2}$ will be negative for some U, and the effective theory will be non-unitary!

(2.) If we were to add a tree-level mass m for the complex scalar Φ (and one of the Weyl fermions as well to preserve an $N = 1$ supersymmetry), then, for $m \gg \Lambda$, Φ can be integrated out leaving a low-energy pure $SU(2)$ $N = 1$ super-YM theory with scale $\widehat{\Lambda}^6 = m^2\Lambda^4$. This theory has two vacua with mass gaps; in particular there are no massless photons. For nonzero $m \ll \Lambda$ by an $N = 1$ nonrenormalization argument one expects this qualitative behavior to persist. In that case our low-energy $N = 2$ theory on the U-plane should be approximately correct, and we should see some way to lift the degenerate vacua and create a mass gap. In particular we need to give the photon a mass, but there are no light charged degrees of freedom to Higgs the photon.

The next subsection will introduce the physical ingredient which resolves these puzzles and allows us to solve for $\tau(U)$.

4.1 Monopoles

The ingredient we need to be aware of is monopoles [28]. Monopoles can be constructed as finite-energy classical solutions of non-Abelian gauge theories spontaneously broken down to Abelian factors [29]. In particular they will occur in the $N = 2$ $SU(2)$ Yang-Mills theory. We illustrate this for simplicity in a (non-supersymmetric) $SU(2)$ theory broken down to $U(1)$ by a real adjoint Higgs:

$$\mathcal{L} = -\frac{1}{4g^2}F^a_{\mu\nu}F^{a\mu\nu} + \frac{1}{2}D^\mu\Phi^a D_\mu\Phi^a - V(\Phi) \tag{40}$$

where V has a minimum on the sphere in field space $\sum_a \Phi^a\Phi^a = v^2$. Different directions on this sphere are gauge-equivalent. In the vacuum $\langle\Phi^a\rangle$ lies on this sphere, Higgsing $SU(2) \to U(1)$ and giving a mass $m_W = gv$ to the $W^\pm$ gauge bosons. The unbroken $U(1)$ has coupling g, so satisfies Gauss's law $\vec{D}\cdot\vec{E} = g^2 j_e^0$, where j_e^μ is the electric current density. Thus the electric charge is computed as $Q_e = \frac{1}{g^2}\int_{S^2_\infty} \vec{E}\cdot d\vec{S}$. In the vacuum, the unbroken $U(1)$ is picked out by the direction of the Higgs vev, so $\vec{E} = \frac{1}{v}\Phi^a \vec{E}^a$. With this normalization of the electic charge, we find that the $W^\pm$ bosons have $Q_e = \pm 1$.

Static, finite-energy configurations must approach the vacuum at spatial infinity. Thus for a finite energy configuration the Higgs field Φ^a, evaluated as $r \to \infty$, provides a map from the S^2 at spatial infinity into the S^2 of the Higgs vacuum. Such maps are characterized by an integer, n_m, which measures the winding of one S^2 around the other. Mathematically, the second homotopy group of S^2 is the integers, $\pi_2(S^2) = \mathbf{Z}$. The winding, n_m, is the magnetic charge of the field configuration. To see this, the total energy from the Higgs field configuration:

$$\text{Energy} = \int d^3x \tfrac{1}{2} D_\mu \Phi^a D^\mu \Phi^a + V(\Phi) \geq \int d^3x \tfrac{1}{2} D_\mu \Phi^a D^\mu \Phi^a. \tag{41}$$

To have finite energy configurations we must therefore ensure that the covariant derivative of Φ^a falls off faster than $1/r$ at infinity. The general solution for the gauge field consistent with this behavior is

$$A^a_\mu \sim -\frac{1}{v^2}\epsilon^{abc}\Phi^b \partial_\mu \Phi^c + \frac{1}{v}\Phi^a A_\mu \tag{42}$$

with A_μ arbitrary. The leading-order behavior of the field strength is then

$$F^{a\mu\nu} = \frac{1}{v}\Phi^a F^{\mu\nu} \tag{43}$$

with

$$F^{\mu\nu} = -\frac{1}{v^3}\epsilon^{abc}\Phi^a \partial^\mu \Phi^b \partial^\nu \Phi^c + \partial^\mu A^\nu - \partial^\nu A^\mu \tag{44}$$

and the equations of motion imply $\partial_\mu F^{\mu\nu} = \partial_\mu * F^{\mu\nu} = 0$. Thus we learn that outside the core of the monopole the non-Abelian gauge field is purely in the direction of Φ^a, that is the direction of the unbroken $U(1)$. The magnetic charge of this field configuration is then computed to be

$$Q_m = \int_{S^2_\infty} \vec{B}\cdot d\vec{S} = \frac{1}{2v^3}\int_{S^2_\infty} \epsilon^{ijk}\epsilon^{abc}\Phi^a \partial^j \Phi^b \partial^k \Phi^c dS^i = 4\pi n_m \tag{45}$$

where n_m is the winding number of the Higgs field configuration, recovering the Dirac quantization condition.[1]

[1] This is actually the Dirac quantization condition only for even values of n_m since in this theory we could add fields in the fundamental **2** representation of $SU(2)$, which would carry electric charge $Q_e = \pm 1/2$.

Note that for such non-singular field configurations, the electric and magnetic charges can be rewritten as

$$\begin{aligned} Q_e &= \frac{1}{g^2}\int_{S^2_\infty} \vec{E}\cdot d\vec{S} = \frac{1}{g^2 v}\int_{S^2_\infty} \Phi^a \vec{E}^a\cdot d\vec{S} = \frac{1}{g^2 v}\int d^3x \vec{E}^a\cdot(\vec{D}\Phi)^a \\ Q_m &= \int_{S^2_\infty} \vec{B}\cdot d\vec{S} = \frac{1}{v}\int_{S^2_\infty} \Phi^a \vec{B}^a\cdot d\vec{S} = \frac{1}{v}\int d^3x \vec{B}^a\cdot(\vec{D}\Phi)^a \end{aligned} \tag{46}$$

using the vacuum equation of motion and the Bianchi identity $\vec{D}\cdot\vec{E}^a = \vec{D}\cdot\vec{B}^a = 0$ and integration by parts.

If we consider a static configuration with vanishing electric field the energy (mass) of the configuration is given by

$$\begin{aligned} m_M &= \int d^3x \left(\frac{1}{2g^2}\vec{B}^a\cdot\vec{B}^a + \frac{1}{2}\vec{D}\Phi^a\cdot\vec{D}\Phi^a + V(\Phi)\right) \geq \int d^3x \left(\frac{1}{2g^2}\vec{B}^a\cdot\vec{B}^a + \frac{1}{2}\vec{D}\Phi^a\cdot\vec{D}\Phi^a\right) \\ &= \frac{1}{2}\int d^3x \left(\frac{1}{g}\vec{B}^a - \vec{D}\Phi^a\right)^2 + \frac{vQ_m}{g}, \end{aligned} \tag{47}$$

giving the BPS bound

$$m_M \geq \left|\frac{vQ_m}{g}\right|. \tag{48}$$

This semi-classical bound can be extended to *dyons* (solitonic states carrying both electric and magnetic charges):

$$m_D \geq gv\left|Q_e + i\frac{Q_m}{g^2}\right|. \tag{49}$$

A theta angle has a non-trivial effect in the presence of magnetic monopoles: it shifts the allowed values of electric charge in the monopole sector of the theory [30]. To see this, consider gauge transformations, constant at infinity, which are rotations in the $U(1)$ subgroup of $SU(2)$ picked out by the Higgs vev, that is, rotations in $SU(2)$ about the axis $\hat{\Phi}^a = \Phi^a/|\Phi^a|$. The action of such an infinitesimal gauge transformation on the field is

$$\delta A^a_\mu = \frac{1}{v}(D_\mu\Phi)^a \tag{50}$$

with Φ the background monopole Higgs field. Let $\mathcal{N}$ denote the generator of this gauge transformation. Then if we rotate by 2π about the $\hat{\Phi}$ axis we must get the identity

$$e^{2\pi i\mathcal{N}} = 1. \tag{51}$$

Including the θ term, it is straightforward to compute $\mathcal{N}$ using the Noether method,

$$\mathcal{N} = \frac{\partial\mathcal{L}}{\partial\partial_0 A^a_\mu}\delta A^a_\mu = Q_e - \frac{\theta Q_m}{8\pi^2}, \tag{52}$$

where we have used the definitions (46) of the electric and magnetic charge operators. This result implies

$$Q_e = n_e + n_m \frac{\theta}{2\pi} \tag{53}$$

where n_e is an arbitrary integer and $n_m = Q_m/4\pi$ determines the magnetic charge of the monopole. We will henceforth label dyons by the integers (n_e, n_m). Note that the BPS bound becomes

$$M_D \geq gv \left| \left(n_e + n_m \frac{\theta}{2\pi} \right) + i n_m \frac{4\pi}{g^2} \right| = gv|n_e + \tau n_m|. \tag{54}$$

This result is classical; quantum mechanically, the coupling τ runs, and gv and $g\tau$ will be replaced by functions of the strong coupling scale Λ and the vevs. In theories with extended supersymmetry the (quantum-corrected) BPS bound can be computed exactly, and states saturating the bound can be identified [20]. For example, in the $N = 2$ $SU(2)$ theory the BPS mass formula becomes [3]

$$M_D = |a(U)n_e + b(U)n_m|, \tag{55}$$

where a and b are holomorphic functions of U and Λ^4 satisfying

$$\frac{\partial b(U)}{\partial a(U)} = \tau(U), \tag{56}$$

with $a(U)$ the same function as appeared in the IREA (38).[2]

4.2 Solution to the SU(2) Theory

Returning to the $N = 2$ $SU(2)$ Yang-Mills theory, we have learned that this theory can have magnetic monopoles. Indeed, one can show that there are BPS solitons with charges $(n_e, n_m) = (0, \pm 1)$ in this theory, and they turn out to lie in hypermultiplets of the supersymmetry algebra. Furthermore, from (39) we see that changing the phase of U shifts the effective theta angle. In particular under the global $\mathbf{Z}_2$: $U \to e^{i\pi}U$, $\tau \to \tau - 1$. From the associated duality transformation on the charges of any massive states (53), we see that there will be $(\mp 1, \pm 1)$ dyons in the spectrum. Repeating this procedure, we find there must be a whole tower of semi-classically stable dyons of charges $(n, \pm 1)$ for arbitrary integers n.

The existence of these dyon states suggests a possible resolution to one of our puzzles: perhaps at some strong coupling point on the moduli space, for example $U = U_0$ with

[2] $b(U)$ is often called $a_D(U)$ in the literature.

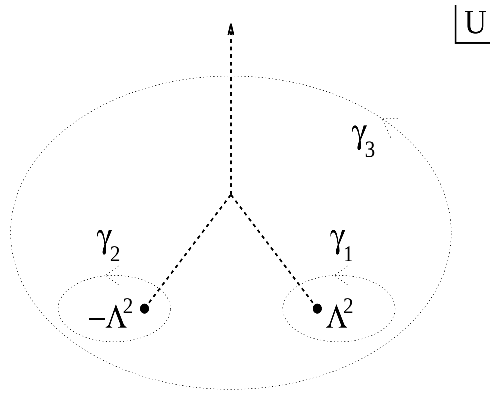

Figure 4: Cut U-plane with three loops. The cuts have been placed in an arbitrary manner connecting the two possible strong-coupling singularities, and a possible singularity at weak coupling ($U = \infty$).

$U_0 \sim \Lambda^2$, one of these dyons becomes massless, thereby providing the light charged scalar fields needed to Higgs the $U(1)$. Since we expect to recover the two gapped vacua of the $N = 1$ $SU(2)$ super-YM theory, and recalling the $\mathbf{Z}_2$ symmetry of the theory, it is natural to assume that there are two points on the U-plane where charged fields become massless, and they are at $U = \pm U_0$. Since Λ is the only scale in the theory, we take $U_0 = \Lambda^2$. (We can take this as the definition of our normalization of Λ, if we like.)

We can check this assumption by examining the behavior of τ as a function of U. Recall the other puzzle we had about the physics on the Coulomb branch: since $\tau(U)$ is holomorphic, $1/g^2 \sim \mathrm{Im}\tau$ is harmonic and therefore unbounded from below, violating unitarity.

This puzzle is resolved by noting that τ is not, in fact, a holomorphic function of U. In particular, by electric-magnetic duality, as we traverse closed loops in the U-plane, τ need not come back to the same value, only one related to it by an $SL(2, \mathbf{Z})$ transformation. Mathematically, this is described by saying that τ is a section of a flat $SL(2, \mathbf{Z})$ bundle. This multi-valuedness of τ can be described by saying that τ is a holomorphic function on a cut U-plane, with cuts emanating from some singularities, and with the jump in τ across the cuts being an element of $SL(2, \mathbf{Z})$. The two points $U = \pm\Lambda^2$ at which we are assuming there are massless charged fields are the natural candidates for the branch points, see Fig. 4. The presence of these cuts allows us to avoid the conclusion that $\mathrm{Im}\tau$ is unbounded.

Upon traversing the various loops γ_i in the above figure, τ will change by the action of an $SL(2,\mathbf{Z})$ element. These elements are called the *monodromies* of τ, and will be denoted $\mathcal{M}_i$.

We first calculate $\mathcal{M}_3$, the monodromy around the weak-coupling singularity at infinity. By taking γ_3 of large enough radius, τ will be accurately given by its one-loop value, the first term in (39). Taking $U \to e^{2\pi i}U$ in this formula gives $\tau \to \tau - 2$, giving for the monodromy at infinity[3]

$$\mathcal{M}_3 = \begin{pmatrix} -1 & 2 \\ 0 & -1 \end{pmatrix}. \tag{57}$$

In order to calculate the $\mathcal{M}_{1,2}$ monodromies, let us first calculate the monodromy we would expect if the field becoming massless at the associated singularity had charge (n_e, n_m). By a duality transformation we can change to a basis where this charge is purely electric: $(\widetilde{n}_e, 0)$. In this basis the physics near the $U = U_0$ singularity is just that of QED with the electron becoming massless. This theory is IR free, so the behavior of the low-energy effective coupling will be dominated by its one-loop expression, at least sufficiently near U_0 where the mass of the charged field $\sim U - U_0$ is arbitrarily small:

$$\widetilde{\tau} = \frac{\widetilde{n}_e^2}{\pi i} \log(U - U_0) + \mathcal{O}(U - U_0)^0. \tag{58}$$

By traversing a small loop around U_0, $(U - U_0) \to e^{2\pi i}(U - U_0)$, we find the monodromy

$$\widetilde{\tau} \to \widetilde{\tau} + 2\widetilde{n}_e^2 \qquad \Longrightarrow \qquad \widetilde{\mathcal{M}} = \begin{pmatrix} 1 & 2\widetilde{n}_e^2 \\ 0 & 1 \end{pmatrix}. \tag{59}$$

Now let us duality-transform this answer back to the basis where the charges are (n_e, n_m). The required $SL(2,\mathbf{Z})$ element will be denoted $\mathcal{N} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, and satisfies

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} n_e \\ n_m \end{pmatrix} = \begin{pmatrix} \widetilde{n}_e \\ 0 \end{pmatrix}, \quad \text{and} \quad ad - bc = 1 \quad \text{with} \quad a, b, c, d \in \mathbf{Z}. \tag{60}$$

The transformed monodromy is then

$$\mathcal{M} = \mathcal{N}\widetilde{\mathcal{M}}\mathcal{N}^{-1} = \begin{pmatrix} 1 + 2n_e n_m & 2n_e^2 \\ -2n_m^2 & 1 - 2n_e n_m \end{pmatrix}. \tag{61}$$

Now, by deforming the γ_i contours in the U-plane, we find that the three monodromies must be related by

$$\mathcal{M}_3 = \mathcal{M}_1 \mathcal{M}_2. \tag{62}$$

[3]This actually only determines the monodromy up to an overall sign. The sign is determined by noting that $U \to e^{2\pi i}U$ has the effect of $\Phi \to -\Phi$ on the elementary Higgs field, so it reverses the sign of the low-energy electromagnetic field which in terms of $SU(2)$ variables is proportional to $\mathrm{tr}(\Phi F)$. Thus it reverses the sign of electric and magnetic charges, giving an "extra" factor of $-\mathbf{1} \in SL(2,\mathbf{Z})$.

Assuming that a field with charges (n_{e1}, n_{m1}) becomes massless at $U = \Lambda^2$, while one with charges (n_{e2}, n_{m2}) does so at $U = -\Lambda^2$, and substituting into (62) using (57) and (61) gives as solutions

$$(n_{e1}, n_{m1}) = \pm(n, 1), \qquad (n_{e2}, n_{m2}) = \pm(n-1, 1), \qquad \text{for all} \quad n \in \mathbf{Z}. \tag{63}$$

This set of charges actually represents a single physical solution. This is because taking $U \to e^{i\pi}U$ takes us to an equivalent theory by the $\mathbf{Z}_2$ symmetry; but this corresponds to shifting the low-energy theta-angle by 2π which in turn shifts all dyon electric charges by their magnetic charges. Repeated applications of this shift can take any of the above solutions to the solution

$$(n_{e1}, n_{m1}) = \pm(0, 1), \qquad (n_{e2}, n_{m2}) = \pm(-1, 1). \tag{64}$$

The plus and minus sign solutions must both be there by anomaly cancellation in the low-energy $U(1)$. We thus learn that there is a consistent solution with a monopole becoming massless at $U = \Lambda^2$ and a charge $(-1, 1)$ dyon becoming massless at $U = -\Lambda^2$. Some progress has been made in weakening the initial assumption that there are just two strong-coupling singularities [31].

With the monodromies around the singularities in hand, we now turn to finding the low-energy coupling τ on the U-plane. The basic idea is that τ is determined by holomorphy and demanding that it match onto the behavior we have determined above at $U = \infty$ and $U = \pm\Lambda^2$. Seeing how to solve this "analytic continuation" problem analytically is not obvious, however. Seiberg and Witten did it by introducing an auxiliary mathematical object: a family of tori varying over the Coulomb branch.

This is a useful construction because the low-energy effective coupling τ has the same properties as the complex structure of a 2-torus. In particular, the complex structure of a torus can be described by its *modulus*, a complex number τ, with $\mathrm{Im}\tau > 0$. In this description, the torus can be thought of as a parallelogram in the complex plane with opposite sides identified, see Fig. 5. Furthermore, the modulus τ of such a torus gives equivalent complex structures modulo $SL(2, \mathbf{Z})$ transformations acting on τ. Therefore, if we associate to each point in the U-plane a holomorphically-varying torus, its modulus will automatically be a holomorphic section of an $SL(2, \mathbf{Z})$ bundle with positive imaginary part, which are just the properties we want for the effective coupling τ.

At $U = \pm\Lambda^2$, magnetically charged states become massless, implying that the effective coupling $\mathrm{Im}\tau \to 0$. (Recall that by $U(1)$ IR freedom, when an electrically charged

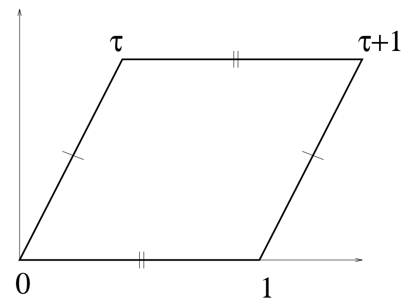

Figure 5: A complex torus as a parallelogram in the complex plane with opposite sides identified.

state becomes massless, the coupling $g \to 0$, implying $\tau \to +i\infty$. Doing the duality transform $\tau \to -1/\tau$ gives the above result for a magnetic charge becoming massless.) From the parallelogram, we see this implies that the torus is degenerating: one of its cycles is vanishing.

Now, a general torus can be described analytically as the Riemann surface which is the solution $y(x)$ to the complex cubic equation

$$y^2 = (x - e_1)(x - e_2)(x - e_3). \tag{65}$$

We can think of this as a double-sheeted cover of the x-plane, branched over the three points e_i and the point at infinity. We let this torus vary over the U-plane by letting the e_i vary: $e_i = e_i(U, \Lambda)$. By choosing the cuts to run between pairs of these branch points, and "gluing" the two sheets together along these cuts, one sees that the Riemann surface is indeed topologically a torus. Furthermore, the condition for a nontrivial cycle on this torus to vanish is that two of the branch points collide. Since we want this to happen at the two points $U = \pm\Lambda^2$, it is natural to choose $e_1 = \Lambda^2$, $e_2 = -\Lambda^2$, and $e_3 = U$:

$$y^2 = (x - \Lambda^2)(x + \Lambda^2)(x - U). \tag{66}$$

Note that this choice has a manifest $U \to -U$ symmetry, under which $x \to -x$ and $y \to \pm iy$.

Given this family of tori, one can compute their moduli as a ratio of line integrals:

$$\tau(U) = \frac{\oint_\beta \omega}{\oint_\alpha \omega}, \tag{67}$$

where ω is the (unique) holomorphic one-form on the Riemann surface,

$$\omega = \frac{dx}{y} = \frac{dx}{\sqrt{(x^2 - \Lambda^4)(x - U)}}, \tag{68}$$

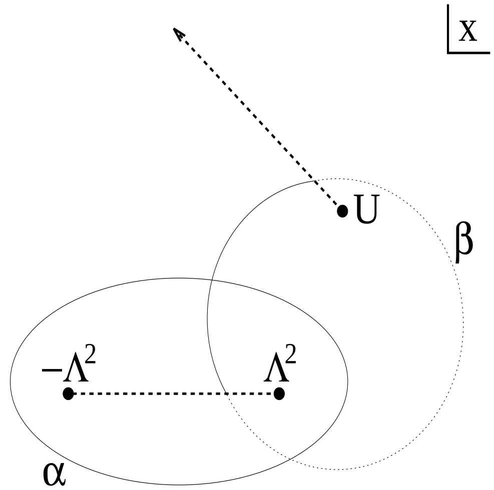

Figure 6: Cut x-plane with α and β cycles.

and α and β are any two non-trivial cycles on the torus which intersect once. For example, we might take α to be a cycle on the x-plane which loops around the branch points at $\pm\Lambda^2$, while β is the one which loops around the branch points at Λ^2 and U. If we chose the cuts on the x-plane to run between $\pm\Lambda^2$ and between U and ∞, then the α cycle would lie all on one sheet, while the β cycle would go onto the second sheet as it passes through the cut, see Fig. 6. Since the integrand in (68) is a closed one form ($d\omega = 0$), the value of τ does not depend on the exact locations of α and β, but only on how they loop around the branch points.

We can now check that our family of tori (66) indeed gives rise to the correct low-energy τ. By taking $U \to \infty$, it is not hard to explicitly evaluate (68) to find agreement with the first term in the weak-coupling expansion (39).[4] Also, without having to explicitly evaluate the integrals in (68), one can check that it reproduces the correct monodromies as U goes around the singularities at $\pm\Lambda^2$ by tracking how the α and β cycles are deformed as U varies. Finally, it turns out that the family of tori (66) is the unique one with these properties [3].

5 Geometry of N=2 Coulomb Branches

We would like to generalize the above arguments to other gauge groups and matter representations. To state the problem clearly:

[4]Though perhaps only up to an $SL(2, \mathbf{Z})$ transformation if I made the "wrong" choice for my α and β cycles.

Given: the field theory data specifying an $N = 2$ supersymmetric gauge theory, namely a gauge group G (not necessarily simple), a matter (hypermultiplet) representation R (not necessarily irreducible), bare masses m for the matter, and UV coupling constant(s) τ or strong coupling scales Λ for the vector multiplets,

Find: the $N = 2$ IREA on the Coulomb branch, namely the $U(1)^n$ couplings τ_{IJ} and the "special coordinates" a^I as functions of the microscopic field theory data and the gauge-invariant coordinates on the Coulomb branch.

We have not emphasized the *special coordinates* above, so let us define them here. Recall from Section 2 that the $N = 2$ IREA on the Coulomb branch has the form

$$\mathcal{L} = \mathrm{Im}\tau_{IJ} \left(\partial a^I \cdot \partial \overline{a}^J + \mathcal{F}^I \cdot \mathcal{F}^J \right), \tag{69}$$

with τ_{IJ} satisfying the conditions

$$\partial_{[I}\tau_{J]K} = 0, \tag{70}$$

where $\partial_I = \partial/\partial a^I$, and τ_{IJ} a holomorphic $Sp(2n, \mathbf{Z})$ section of the vector multiplet scalars a^I. Clearly we could make a non-singular field redefinition on the scalars, effectively changing the coordinates we use to describe the Coulomb branch, and changing the form of the IREA (69). The choice of scalar fields such that the IREA has the above form where τ_{IJ} plays the role of both the Coulomb branch metric and the $U(1)^n$ effective couplings, are called special coordinates. In general the special coordinates can become singular, as they do at the monopole and dyon points in the $SU(2)$ example, so it is useful to choose well-behaved global coordinates on the Coulomb branch—the s_n. At weak coupling the special coordinates and the global coordinates are related by (33), but at strong coupling no such simple relation need exist. The special coordinates also appear in the BPS mass formula

$$M = |n_{e,I}a^I + n_m^J b_J|, \tag{71}$$

where the b_J are defined by

$$\partial_I b_J = \tau_{IJ}, \tag{72}$$

and exist by virtue of the integrability condition for this equation, (70).

This problem of determining the Coulomb branch IREA given the UV field theory data has not been solved, though many infinite series of solutions are known. Most of the known solutions were found essentially by (educated) guessing. In section 6 we will discuss one method which, although it is not known how to use it to solve the general problem, permits a *derivation* of the solutions when it works. In order to get to the point where we can discuss this method, we first need to reformulate the geometry of the vector multiplet manifolds; this is of interest also for the light it sheds on the general problem.

The Coulomb branch moduli space of the $N = 2$ IREA (69) satisfying condition (70) and the $Sp(2n, \mathbf{Z})$ properties of τ_{IJ} defines a *rigid special Kähler* (RSK) manifold [32]. Abstracting away from the IREA, we can thus define an RSK manifold as an n-complex-dimensional manifold $\mathcal{M}$ with certain properties. Choose some global complex coordinates s_K, $K = 1, \ldots, n$ on $\mathcal{M}$.[5] Then an RSK manifold has "special coordinates" $a^I(s_K)$, $I = 1, \ldots, n$, which are local holomorphic coordinates almost everywhere on $\mathcal{M}$, and a symmetric, holomorphic section τ_{IJ} of an $Sp(2n, \mathbf{Z})$ bundle on $\mathcal{M}$,[6] such that the metric in special coordinates is $g_{I\overline{J}} = \mathrm{Im}\tau_{IJ}$ and $\partial_{[I}\tau_{J]K} = 0$, where the derivative is with respect to the special coordinates. Note that $\mathrm{Im}\tau_{IJ}$ must be positive definite for the metric to be non-singular.

Several properties of RSK manifolds can immediately be deduced from this definition. The first is the existence of the "dual" special coordinates b_I, satisfying (72). Then, defining $\mathcal{K} = i(a^I\overline{b_I} - \overline{a^I}b_I)$, it is easy to check that $g_{I\overline{J}} = \partial_I\overline{\partial_J}\mathcal{K}$, which is the defining condition for a Kähler manifold. Defining the $2n$-component column vector $\mathbf{c}$ by

$$\mathbf{c} = \begin{pmatrix} b_I \\ a^I \end{pmatrix}, \tag{73}$$

the expression for the Kähler potential can be written compactly as $\mathcal{K} = \langle \mathbf{c}, \overline{\mathbf{c}} \rangle$, where the brackets denote the symplectic inner product

$$\langle \mathbf{c}, \mathbf{d} \rangle = \mathbf{c}^T \cdot J \cdot \mathbf{d} \qquad \text{with} \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{74}$$

Under transformations $M \in Sp(2n, \mathbf{Z})$ it is not hard to see that $\mathbf{c}$ transforms in the $2n$-dimensional representation:[7]

$$\mathbf{c} \to M \cdot \mathbf{c}, \tag{75}$$

and so the special coordinates are really part of a holomorphic $Sp(2n, \mathbf{Z})$ bundle $\mathcal{M}$ in the fundamental representation.

We will now describe three reformulations of RSK geometry. The first will be to show that RSK geometry is equivalent to having a family of algebraic varieties varying

[5] Or a patch of $\mathcal{M}$; this definition can be applied patch by patch to an atlas covering $\mathcal{M}$.

[6] With the usual action (6) on τ_{IJ}.

[7] Actually, the $\mathbf{c}$'s can in general also shift by constants under electric-magnetic duality transformations, which is important when there are hypermultiplet masses in the problem [22].

holomorphically with the s_K along with some extra structures; these are in turn equivalent to algebraically completely integrable Hamiltonian systems [33]. The second will be to show that a class of RSK manifolds are described by n-complex-dimensional families of Riemann surfaces of genus n with certain meromorphic one-forms [3]. The third will be to show that at least a subset of the RSK manifolds described in the second way can also be described by families of Riemann surfaces embedded in hyperKähler manifolds [4].

Though it is not known whether all RSK manifolds can be described in the second or third ways, in fact all RSK manifolds that have been found as Coulomb branches of $N = 2$ gauge theories do fall into the third category.

5.1 RSK and Families of Abelian Varieties

A straightforward generalization of the complex torus construction introduced in our discussion of the $SU(2)$ theory where the Coulomb branch was one-complex-dimensional to the case where the Coulomb branch is n-complex-dimensional, is to think of τ_{ij} as specifying the complex structure of an n-complex-dimensional torus [33]. Such a torus is specified by $2n$ linearly independent vectors forming the basis of a lattice Γ in $\mathbf{C}^n$. so the torus is $T^{2n} = \mathbf{C}^n/\Gamma$. Global linear complex changes of variables on $\mathbf{C}^n$ do not change the complex structure of T^{2n}, and can be used to set half of the basis vectors of Γ to real unit vectors. Thus the complex structure of T^{2n} is encoded in the $n \times n$ complex matrix, τ_{ij}, of coordinates of the remaining n basis vectors. It is easy to check that this τ_{ij} is really only defined up to $GL(2n, \mathbf{Z})$ fractional linear transformations reflecting the ability to choose a different set of n lattice vectors to set to the real unit vectors.

The τ_{ij}'s describing RSK geometry have four further constraints, however: they are symmetric, have positive definite imaginary part, are a section of an $Sp(2n, \mathbf{Z})$ bundle, and satisfy the integrability condition (70). The third constraint can be encoded in the geometry of T^{2n} by introducing an extra structure, a *polarization*, which is a non-degenerate $(1, 1)$-form t on T^{2n} with integral periods, and can be thought of as defining a symplectic inner product on the periods of 1-cycles on the torus as in (74). Complex tori obeying the first three conditions are known as *Abelian varieties*, which are essentially tori that can be described by algebraic equations involving generalized theta functions.

The fourth condition can be incorporated as the additional structure of a meromorphic 1-form, λ, on T^{2n} with the property that

$$\frac{\partial}{\partial s_K}\lambda = \omega^K, \tag{76}$$

up to total derivatives, where ω^K are a basis of n holomorphic one-forms on T^{2n}. This is related to τ_{ij} as follows. Choose a symplectic (or canonical) homology basis of one-cycles on T^{2n}. This is a basis of $2n$ one-cycles $\{\beta_I, \alpha^J\}$ such that

$$\int_{\alpha^I\wedge\alpha^J} t = \int_{\beta^I\wedge\beta^J} t = 0, \qquad \int_{\alpha^I\wedge\beta^J} t = \delta^I_J, \tag{77}$$

where the "wedge product" of one-cycles refers to the two-cycle spanned by them. Then the periods of λ (the integrals of λ over this basis of one-cycles) is the $2n$-component vector $\mathbf{c}$ introduced in (73).[8]

This reformulation of RSK geometry as complex manifolds with a family of Abelian varieties with meromorphic one form is quite general. Furthermore, the exterior derivative of the one-form on the total space of the RSK manifold plus its T^{2n} fibers is a symplectic two-form of a complex integrable system [33]. This equivalence of RSK geometry to integrable systems has led to the solution of many $N = 2$ IREAs [34, 33, 35]. However the procedure essentially involves matching an integrable system to the appropriate $N = 2$ field theory data, and no systematic way is known to do this matching.

5.2 RSK and Families of Riemann Surfaces

More systematic control over the construction of RSK geometries is obtained by specializing to classes of RSK manifolds whose geometry can be naturally encoded in simpler structures. One such specialization is to RSK manifolds whose associated Abelian variety T^{2n} can be realized as the Jacobian variety of a genus-n Riemann surface. For $n \geq 4$, these varieties form a subset of measure zero in the space of all Abelian varieties. Whether all families of Abelian varieties admitting the existence of an appropriate meromorphic one-form (to describe RSK geometry) are actually Jacobian varieties is an open question. In any case, to date all known constructions of RSK geometry are in terms of families of Jacobian varieties.

The connection between genus-n Riemann surfaces, Σ_n, and Jacobian varieties,

[8] For these periods to depend only on the homology class of the cycles, the one-form λ must have vanishing residues. Actually, λ's with non-vanishing residues are allowed, and are interpreted physically as bare hypermultiplet masses [22]. In what follows we will assume zero bare masses.

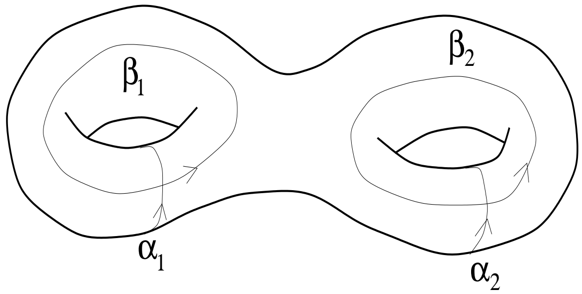

Figure 7: A genus 2 Riemann surface with a canonical homology basis of α and β cycles.

$Jac(\Sigma_n) \simeq T^{2n}$, is through the Jacobian map,

$$P \in \Sigma_n \longrightarrow \left\{ \int_{P_0}^{P} \omega^1, \ldots, \int_{P_0}^{P} \omega^n \right\} \quad \text{mod periods,} \tag{78}$$

where P_0 is some argbitrary fixed base point and ω^K is a basis of the n holomorphic one-forms on Σ_n [36]. Under this map one-cycles on Σ_n are pushed forward to one-cycles on T^{2n}, the symplectic inner product (polarization) two-form t is pulled back to the intersection form on Σ_n, and the basis of holomorphic one forms ω^K and the meromorphic one-form λ on T^{2n} are pulled back to one-forms on Σ_n (which we call by the same names).

Thus τ_{IJ} is just the period matrix of the Riemann surface, and is given by

$$\tau_{IJ} = \left(\int_{\beta_I} \omega^K \right) \left(\int_{\alpha^J} \omega^K \right)^{-1} \tag{79}$$

where the second factor is to be interpreted as a matrix inverse on the JK indices and K is to be summed over. The symmetry and positive-definiteness conditions on τ_{IJ} follow from the Riemann bilinear relations, while the $Sp(2n, \mathbf{Z})$ structure follows from the intersection form on Riemann surfaces. In particular, one can always choose a canonical homology basis of $2n$ one-cycles $\{\alpha^I, \beta_J\}$ such that their intersections obey $\alpha^I \cdot \alpha^J = \beta_I \cdot \beta_J = 0$ and $\alpha^I \cdot \beta_J = \delta^I_J$; see Fig. 7.

To summarize, we have encoded the RSK geometry of an n complex-dimensional Coulomb branch in a family of genus-n Riemann surfaces varying holomorphically over the Coulomb branch and endowed with a meromorphic one-form λ satisfying (76). This formulation has been used to solve for many $N = 2$ IREAs essentially by guessing a form for the family of Riemann surfaces and matching to $N = 2$ field theory data [37, 23, 24]. Again, as with the integrable system formulation, this matching procedure has not been made systematic.

5.3 RSK and Riemann Surfaces in HyperKähler Manifolds

We now turn to one further reformulation of RSK geometry which, when combined with some string theory ideas, has allowed a more (though not completely) systematic approach.

The previous encoding of RSK geometry in a family of Riemann surfaces failed to "geometrize" the meromorphic one-form λ. This geometrization can be performed as follows [4, 38]. Suppose the family of Riemann surfaces Σ_n can be embedded in a fixed (independent of the s_K) hyperKähler 4-manifold Q. Now a hyperKähler manifold is a manifold Kähler with respect to three complex structures, I, J, and K, satisfying the quaternion algebra

$$I^2 = J^2 = K^2 = -1, \qquad IJ = -JI = K, \quad \text{and cyclic permutations.} \tag{80}$$

Each complex structure can be thought of either as a rank-2 tensor acting on the tangent space to the manifold, *e.g.* $I = I^i_j$, or, using the metric to lower one of the indices, as an antisymmetric 2-tensor (a 2-form) on Q. Furthermore, the Kähler condition implies that these 2-forms are closed. Then, with respect to the complex structure I, $\omega \equiv J + iK$ is a closed holomorphic 2-form. Thus locally $\omega = d\lambda$ and λ pulls back to a meromorphic one-form on Σ_n. Because the family of Σ_n obtained as we vary the s_K are all embedded holomorphically in the fixed manifold Q, the RSK condition (76) on λ is automatically satisfied.

RSK manifolds which can be described in this way are clearly a subset of those that can be described just in terms of those described in terms of a family of Reimann surfaces and a one-form λ. But this restricted class has the great advantage that everything appears geometrically, requiring only a choice of a fixed "background" hyperKähler 4-manifold Q.

6 M-theory 5-Brane Construction of the Coulomb Branch

In this section we will outline the construction of solutions to $N = 2$ IREAs (Coulomb branch geometries) using the encoding of RSK geometry in a holomorphically varying family of Riemann surfaces Σ_n embedded in a hyperKähler four-manifold Q, following [4]. First we will outline an argument using the M-theory/IIA string theory equivalence to identify Q and some gross topological properties of the embedding corresponding to some $N = 2$ field theory data. Then we will show, in an example, how easy it is to solve for the specific family of embedded surfaces given this data, thus solving for the $N = 2$ IREA on the Coulomb branch of $SU(n+1)$ Yang-Mills theory.

6.1 5-Branes in M-Theory

The basic idea [4] is to interpret the geometrical objects Q and Σ_n as physical objects in a supergravity theory such that at energies far below the Planck scale where the gravity decouples we are left with an $N = 2$ field theory.

Choose the supergravity theory to be the unique 11-dimensional supergravity theory, which is the low energy effective theory of M-theory [39]. This theory has 32 supercharges, corresponding to $N = 8$ supergravity in four dimensions. A consistent background to this theory is $\mathbf{R}^{6+1} \times Q$; if Q is hyperKähler this background breaks half the supersymmetries to $N = 4$ in four dimensions.

Now M theory has 5-branes, which are excitations of the theory extended in $5 + 1$ dimensions. On length scales much larger than the Planck scale of this theory, we can think of the 5-brane as a mathematical 6-manifold embedded in the 11-dimensional space-time. The 5-brane has field theory degrees of freedom which are constrained to propagate only on the brane. Furthermore, as long as the 5-brane is holomorphically embedded in the background 11-dimensional space, a solution to the supergravity equations of motion are obtained with only half of the remaining supersymmetries broken. Thus the 5-brane (six-dimensional) field theory has 8 conserved supercharges, corresponding to $N = 2$ supersymmetry in four dimensions. In the limit that curvature length scales of the brane are much greater than the Planck scale, the "bulk" 11-dimensional supergravity degrees of freedom decouple, leaving a unitary six-dimensional field theory on the brane.

The final step in this M-theory construction is to interpret the embedded Riemann surface Σ_n as part of the 5-brane world-volume. In particular, take the 5-brane world-volume to be the manifold $\mathbf{R}^{3+1} \times \Sigma_n$ with $\mathbf{R}^{3+1} \subset \mathbf{R}^{6+1}$ and $\Sigma_n \subset Q$. Then on distance scales large compared to the size of Σ_n the brane field theory is effectively an $N = 2$ four-dimensional field theory.

Thus we have incorporated all the mathematical ingredients needed to describe the RSK Coulomb branch geometry together with the associated physical degrees of freedom (the four-dimensional field theory) in a single supergravity configuration.

6.2 IIA/M-Theory Duality

To use this to solve a concrete $N = 2$ theory, we need to choose the background hyperKähler manifold Q. The simplest choice would be a flat four-manifold $Q = \mathbf{R}^4$, but this does not work. To see this, pick some complex structure on Q, say coordinates $v = x_1 + ix_2$ and $s = x_3 + ix_4$. Then any holomorphically embedded Riemann surface Σ_n will be described by some complex analytic equation in s and v: $F(s, v) = 0$. But, by the properties of analytic maps, the surface described by such an equation cannot be compact—it must extend to infinity in Q. Thus it would seem that we are really describing in this way a six-dimensional field theory on some curved background.

The key to connecting this construction to a four-dimensional interpretation is to use the equivalence of the (ten-dimensional) type IIA string theory to M theory compactified on a circle. In this equivalence, the string coupling g_s of the IIA theory is related to the radius of the compactified circle, R, by

$$g_s = (R/\ell_p)^{3/2}, \tag{81}$$

where ℓ_p is the 11-dimensional Planck length, and the defining string scale ℓ_s (related to the fundamental string tension) satisfies

$$g_s \ell_s = R. \tag{82}$$

Thus when the 11-dimensional supergravity description is good, that is when $\ell_p \ll R$, we have $g_s \gg 1$, so the string description is strongly coupled, and *vice versa*.

The connection to a four-dimensional description comes from taking the limit as the compactification radius $R \to 0$, so that the ten-dimensional string description becomes weakly coupled. In that limit the 5-brane reduces to either a 4-brane or a 5-brane in the ten-dimensional theory, depending on whether the M-theory 5-brane is or is not wrapped around the shrinking circle. The eventual brane configuration in ten dimensions will look like that shown in Fig. 8, with short 4-brane segments suspended between infinite 5-branes.

Now, at weak coupling IIA 5-branes are much heavier than 4-branes, and so can be considered as fixed objects, with any field theory degrees of freedom propagating on the 4-branes. Indeed, the typical length scales (inverse of the mass scales) of NS5-branes (ℓ_5) and D4-branes (ℓ_4) are

$$\ell_4 = g_s \ell_s \qquad \text{and} \qquad \ell_5 = g_s^2 \ell_s. \tag{83}$$

Furthermore, since the extent of the 4-branes is finite in one dimension, at long distances the 4-brane field theory will be effectively four-dimensional. Thus we recover

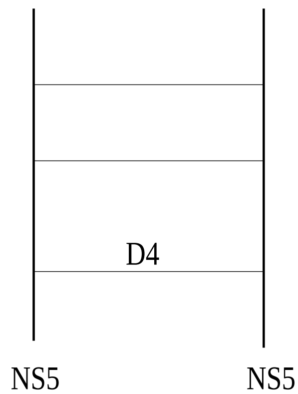

Figure 8: Three D4-branes suspended between two NS5-branes in Type IIA string theory. Only two of the ten dimensions are shown; the 4- and 5-branes are all parallel along an additional $3+1$ dimensions.

the four-dimensional $N=2$ field theory. Finally, an important property of D-branes in string theory (of which the 4-branes are examples) is that the field theory degrees of freedom living on $n+1$ parallel D-branes are described by an $SU(n+1)$ theory Higgsed to $U(1)$'s, *i.e.* a Yang-Mills theory on its Coulomb branch [40]. The size of the vevs Higgsing the gauge group (*i.e.* the Coulomb branch coordinates) are proportional to the separations of the 4-branes. Thus we have learned that in order to describe the $SU(n+1)$ $N=2$ Yang-Mills Coulomb branch we should choose as our M-theory background $Q = \mathbf{R}^3 \times S^1$.

Before turning to the explicit construction of the $SU(n+1)$ IREA, there is an important question to address in this construction, namely, why is an essentially classical 11-dimensional supergravity construction at all reliable to describe a field theory we only see in the $R \to 0$ limit, where the M-theory description should be strongly coupled? The answer lies in a supersymmetric selection rule. Denote by L a typical length scale of the brane configuration shown in Fig. 8, say the distance between some 4-branes or between the 5-branes. Now the typical length scale of the 4-brane dynamics is, from (83) and (82), $\ell_4 = g_s \ell_s = R$. So the relevant scales on the Coulomb branch are measured by the ratios L/R. In terms of the 11-dimensional picture, these ratios determine the shape (complex structure) of Σ_n but not its overall size—which is just as expected since only the complex structure of Σ_n encoded the RSK geometry. Furthermore, the overall size parameter enters as the vev of a hypermultiplet in the supergravity theory. By the $N=2$ selection rule described in Section 2.1, hypermultiplet vevs do not affect the vector multiplet vevs (the Coulomb branch). Thus we learn that the size of R (or equivalently of the string coupling g_s) has no effect on the complex sructure of Σ_n,

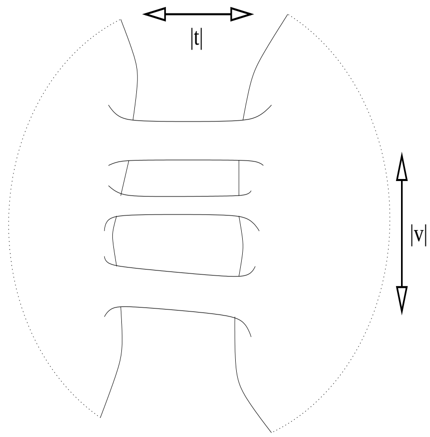

Figure 9: Two dimensions of an M-theory 5-brane embedded in the Q manifold. The three tubes wrap around the S^1 and extend along the $|t|$ direction. The two sheets extend to infinty in Q along the complex v direction far from the tubes.

which can therefore be computed in whatever limit is convenient. In physical terms, this argument shows that R is an irrelevant parameter in the Coulomb branch vacua of the 4-brane field theories.

6.3 The SU(n+1) Coulomb Branch

Let us choose complex coordinates on our hyperKähler 4-manifold $Q = \mathbf{R}^3 \times S^1$ to be $v = (x^1 + ix^2)/R$ and $s = (x^3 + iy)/R$ where y is a periodic coordinate along the S^1, $y \simeq y + 2\pi R$. Good global complex coordinates on Q can then be taken to be v and $t = e^s$. A holomorphically embedded Riemann surface Σ_n will be described by some complex analytic equation in t and v: $F(t, v) = 0$. Since upon shrinking the circle, the surface is supposed to reproduce the IIA brane configuration of Fig. 8, we expect that Σ_n will look globally something like two sheets connected by $n + 1$ tubes as in Fig. 9. Since the tubes are to collapse to D4-branes, they must be wrapped around the S^1, which is the phase of t, and extend along the modulus of t. The two sheets are to become NS5-branes so do not wrap the S^1; thus they should extend to infinity along the complex v direction.

Since this surface wraps $n+1$ times around the S^1 at intermediate $|t|$, by conservation of this winding number, it must also do so as $t \to 0, \infty$. The simplest way of satisfying this constraint is to demand that

$$t \sim v^{n+1} \qquad \text{as} \qquad t \to \infty, \tag{84}$$

and

$$t \sim v^{-n-1} \qquad \text{as} \qquad t \to 0. \tag{85}$$

(Other choices can also satisfy this constraint, but turn out to lead to $SU(n+1)$ $N = 2$ theories with hypermultiplets, and correspond in the IIA picture to configurations with ssemi-infinite D4-branes extending to the left or right of the NS5-branes in Fig. 8.)

Now we can write determine the holomorphic equation $F(v, t) = 0$ for Σ_n. Since at fixed v there are two values of the t coordinate that lie on the surface in Fig. 9, F should be at most quadratic in t:

$$0 = F = A(v)t^2 + B(v)t + C(v). \tag{86}$$

Furthermore since at generic fixed t we found $n+1$ values of v on the surface, we see that A, B, and C can be at most $(n+1)$th order polynomials in v. Suppose the highest powers of v in A, B, and C are n_A, n_B, and n_C, respectively, with

$$0 \leq n_A, n_B, n_C \leq n + 1. \tag{87}$$

Then the leading terms in (86) as $t \to \infty$ according to (84) give

$$v^{n_A+2n+2} + v^{n_B+n+1} + v^{n_C} = 0. \tag{88}$$

This has a solution as $v \to \infty$ with the $n_{A,B,C}$ in the range (87) only if

$$n_A = 0 \qquad \text{and} \qquad n_B = n + 1. \tag{89}$$

A similar argument using (85) as $t \to 0$ gives

$$n_C = 0 \qquad \text{and} \qquad n_B = n + 1. \tag{90}$$

Thus the equation for Σ_n must have the form $0 = \alpha t^2 + \beta(v^{n+1} + a_1 v^n + \cdots + a_n)t + \gamma$ with α, β, γ, and the a_i complex constants. Under holomorphic changes of variables which do not affect the asymptotic behavior of v and t, namely $t \to at$ and $v \to bv + c$, as well as an overall rescaling of F, we can finally put Σ_n in the form

$$t^2 + \frac{1}{\Lambda^{n+1}} \left(v^{n+1} + s_2 v^{n-1} + s_3 v^{n-2} + \cdots + s_n \right) t + 1 = 0. \tag{91}$$

We have identified the coefficients with the strong-coupling scale Λ of the Yang-Mills theory, and the gauge-invariant coordinates s_K on the $SU(n+1)$ Coulomb branch. This makes it clear that this curve indeed has precisely the right number of parameters to describe the Coulomb branch of the $SU(n+1)$ Yang-Mills theory. Furthermore, they can be assigned the right dimensions as well, by assigning v dimensions of energy. Many more detailed checks on this answer can be made by taking the s_K vevs large and comparing the resulting complex structure τ_{IJ} of this curve to that computed from loops in perturbation theory and instanton contributions semiclassically. Finally, it is easy to compute the meromorphic one-form from this data to be $\lambda = (v/t)dt$, thus allowing the computation of BPS masses.

To summarize, we have seen how interpreting geometrical structures in the RSK geometry of $N = 2$ Coulomb branches as physical objects in M-theory together with the type IIA/M-theory equivalence has allowed us to solve for the Coulomb branch IREA associated to particular field theory data in a simple algebraic way. This approach has been extended to solve for the IREAs of many infinite series of $N = 2$ field theory data. It is an open question whether all $N = 2$ field theory IREAs can be solved in this way.

Acknowledgments

It is a pleasure to thank the organizers and participants of the Feza Gürsey Institute Summer School for a pleasant and stimulating visit. This work was supported in part by NSF grant PHY-9513717 and an A.P. Sloan Foundation fellowship

References

References

[1] For an introduction to the $N = 1$ and 2 supersymmetric results, see *e.g.* K. Intriligator, N. Seiberg, *Nucl. Phys. Suppl.* BC **45**, 1 (1996), hep-th/9509066; M. Peskin, hep-th/9702094.

[2] P. Argyres, *Supersymmetric Effective Actions in Four Dimensions,* to appear in the proceedings of the Trieste Spring School, March 1998.

[3] N. Seiberg, E. Witten, *Nucl. Phys.* B **426**, 19 (1994), hep-th/9407087.

[4] E. Witten, *Nucl. Phys.* B **500**, 3 (1997), hep-th/9703166.

[5] See *e.g.* S. Minwalla, hep-th/9712074.

[6] See *e.g.* P. Argyres, R. Plesser, N. Seiberg, E. Witten, *Nucl. Phys.* B **461**, 71 (1996), hep-th/9511154.

[7] See *e.g.* J. Wess, J. Bagger, *Supersymmetry and Supergravity* (2nd edition, Princeton University Press, Princeton NJ, 1992); P. West, *Introduction to Supersymmetry and Supergravity* (2nd edition, World Scientific, Singapore, 1990); M. Sohnius, *Phys. Rep.* **128**, 39 (1985).

[8] See *e.g.* N. Seiberg in *Proceedings of the Trieste Spring School* (1997), hep-th/9705117.

[9] See, however, J. Polchinski, *Nucl. Phys.* B **303**, 226 (1988).

[10] See, however, J. Maldacena, *Adv. Theor. Math. Phys.* **2**, 231 (1998), hep-th/9711200; S. Gubser, I. Klebanov, A. Polyakov, *Phys. Lett.* B **428**, 105 (1998), hep-th/9802109; E. Witten, *Adv. Theor. Math. Phys.* **2**, 253 (1998), hep-th/9802150.

[11] S. Coleman, D. Gross, *Phys. Rev. Lett.* **31**, 851 (1973).

[12] P. Dirac, *Proc. R. Soc.* A **133**, 60 (1931); J. Schwinger, *Phys. Rev.* **144**, 1087 (1966), **173**, 1536 (1968); D. Zwanziger, *Phys. Rev.* **176**, 1480,1489 (1968).

[13] E. Witten, *Phys. Lett.* B **86**, 283 (1979).

[14] A. Shapere, F. Wilczek, *Nucl. Phys.* B **320**, 669 (1989).

[15] In the $N = 1$ context, K. Intriligator, N. Seiberg, *Nucl. Phys.* B **431**, 551 (1994), hep-th/9408155.

[16] P. Argyres, R. Plesser, N. Seiberg, *Nucl. Phys.* B **471**, 159 (1996), hep-th/9603042.

[17] P. Argyres, R. Plesser, A. Shapere, *Nucl. Phys.* B **483**, 172 (1997), hep-th/9608129.

[18] N. Seiberg, *Phys. Lett.* B **206**, 75 (1988); N. Seiberg, *Phys. Lett.* B **318**, 469 (1993), hep-ph/9309335.

[19] C. Montonen, D. Olive, *Phys. Lett.* B **72**, 117 (1977); H. Osborn, *Phys. Lett.* B **83**, 321 (1979).

[20] E. Witten, D. Olive, *Phys. Lett.* B **78**, 97 (1978);

[21] A. Sen, *Phys. Lett.* B **329**, 217 (1994), hep-th/9402032.

[22] N. Seiberg, E. Witten, *Nucl. Phys.* B **431**, 484 (1994), hep-th/9408099.

[23] P. Argyres, R. Plesser, A. Shapere, *Phys. Rev. Lett.* **75**, 1699 (1995), hep-th/9505100.

[24] P. Argyres, A. Shapere, *Nucl. Phys.* B **461**, 437 (1996), hep-th/9509175.

[25] P. Argyres, *Adv. Theor. Math. Phys.* **2**, 293 (1998), hep-th/9706095.

[26] P. Argyres and A. Buchel, *Phys. Lett.* B **431**, 317 (1998), hep-th/9804007.

[27] N. Dorey, V. Khoze, and M. Mattis, *Phys. Lett.* B **388**, 324 (1996), hep-th/9607066.

[28] J. Harvey, in *1995 Summer School in High-Energy Physics and Cosmology: Proceedings*, edited by E. Gava *et. al.* (World Scientific, 1997), hep-th/9603086.

[29] G. 't Hooft, *Nucl. Phys.* B **79**, 276 (1974); A. Polyakov, *J.E.T.P. Lett.* **20**, 194 (1974).

[30] E. Witten, *Phys. Lett.* B **86**, 283 (1979).

[31] R. Flume, M. Magro, L. O'Raifeartaigh, I. Sachs, and O. Schnetz, *Nucl. Phys.* B **494**, 331 (1997), hep-th/9611123.

[32] For discussions of the various definitions of rigid special Kähler geometry with references, see for example P. Fré, *Nucl. Phys. Proc. Suppl.* BC **45**, 59 (1996), hep-th/9512043; I. Antoniadis, B. Pioline, *Int. J. Mod. Phys.* A **12**, 4907 (1997), hep-th/9607058; B. Craps, F. Roose, W. Troost, A. Van Proeyen, *Nucl. Phys.* B **503**, 565 (1997), hep-th/9703082; D. Freed, hep-th/9712042.

[33] R. Donagi, E. Witten, *Nucl. Phys.* B **460**, 299 (1996), hep-th/9510101.

[34] For the first such papers, see A. Gorsky, I. Krichever, A. Marshakov, A. Mironov, and A. Morozov, *Phys. Lett.* B **355**, 466 (1995), hep-th/9505035; E. Martinec and N. Warner, *Phys. Lett.* B **549**, 97 (1996), hep-th/9509161; T. Nakatsu and K. Takasaki, *Mod.Phys.Lett.* A **11**, 157 (1996), hep-th/9509162; E. Martinec, *Phys. Lett.* B **367**, 91 (1996), hep-th/9510204.

[35] For a review of these constructions, see *e.g.* R. Donagi in *Surveys in Differential Geometry*, ed. S.T. Yau, alg-geom/9705010.

[36] For a review of Riemann surfaces, see for example, D. Mumford, *Tata Lectures on Theta I* (Birkhäuser, 1983).

[37] P. Argyres and A. Faraggi, *Phys. Rev. Lett.* **74**, 3931 (1995), hep-th/9411057; A. Klemm, W. Lerche, S. Theisen, and S. Yankielowicz, *Phys. Lett.* B **344**, 169 (1995), hep-th/9411048; U. Danielsson and B. Sundborg, *Phys. Lett.* B **358**, 273 (1995), hep-th/9504102; A. Hanany and Y. Oz, *Nucl. Phys.* B **452**, 283 (1995), hep-th/9505075; A. Brandhuber and K. Landsteiner, *Phys. Lett.* B **358**, 73 (1995), hep-th/9507008.

[38] A. Klemm, W. Lerche, P. Mayr, C. Vafa, N. Warner, *Nucl. Phys.* B **477**, 746 (1996), hep-th/9604034.

[39] For an introduction to the string/M theory properties used in this subsection, see for example Chapter 14 of J. Polchinksi, *String Theory*, vol. II (Cambridge, 1998).

[40] E. Witten, *Nucl. Phys.* B **460**, 335 (1996), hep-th/9510135.