COMMUNICATIONS IN
NUMBER THEORY AND PHYSICS
Volume 1, Number 1, 1–236, 2007

Electric-Magnetic Duality And The Geometric Langlands Program

Anton Kapustin
Department of Physics, California Institute of Technology, Pasadena, CA 91125

and

Edward Witten
School of Natural Sciences, Institute for Advanced Study, Princeton, New Jersey 08540

The geometric Langlands program can be described in a natural way by compactifying on a Riemann surface C a twisted version of $\mathcal{N} = 4$ super Yang-Mills theory in four dimensions. The key ingredients are electric-magnetic duality of gauge theory, mirror symmetry of sigma-models, branes, Wilson and 't Hooft operators, and topological field theory. Seemingly esoteric notions of the geometric Langlands program, such as Hecke eigensheaves and $\mathcal{D}$-modules, arise naturally from the physics.

Received January 18, 2007

Contents

1. Introduction

The Langlands program for number fields [1] unifies many classical and contemporary results in number theory and is a vast area of research. It has an analog for curves over a finite field, which has also been the subject of much celebrated work [2,3]. In addition, a geometric version of the Langlands program for curves has been much developed [4−8], both for curves over a field of characteristic p and for ordinary complex Riemann surfaces. For a survey that is relatively readable for physicists, with numerous references, see [9].

Our focus in the present paper is on the geometric Langlands program for complex Riemann surfaces. We aim to show how this program can be understood as a chapter in quantum field theory. No prior familiarity with the Langlands program is assumed; instead, we assume a familiarity with subjects such as supersymmetric gauge theories, electric-magnetic duality, sigma-models, mirror symmetry, branes, and topological field theory. The theme of the paper is to show that when these familiar physical ingredients are applied to just the right problem, the geometric Langlands program arises naturally. Seemingly esoteric notions such as Hecke eigensheaves, $\mathcal{D}$-modules, and so on, appear spontaneously in the physics, with new insights about their properties.

The first hints of a connection between the Langlands program and quantum field theory came from the work of Goddard, Nuyts, and Olive (GNO), who showed in 1976 [10] that in gauge theories, though electric charge takes values in the weight lattice of the gauge group, magnetic charge takes values in the weight lattice of a dual group. Magnetic charges for general compact Lie groups had been first analyzed by Englert and Windey [11]. The GNO analysis motivated the Montonen-Olive electric-magnetic duality conjecture [12] according to which a specific gauge theory based on a given gauge group is equivalent to a similar theory with the coupling constant inverted and the gauge group replaced by its dual.

Table 1: Examples of the correspondence between a Lie group G and its Langlands or GNO dual ${}^LG$.

G	${}^LG$
$U(N)$	$U(N)$
$SU(N)$	$PSU(N) = SU(N)/\mathbb{Z}_N$
$Spin(2n)$	$SO(2n)/\mathbb{Z}_2$
$Sp(n)$	$SO(2n+1)$
$Spin(2n+1)$	$Sp(n)/\mathbb{Z}_2$
G_2	G_2
E_8	E_8

For a gauge group G, the GNO dual group is actually the same as the Langlands dual group ${}^LG$, which plays an important role in formulating the Langlands conjectures. (For some examples of the correspondence between G and ${}^LG$, see Table 1.) This was observed by Atiyah, who suggested to the second author at the end of 1977 that the Langlands program is related to quantum field theory and recommended the two papers [10,12]. There resulted a further development [13] in which it was understood that Montonen-Olive duality is more natural in supersymmetric gauge theory. It was later understood that $\mathcal{N} = 4$ supersymmetry (i.e., the maximal possible supersymmetry in four dimensions) is the right context for this duality [14], and that the $\mathbb{Z}_2$ duality originally proposed has a natural extension to $SL(2,\mathbb{Z})$ [15,16] when the theta angle of the gauge theory is included.

In the early 1990's, extensions of Montonen-Olive duality to string theory were conjectured [17]. Subsequently, Montonen-Olive duality and its generalizations were studied from many new points of view and were recognized as a crucial, though still mysterious, ingredient in understanding field theory and string theory. These developments were far too extensive to be reviewed here, but one observation of that period, though a sideline at the time, is a starting point for the present paper. Compactification of $\mathcal{N} = 4$ super Yang-Mills theory from four dimensions to two dimensions was studied [18,19] and was shown to lead at low energies to a two-dimensional supersymmetric sigma-model in which the target space is a hyper-Kahler manifold that is Hitchin's moduli space $\mathcal{M}_H$ of stable Higgs bundles [20]. Electric-magnetic duality in four dimensions reduces in two dimensions to T-duality of the sigma-model. This particular T-duality was subsequently used mathematically [21] to show (for $SU(N)$) that the Hodge numbers of the Higgs bundle moduli space of a gauge group G are equal to those of ${}^LG$. The geometry

underlying this T-duality was investigated in [22] and subsequently in [23] for any semi-simple Lie group G.

Other clues about the relation of the geometric Langlands program to quantum field theory have come from relatively recent mathematical work. The approach of Beilinson and Drinfeld to the geometric Langlands program is based on quantization of $\mathcal{M}_H$, as the title of their paper implies [5]. The T-duality of $\mathcal{M}_H$, understood mathematically as a Fourier-Mukai transform, has been interpreted as a sort of semiclassical approximation to the geometric Langlands program. This point of view underlies the paper [24]. We understand that there have also been important unpublished contributions by other mathematicians, including Donagi and Pantev. The second author learned of this interpretation of the T-duality of $\mathcal{M}_H$ from a lecture by D. Ben-Zvi at a conference on the geometric Langlands program, held at the IAS in the spring of 2004. This was an extremely strong hint that it must be possible to understand the geometric Langlands program using four-dimensional electric-magnetic duality (which leads to this particular T-duality) and branes (the natural quantum field theory setting for interpreting T-duality as a Fourier-Mukai transform). This hint was the starting point for the present paper.

To summarize this paper in the briefest possible terms, we will develop six main ideas. The first is that from a certain twisted version of $\mathcal{N} = 4$ supersymmetric Yang-Mills theory in four dimensions, one can construct a family of four-dimensional topological field theories in four dimensions. After reviewing some of the background in section 2, we explain this construction in section 3. The twisting procedure is formally just analogous to the construction by which Donaldson theory can be obtained [25] from $\mathcal{N} = 2$ super Yang-Mills theory. The second main idea, developed in sections 4 and 5, is that, extending the insights of [18,19], compactification on a Riemann surface C gives in two dimensions a family of topological sigma-models, with target $\mathcal{M}_H$, which are "generalized B-models." Moreover, for a special value of the parameter, four-dimensional S-duality acts as two-dimensional mirror symmetry. The third main idea, developed in section 6, is that Wilson and 't Hooft line operators are topological operators that act on the branes of the two-dimensional sigma-model in a natural fashion. Here we consider an operator that maps one brane to another (or roughly speaking, one theory to another), not the more familiar sort of operator in Hilbert space that maps one state to another. A brane that is mapped by the Wilson or 't Hooft operators to, roughly speaking, a multiple of itself is what we call an electric or magnetic eigenbrane. S-duality will automatically exchange the electric and magnetic eigenbranes. The fourth main idea, explained in section 8, is

that, in the right context, electric eigenbranes are in natural correspondence with homomorphisms of $\pi_1(C)$ to the complexification of the Langlands dual group ${}^L G_{\mathbb{C}}$. The fifth main idea, developed in sections 9 and 10, is that 't Hooft operators correspond naturally to geometric Hecke operators similar to those of the geometric Langlands program but acting on Higgs bundles instead of ordinary G-bundles. It takes one more important idea, developed in section 11, to make contact with the usual formulation of the geometric Langlands correspondence. We show that, because of the existence of a canonical coisotropic brane on $\mathcal{M}_H$, the magnetic eigenbranes in our sense are naturally associated to $\mathcal{D}$-modules (modules for the sheaf of differential operators) on the moduli space $\mathcal{M}(G,C)$ of holomorphic G-bundles on C.

Sections 7, 8, 9, and 10 of this paper and some expository portions of other sections are primarily adapted from a forthcoming book [26] which will also contain some additional results. This material is included here to make the paper more comprehensible.

One obvious gap in our analysis is that we consider only the unramified case of the geometric Langlands correspondence. We expect that it is possible to apply somewhat similar ideas to the ramified case. A second major gap is that we do not shed light on the utility of two-dimensional conformal field theory for the geometric Langlands program [5,9,27–30] . The last of these references applies conformal field theory to the ramified case. Hopefully it will prove possible to deduce the conformal field theory approach from the gauge theory approach of this paper. In fact, there is an analogy even at a naive level [31] between conformal field theory and the theory of automorphic representations, which is the basis of the Langlands program. Finally, though we have nothing to contribute about this here, an additional clue about the relation of the Langlands program with physics presumably comes from the diverse ways that automorphic forms enter string theory. For a tiny sampling of this, see [32–40].

A.K. would like to thank D. Arinkin, R. Bezrukavnikov, and D. Orlov for useful conversations. In particular, Orlov's explanations in 2002 about the abelian case of the geometric Langlands program partially motivated the paper [41], which will enter our story in section 11.

E.W. would like to thank the many mathematicians who over the years have explained matters relevant to the Langlands program, including A. Beilinson, P. Deligne, V. Drinfeld, and K. Vilonen, and especially M. F. Atiyah, D. Ben-Zvi, R. Donagi, E. Frenkel, and D. Kazhdan, and most recently M. Goresky and R. MacPherson. In addition, among others, T. Hausel, N. Hitchin, M. Hopkins, P. Kronheimer, L. Jeffrey, J. Morgan, G. Moore, D. Morrison, N. Nekrasov, M. Thaddeus, C. Vafa, and E. J. Weinberg

clarified some points relevant to the present paper, and many of the physicists at the IAS, including S. Hellerman, K. Intriligator, J. Maldacena, N. Seiberg, and J. Walcher, made helpful comments.

2. $\mathcal{N} = 4$ Super Yang-Mills Theory And S-Duality

In this section, we recall a few properties of $\mathcal{N} = 4$ super Yang-Mills theory and its S-duality.

2.1. Review Of $\mathcal{N} = 4$ Super Yang-Mills Theory

As in the original work [42], $\mathcal{N} = 4$ super Yang-Mills is most easily constructed by dimensional reduction from ten dimensions. Ten dimensions is the maximum possible dimension for supersymmetric Yang-Mills theory by virtue of Nahm's theorem [43], and for given gauge group G, there is a very simple supersymmetric Lagrangian which moreover is unique up to the choice of a few coupling parameters if we ask for a Lagrangian quadratic in the curvature. In this paper, we always assume G to be compact and denote its complexification as $G_{\mathbb{C}}$. This differs from most other expositions of the geometric Langlands program.

Spacetime Conventions We begin by describing some conventions. We will work with Lorentz signature $-++\ldots+$ or Euclidean signature $++++\ldots+$. Basically, when emphasizing questions of topological field theory, we will use Euclidean signature, but when we want to stress that the constructions are natural in physically sensible, unitary quantum field theory, we use Lorentz signature.

We write the metric of ten-dimensional Minkowski space $\mathbb{R}^{1,9}$ or Euclidean space $\mathbb{R}^{10}$ as $ds^2 = \sum_{I,J=0}^{9} g_{IJ}\, dx^I\, dx^J = \mp(dx^0)^2 + (dx^1)^2 + \ldots + (dx^9)^2$. These have symmetry groups $SO(1,9)$ or $SO(10)$, whose spin representations $\mathcal{S}^+$ and $\mathcal{S}^-$ are of rank 16. They are real (and dual to one another) in Lorentz signature, while in Euclidean signature they are complex conjugate representations.

The gamma matrices Γ_I, $I = 0, \ldots, 9$ (which in Lorentz signature are real) reverse the chirality, mapping $\mathcal{S}^\pm$ to $\mathcal{S}^\mp$, and obey the Clifford algebra $\{\Gamma_I, \Gamma_J\} = 2g_{IJ}$. Moreover, the operator

$$\overline{\Gamma} = \Gamma_0\Gamma_1\cdots\Gamma_9 \tag{2.1}$$

acts on $\mathcal{S}^+$ or $\mathcal{S}^-$ as multiplication by 1 or -1. Because $\mathcal{S}^+$ and $\mathcal{S}^-$ are dual, we can regard the Γ's not as maps from $\mathcal{S}^\pm$ to $\mathcal{S}^\mp$ but as bilinear maps

$\Gamma_I : \mathcal{S}^+ \otimes \mathcal{S}^+ \to \mathbb{R}$ or $\Gamma_I : \mathcal{S}^- \otimes \mathcal{S}^- \to \mathbb{R}$. (In this paragraph, we assume Lorentz signature; in Euclidean signature, all such maps are to $\mathbb{C}$, since the spinors are complex.) If $\beta, \gamma \in \mathcal{S}^+$, it is conventional to write $\overline{\beta}\Gamma_I\gamma$ for the bilinear map $\Gamma_I(\beta,\gamma)$, which can also be written in components as $\sum_{a,b=1}^{16} \Gamma_{I\,ab}\beta^a\gamma^b$. (The bar in $\overline{\beta}$ is conventional for spinors and should perhaps be read as transpose, not complex conjugation.) A standard convention is to define $\Gamma_{I_1 I_2 \dots I_k}$ to be zero if the indices $I_1, I_2, \dots, I_k$ are not pairwise distinct and otherwise to equal the product $\Gamma_{I_1}\Gamma_{I_2}\cdots\Gamma_{I_k}$. So $\Gamma_{I_1 I_2 \dots I_k}$ reverses the chirality if k is odd, and again can be regarded as a bilinear map $\mathcal{S}^+ \otimes \mathcal{S}^+ \to \mathbb{R}$ (or $\mathcal{S}^- \otimes \mathcal{S}^- \to \mathbb{R}$). These maps are symmetric for $k = 1, 5, 9$ and antisymmetric for $k = 3, 7$. For $\beta, \gamma \in \mathcal{S}^+$, we again write $\overline{\beta}\Gamma_{I_1 \dots I_k}\gamma$ for these bilinear maps $\Gamma_{I_1 \dots I_k}(\beta,\gamma)$. For k even, we have $\Gamma_{I_1 I_2 \dots I_k} : \mathcal{S}^\pm \to \mathcal{S}^\pm$, or equivalently we have bilinear maps $\Gamma_{I_1 I_2 \dots I_k} : \mathcal{S}^+ \otimes \mathcal{S}^- \to \mathbb{R}$, again denoted $\overline{\beta}\Gamma_{I_1 \dots I_k}\gamma$.

Fields, Transformation Laws, And Lagrangian The fields of ten-dimensional super Yang-Mills theory are the gauge field A, which is a connection on a G-bundle E, and a fermion field λ that is a section of $\mathcal{S}^+ \otimes \mathrm{ad}(E)$; in other words, λ is a positive chirality spinor field with values in the adjoint representation of G. In Lorentz signature, λ is real, since the bundles $\mathcal{S}^+$ and $\mathcal{S}^-$ are real. In Euclidean signature, λ is not real but its complex conjugate does not enter the formalism. The covariant derivative is $D = d + A$ and the curvature of A is $F = D^2 = dA + A \wedge A$.

We consider A and λ to take values in the real Lie algebra of G, which has real structure constants. This means that, in a unitary representation of G, A and λ take values in antihermitian matrices. This is opposite to the usual physics convention, but is in accord with the math literature. (Taking the fields to be antihermitian may look unnatural for $G = U(1)$, which is the reason for the usual physics convention, but it avoids unnatural factors of i for general G.) In a standard set of physics conventions (see p. 4 of [44]), the covariant derivative is $D = d - iA'$ with a hermitian gauge field A'. The relation between our antihermitian A and this hermitian gauge field A' is thus

$$A = -iA'. \tag{2.2}$$

The curvature of A' is defined as $F' = i(D)^2 = dA' - iA' \wedge A'$, so

$$F = -iF'. \tag{2.3}$$

We use the symbol "Tr" to denote an invariant and negative definite quadratic form on the Lie algebra of G. As we assume G to be simple, such

a quadratic form is unique up to a constant multiple; we normalize it so that for $M = S^4$, the characteristic class $\frac{1}{8\pi^2}\int_M \mathrm{Tr}\, F \wedge F$ takes arbitrary integer values. For $G = SU(N)$, this quadratic form can be obtained as a trace in the N-dimensional representation, which motivates denoting it as Tr. Since our Lie algebras are generated by antihermitian matrices, Tr is *negative*-definite.

With this understood, the supersymmetry transformation laws and the Lagrangian of ten-dimensional super Yang-Mills theory can be described as follows. The generator of supersymmetry is a constant spinor ϵ that takes values in $\mathcal{S}^+$, and hence obeys

$$\overline{\Gamma}\epsilon = \epsilon. \tag{2.4}$$

($\overline{\Gamma}$ was defined in (2.1); we take ϵ to be bosonic.) The supersymmetry transformation generated by ϵ is

$$\begin{aligned} \delta_S A_I &= i\overline{\epsilon}\Gamma_I\lambda \\ \delta_S\lambda &= \frac{1}{2}\Gamma^{IJ}F_{IJ}\epsilon. \end{aligned} \tag{2.5}$$

For any field Φ, the symbol $\delta_S\Phi$ is short for[1] $\sum_{a=1}^{16}[\epsilon^a Q_a, \Phi\}$, where Q_a, taking values in $\mathcal{S}^-$, are the sixteen supersymmetries. The symbol δ_S stands for supersymmetric variation. The invariant action is[2]

$$\boldsymbol{I}_{10} = \frac{1}{e^2}\int d^{10}x\, \mathrm{Tr}\left(\frac{1}{2}F_{IJ}F^{IJ} - i\overline{\lambda}\Gamma^I D_I\lambda\right), \tag{2.6}$$

with an arbitrary constant e, the gauge coupling. The verification of supersymmetry is described in [42]. Finally, the conserved supercurrent that generates the supersymmetries (2.5) is

$$J^I = \frac{1}{2}\mathrm{Tr}\,\Gamma^{JK}F_{JK}\Gamma^I\lambda. \tag{2.7}$$

The bosonic symmetry of this theory is not just $SO(1,9)$ but the "Poincaré group" $\mathcal{P}$, which is an extension of $SO(1,9)$ by the "translation" group of $\mathbb{R}^{1,9}$. This translation group is isomorphic to $\mathbb{R}^{1,9}$ itself (regarded

[1] In a $\mathbb{Z}_2$-graded algebra, the symbol $[A,B\}$ denotes $AB - (-1)^{AB}BA$.

[2] This is written in Lorentz signature, with the usual convention in which the kinetic energy is positive. For Euclidean signature, one must change the overall sign of the Lagrangian to make the bosonic part of the action positive definite.

as an abelian group), and $\mathcal{P}$ is an extension

$$0 \longrightarrow \mathbb{R}^{1,9} \longrightarrow \mathcal{P} \longrightarrow SO(1,9) \longrightarrow 1. \tag{2.8}$$

The generators of $\mathbb{R}^{1,9}$ are called the "momentum operators" P_I. The algebra obeyed by the conserved supercharges Q_a, $a = 1, \dots, 16$ is

$$\{Q_a, Q_b\} = \sum_{I=1}^{10} \Gamma^I_{ab} P_I. \tag{2.9}$$

In addition, the Q_a commute with P_I and transform under $SO(1,9)$ as $\mathcal{S}^-$.

Dimensional Reduction To Four Dimensions To reduce to four dimensions, we simply take all fields to be independent of the coordinates $x^4, \dots, x^9$. The components A_I, $I = 0, \dots, 3$ describe the four-dimensional gauge field $A = \sum_{\mu=0}^{3} A_\mu dx^\mu$, while the components A_I, $I \geq 4$, become four-dimensional scalar fields $\phi_i = A_{i+4}$, $i = 0, \dots, 5$. The ten-dimensional curvature F_{IJ} has three types of contribution; depending on whether the number of indices I, J in the range $4, \dots, 9$ is zero, one, or two, we get a four-dimensional curvature $F_{\mu\nu}$, a derivative $D_\mu \phi_i$ of a scalar field, or a commutator $[\phi_i, \phi_j]$ of scalar fields.

The bosonic part of the action, in four dimensions, has all three types of contribution and becomes

$$\boldsymbol{I}_4 = \frac{1}{e^2} \int d^4x \, \mathrm{Tr} \left(\frac{1}{2} \sum_{\mu,\nu=0}^{3} F_{\mu\nu} F^{\mu\nu} + \sum_{\mu=0}^{3} \sum_{i=1}^{6} D_\mu \phi_i D^\mu \phi_i + \frac{1}{2} \sum_{i,j=1}^{6} [\phi_i, \phi_j]^2 \right). \tag{2.10}$$

Together with the part of the action involving fermions, which can be similarly written in four-dimensional terms, though we will not do so, this is the essentially unique four-dimensional gauge theory with the maximal possible supersymmetry. If G is simple and if we want a Lagrangian quadratic in derivatives (the case that leads to a sensible quantum theory), the action is unique except for the choice of parameter e and, in four dimensions, another possible parameter that measures the topology of the G-bundle E:

$$\boldsymbol{I}_\theta = -\frac{\theta}{8\pi^2} \int \mathrm{Tr}\, F \wedge F. \tag{2.11}$$

This last term is θ times the second Chern class or instanton number of the bundle. The parameters e and θ combine into a complex parameter

$$\tau = \frac{\theta}{2\pi} + \frac{4\pi i}{e^2}. \tag{2.12}$$

As long as we are on $\mathbb{R}^4$ or any four-manifold M with $H^2(M,\mathbb{Z}) = 0$ (for the generalization, in which $\pi_1(G)$ comes into play, see [45,26]), there is an elementary symmetry $\tau \to \tau + 1$, which expresses the fact that $(1/8\pi^2)\int \mathrm{Tr} F \wedge F$ is integer-valued, and that in quantum mechanics one only cares about the action modulo an integer multiple of 2π. Equivalently, θ is an angular variable, with $\theta \cong \theta + 2\pi$.

The $SO(1,9)$ (or $SO(10)$) symmetry in ten dimensions becomes, after dimensional reduction to four dimensions, $SO(1,3) \times SO(6)$ (or $SO(4) \times SO(6)$). Allowing for the presence of spinors in the theory, the symmetry is really $Spin(1,9)$ reduced to $Spin(1,3) \times Spin(6)$ (or $Spin(10)$ reduced to $Spin(4) \times Spin(6)$). The group $Spin(6)$ is isomorphic to $SU(4)$ and is known as the "R symmetry group" of the theory. We will call it $SU(4)_{\mathcal{R}}$.

The chirality condition $\overline{\Gamma}\epsilon = \epsilon$ in ten dimensions becomes in four dimensions

$$\widehat{\Gamma}\Gamma'\epsilon = \epsilon, \tag{2.13}$$

where $\widehat{\Gamma} = \Gamma_0\Gamma_1\Gamma_2\Gamma_3$ measures the $Spin(1,3)$ chirality and $\Gamma' = \Gamma_4\Gamma_5\dots\Gamma_9$ measures the $Spin(6)$ chirality. $\widehat{\Gamma}$ and Γ' have eigenvalues $\pm i$; (2.13) means that the eigenvalue of Γ' is minus that of $\widehat{\Gamma}$. The two eigenvalues of $\widehat{\Gamma}$ distinguish the two spin representations of $Spin(1,3)$, while the eigenvalues of Γ' similarly label the spin representations of $Spin(6)$. The complexification of $Spin(1,3)$ is $SL(2) \times SL(2)$ and the two spin representations correspond to the representations $(\mathbf{2},\mathbf{1})$ and $(\mathbf{1},\mathbf{2})$ of $SL(2) \times SL(2)$ (here $(\mathbf{2},\mathbf{1})$ is the two-dimensional representation of the first $SL(2)$ tensored with the trivial one-dimensional representation of the second $SL(2)$, and vice-versa for $(\mathbf{1},\mathbf{2})$). The spin representations of $Spin(6)$ are the defining four-dimensional representation of $SU(4)_{\mathcal{R}}$ and its dual; we denote them as $\mathbf{4}$ and $\overline{\mathbf{4}}$. We pick orientations so that $\widehat{\Gamma}$ acts as i^{-1} or i on the $(\mathbf{2},\mathbf{1})$ and $(\mathbf{1},\mathbf{2})$, respectively, and Γ' acts as i and i^{-1} on the $\overline{\mathbf{4}}$ and $\mathbf{4}$. So (2.13) means that the four-dimensional supersymmetries transform under $Spin(1,3) \times Spin(6) \sim SL(2) \times SL(2) \times Spin(6)$ as

$$(\mathbf{2},\mathbf{1},\overline{\mathbf{4}}) \oplus (\mathbf{1},\mathbf{2},\mathbf{4}). \tag{2.14}$$

The fermion fields λ transform the same way under $Spin(1,3) \times Spin(6)$.

If we write $\overline{Q}_{AX}$, $A = 1, 2$, $X = 1, \dots, 4$ for the supersymmetries of type $(\mathbf{2}, \mathbf{1}, \overline{\mathbf{4}})$, and similarly $Q^Y_{\dot{A}}$, $\dot{A} = 1, 2$, $Y = 1, \dots, 4$ for those of type $(\mathbf{1}, \mathbf{2}, \mathbf{4})$, then the algebra generated by the supersymmetries is the reduction of (2.9),

$$\begin{aligned} \{\overline{Q}_{AX}, Q^Y_{\dot{A}}\} &= \delta_X{}^Y \sum_{\mu=0}^{3} \Gamma^\mu_{A\dot{A}} P_\mu \\ \{Q, Q\} = \{\overline{Q}, \overline{Q}\} &= 0, \end{aligned} \tag{2.15}$$

where now the four momentum operators P_μ generate the translations of $\mathbb{R}^{1,3}$. With suitable boundary conditions, additional terms appear [13] on the right hand side of (2.15) (they are related to the extra six momenta that were dropped in going to four dimensions, and their magnetic duals). They will make a brief appearance at the end of this section.

2.2. *S*-Duality

A review of S-duality is unfortunately beyond our scope in this paper. We will just mention a few relevant facts.

Since its imaginary part is positive, $\tau = \theta/2\pi + 4\pi i/e^2$ defines a point in the upper half plane $\mathcal{H}$. The group $SL(2, \mathbb{R})$ acts on $\mathcal{H}$ in the standard fashion $\tau \to (a\tau + b)/(c\tau + d)$, with $ad - bc = 1$.[3] The transformation $T : \tau \to \tau + 1$ is simply a 2π shift of the angle θ and thus a classical symmetry of the theory on $\mathbb{R}^4$, for any gauge group G. The S-duality conjecture asserts that there exists an additional quantum symmetry that inverts τ, exchanges G with ${}^L G$, and exchanges electric and magnetic charges. Moreover, this symmetry, which we will call S, combines with the classical symmetry $T : \tau \to \tau + 1$ to generate an infinite discrete subgroup Γ of $SL(2, \mathbb{R})$.

The most familiar case is the case that G is simply-laced. Then

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{2.16}$$

acts as $\tau \to -1/\tau$, and together with T generates the group $SL(2, \mathbb{Z})$.

If G is not simply-laced, then the S-transformation is not $\tau \to -1/\tau$. Rather, it is $\tau \to -1/n_{\mathfrak{g}}\tau$, where $n_{\mathfrak{g}}$ is 2 for F_4 and 3 for G_2. This transformation can be achieved by the $SL(2, \mathbb{R})$ transformation

$$S = \begin{pmatrix} 0 & 1/\sqrt{n_{\mathfrak{g}}} \\ -\sqrt{n_{\mathfrak{g}}} & 0 \end{pmatrix}. \tag{2.17}$$

[3] The group that acts faithfully on $\mathcal{H}$ is the quotient $PSL(2, \mathbb{R})$, but in application to four-dimensional gauge theory, one really needs the double cover.

This presence of a factor of 2 or 3 in the action of S on τ can be seen [46,47] by examining the BPS mass formulas for electric and magnetic charges and reflects the relation between roots and coroots for these groups. It also can be extracted from a string-theoretic explanation of S-duality for non-simply-laced groups [48]. The factor of 2 or 3 means that the duality groups for G_2 or F_4 are not simply $SL(2,\mathbb{Z})$, but certain infinite discrete subgroups of $SL(2,\mathbb{R})$ that are known as Hecke groups.

The remaining simple Lie groups $Sp(k)$ and $Spin(2k+1)$ (and their respective quotients by $\mathbb{Z}_2$) require some special comment, because these are the only simple Lie groups such that G and ${}^LG$ have non-isomorphic Lie algebras. The Yang-Mills Lagrangian and therefore the definition of the τ parameter depend only on the Lie algebra $\mathfrak{g}$ of the gauge group. Hence, whenever G and ${}^LG$ have the same Lie algebra, we can discuss how the S-transformation acts on τ without distinguishing whether we have in mind τ of a theory with gauge group G or τ of a theory with gauge group ${}^LG$. This indeed is what we have implicitly done so far.

For the pair $Sp(k)$ and $Spin(2k+1)$, however, there is no equally direct way to compare the two τ parameters. Hence, one may introduce separate gauge coupling parameters, say τ for $Sp(k)$ or $Sp(k)/\mathbb{Z}_2$ gauge theory, and τ' for $Spin(2k+1)$ or $Spin(2k+1)/\mathbb{Z}_2 = SO(2k+1)$ gauge theory. If one normalizes the respective τ parameters so that the respective T-transformations act by $\tau \to \tau+1$ and $\tau' \to \tau'+1$, then S acts by $\tau = -1/2\tau'$, $\tau' = -1/2\tau$, just as for F_4. (Thus we set $n_{\mathfrak{g}} = 2$ for $Spin(2k+1)$ and $Sp(k)$.)

This normalization is natural, since it leads to the most uniform gauge theory formulas for arbitrary gauge groups. A slight complication is that there is for the same groups a second normalization that one might also consider natural. A useful string theory realization of the $Sp(k)/Spin(2k+1)$ duality, involving orientifold threeplanes [49], actually motivates a different normalization. In this normalization, the $Sp(k)$ theory is described by a coupling parameter $\widetilde{\tau} = 2\tau$, so that S acts simply by $\widetilde{\tau} = -1/\tau'$, but instead T acts by $\widetilde{\tau} \to \widetilde{\tau} + 2$. (In figure 3 of [49], the $Sp(k)$ theory appears twice, precisely because the $Sp(k)$ coupling parameter $\widetilde{\tau}$ is normalized in that paper so that the T-symmetry acts by $\widetilde{\tau} \to \widetilde{\tau} + 2$.)

A final comment on this is that the assertion that Montonen-Olive duality exchanges G and ${}^LG$ is not quite the whole story. On $\mathbb{R}^4$, this is an adequate description, but on a general four-manifold, the full story is somewhat more elaborate. A G-bundle on a four-manifold has a characteristic class $\xi \in H^2(M, \pi_1(G))$, studied in this paper in section 7. A path integral can be defined for each value of ξ, and the resulting partition functions Z_ξ transform as a unitary representation of the duality group [45,26].

How much of this is important for the geometric Langlands program? The basic geometric Langlands duality is the transformation S that acts as $\tau \to -1/n_{\mathfrak{g}}\tau$. It acts on the canonical parameter Ψ (introduced in section 3.5) by $\Psi \to -1/n_{\mathfrak{g}}\Psi$. The basic geometric Langlands duality involves comparing $\Psi = \infty$ to $\Psi = 0$. These are exchanged by S regardless of the value of $n_{\mathfrak{g}}$. So the value of $n_{\mathfrak{g}}$ is not very important for the basic geometric Langlands duality. In section 11.3, we come to a generalization of the geometric Langlands duality to arbitrary Ψ. Here the precise duality group is important and, therefore, the value of $n_{\mathfrak{g}}$ does play a role.

Transformation of Supersymmetries There is one question about Montonen-Olive duality that actually will play a bigger role in this paper: How does it act on the supersymmetries?

We cannot the answer this question by inspection of the classical Lagrangian, because, apart from the subgroup generated by $\tau \to \tau + 1$, Γ does not consist of symmetries of the classical theory. This after all is what makes S-duality interesting. So obtaining the answer will require a more subtle reasoning.

Before determining the answer, let us ask to what extent the answer is unique. Consider an element $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ of the duality group Γ generated by S and T. Such an element acts on τ by $\tau \to (a\tau + b)/(c\tau + d)$ and on the supersymmetry algebra by an automorphism. This automorphism is not uniquely defined, since it could be combined with a symmetry of the $\mathcal{N} = 4$ super Yang-Mills theory (at a fixed value of τ). An important simplification is that, according to the Montonen-Olive conjecture, γ commutes with the Poincaré group. Moreover, one can define it to commute with the global R-symmetry group $SU(4)_{\mathcal{R}}$.[4] Combining these facts, it follows that γ acts as a scalar multiplication $\exp(i\widehat{\phi})$ on the supersymmetries that transform as the $\overline{\mathbf{4}}$ (that is, the $\overline{Q}_{AX}$) and as $\exp(-i\widehat{\phi})$ on those that transform as $\mathbf{4}$ (the $Q^Y_{\dot{A}}$). Moreover, $\widehat{\phi}$ must be real to preserve the real structure of the algebra (with respect to which, in Lorentz signature, P is hermitian and $\overline{Q}$ is the hermitian adjoint of Q). We will call these symmetries $U(1)$ chiral rotations. The action of γ is defined up to an element of the center of $SU(4)_{\mathcal{R}}$.

[4] To show this, we first observe that conjugation by γ generates an automorphism of $SU(4)_{\mathcal{R}}$. This automorphism is necessarily inner, as the classical theory has no symmetry that acts trivially on spacetime and by an outer automorphism of $SU(4)_{\mathcal{R}}$. Finally, given that γ generates an inner automorphism of $SU(4)_{\mathcal{R}}$, we can combine it with an $SU(4)_{\mathcal{R}}$ element that generates the inverse automorphism to get a duality symmetry that commutes with $SU(4)_{\mathcal{R}}$.

The center is generated by an element $\mathcal{I}$ that acts as i on the $\mathbf{4}$ of $SU(4)_{\mathcal{R}}$ and $-i$ on the $\overline{\mathbf{4}}$. Thus $\widehat{\phi}(\gamma)$ is defined up to $\widehat{\phi}(\gamma) \to \widehat{\phi}(\gamma) + \pi/2$.

To determine $\widehat{\phi}(\gamma)$, we can compute the action of γ in any convenient state. We choose to perform the computation on the Coulomb branch of the theory, where the gauge group is broken to $U(1)^r$ and the supersymmetry algebra is centrally extended by electric and magnetic charges. For an abelian gauge group such as $U(1)$ or $U(1)^r$, one can calculate everything explicitly and thereby determine how γ acts on the supersymmetries. Or one can use the realization of $\mathcal{N} = 4$ super Yang-Mills theory as the gauge theory on a $D3$-brane of Type IIB superstring theory; this gives a geometrical way to determine the action of γ on supercharges for $G = U(1)$ (and this approach also extends directly to classical groups such as $U(N)$).

We will follow the alternative approach of determining the result by examining the mass formula for BPS states. To simplify notation, we focus on an $\mathcal{N} = 2$ subalgebra of the supersymmetry algebra, which has a pair of right-handed supercharges $Q^i_{\dot{A}}$, $i = 1, 2$ and a single central charge Z. They satisfy

$$\{Q^i_{\dot{A}}, Q^j_{\dot{B}}\} = \epsilon_{\dot{A}\dot{B}}\epsilon^{ij} Z. \tag{2.18}$$

The complex scalar which is the $\mathcal{N} = 2$ superpartner of the massless gauge fields takes values in the Cartan subalgebra $\mathfrak{t}$ of the Lie algebra $\mathfrak{g}$. We normalize this scalar so that its kinetic term is τ-independent and denote its expectation value by $\vec{\phi}$. The electric and magnetic charges, denoted by $\vec{\mathfrak{n}}$, $\vec{\mathfrak{m}}$, take values in the weight and coweight lattices of $\mathfrak{g}$. The vectors $\vec{\phi}, \vec{\mathfrak{n}}, \vec{\mathfrak{m}}$ are defined up to an action of the Weyl group.

To make our conventions clear, we will give the explicit definition of $\mathfrak{n}$ and $\mathfrak{m}$ for $G = U(1)$. Physically, one usually describes $U(1)$ gauge theory with a real connection A' of curvature $F' = dA'$. We write Minkowski spacetime as $\mathbb{R}^{1,3} = \mathbb{R} \times \mathbb{R}^3$ where the first factor parametrizes time and the second parametrizes space. To agree with Maxwell's equations (see, for example, p. 42 of [51]), one defines the electric and magnetic fields by $E_i = F'^{0i}$ and $B_i = \frac{1}{2}\epsilon_{ijk}F'_{jk}$ (here ϵ_{ijk} is the antisymmetric tensor on $\mathbb{R}^3$ with $\epsilon_{123} = 1$). Then $\mathfrak{n}$ and $\mathfrak{m}$ are defined by

$$\begin{aligned} \mathfrak{n} &= \frac{1}{4\pi}\int_Y d\Omega\, \vec{n}\cdot\vec{E} \\ \mathfrak{m} &= \frac{1}{4\pi}\int_Y d\Omega\, \vec{n}\cdot\vec{B} \end{aligned} \tag{2.19}$$

where Y is a large sphere at infinity in $\mathbb{R}^3$ with volume-form $d\Omega$ and normal vector $\vec{n}$. In terms of the antihermitian gauge field A related to A' by $A = -iA'$ and $F = -iF'$ (recall eqns. (2.2) and (2.3)), we have

$$\begin{aligned} \mathfrak{n} &= \frac{i}{4\pi}\int_Y d\Omega\, n^i F^{0i} \\ \mathfrak{m} &= \frac{i}{4\pi}\int_Y d\Omega\, \frac{1}{2}\epsilon_{ijk} n^i F^{jk}. \end{aligned} \tag{2.20}$$

In terms of differential forms, the second formula is

$$\mathfrak{m} = \frac{i}{2\pi}\int_Y F = \int_Y c_1(\mathcal{L}), \tag{2.21}$$

where $\mathcal{L}$ is a line bundle on which A is a connection.

The central charge Z is given by

$$Z = \sqrt{\frac{2}{\mathrm{Im}\,\tau}}\ (\vec{\mathfrak{m}}\ \vec{\mathfrak{n}})\cdot\begin{pmatrix}\tau\vec{\phi}\\ \vec{\phi}\end{pmatrix} = \sqrt{\frac{2}{\mathrm{Im}\,\tau}}\ \vec{\phi}\cdot(\vec{\mathfrak{n}} + \tau\vec{\mathfrak{m}}) \tag{2.22}$$

For simply-laced G, the low energy abelian gauge theory on the Coulomb branch admits an $SL(2,\mathbb{Z})$ group of duality symmetries such that a general element $\mathsf{M} = \begin{pmatrix} a & b \\ c & d\end{pmatrix}$, which acts on τ by $\tau \to (a\tau + b)/(c\tau + d)$, acts on $\vec{\phi}$, $\vec{\mathfrak{m}}$, $\vec{\mathfrak{n}}$ by

$$\begin{aligned} \vec{\phi} &\longrightarrow \vec{\phi} \\ (\,\vec{\mathfrak{m}}\quad \vec{\mathfrak{n}}\,) &\longrightarrow (\,\vec{\mathfrak{m}}\quad \vec{\mathfrak{n}}\,)\,\mathsf{M}^{-1} = (\,\vec{\mathfrak{m}}\quad \vec{\mathfrak{n}}\,)\begin{pmatrix} d & -b \\ -c & a\end{pmatrix}. \end{aligned} \tag{2.23}$$

The S-duality conjecture asserts that this symmetry of the low energy theory actually extends to a symmetry of the full theory, necessarily acting in the same way. That the generator $T : \tau \to \tau + 1$ of the duality group Γ extends to the full theory is clear (since it is just a 2π shift in the θ angle), so the conjecture really is that $S : \tau \to -1/\tau$ similarly extends to the full theory. The action of T claimed in (2.23) can be seen in the full theory by a direct computation [50]. This direct computation is valid for all G, simply-laced or not; a variant of it will be presented in section 6.2. From (2.22) and (2.23), one deduces the transformation of the central charge:

$$Z \to \frac{|c\tau + d|}{c\tau + d}\, Z. \tag{2.24}$$

Then the centrally-extended supersymmetry algebra implies that the right-handed supersymmetries get multiplied by

$$\exp(-i\widehat{\phi}(\gamma)) = \left(\frac{|c\tau + d|}{c\tau + d}\right)^{1/2}. \tag{2.25}$$

The square root means that the group that acts on the supersymmetries is a double cover of the duality group Γ. The extension is by a symmetry $(-1)^F$ of order two that changes the sign of all fermions. This symmetry is the *square* of the generator $\mathcal{I}$ of the center of $SU(4)_{\mathcal{R}}$. We avoided getting an extension of Γ by the full center by taking $\vec{\phi}$ to be invariant under Γ rather than invariant up to sign.

In case G is not simply-laced, this argument needs to be phrased more carefully. The duality group is generated by $T : \tau \to \tau + 1$ and $S : \tau \to -1/n_{\mathfrak{g}}\tau$. T leaves fixed the supersymmetries, so $\widehat{\phi} = 0$ for T, in keeping with (2.25). As for S, its action is now

$$\begin{aligned} \vec{\phi} &\longmapsto \mathfrak{R}\cdot\vec{\phi} \\ (\,\vec{\mathfrak{m}} \quad \vec{\mathfrak{n}}\,) &\longmapsto (\,\mathfrak{R}\cdot\vec{\mathfrak{m}} \quad \mathfrak{R}\cdot\vec{\mathfrak{n}}\,)\begin{pmatrix} 0 & -1/\sqrt{n_{\mathfrak{g}}} \\ \sqrt{n_{\mathfrak{g}}} & 0 \end{pmatrix}. \end{aligned} \tag{2.26}$$

Here $\mathfrak{R}$ is an orthogonal transformation of $\mathfrak{t}$ which for simply-laced groups can be taken to be trivial, and for non-simply-laced groups is described in [47]. The square of $\mathfrak{R}$ belongs to the Weyl group, so the square of S acts trivially on the moduli space. $\mathfrak{R}$ does not affect the computation of the transformation of the central charge, and a small computation shows that $n_{\mathfrak{g}}$ cancels out, so (2.25) remains valid in the general case.

3. Topological Field Theory From $\mathcal{N} = 4$ Super Yang-Mills Theory

Our next goal is to describe how to construct a four-dimensional topological quantum field theory (TQFT) on any four-manifold[5] M by twisting of $\mathcal{N} = 4$ super Yang-Mills theory. In constructing TQFT's, it is most natural to use

[5] We generally assume M to be oriented, as the construction is more interesting in this case. However, if we specialize to the case that the theta angle of eqn. (2.11) vanishes and the canonical parameter Ψ of section 3.5 equals 0 or ∞, then the construction makes sense even for unorientable M.

Euclidean signature. Topological field theory is most naturally related to Euclidean signature, and in any event the twisting we use does not work well in Lorentz signature. When we specialize to $M = \mathbb{R} \times W$ or $\mathbb{R}^2 \times C$ for a three-manifold W or a two-manifold C, we can usefully return to Lorentz signature by taking Lorentz signature on $\mathbb{R}$ or $\mathbb{R}^2$.

One important change in going to Euclidean signature is that the spin representations $(\mathbf{2},\mathbf{1})$ and $(\mathbf{1},\mathbf{2})$ of $Spin(4)$ are pseudoreal, while for $Spin(1,3)$ these representations are complex conjugates of each other. Correspondingly, the operator $\Gamma_0\Gamma_1\Gamma_2\Gamma_3$ that distinguishes the two spin representations now squares to $+1$; to avoid confusion, we will call it $\widehat{\Gamma}_E$ when using Euclidean signature and restrict $\widehat{\Gamma}$ for the Lorentz signature case. Similarly, the two spin representations $\mathcal{S}^{\pm}$ are complex conjugates of each other in Euclidean signature, and the operator $\Gamma_0\Gamma_1\dots\Gamma_9$ that distinguishes them, which we now call $\overline{\Gamma}_E$, has eigenvalues $\mp i$ on $\mathcal{S}^{\pm}$.

To determine the sign, letting a subscript L or E refer to Lorentz or Euclidean signature, we make the Wick rotation from Lorentz to Euclidean signature by $ix^0_L = x^0_E$, so $\partial/\partial x^0_L = i\partial/\partial x^0_E$ and $\Gamma_{0\,L} = i\Gamma_{0\,E}$. Hence, $\overline{\Gamma}_L = i\overline{\Gamma}_E$, and as $\overline{\Gamma} = \overline{\Gamma}_L$ acts on $\mathcal{S}^{\pm}$ as multiplication by ± 1, $\overline{\Gamma}_E$ acts by multiplication by $\mp i$. In particular, if ϵ is a supersymmetry generator, then

$$\overline{\Gamma}_E\epsilon = -i\epsilon. \tag{3.1}$$

3.1. Twisting $\mathcal{N} = 4$ Super Yang-Mills

General Idea Of Twisting $\mathcal{N} = 2$ super Yang-Mills theory can be twisted to make a quantum field theory realization of Donaldson theory [25]. Similarly [45], $\mathcal{N} = 4$ super Yang-Mills theory can be twisted in three ways to make a topological field theory. Two of the twisted theories, including one that was investigated in detail in [45], are closely analogous to Donaldson theory in the sense that they lead to instanton invariants which, like the Donaldson invariants of four-manifolds, can be expressed in terms of the Seiberg-Witten invariants. The third twist, which was mentioned in [52] and has been investigated in [53–57], has had no applications until now. It turns out to be the twist relevant to the geometric Langlands program, and we will call it the GL twist.

To give an inevitably very brief explanation of the notion of twisting, we first take $M = \mathbb{R}^4$. This has rotational symmetry group $Spin(4)$, of course, while the $\mathcal{N} = 4$ theory has the larger symmetry $Spin(4) \times Spin(6)$. "Twisting" means replacing $Spin(4)$ by a different subgroup $Spin'(4)$ of $Spin(4) \times Spin(6)$ that is isomorphic to $Spin(4)$ and acts on $\mathbb{R}^4$ the same

way that $Spin(4)$ does, but acts differently on the $\mathcal{N} = 4$ gauge theory. To accomplish this, we pick a homomorphism $\varkappa : Spin(4) \to Spin(6)$ and set $Spin'(4) = (1 \times \varkappa)(Spin(4)) \subset Spin(4) \times Spin(6)$.

We also want to pick $\varkappa$ such that the action of $Spin'(4)$ on $\mathcal{S}^+$ has a non-zero invariant vector. Since the supersymmetry generator ϵ takes values in $\mathcal{S}^+$, a choice of an invariant vector in $\mathcal{S}^+$ will give us a $Spin'(4)$-invariant supersymmetry that we will call Q. It will automatically obey $Q^2 = 0$, by virtue of (2.15). (There is no $Spin'(4)$ invariant operator on the right hand side of (2.15), so as $Q^2 = \frac{1}{2}\{Q, Q\}$ is $Spin'(4)$-invariant, it must vanish.) Once Q is chosen, we change the physical interpretation of the theory to say that we are only interested in Q-invariant path integrals, operators, and states, and that we consider anything of the form $[Q, \mathcal{O}\}$, for any operator $\mathcal{O}$, to be trivial. So henceforth, the interesting operators or states lie in suitable cohomology groups of Q.

It turns out that theories obtained this way are, loosely speaking, topological field theories. (They may depend on the smooth structure of M, which goes into the definition of the quantum field theory, before or after twisting.) This is proved by showing that the definition of the theory on flat $\mathbb{R}^4$ can be extended to any four-manifold in such a way that the Q symmetry is retained and the choice of metric is irrelevant modulo Q-exact terms. For the GL twist of $\mathcal{N} = 4$ super Yang-Mills theory, one approach to this can be found in section 7 of [56]. In the present paper, we will take another route; once we have worked out the topological equations and the vanishing theorems, we will see directly, in section 3.4, what action on a curved four-manifold has the right properties.

Description Of Twist Now we will describe the GL twist. The homomorphism $\varkappa : Spin(4) \to Spin(6)$ is chosen so that the $\mathbf{4}$ of $Spin(6) = SU(4)_{\mathcal{R}}$ transforms as $(\mathbf{2}, \mathbf{1}) \oplus (\mathbf{1}, \mathbf{2})$ of $Spin(4) = SU(2) \times SU(2)$, which we will refer to as $SU(2)_\ell \times SU(2)_r$.[6] (The $\overline{\mathbf{4}}$ of $Spin(6)$, which is the complex conjugate of the $\mathbf{4}$, transforms the same way under $Spin(4)$, since the $(\mathbf{2}, \mathbf{1})$ and $(\mathbf{1}, \mathbf{2})$ of $Spin(4)$ are pseudoreal.)

This choice of $\varkappa$ amounts to embedding $SU(2)_\ell \times SU(2)_r$ in $Spin(6) = SU(4)_{\mathcal{R}}$ in the following way:

$$\begin{pmatrix} SU(2)_\ell & 0 \\ 0 & SU(2)_r \end{pmatrix}. \tag{3.2}$$

[6]The "ℓ" and "r" in $SU(2)_\ell$ and $SU(2)_r$ are usually read as "left" and "right," referring to the spin of physical massless particles.

This embedding obviously commutes with an additional $U(1)$ group,[7] whose generator $\mathcal{K}$ we can take to have the form

$$\mathcal{K} = i \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{3.3}$$

in 2×2 blocks. So our embedding is such that the $\mathbf{4}$ of $Spin(6)$ transforms under $SU(2)_\ell \times SU(2)_r \times U(1)$ as $(\mathbf{2},\mathbf{1})^1 \oplus (\mathbf{1},\mathbf{2})^{-1}$, where now the exponent is the eigenvalue of $-i\mathcal{K}$, which we will call $\mathcal{K}$. The $\overline{\mathbf{4}}$ transforms as the complex conjugate of this, or $(\mathbf{2},\mathbf{1})^{-1} \oplus (\mathbf{1},\mathbf{2})^1$.

This twist can also be conveniently described in terms of SO groups rather than $Spin$ groups. To do so, we use the fact that the fundamental, six-dimensional vector representation $\mathbf{6_v}$ of $SO(6)$ is, in terms of $Spin(6) = SU(4)_{\mathcal{R}}$, the same as $\wedge^2\mathbf{4}$, the skew-symmetric part of $\mathbf{4} \otimes \mathbf{4}$. So $\mathbf{6_v} = \wedge^2((\mathbf{2},\mathbf{1})^1 \oplus (\mathbf{1},\mathbf{2})^{-1}) = (\mathbf{2},\mathbf{2})^0 \oplus (\mathbf{1},\mathbf{1})^2 \oplus (\mathbf{1},\mathbf{1})^{-2}$. Here $(\mathbf{2},\mathbf{2})$ is the same as the vector representation $\mathbf{4_v}$ of $SO(4)$. So the $\mathbf{6_v}$ is the sum of a vector $\mathbf{4_v}$ of $SO(4)$ with $\mathcal{K} = 0$, and two $SO(4)$ scalars with $\mathcal{K} = \pm 2$. This corresponds to the obvious homomorphism of $SO(4) \times U(1) = SO(4) \times SO(2)$ to $SO(6)$:[8]

$$\begin{pmatrix} SO(4) & 0 \\ 0 & SO(2) \end{pmatrix}. \tag{3.4}$$

The six spin zero fields ϕ_i, $i = 0, \dots, 5$ of $\mathcal{N} = 4$ super Yang-Mills theory transform as $\mathbf{6_v}$ of $SO(6)$, so this analysis applies to them. We can pick coordinates so the first four fields $\phi_0, \dots, \phi_3$ form a $\mathbf{4_v}$ of $SO(4)$, while ϕ_4 and ϕ_5 are the $SO(4)$ scalars, which are rotated by $SO(2)$. Moreover, we can label the scalars so that the linear combinations $\sigma = (\phi_4 - i\phi_5)/\sqrt{2}$ and $\overline{\sigma} = (\phi_4 + i\phi_5)/\sqrt{2}$ have $\mathcal{K} = 2$ and $\mathcal{K} = -2$. The fields $\phi_0, \dots, \phi_3$ can then be interpreted geometrically as the components of an adjoint-valued one-form, while σ is a scalar field or zero-form with values in the complexification of the Lie algebra. So the bosonic fields of the theory are a gauge field, which locally is an adjoint-valued one-form $A = A_\mu\, dx^\mu$, along with a second adjoint-valued one-form $\phi = \phi_\mu\, dx^\mu$, and the complex scalars σ and $\overline{\sigma}$.

[7] The global structure of the combined group is actually not a product $SU(2)_\ell \times SU(2)_r \times U(1)$, but rather the quotient $(SU(2)_\ell \times SU(2)_r \times U(1))/\mathbb{Z}_2$.

[8] The $U(1)$ in $(SU(2) \times SU(2) \times U(1))/\mathbb{Z}_2 \subset SU(4)$ is a double cover of the $U(1)$ in $SO(4) \times U(1) \subset SO(6)$.

Transformation Of The Supersymmetries We can likewise analyze how the supersymmetries transform under $Spin'(4)$. The $\overline{\mathbf{4}}$ of $Spin(6)$ transforms as $(\mathbf{2},\mathbf{1})^{-1}\oplus(\mathbf{1},\mathbf{2})^{1}$ of $Spin'(4)\times U(1)$, and the $\mathbf{4}$ as $(\mathbf{2},\mathbf{1})^{1}\oplus(\mathbf{1},\mathbf{2})^{-1}$. So, using (2.14), the supersymmetries that transform as $(\mathbf{2},\mathbf{1})$ of $Spin(4)$ transform under $Spin'(4)\times U(1)$ as

$$(\mathbf{2},\mathbf{1})^{0}\otimes\left((\mathbf{2},\mathbf{1})^{-1}\oplus(\mathbf{1},\mathbf{2})^{1}\right)=(\mathbf{1},\mathbf{1})^{-1}\oplus(\mathbf{3},\mathbf{1})^{-1}\oplus(\mathbf{2},\mathbf{2})^{1}. \tag{3.5}$$

And the supersymmetries that transform as $(\mathbf{1},\mathbf{2})$ of $Spin(4)$ transform under $Spin'(4)\times U(1)$ as

$$(\mathbf{1},\mathbf{2})^{0}\otimes\left((\mathbf{2},\mathbf{1})^{1}\oplus(\mathbf{1},\mathbf{2})^{-1}\right)=(\mathbf{1},\mathbf{1})^{-1}\oplus(\mathbf{1},\mathbf{3})^{-1}\oplus(\mathbf{2},\mathbf{2})^{1}. \tag{3.6}$$

Next, we can find the $Spin'(4)$-invariant supersymmetries. From (3.5) and (3.6), there is one invariant supersymmetry generator ϵ_ℓ that is left-handed in the four-dimensional sense (transforms as $(\mathbf{2},\mathbf{1})$ under the original $Spin(4)$), and one such generator ϵ_r that is right-handed (transforms as $(\mathbf{1},\mathbf{2})$). With our conventions, they are distinguished by

$$\begin{aligned}\widehat{\Gamma}_E\epsilon_\ell&=-\epsilon_\ell\\ \widehat{\Gamma}_E\epsilon_r&=\epsilon_r.\end{aligned} \tag{3.7}$$

A choice of ϵ_ℓ determines a natural choice of ϵ_r, namely

$$\epsilon_r=N\epsilon_\ell, \tag{3.8}$$

where $N=\frac{1}{4}\sum_{\mu=0}^{3}\Gamma_{\mu+4}\Gamma_\mu$. This idea here is that N commutes with $Spin'(4)$ but anticommutes with $\widehat{\Gamma}_E$, so the definition (3.8) makes sense. We have normalized N so that $N^2\epsilon=-\epsilon$ if ϵ is $Spin'(4)$-invariant. Hence we have also

$$\epsilon_\ell=-N\epsilon_r. \tag{3.9}$$

By exploiting the $Spin'(4)$ symmetry, one can show that for $\mu=0,\dots,3$, we have

$$\Gamma_{\mu+4}\epsilon_\ell=-\Gamma_\mu\epsilon_r,\ \Gamma_{\mu+4}\epsilon_r=\Gamma_\mu\epsilon_\ell. \tag{3.10}$$

The generators of $Spin'(4)$ are $\Gamma_{\mu\nu}+\Gamma_{\mu+4,\nu+4}$, for $\mu,\nu=0,\dots,3$, so a $Spin'(4)$-invariant spinor ϵ obeys $(\Gamma_{\mu\nu}+\Gamma_{\mu+4,\nu+4})\epsilon=0$. Setting $\mu,\nu=0,1$,

we learn that $\Gamma_{0145}\epsilon = \epsilon$, and setting $\mu, \nu = 2, 3$, we learn that $\Gamma_{2367}\epsilon = \epsilon$. So $\Gamma_0\Gamma_1 \dots \Gamma_7\epsilon = \epsilon$, and using also the chirality condition $\Gamma_0\Gamma_1 \dots \Gamma_9\epsilon = -i\epsilon$ of eqn. (3.1), we learn that

$$\Gamma_{89}\epsilon = -i\epsilon. \tag{3.11}$$

It follows in particular that $\mathcal{K} = \Gamma_{89}$ acts with the same eigenvalue $\mathcal{K} = -i\mathcal{K} = -1$ on ϵ_ℓ and ϵ_r. It also follows that $\Gamma_{8-i9}\epsilon = 0$, where $\Gamma_{8-i9} = (\Gamma_8 - i\Gamma_9)/\sqrt{2}$. These facts are helpful in verifying the supersymmetry transformation laws presented below.

We now can see why this construction gives a family of topological field theories parametrized by $\mathbb{CP}^1$ (Figure 1). We pick any nonzero complex linear combination of ϵ_ℓ and ϵ_r

$$\epsilon = u\epsilon_\ell + v\epsilon_r, \tag{3.12}$$

and take this to be the generator of the topological symmetry. An overall scaling of ϵ would not matter, so the possible choices for the topological symmetry are parametrized by $\mathbb{CP}^1$. Because there is a natural choice of ϵ_r once ϵ_ℓ is given, this has much more structure than an abstract copy of $\mathbb{CP}^1$; it has a natural affine parameter $t = v/u$.

We can use (2.25) to determine the transformation of u, v, and t under S-duality. Under an element γ of the duality group Γ, we have $\epsilon_\ell \to \exp(i\widehat{\phi}(\gamma))\epsilon_\ell$, $\epsilon_r \to \exp(-i\widehat{\phi}(\gamma))\epsilon_r$. Hence, to ensure invariance of $\epsilon = u\epsilon_\ell + v\epsilon_r$, we have $u \to \exp(-i\widehat{\phi}(\gamma))u$, $v \to \exp(i\widehat{\phi}(\gamma))v$, and finally the

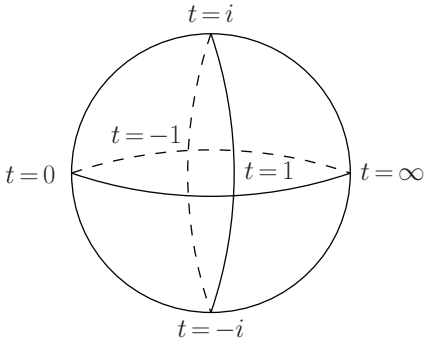

Figure 1: A family of topological field theories parametrized by the variable t.

transformation of $t = v/u$ is

$$t \longrightarrow \exp(2i\widehat{\phi}(\gamma))t = \frac{c\tau + d}{|c\tau + d|}t. \tag{3.13}$$

Similarly, we can determine the action on t of the center of $SU(4)_{\mathcal{R}}$. The generator $\mathcal{I}$ of the center acts by $\epsilon_\ell \to i\epsilon_\ell$, $\epsilon_r \to i^{-1}\epsilon_r$, so by a similar reasoning to the above, we have

$$\mathcal{I}(t) = -t. \tag{3.14}$$

Thus, the topological field theories with parameter t or $-t$ are equivalent. This is a trivial equivalence in the sense that it follows from a symmetry of the classical action – as opposed to a non-trivial S-duality. The $\mathbb{Z}_2$ symmetry (3.14) is the only trivial equivalence among the twisted topological field theories parametrized by t. In making this assertion, it is essential that ϵ_ℓ and ϵ_r have the same eigenvalue of $\mathcal{K}$. Otherwise, a transformation in the group $\mathbb{C}^*$ generated by $\mathcal{K}$ would act non-trivially on t and would cause the whole $\mathbb{CP}^1$ family of topological field theories to be trivially equivalent. Of course, it is only because there are (almost) no trivial equivalences that it is possible for the non-trivial equivalences coming from S-duality to generate something interesting like the geometric Langlands program.

We write Q_ℓ and Q_r for the supersymmetries generated by ϵ_ℓ and ϵ_r. The supersymmetry generated by the linear combination $\epsilon = u\epsilon_\ell + v\epsilon_r$ is then

$$Q = uQ_\ell + vQ_r, \tag{3.15}$$

and this is the topological symmetry of the twisted theory. It is also convenient to write

$$\delta_T = \epsilon_\ell\delta_\ell + \epsilon_r\delta_r, \tag{3.16}$$

where for any field Φ,

$$\delta_T\Phi = [Q, \Phi\}, \quad \delta_\ell\Phi = [Q_\ell, \Phi\}, \quad \delta_r\Phi = [Q_r, \Phi\}. \tag{3.17}$$

(T stands for twisted or topological.) Since ϵ_ℓ and ϵ_r have $\mathcal{K} = -1$, it follows that Q_ℓ, Q_r, and Q all have $\mathcal{K} = 1$. Indeed, the supersymmetry transformation laws presented below show explicitly that δ_T increases $\mathcal{K}$ by 1 for every field. Thus in mathematical language, the space of gauge-invariant operators or states is a complex, $\mathbb{Z}$-graded by $\mathcal{K}$, with differential Q.

The Fermion Fields And The Supersymmetry Transformations The fermion fields λ transform the same way as the supersymmetry generators ϵ, so using (3.5) and (3.6), we can identify them in four-dimensional terms. The fields with $\mathcal{K}=1$ are two copies of $(\mathbf{2},\mathbf{2})^1$, corresponding to two one-forms ψ and $\widetilde{\psi}$. The fields transforming as $(\mathbf{3},\mathbf{1})^{-1}$ and $(\mathbf{1},\mathbf{3})^{-1}$ are the selfdual and anti-selfdual parts $\chi^{\pm}$ of a two-form χ of $\mathcal{K}=-1$. Finally, there are two zero-forms η and $\widetilde{\eta}$ of $\mathcal{K}=-1$.

It is helpful to define these four-dimensional fields more precisely by specifying the following expansion of the ten-dimensional spinor field λ in terms of four-dimensional fermion fields:[9]

$$\begin{aligned}\lambda = &\left(\eta+\sum_\mu \Gamma^\mu\psi_\mu\Gamma_{8+i9}+\sum_{\mu<\nu}\Gamma^{\mu\nu}\chi^+_{\mu\nu}\right)\epsilon_\ell\\ &+\left(\widetilde{\eta}+\sum_\mu \Gamma^\mu\widetilde{\psi}_\mu\Gamma_{8+i9}+\sum_{\mu<\nu}\Gamma^{\mu\nu}\chi^-_{\mu\nu}\right)\epsilon_r.\end{aligned}\tag{3.18}$$

This formula uniquely determines sixteen fields (η and $\widetilde{\eta}$, four components each of ψ and $\widetilde{\psi}$, and six components of χ) in terms of the sixteen components of λ. Note that, because $\widehat{\Gamma}_E\epsilon_\ell=-\epsilon_\ell$, $\Gamma_{\mu\nu}\epsilon_\ell$ is selfdual, so $\Gamma^{\mu\nu}\chi_{\mu\nu}\epsilon_\ell=\Gamma^{\mu\nu}\chi^+_{\mu\nu}\epsilon_\ell$. Likewise $\Gamma^{\mu\nu}\chi_{\mu\nu}\epsilon_r=\Gamma^{\mu\nu}\chi^-_{\mu\nu}\epsilon_r$. So one could equivalently replace $\chi^\pm$ by χ in (3.18).

It is convenient to normalize ϵ_ℓ up to sign so that

$$\overline{\epsilon}_\ell\Gamma_{8+i9}\epsilon_\ell=1.\tag{3.19}$$

Then it follows from $\epsilon_r=N\epsilon_\ell$ that likewise

$$\overline{\epsilon}_r\Gamma_{8+i9}\epsilon_r=1.\tag{3.20}$$

From considerations of four-dimensional chirality, it follows that

$$\overline{\epsilon}_\ell\Gamma_{8+i9}\epsilon_r=\overline{\epsilon}_r\Gamma_{8+i9}\epsilon_\ell=0.\tag{3.21}$$

The next step is to describe how the fields transform under the topological symmetries, that is, the supersymmetries generated by ϵ_ℓ and ϵ_r. For

[9] Since λ, ϵ_ℓ, and ϵ_r are all of positive chirality, this expansion must be made using only even elements of the Clifford algebra. To ensure this, certain factors of $\Gamma_{8+i9}=\frac{1}{\sqrt{2}}(\Gamma_8+i\Gamma_9)$ are included in the expansion.

example, to transform the bosons, we start with $\delta_S A_I = i\overline{\epsilon}\Gamma_I\lambda$. Upon setting $\epsilon = u\epsilon_\ell + v\epsilon_r$, we get for the four-dimensional gauge field

$$\delta_T A_\mu = iu\overline{\epsilon}_\ell\Gamma_\mu\lambda + iv\overline{\epsilon}_r\Gamma_\mu\lambda. \tag{3.22}$$

After some Γ-matrix gymnastics, one finds the transformation of the four-dimensional gauge field under the topological supersymmetry:

$$\delta_T A_\mu = iu\psi_\mu + iv\widetilde{\psi}_\mu. \tag{3.23}$$

Similarly, for the other bosonic fields we find

$$\delta_T \phi_\mu = iv\psi_\mu - iu\widetilde{\psi}_\mu. \tag{3.24}$$

$$\delta_T \sigma = 0$$

$$\delta_T \overline{\sigma} = iu\eta + iv\widetilde{\eta}. \tag{3.25}$$

The vanishing of $\delta_T\sigma$ actually follows merely from the fact that σ has $\mathcal{K} = 2$, and there is no appropriate field of $\mathcal{K} = 3$.

Likewise, we can find the transformations of the four-dimensional fermionic fields starting with the underlying transformation law

$$\delta_T\lambda = \frac{1}{2}\sum_{I,J=0}^{9} F_{IJ}\Gamma^{IJ}\epsilon, \tag{3.26}$$

where $\epsilon = u\epsilon_\ell + v\epsilon_r$. The fermionic fields with $\mathcal{K} = -1$ transform by

$$\begin{aligned}
\delta_T\chi^+ &= u(F - \phi\wedge\phi)^+ + v(D\phi)^+ \\
\delta_T\chi^- &= v(F - \phi\wedge\phi)^- - u(D\phi)^- \\
\delta_T\eta &= vD^*\phi + u[\overline{\sigma},\sigma] \\
\delta_T\widetilde{\eta} &= -uD^*\phi + v[\overline{\sigma},\sigma].
\end{aligned} \tag{3.27}$$

Those of $\mathcal{K} = 1$ transform by

$$\begin{aligned}
\delta_T\psi &= u\,D\sigma + v[\phi,\sigma] \\
\delta_T\widetilde{\psi} &= v\,D\sigma - u[\phi,\sigma].
\end{aligned} \tag{3.28}$$

In these formulas, $D = d + [A,\,\cdot\,]$ is the covariant derivative with respect to the connection A; a Lie bracket is implicit in the adjoint-valued two-form $\phi\wedge\phi$; and $D^*\phi = \star D\star\phi = D_\mu\phi^\mu$, with $\star$ the Hodge star.

3.2. A Family Of A-Models

We are in the following general situation. We have a quantum theory with a fermionic symmetry $Q = uQ_\ell + vQ_r$ that obeys $Q^2 = 0$ (modulo a gauge transformation). In such a situation, in general, the path integral localizes on fields that are invariant under Q. When the space Y of Q-invariant fields is smooth and the gauge group acts freely on it, the path integral can be evaluated by a Gaussian or one-loop approximation, expanding around Y. More generally, one might have to go to higher order, but the path integral is always determined by the structure of a finite order infinitesimal neighborhood of Y.

It suffices to identify the bosonic fields that are invariant under Q, since fermions are in any case infinitesimal, and automatically are incorporated in perturbation theory. The condition for a bosonic field to be invariant under Q is that $\delta_T\Psi$ vanishes for every fermionic field Ψ. In other words, Y is defined by setting to zero the right hand sides of (3.27) and (3.28).

For these expressions to vanish is a condition that is invariant under scaling of u and v and so only depends on $t = v/u$. Vanishing of the right hand side of (3.27) gives[10]

$$\begin{aligned} (F - \phi\wedge\phi + tD\phi)^+ &= 0 \\ (F - \phi\wedge\phi - t^{-1}D\phi)^- &= 0 \\ D^*\phi &= 0. \end{aligned} \tag{3.29}$$

The most elementary interpretation of these equations is for $-t^{-1} = t$, or equivalently $t = \pm i$. For these values of t, which will be important for the geometric Langlands program, let $\mathcal{A}$ be the $G_{\mathbb{C}}$-valued connection $\mathcal{A} = A + i\phi$, with curvature $\mathcal{F} = d\mathcal{A} + \mathcal{A}\wedge\mathcal{A}$. Then the first two equations in (3.29) are equivalent to $\mathcal{F} = 0$, so that a solution of (3.29) at $t = \pm i$ determines a complex-valued flat connection and hence a homomorphism $\pi_1(M) \to G_{\mathbb{C}}$. Eqn. (3.29) instructs us to impose the further condition $D^*\phi = 0$ and to divide by G-valued gauge transformations. The combined operation constructs the moduli space $\mathcal{Y}$ of homomorphisms $\vartheta : \pi_1(M) \to G_{\mathbb{C}}$ (we refer to this space as $\mathcal{Y}_M$ or $\mathcal{Y}_M(G)$ if greater clarity is needed). This statement depends on a theorem of Corlette [58] (the two-dimensional case was

[10] To get the third equation, we take a t-dependent linear combination of the conditions $\delta_T\eta = 0$ and $\delta_T\widetilde{\eta} = 0$, if $t \neq \pm i$. What happens if $t = \pm i$ is discussed below.

also proved by Donaldson in the appendix to [20]): dividing by G-valued gauge transformations on the pair A, ϕ, requiring that $\mathcal{F} = 0$ and imposing $D^*\phi = 0$ is equivalent to simply dropping that last equation, imposing $\mathcal{F} = 0$ and a condition of stability[11] and dividing by the group of $G_{\mathbb{C}}$-valued gauge transformations.

More generally, if t is not real, then the equations (3.29) for fields A, ϕ valued in the real Lie algebra of G are overdetermined, rather than elliptic. Setting to zero separately the hermitian and antihermitian parts of these equations, we learn that they imply $\mathcal{F} = 0$ for any non-real t.

We want to understand, however, what happens when t is real. (The most important values of t for the geometric Langlands program are $t = \pm i$ and $t = \pm 1$.) In this case, the equations are real and we get a family of real elliptic equations parametrized by t. As written, these equations are regular for $t \neq 0, \infty$. However, as is clear from the homogeneous expressions on the right hand side of (3.27), the equations can be extended over $t = 0$ or $t = \infty$, by multiplying the second equation by t or the first by t^{-1}. Thus, at $t = 0$, we should replace the second equation by $(D\phi)^- = 0$, and at $t = \infty$, we replace the first by $(D\phi)^+ = 0$. So we get a family of real elliptic equations parametrized by $\mathbb{RP}^1$.

We can similarly understand the conditions for unbroken supersymmetry, which come from (3.28):

$$\begin{aligned} D\sigma + t[\phi, \sigma] &= 0 \\ D\sigma - t^{-1}[\phi, \sigma] &= 0. \end{aligned} \tag{3.30}$$

For any t other than $\pm i$, these equations imply that $D\sigma = [\phi, \sigma] = 0$, a condition which means that the gauge symmetry generated by σ leaves invariant the given solution of (3.29). In addition, for $t \neq \pm i$, a linear combination of the equations $\delta_T \eta = 0$ and $\delta_T \widetilde{\eta} = 0$ gives

$$[\sigma, \overline{\sigma}] = 0. \tag{3.31}$$

[11] As is usual for moduli problems involving non-compact symmetry groups such as $G_{\mathbb{C}}$, the moduli space $\mathcal{Y}_M(G)$ does not literally parametrize all conjugacy classes of homomorphisms $\pi_1(M) \to G_{\mathbb{C}}$. To get a good moduli space, one must drop certain "unstable orbits," corresponding in this problem to homomorphisms that can be reduced to a triangular form but not to a direct sum. That setting $D^*\phi = 0$ and dividing by G-valued gauge transformations has the effect of dropping such unstable orbits was shown in [58]. We explain the role of stability more fully in section 4.2.

As long as we only consider supersymmetric fields that are irreducible – or more generally, those for which the automorphism group is finite – (3.30) implies that $\sigma = 0$. In this paper, we generally consider only supersymmetric fields that are irreducible. Some simple remarks on the reducible case will appear in [26].

At $t = \pm i$, the conclusion is similar although the details are different. In this case, (3.30) says that σ generates a symmetry of the flat $G_{\mathbb{C}}$-valued connection $\mathcal{A}$ or $\overline{\mathcal{A}}$. Again, as long as we only consider supersymmetric fields that are irreducible, we can assume that $\sigma = 0$. (For $t = \pm i$, in case of a reducible supersymmetric configuration with $\sigma \neq 0$, we do not get separate conditions $D^*\phi = [\sigma, \overline{\sigma}] = 0$, but only the linear combination $D^*\phi \pm i[\sigma, \overline{\sigma}] = 0$.)

For real t, the theory discussed here fits in the general framework of cohomological field theories [59], like Donaldson theory in four dimensions or A-models in two dimensions. The fields are the bosonic fields A, ϕ of $\mathcal{K} = 0$; the equations are (3.29); and the symmetries are simply the gauge symmetries. As in any such theory, the path integral in the absence of operator insertions can be evaluated by counting, with signs, the number of solutions of the equations, as long as those solutions are isolated and irreducible. The sign with which a given solution contributes is simply the sign of the corresponding one-loop fermion determinant. A smooth compact family of irreducible solutions, parametrized by a space Y_0 (which is a component of Y), makes a contribution that is plus or minus the Euler characteristic of Y_0, with the sign again coming from the fermion determinant. Contributions of singularities in Y and of reducible solutions – such as the trivial solution with $A = \phi = 0$ – are more subtle to evaluate and will not be considered here.

Since the number of solutions, weighted by sign, is invariant under continuous deformation of a family of elliptic equations, the partition function on a closed four-manifold without operator insertions must be independent of t, at least for real t. By holomorphy, this is also true for complex t. That the partition function on a closed four-manifold is independent of t can be seen more directly from the path integral, as we explain later. From a mathematical point of view, the vanishing theorems of section 3.3 will give a strong statement about t-independence: the space of solutions of (3.29) is actually independent of t except at $t = 0, \infty$.

The importance of the parameter t is that it is possible to introduce operators or, in case M has a boundary, boundary conditions, that preserve the topological symmetry only for a particular value of t. Like Donaldson theory, the TQFT considered here has local operators that preserve the topological symmetry. They have been discussed in [53-57] and will be further analyzed elsewhere [26]. However, they are not the most important

operators for the geometric Langlands program. The important operators, introduced in section 6, will be the Wilson and 't Hooft line operators, which have no obvious close analogs in more familiar cohomological field theories.

If our theory for real t is somewhat analogous to a two-dimensional A-model, what does it correspond to for complex t? The theory at $t = \pm i$ is analogous to a two-dimensional B-model. In the B-model that describes maps $\Phi : \Sigma \to X$ from a Riemann surface Σ to a complex manifold X, the supersymmetric fields are the constant maps of Σ to X. The obvious similarity between the two cases is that the constant maps or the flat $G_{\mathbb{C}}$-valued connections are relatively elementary to describe by using over-determined rather than elliptic equations (namely $d\Phi = 0$ in one case, $\mathcal{F} = 0$ in the other) in which the complex structure is obvious. A deeper analogy will be apparent when we consider dimensional reduction in section 4.

3.3. Vanishing Theorems

Let $\mathcal{V}^+(t) = (F - \phi \wedge \phi + tD\phi)^+$, $\mathcal{V}^-(t) = (F - \phi \wedge \phi - t^{-1}D\phi)^-$, and $\mathcal{V}_0 = D^*\phi$, so the supersymmetric equations are $\mathcal{V}^+(t) = \mathcal{V}^-(t) = \mathcal{V}_0 = 0$. In this section, we concentrate on the case of real t, so that the supersymmetric equations are elliptic.

Like many first order equations associated with supersymmetry, such as the equations for a holomorphic curve, the Yang-Mills instanton equations, the Seiberg-Witten equations, or the equations studied in [45], these are subject to unusual vanishing theorems. In the case at hand, we have:

Vanishing Theorem 1 Let M be a compact four-manifold without boundary, and E a G-bundle over M with nonzero Pontryagin class, $\int_M \mathrm{Tr}\, F \wedge F \neq 0$. Then for any $t \neq 0, \infty$, the supersymmetric equations $\mathcal{V}^+(t) = \mathcal{V}^-(t) = \mathcal{V}_0 = 0$ have no solutions.

Vanishing Theorem 2 On such an M, any field that obeys those equations for one value of t other than $0, \infty$ obeys them for all t and hence is given by a flat $G_{\mathbb{C}}$-valued connection.

The shortest proof of the first vanishing theorem comes from the following identity:

$$\int_M \mathrm{Tr}\, F \wedge F = \int_M \mathrm{Tr}\left(\mathcal{V}^+(t) \wedge \mathcal{V}^+(-t^{-1}) + \mathcal{V}^-(t) \wedge \mathcal{V}^-(-t^{-1})\right). \tag{3.32}$$

The proof of the identity is straightforward and uses integration by parts. Clearly, this identity implies that if for some t, $\mathcal{V}^{\pm}(t) = 0$, then $\int_M \mathrm{Tr}\, F \wedge F = 0$.

This proof fails if $t = 0, \infty$ because of the poles in $\mathcal{V}^{\pm}$. In fact, the result is false for $t = 0, \infty$; the equations at $t = 0$ or $t = \infty$ have at least the obvious solutions given by an instanton or anti-instanton with $\phi = 0$.

For the second vanishing theorem, we need a similar but more intricate identity:

(3.33)
$$\begin{aligned} &-\int_M d^4x\sqrt{g}\,\mathrm{Tr}\,\left(\frac{t^{-1}}{t+t^{-1}}\mathcal{V}^+(t)_{\mu\nu}\mathcal{V}^+(t)^{\mu\nu} + \frac{t}{t+t^{-1}}\mathcal{V}^-(t)_{\mu\nu}\mathcal{V}^-(t)^{\mu\nu} + \mathcal{V}_0^2\right) \\ &\quad = -\int d^4x\sqrt{g}\,\mathrm{Tr}\,\left(\frac{1}{2}F_{\mu\nu}F^{\mu\nu} + D_\mu\phi_\nu D^\mu\phi^\nu + R_{\mu\nu}\phi^\mu\phi^\nu + \frac{1}{2}[\phi_\mu,\phi_\nu]^2\right) \\ &\qquad + \frac{t-t^{-1}}{(t+t^{-1})}\int_M \mathrm{Tr}\, F\wedge F. \end{aligned}$$

What is surprising about this identity is that apart from the topological invariant proportional to $\int \mathrm{Tr}\, F \wedge F$, the right hand side is independent of t. The proof is similar to the proof of (3.32), but more elaborate. $R_{\mu\nu}$ is the Ricci tensor of M, which enters in integrating by parts to relate $\int_M d^4x\sqrt{g}D_\mu\phi_\nu D^\nu\phi^\mu$ to $\int d^4x\sqrt{g}(D_\mu\phi^\mu)^2$. Unlike (3.32), (3.33) has a natural limit as $t \to 0, \infty$. One simply replaces $t(\mathcal{V}^-)^2/(t+t^{-1})$ by $\lim_{t\to 0}(t\mathcal{V}^-)^2$ or $t^{-1}(\mathcal{V}^+)^2/(t+t^{-1})$ by $\lim_{t\to\infty}(t^{-1}\mathcal{V}^+)^2$. So arguments based on (3.33) are valid at $t = 0, \infty$.

In section 3.4, we will understand the physical meaning of (3.33). For now, we simply use it to complete the proof of the vanishing theorems.

To prove the second vanishing theorem, we must show that if the bundle E admits a pair A, ϕ for which $\mathcal{V}^{\pm}(t) = \mathcal{V}_0 = 0$ for some value of t other than $0, \infty$, then these equations hold for all t. We already know that in such a solution, $\int_M \mathrm{Tr}\, F \wedge F = 0$, whence the right hand side of (3.33) is independent of t. So if the left hand side vanishes for one value of t, it vanishes for all t. Moreover, the left hand side is a sum of squares, so it vanishes for given t if and only if $\mathcal{V}^{\pm}(t) = \mathcal{V}_0 = 0$. Moreover, as we have already noted, the vanishing of $\mathcal{V}^{\pm}(t)$ and $\mathcal{V}_0$ for all t is equivalent to the vanishing of the curvature $\mathcal{F} = d\mathcal{A} + \mathcal{A}\wedge\mathcal{A}$ of the $G_{\mathbb{C}}$-valued connection $\mathcal{A} = A + i\phi$. This completes the argument.

Finally, we can use (3.33) to prove a slightly stronger version of the first vanishing theorem:

Vanishing Theorem 1′ For $\int_M \mathrm{Tr}\, F\wedge F>0$, the supersymmetric equations have no solutions except at $t=0$, and for $\int_M \mathrm{Tr}\, F\wedge F<0$, they have no solutions except at $t=\infty$.

Let $\mathsf{B}(t)$ denote the left hand side of (3.33), and let $f(t)=(t-t^{-1})/(t+t^{-1})$. We have for any t,u

$$\mathsf{B}(u)-\mathsf{B}(t)=(f(u)-f(t))\int_M \mathrm{Tr}\, F\wedge F. \tag{3.34}$$

If $\mathcal{V}^{\pm}(t)=\mathcal{V}_0=0$, then $\mathsf{B}(t)=0$, and as the function B is positive semidefinite, the left hand side of (3.34) is positive semidefinite. To make the right hand side positive semidefinite, t must be a minimum of the function f if $\int_M \mathrm{Tr}\, F\wedge F>0$, and a maximum if $\int_M \mathrm{Tr}\, F\wedge F<0$. But the only minimum is at $t=0$, and the only maximum is at $t=\infty$.

A Note On Compactification A special case of the vanishing theorems is relevant to compactification. We take $M=\Sigma\times C$, where Σ and C are two compact Riemann surfaces. In our applications, C is the Riemann surface on which we study the geometric Langlands program, and we think of Σ as being much larger than C. However, in topological field theory, the metrics on Σ and C are inessential.

For $t\neq 0,\infty$, the second vanishing theorem says that any solution of the supersymmetric equations on M is given by a flat $G_{\mathbb{C}}$-valued connection. A flat bundle $E\to M=\Sigma\times C$ has a decomposition[12] $E=\oplus_{i=1}^n E_i'\otimes E_i''$, where the E_i' are flat bundles pulled back from Σ and the E_i'' are flat bundles pulled back from C.

If we restrict to a point $x\in\Sigma$, then the E_i' become trivial. If d_i is the rank of E_i', then E reduces to $E_x=\oplus_{i=1}^n d_i E_i''$, a direct sum of flat bundles pulled back from C. E_x is a flat bundle over C, and as such defines a point in Hitchin's moduli space $\mathcal{M}_H$, which we describe more thoroughly in section 4. For now, we can just think of $\mathcal{M}_H$ as the moduli space of flat $G_{\mathbb{C}}$-bundles over C. Thus any supersymmetric field on $\Sigma\times C$ is given by a map from Σ to $\mathcal{M}_H$, namely $x\to E_x$.

The moduli space $\mathcal{M}_H$ has singularities that correspond to reducible flat bundles. In the present construction, the bundle E_x is reducible if $n>1$. For $n=1$, we have $E=E'\otimes E''$, where the two factors are pullbacks from

[12] In this discussion, for simplicity, we use the language of vector bundles, as if $G=SU(N)$. To extend the argument to any G, one simply formulates it in terms of subgroups of G rather than subbundles of E.

Σ and C, respectively. E_x is still reducible unless E' has rank 1, in which case, for $G = SU(N)$, it must be trivial. Thus, the case in which the singularities of $\mathcal{M}_H$ are avoided is precisely the case that E is the pullback of an irreducible flat bundle $E'' \to C$.

The supersymmetric fields are also the configurations that minimize the action; this will become clear in section 3.4, when we construct the supersymmetric Lagrangian. In the above reasoning, we took Σ and C to be compact. In our applications, we are also interested in the case that Σ is a complete but noncompact two-manifold, such as $\mathbb{R}^2$, $\mathbb{R}^2_+$, or $\mathbb{R} \times I$ ($\mathbb{R}^2_+$ is a halfspace and I is a closed interval). Given appropriate boundary conditions and asymptotic behavior at infinity to justify the proof of the second vanishing theorem, one still has in these cases the decomposition of supersymmetric fields $E = \oplus_{i=1}^n E'_i \otimes E''_i$. (For Σ flat and simply-connected, by E'_i we mean a bundle with trivial connection and covariantly constant Higgs field.) This leads to the same conclusion that we had for compact Σ: a supersymmetric field that avoids singularities of $\mathcal{M}_H$ is a pullback from C.

3.4. The Topological Lagrangian

Our goal now is to find a Lagrangian that possesses the topological symmetry for any value of t and reduces when M is flat to the Lagrangian of the underlying $\mathcal{N} = 4$ super Yang-Mills theory. We could do this by hand, starting with (2.6) for flat M and asking what curvature dependent terms are needed to maintain the topological symmetry when M is not flat. We will follow a different approach.

A very useful first step is to compute the algebra generated on the fields by the the supersymmetries Q_ℓ and Q_r, or equivalently by $Q = uQ_\ell + vQ_r$. For the fields $A, \phi, \psi, \widetilde{\psi}$, and σ, we compute

$$\begin{aligned}
{\delta_T}^2 A &= -i(u^2+v^2)(-D\sigma) \\
{\delta_T}^2 \phi &= -i(u^2+v^2)[\sigma,\phi] \\
{\delta_T}^2 \psi &= -i(u^2+v^2)[\sigma,\psi] \\
{\delta_T}^2 \widetilde{\psi} &= -i(u^2+v^2)[\sigma,\widetilde{\psi}] \\
{\delta_T}^2 \sigma &= 0.
\end{aligned} \tag{3.35}$$

These results can be summarized by saying that if Φ is any field of $\mathcal{K} \geq 0$, then ${\delta_T}^2 \Phi = -i(u^2+v^2)\pounds_\sigma(\Phi)$, where $\pounds_\sigma(\Phi)$ is the first order change in Φ in a gauge transformation generated by σ (so $\pounds$ stands for Lie derivative).

For example, $\pounds_\sigma(A) = -D\sigma$, and for the other fields here, $\pounds_\sigma(\Phi) = [\sigma, \Phi]$. The relation ${\delta_T}^2\Phi = -i(u^2 + v^2)\pounds_\sigma(\Phi)$ can be expanded in powers of u and v and is equivalent to the following:

$$\begin{aligned} \delta_\ell^2\Phi = \delta_r^2\Phi &= -i\pounds_\sigma(\Phi) \\ \{\delta_\ell, \delta_r\}\Phi &= 0. \end{aligned} \tag{3.36}$$

If, however, one computes ${\delta_T}^2\Phi$ for a field of $\mathcal{K} < 0$, one does not get $-i(u^2 + v^2)\pounds_\sigma(\Phi)$ in an obvious way. In fact, the relation ${\delta_T}^2\Phi = -i(u^2 + v^2)\pounds_\sigma(\Phi)$ does hold for all Φ in the twisted $\mathcal{N} = 4$ super Yang-Mills theory, but for some fields, it only holds upon using the equations of motion.

Construction of supersymmetric Lagrangians is generally much easier if one can introduce "auxiliary fields," which are extra fields that can ultimately be eliminated by their equations of motion if one so chooses, such that the algebra, in the present case ${\delta_T}^2\Phi = -i(u^2 + v^2)\pounds_\sigma(\Phi)$, is satisfied without using equations of motion. In theories with a great deal of supersymmetry, such as $\mathcal{N} = 4$ super Yang-Mills theory, it can be very difficult to find suitable auxiliary fields. The present example, as first noted in [53], is a case in which this problem arises.

However, the problem can be greatly alleviated by introducing an auxiliary field P, a zero-form with values in the Lie algebra, so as to close the algebra for η and $\widetilde{\eta}$ (but not χ). One takes the transformation laws of $\overline{\sigma}, \eta, \widetilde{\eta}$, and P to be

$$\begin{aligned} \delta_T\overline{\sigma} &= iu\eta + iv\widetilde{\eta} \\ \delta_T\eta &= vP + u[\overline{\sigma}, \sigma] \\ \delta_T\widetilde{\eta} &= -uP + v[\overline{\sigma}, \sigma] \\ \delta_T P &= -iv[\sigma, \eta] + iu[\sigma, \widetilde{\eta}]. \end{aligned} \tag{3.37}$$

These equations will reduce to those in section 3.0 once we impose the equations of motion – which notably will set $P = D^*\phi$.

Once we have a closed algebra on a set of fields – in this case all fields in the theory including P but excluding χ, since we have not closed the algebra on χ – it is generally a simple matter to find the possible invariant Lagrangians. In the present case, we would like to find a partial action $\boldsymbol{I}_0$ for the the fields on which we have closed the algebra which obeys $\delta_\ell\boldsymbol{I}_0 = \delta_r\boldsymbol{I}_0 = 0$, and hence, for any u, v, obeys $\delta_T\boldsymbol{I}_0 = 0$ where $\delta_T = u\delta_\ell + v\delta_r$. These properties will hold if $\boldsymbol{I}_0 = \delta_\ell\delta_r V$ for some gauge-invariant V. Indeed, if V

is gauge-invariant, then $\pounds_\sigma(V) = 0$ and (3.36) reduces to $\delta_\ell^2 V = \{\delta_\ell, \delta_r\}V = \delta_r^2 V = 0$. So $\delta_\ell\delta_r V$ will automatically be annihilated by δ.

In addition, we want to pick V so that the equation of motion for P is $P = D^*\phi$, so as to be compatible with the formulas of section 3.1. So we take

$$V = \frac{2}{e^2}\int_M d^4x\sqrt{g}\left(-\frac{1}{2}\mathrm{Tr}\,\eta\widetilde{\eta} - i\mathrm{Tr}\,\overline{\sigma}D^*\phi\right), \tag{3.38}$$

and compute that $\boldsymbol{I}_0 = \delta_\ell\delta_r V$ is

$$\begin{aligned}\boldsymbol{I}_0 = \frac{2}{e^2}\int_M d^4x\sqrt{g}\mathrm{Tr}\,\Big(&\frac{1}{2}P^2 - PD^*\phi + \frac{1}{2}[\overline{\sigma},\sigma]^2 - D_\mu\overline{\sigma}\,D^\mu\sigma - [\phi_\mu,\sigma][\phi^\mu,\overline{\sigma}]\\ &+ i\widetilde{\eta}D_\mu\widetilde{\psi}^\mu + i\eta D_\mu\psi^\mu - i\widetilde{\eta}[\psi_\mu,\phi^\mu] + i\eta[\widetilde{\psi}_\mu,\phi^\mu]\\ &-\frac{i}{2}[\sigma,\widetilde{\eta}]\widetilde{\eta} - \frac{i}{2}[\sigma,\eta]\eta + i[\overline{\sigma},\psi_\mu]\psi^\mu + i[\overline{\sigma},\widetilde{\psi}_\mu]\widetilde{\psi}^\mu\Big).\end{aligned} \tag{3.39}$$

The Euler-Lagrange equation for P is $P = D^*\phi$, as desired, and one can make the replacement

$$\mathrm{Tr}\left(P^2 - 2PD^*\phi\right) \to -\mathrm{Tr}(D^*\phi)^2, \tag{3.40}$$

if one wishes.

$\boldsymbol{I}_0$ manifestly possesses the topological symmetry for any value of $t = v/u$. Moreover, as it is of the form $\delta_\ell\delta_r V$, where the metric of M does not enter in the definition of δ_ℓ and δ_r but only in the choice of V, it also has the key property that leads to a topological field theory: its dependence on the metric of M is of the form $\delta_\ell\delta_r(\ldots)$.

The only reason that $\boldsymbol{I}_0$ is not a satisfactory action is that it is degenerate; it does not contain the kinetic energy for the gauge fields or any terms containing χ. It does not seem to be possible to complete the construction of the action with a construction as simple as that above. Though it is possible to add auxiliary fields so as to close the algebra on χ, it does not seem to be possible to do this in a way that incorporates both δ_ℓ and δ_r and is useful for understanding the appropriate twisted $\mathcal{N} = 4$ action. (The fact that a t-dependent topological invariant occurs on the right hand side of

(3.33) appears to be an obstruction to this.) Instead, we will fix a particular value of t and consider only the differential $\delta_t = \delta_\ell + t\delta_r$. We add an auxiliary field H, which is to be a two-form with values in the Lie algebra, and postulate

$$\begin{aligned} \delta_t \chi &= H \\ \delta_t H &= -i(1+t^2)[\sigma, \chi]. \end{aligned} \tag{3.41}$$

This is compatible with

$$\delta_t^2(\Phi) = -i(1+t^2)\pounds_\sigma(\Phi), \tag{3.42}$$

for $\Phi = \chi, H$; this is the specialization of $\delta_T^2 = -i(u^2+v^2)\pounds_\sigma$ for $(u,v) = (1,t)$. For (3.41) to agree with the transformation of χ found in section 3.1, the equations of motion will have to give $H^+ = \mathcal{V}^+(t)$, $H^- = t\mathcal{V}^-(t)$. We will construct the action to ensure this.

As before, an action annihilated by δ_t can be $\boldsymbol{I}_1 = \delta_t V_1$, for any gauge-invariant V_1. We pick

$$\begin{aligned} V_1 = \frac{2}{e^2}\int_M d^4x\sqrt{g}\frac{1}{(1+t^2)}\mathrm{Tr}\Big(\chi^+_{\mu\nu}\Big(\frac{1}{2}H^{+\mu\nu} - \mathcal{V}^+(t)^{\mu\nu}\Big) \\ + \chi^-_{\mu\nu}\Big(\frac{1}{2}H^{-\mu\nu} - t\mathcal{V}^-(t)^{\mu\nu}\Big)\Big). \end{aligned} \tag{3.43}$$

Then

$$\begin{aligned} \boldsymbol{I}_1 = \frac{1}{e^2}\int_M d^4x\sqrt{g}\Big(\frac{2}{(1+t^2)}\mathrm{Tr}\Big(\frac{1}{2}H^+_{\mu\nu}H^{+\mu\nu} - H^+_{\mu\nu}\mathcal{V}^+(t)^{\mu\nu}\Big) \\ + \frac{2}{(1+t^2)}\mathrm{Tr}\Big(\frac{1}{2}H^-_{\mu\nu}H^{-\mu\nu} - tH^-_{\mu\nu}\mathcal{V}^-(t)^{\mu\nu}\Big) \\ + 2\mathrm{Tr}\Big(\chi^+_{\mu\nu}\Big(iD\psi + i[\widetilde{\psi},\phi]\Big)^{\mu\nu} + \chi^-_{\mu\nu}\Big(iD\widetilde{\psi} - i[\psi,\phi]\Big)^{\mu\nu}\Big) \\ + \mathrm{Tr}\left(i\chi^+_{\mu\nu}[\sigma,\chi^{+\mu\nu}] + i\chi^-_{\mu\nu}[\sigma,\chi^{-\mu\nu}]\right)\Big). \end{aligned} \tag{3.44}$$

Upon integrating $H^\pm$ out of this, we can write the equivalent action

$$\begin{aligned}\boldsymbol{I}_1 = \frac{1}{e^2}\int_M d^4x\sqrt{g}\,\Bigg(&-\frac{t^{-1}}{(t+t^{-1})}\mathrm{Tr}\,\mathcal{V}^+(t)_{\mu\nu}\mathcal{V}^+(t)^{\mu\nu}\\ &-\frac{t}{(t+t^{-1})}\mathrm{Tr}\,\mathcal{V}^-(t)_{\mu\nu}\mathcal{V}^-(t)^{\mu\nu}\Bigg)\\ &+2\mathrm{Tr}\left(\chi^+_{\mu\nu}\left(iD\psi+i[\widetilde{\psi},\phi]\right)^{\mu\nu}+\chi^-_{\mu\nu}\left(iD\widetilde{\psi}-i[\psi,\phi]\right)^{\mu\nu}\right)\\ &+\mathrm{Tr}(i\chi^+_{\mu\nu}[\sigma,\chi^{+\mu\nu}]+i\chi^-_{\mu\nu}[\sigma,\chi^{-\mu\nu}])\Bigg).\end{aligned}\tag{3.45}$$

Finally, the useful identity (3.33) enables us to write

$$\begin{aligned}\boldsymbol{I}_1 = -\frac{1}{e^2}\int_M d^4x\sqrt{g}\,\Bigg(&\mathrm{Tr}\left(\frac{1}{2}F_{\mu\nu}F^{\mu\nu}+D_\mu\phi_\nu D^\mu\phi^\nu+R_{\mu\nu}\phi^\mu\phi^\nu\right.\\ &\left.+\frac{1}{2}[\phi_\mu,\phi_\nu]^2-(D^*\phi)^2\right)-2\mathrm{Tr}\left(\chi^+_{\mu\nu}\left(iD\psi+i[\widetilde{\psi},\phi]\right)^{\mu\nu}\right.\\ &\left.+\chi^-_{\mu\nu}\left(iD\widetilde{\psi}-i[\psi,\phi]\right)^{\mu\nu}\right)-\mathrm{Tr}\left(i\chi^+_{\mu\nu}[\sigma,\chi^{+\mu\nu}]+i\chi^-_{\mu\nu}[\sigma,\chi^{-\mu\nu}]\right)\Bigg)\\ &+\frac{t-t^{-1}}{e^2(t+t^{-1})}\int_M \mathrm{Tr}\,F\wedge F.\end{aligned}\tag{3.46}$$

Apart from the topological term, the part of $\boldsymbol{I}_0+\boldsymbol{I}_1$ that depends only on A and ϕ can also [53] be written

$$\boldsymbol{I}^{(A,\phi)} = -\frac{1}{e^2}\int d^4x\,\sqrt{g}\,\mathrm{Tr}\left(\frac{1}{2}\mathcal{F}_{\mu\nu}\overline{\mathcal{F}}^{\mu\nu}+(D^*\phi)^2\right),\tag{3.47}$$

as one can show with some integration by parts similar to what is used in proving the vanishing theorems. The analogous terms involving σ can be found in (3.39) and are

$$\boldsymbol{I}^{\sigma} = \frac{2}{e^2}\int d^4x\sqrt{g}\,\mathrm{Tr}\left(\frac{1}{2}[\overline{\sigma},\sigma]^2-D_\mu\overline{\sigma}\,D^\mu\sigma-[\phi_\mu,\sigma][\phi^\mu,\overline{\sigma}]\right).\tag{3.48}$$

The key point – which we already exploited in proving the vanishing theorems – is that $\boldsymbol{I}_1$ is independent of t except for the last term. But that last term, being a topological invariant, is automatically annihilated by δ_t,

and indeed by both δ_ℓ and δ_r, all by itself. So without spoiling the topological symmetry, we can add another term to the action,

$$\boldsymbol{I}_2 = -\left(\frac{t-t^{-1}}{e^2(t+t^{-1})} - i\frac{\theta}{8\pi^2}\right)\int_M \mathrm{Tr}\, F\wedge F. \tag{3.49}$$

We have simply chosen the coefficient to cancel the t-dependence in $\boldsymbol{I}_1$, leaving us with the standard θ parameter multiplying this term in the action.

Finally, then, the overall action is $\boldsymbol{I} = \boldsymbol{I}_0 + \boldsymbol{I}_1 + \boldsymbol{I}_2$. The construction makes it manifest that $\boldsymbol{I}$ is annihilated by δ_t for the specific value of t used in constructing $\boldsymbol{I}_1$. But since in fact $\boldsymbol{I}$ is independent of t, it is annihilated by δ_t for all t.

Moreover, $\boldsymbol{I}$ is the action of a topological field theory. The change in $\boldsymbol{I}$ in a change in the metric of M is of the form $\delta_t(\ldots)$, since $\boldsymbol{I}_0$ and $\boldsymbol{I}_1$ are of this form even before varying the metric, and $\boldsymbol{I}_2$ does not depend on the metric of M at all.

We have accomplished our goal of constructing an action that has the full $\mathbb{CP}^1$ family of topological symmetries – making it clear, among other things, that the partition function of this theory on a closed four-manifold M without operator insertions is independent of t. One can readily verify that (after eliminating P via (3.40), whereupon the $(D^*\phi)^2$ terms cancel), the bosonic part of $\boldsymbol{I}$ reduces for $M=\mathbb{R}^4$ to (2.10). The topological symmetry implies that the fermionic terms also agree. So this theory is a generalization of $\mathcal{N}=4$ super Yang-Mills theory to a curved four-manifold.

3.5. The Canonical Parameter

So far, we have obtained a family of topological field theories that depend on a coupling parameter $\tau = \theta/2\pi + 4\pi i/e^2$, and a twisting parameter t. But these parameters are not really all independent. We will now show that the theory really only depends on a certain combination Ψ of t and τ, which we will call the canonical parameter.

Let us write the coefficient of the topological term in $\boldsymbol{I}_2$, which takes the form $-\left((t-t^{-1})/e^2(t+t^{-1}) - i\theta/8\pi^2\right)$, as $-\Psi/4\pi i$, where

$$\Psi = \frac{\tau+\overline{\tau}}{2} + \frac{\tau-\overline{\tau}}{2}\left(\frac{t-t^{-1}}{t+t^{-1}}\right). \tag{3.50}$$

We will call Ψ the canonical parameter.

We claim that any observable $\mathcal{O}(\tau,\overline{\tau},t)$ really depends only on the canonical parameter Ψ. Indeed, the action of the theory is of the form

$$\boldsymbol{I} = \delta_t V + \frac{i\Psi}{4\pi}\int_M \mathrm{Tr}\, F\wedge F. \tag{3.51}$$

To the extent that the term $\delta_t V$ is irrelevant, the theory naturally only depends on the coefficient Ψ.

At first sight, there is a fallacy in this argument. In addition to whatever t-dependence may be explicit in $\boldsymbol{I}$ (we have arranged so that there is none, though individual terms are separately t-dependent), there might be t-dependence coming from the fact that the definition of δ_t depends on t. The definition of the theory, after all, depends on the topological symmetry δ_t as well as on the action. But in fact, the t-dependence of δ_t can be eliminated (as long as $t\neq\pm i$) by redefining the fields. One rescales δ_t to $\delta_t' = \delta_t/\sqrt{1+t^2}$, so that the algebra becomes $(\delta_t')^2 = -i\pounds_\sigma$, independent of t. Then, with the auxiliary fields P and H included, after some small redefinitions of the fermions, every multiplet takes the form $\delta_t' U = V$, $\delta_t' V = -i\pounds_\sigma(U)$, for some pair of fields U, V. (This is a standard type of multiplet in cohomological field theories, related mathematically to the Cartan model of equivariant cohomology.) Once this is done, the transformation laws of the multiplets are completely independent of t.

Some Special Values Some special cases of the relation between Ψ, t, and τ are worthy of note. We focus especially on values that have a simple behavior under S-duality and a simple interpretation after compactification to two dimensions.

If $t=\pm 1$, we have $\Psi = \mathrm{Re}\,\tau$. Thus, any real value of Ψ can be achieved at $t=\pm 1$, simply by varying the θ angle of the gauge theory, but only real values are possible. More generally, if $|t|=1$, then Ψ is real.

To make Ψ complex, we must get away from $|t|=1$. All values of Ψ in the upper or lower half plane are possible, respectively, for t real and of modulus greater than or less than 1.

If $t=\pm i$, then $\Psi=\infty$, independent of τ. Thus, at $t=\pm i$, the value of τ is completely irrelevant. This has essentially been demonstrated directly by Marcus [53].

A final observation is that all values of Ψ can be realized at arbitrarily weak coupling, that is arbitrarily large $\mathrm{Im}\,\tau$. Indeed, inverting the definition of Ψ, we have

$$t^2 = -\frac{\Psi-\overline{\tau}}{\Psi-\tau}, \tag{3.52}$$

showing that if we keep Ψ and $\mathrm{Re}\,\tau$ fixed and take $\mathrm{Im}\,\tau \to \infty$, then $t \to \pm 1$. The limit $\mathrm{Im}\,\tau \to \infty$ is the weak coupling limit where the theory should be understandable.

Action Of S-Duality On The Canonical Parameter Now we want to determine how Ψ transforms under S-duality. We claim that a transformation that maps τ by $\tau \to (a\tau + b)/(c\tau + d)$ acts likewise on Ψ:

$$\Psi \longrightarrow \frac{a\Psi + b}{c\Psi + d}. \tag{3.53}$$

Clearly, under $T : \tau \to \tau + 1$, which leaves t invariant, we have $\Psi \to \Psi + 1$. We also have to check the behavior under $S : \tau \to -1/n_{\mathfrak{g}}\tau$. This mapping transforms t to $\pm t\tau/|\tau|$ according to (3.13). (The sign depends on how S is lifted to $SL(2,\mathbb{R})$, and does not matter as Ψ is invariant under $t \to -t$.) So Ψ transforms by

$$\Psi \to -\frac{1}{2}\left(\frac{1}{n_{\mathfrak{g}}\tau} + \frac{1}{n_{\mathfrak{g}}\overline{\tau}}\right) - \frac{1}{2}\left(\frac{1}{n_{\mathfrak{g}}\tau} - \frac{1}{n_{\mathfrak{g}}\overline{\tau}}\right)\left(\frac{t^2(\tau/\overline{\tau}) - 1}{t^2(\tau/\overline{\tau}) + 1}\right) = -\frac{1}{n_{\mathfrak{g}}\Psi}, \tag{3.54}$$

where (3.52) has been used.

Orientation Reversal It is a little unusual to find an action of a discrete group such as $SL(2,\mathbb{Z})$ on a complex parameter Ψ which, as here, takes values in $\mathbb{CP}^1$, not just in the upper half of the complex plane. This being the case, it is conceivable to extend $SL(2,\mathbb{Z})$ to $GL(2,\mathbb{Z})$ by adding an extra symmetry that we can take to be

$$\mathcal{T} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.55}$$

Such an element will act on Ψ by $\Psi \to -\Psi$, exchanging the upper half plane with the lower half plane.

In fact, such a symmetry exists precisely when the four-manifold M possesses a symmetry $\mathcal{T} : M \to M$ that reverses its orientation. Such a transformation leaves fixed the gauge coupling e^2 and hence leaves fixed $\mathrm{Im}\,\tau = 4\pi/e^2$. On the other hand, it reverses the sign of the instanton number and hence reverses the sign of $\mathrm{Re}\,\tau$. Looking back at the definition of the canonical parameter in (3.50), we see that we will get $\Psi \to -\Psi$ if $\mathcal{T}$ acts by

$$t \longrightarrow \frac{1}{t}. \tag{3.56}$$

The points $t = \pm 1$ are fixed points of this transformation.

We will have a symmetry of the topological equations (3.29) if we take $\mathcal{T}$ to act on the bosons (A, ϕ) by

$$(A, \phi) \longrightarrow (\mathcal{T}^* A, -\mathcal{T}^* \phi). \tag{3.57}$$

The action on fermions is

$$\begin{aligned} (\psi, \widetilde{\psi}) &\longrightarrow (\mathcal{T}^* \widetilde{\psi}, \mathcal{T}^* \psi) \\ (\chi^+, \chi^-) &\longrightarrow (\mathcal{T}^* \chi^-, \mathcal{T}^* \chi^+) \\ (\eta, \widetilde{\eta}) &\longrightarrow (\mathcal{T}^* \widetilde{\eta}, \mathcal{T}^* \eta). \end{aligned} \tag{3.58}$$

This gives a manifest symmetry of the action. The supersymmetry generators transform by $(u, v) \to (v, u)$ and hence $\delta_t \to t\delta_{t^{-1}}$.

An important application, as we will see, is to the case $M = \mathbb{R} \times W$, with $\mathcal{T}$ being a "time-reversal" symmetry that acts trivially on W and acts on $\mathbb{R}$ by multiplication by -1.

4. Compactification And The Geometry Of Hitchin's Moduli Space

We now consider compactification to two dimensions. We take our four-manifold to be $M = \Sigma \times C$, a product of two two-manifolds. We take C to be a compact Riemann surface on which we will study the geometric Langlands program. Generally, we assume that the genus of C is greater than one, so that for simple gauge group G, the generic flat connection on C is irreducible. Σ may be either a complete but noncompact two-manifold such as $\mathbb{R}^2$, or a second compact Riemann surface. In the latter case, we assume that the size of Σ is much larger than that of C. We aim to reduce the four-dimensional gauge theory on $M = \Sigma \times C$ to an alternative description by an effective field theory on Σ.

To find the effective physics on Σ, we must find the configurations on M that minimize or nearly minimize the action (Euclidean signature) or the energy (Lorentz signature). Either way, one arrives at the conclusion obtained in [18], with closely related results in [19]: the four-dimensional twisted $\mathcal{N} = 4$ supersymmetric gauge theory reduces on Σ to a sigma-model

of maps $\Phi : \Sigma \to \mathcal{M}_H$, where $\mathcal{M}_H$ is the moduli space of solutions on C of Hitchin's equations with gauge group G.

This follows from our results in section 3.4. Apart from the topological term, which does not affect the equations of motion or the energy of a classical configuration, the bosonic action of the twisted $\mathcal{N} = 4$ super Yang-Mills theory is a sum of squares. As we see from (3.47) and (3.48), it is minimized precisely if the topological equations are obeyed for all t, or in other words if $\mathcal{F} = D^*\phi = 0$, along with $D\sigma = [\phi, \sigma] = [\sigma, \overline{\sigma}] = 0$. One simple type of solution of these equations is obtained by taking A and ϕ to be pullbacks from C, with $\sigma = 0$. In fact, the equations $\mathcal{F} = D^*\phi = 0$ for a connection A and an adjoint-valued one-form ϕ on the Riemann surface C are one way to present Hitchin's equations [20]. By taking the real and imaginary parts of $\mathcal{F}$, these equations can alternatively be written

$$\begin{aligned} F - \phi \wedge \phi &= 0 \\ D\phi = D^*\phi &= 0. \end{aligned} \tag{4.1}$$

We will write $\mathcal{M}_H$, or in more detail $\mathcal{M}_H(G, C)$, for the moduli space of solutions of these equations on C up to gauge transformations. If the genus g of C is greater than one, then $\mathcal{M}_H(G, C)$ is a complex manifold of dimension $(2g - 2)\dim G$. It has singularities corresponding to reducible solutions of Hitchin's equations, that is, solutions with continuous unbroken gauge symmetries.

As we explained at the end of section 3.2, any solution of the equations $\mathcal{F} = D^*\phi = 0$ on M is associated with a map from Σ to $\mathcal{M}_H$. Moreover, if the singularities of $\mathcal{M}_H$ are avoided, the solution on M is actually a pullback from a solution on C. In this case, the conditions on σ require $\sigma = 0$. So a field of zero energy or action that avoids singularities of $\mathcal{M}_H$ is a constant map from Σ to a smooth point in $\mathcal{M}_H$.

Accordingly, a field of almost zero energy or action comes from a slowly varying map from Σ to $\mathcal{M}_H$. (A map is slowly varying if it varies significantly only over a length scale in Σ that is much greater than the size of C.) So, as long as we do not encounter fields on C with continuous gauge symmetries, the effective physics at long distances on Σ is given by a supersymmetric sigma-model in which the target space is $\mathcal{M}_H(G, C)$.

The importance of the assumption that Σ is very large compared to C is that it ensures that nonconstant but slowly varying maps $\Sigma \to \mathcal{M}_H(G, C)$ can exist, and moreover that a pair of fields (A, ϕ) that determine a slowly

varying map $\Phi : \Sigma \to \mathcal{M}_H(G,C)$ has very small action.[13] If alternatively we assume that C is much larger than Σ, a pair (A,ϕ) representing a generic nonconstant map $\Phi : \Sigma \to \mathcal{M}_H(G,C)$ has large action; instead, we could get a good description involving slowly varying maps $\widetilde{\Phi} : C \to \mathcal{M}_H(G,\Sigma)$.

The above analysis assumes that G is semi-simple. If $G = U(1)$ or $U(N)$, some modification is required, because every solution of Hitchin's equations has a $U(1)$ group of symmetries. As a result, the low energy description involves the product of a sigma-model and a supersymmetric $U(1)$ gauge theory. Moreover, for any G, the low energy description breaks down at singularities of $\mathcal{M}_H$ corresponding to solutions of Hitchin's equations that have continuous gauge symmetries (or extra continuous gauge symmetries for $G = U(1)$ or $U(N)$). At such points, additional degrees of freedom come into play: a field of zero energy or action can have $\sigma \neq 0$ and need not be a pullback from C. (It has a more general decomposition $E = \oplus_{i=1}^n E_i' \otimes E_i''$ described at the end of section 3.2.) For semi-simple G, the generic solution of Hitchin's equations is irreducible if the genus of C is at least 2. In this paper, we will keep away from the singularities of $\mathcal{M}_H$ (by considering the Langlands correspondence for irreducible flat ${}^L G_{\mathbb{C}}$ bundles, for example). Some simple generalizations will be considered in [26].

The sigma-model with target $\mathcal{M}_H(G,C)$ will prove to be a powerful tool for understanding the geometric Langlands program. But its relation to the underlying gauge theory is important. For one thing, key actors in the story are the Wilson and 't Hooft line operators, which are most naturally constructed in the underlying four-dimensional gauge theory. In addition, the underlying gauge theory is an important tool when the sigma-model breaks down because of singularities of $\mathcal{M}_H$. But many things can be understood directly in the sigma-model.

The rest of this section is devoted to a review of the classical geometry of $\mathcal{M}_H$, focussing on just those topics that we need for the present paper.

4.1. $\mathcal{M}_H$ As A Hyper-Kahler Quotient

For us the fact of central importance is that $\mathcal{M}_H$ can be regarded as a hyper-Kahler quotient and therefore is a hyper-Kahler manifold. Our first goal is to describe how this comes about.

[13] There is also a converse, which we explain in section 8.1. Given a slowly varying map $\Phi : \Sigma \to \mathcal{M}_H(G,C)$, one can reconstruct a pair of fields (A,ϕ) on M that determine this map and have very small action.

We consider a problem in gauge theory on a Riemann surface C, the fields being a gauge field A and a one-form ϕ that takes values in the adjoint representation. (These arise by restricting to C the fields A, ϕ that appeared in section 3 in constructing the twisted gauge theory on a general four-manifold M.) We think of the space of all fields A, ϕ as an infinite-dimensional affine space $\mathcal{W}$. We write δ for the exterior derivative on $\mathcal{W}$ (to denote this as d would invite confusion with the exterior derivative on C; hopefully, there will be no confusion with the use of the symbol δ_S in section 3 to denote supersymmetric variation).

We define a flat metric on $\mathcal{W}$ by

$$\begin{aligned} ds^2 = -\frac{1}{4\pi}\int_C |d^2z|\,\mathrm{Tr}\,(&\delta A_z\otimes\delta A_{\overline{z}} + \delta A_{\overline{z}}\otimes\delta A_z \\ &+ \delta\phi_z\otimes\delta\phi_{\overline{z}} + \delta\phi_{\overline{z}}\otimes\delta\phi_z)\,. \end{aligned} \tag{4.2}$$

Here for a local complex coordinate z on C, we write $A = dz\,A_z + d\overline{z}A_{\overline{z}}$, $\phi = dz\,\phi_z + d\overline{z}\,\phi_{\overline{z}}$, and $|d^2z|$ is the measure corresponding to the $(1,1)$-form $idz\wedge d\overline{z}$. This metric on $\mathcal{W}$, in which $\mathcal{W}$ is an infinite-dimensional Euclidean space, is determined by the complex structure on C (and does not depend on the local coordinate z that we used in writing it).

We want to extend this flat metric on $\mathcal{W}$ to a flat hyper-Kahler structure. To do this, we introduce three complex structures I, J, and K on $\mathcal{W}$ obeying certain axioms. They will all be translation-invariant on the affine space $\mathcal{W}$, so they can be defined by (constant) linear transformations of the space of one-forms. Any linear transformation of square -1 will define a complex structure; integrability is automatic.

A complex structure such as I, J, or K is usually defined as a linear transformation of the tangent space, but instead we will write down the transposed matrices I^t, J^t, and K^t acting on the cotangent space. We define I^t by the formulas

$$\begin{aligned} I^t(\delta A_{\overline{z}}) &= i\delta A_{\overline{z}} \\ I^t(\delta\phi_z) &= i\delta\phi_z \\ I^t(\delta A_z) &= -i\delta A_z \\ I^t(\delta\phi_{\overline{z}}) &= -i\delta\phi_{\overline{z}}, \end{aligned} \tag{4.3}$$

which do ensure $I^2 = (I^t)^2 = -1$. A translation-invariant complex structure on an affine space $\mathcal{W}$ can also be characterized by saying which linear functions on $\mathcal{W}$ are holomorphic. For complex structure I, the linear holomorphic functions are $A_{\overline{z}}$ and ϕ_z, evaluated at any point on C.

Similarly, we define a complex structure J by

$$\begin{aligned} J^t(\delta A_{\overline{z}}) &= -\delta\phi_{\overline{z}} \\ J^t(\delta A_z) &= -\delta\phi_z \\ J^t(\delta\phi_{\overline{z}}) &= \delta A_{\overline{z}} \\ J^t(\delta\phi_z) &= \delta A_z. \end{aligned} \tag{4.4}$$

The linear holomorphic functions are $\mathcal{A}_{\overline{z}} = A_{\overline{z}} + i\phi_{\overline{z}}$ and $\mathcal{A}_z = A_z + i\phi_z$, or more succinctly $\mathcal{A} = A + i\phi$.

Finally, we define complex structure K by

$$\begin{aligned} K^t(\delta A_{\overline{z}}) &= -i\delta\phi_{\overline{z}} \\ K^t(\delta\phi_{\overline{z}}) &= -i\delta A_{\overline{z}} \\ K^t(\delta A_z) &= i\delta\phi_z \\ K^t(\delta\phi_z) &= i\delta A_z. \end{aligned} \tag{4.5}$$

The linear holomorphic functions are $A_{\overline{z}} - \phi_{\overline{z}}$ and $A_z + \phi_z$.

These linear transformations of the cotangent space of $\mathcal{W}$ obey $I^tJ^t = -J^tI^t = -K^t$, and cyclic permutations thereof. Taking the transpose, we get the usual form of the quaternion relations $IJ = -JI = K$, etc.

Beyond the quaternion relations, the definition of a hyper-Kahler structure requires that the metric ds^2 should be of type $(1,1)$ in each complex structure. Equivalently, we want $ds^2 = I^t \otimes I^t(ds^2) = J^t \otimes J^t(ds^2) = K^t \otimes K^t(ds^2)$. This can be readily verified. Because $\mathcal{W}$ is an affine space and our complex structures are linear, there is nothing else to check; we have obtained a flat hyper-Kahler structure on $\mathcal{W}$.

A hyper-Kahler manifold has (Figure 2) a family of complex structures parametrized by $\mathbb{CP}^1$, obtained by taking linear combinations of I, J, and K. In the case at hand, we can describe this family simply. For any $w \in \mathbb{C}^*$, we consider the complex structure I_w in which $A_{\overline{z}} - w\phi_{\overline{z}}$ and $A_z + w^{-1}\phi_z$ are holomorphic. Thus, $w = -i$ corresponds to complex structure J, and $w = 1$ corresponds to K. We extend the definition to include $w = 0, \infty$ by setting $I_0 = I$, and $I_\infty = -I$ (here $-I$ is the complex structure opposite to I, in which A_z and $\phi_{\overline{z}}$ are holomorphic), as these are the limits of I_w for $w \to 0, \infty$. So we get the expected family of complex structures, parametrized by a copy of $\mathbb{CP}^1$ that we will call $\mathbb{CP}^1_h$ (see Figure 1). An explicit formula

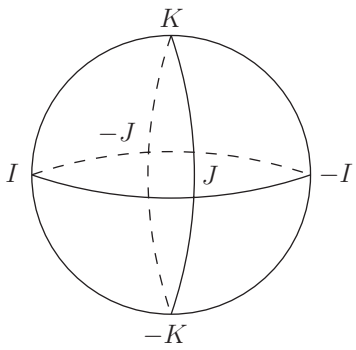

Figure 2: A family of complex structures on the hyper-Kahler manifold $\mathcal{M}_H$.

for I_w is

$$I_w = \frac{1-\overline{w}w}{1+\overline{w}w}I + \frac{i(w-\overline{w})}{1+\overline{w}w}J + \frac{w+\overline{w}}{1+\overline{w}w}K. \tag{4.6}$$

Of course, this is of the form $aI + bJ + cK$ with $a^2 + b^2 + c^2 = 1$.

We introduce a group $\mathcal{U}_1 \cong U(1)$ that acts on ϕ by $\phi_z \to \lambda\phi_z$, $\phi_{\overline{z}} \to \lambda^{-1}\phi_{\overline{z}}$, $|\lambda| = 1$, while leaving A fixed. $\mathcal{U}_1$ preserves the hyper-Kahler metric and acts on $\mathbb{CP}^1_h$ by

$$I_w \longrightarrow I_{\lambda^{-1}w}; \tag{4.7}$$

it leaves fixed the two points $w = 0, \infty$. We explain at the end of section 4.2 in what sense the action of $\mathcal{U}_1$ extends to an action of its complexification $\mathcal{U} \cong \mathbb{C}^*$.

Most of the complex structures on $\mathcal{W}$ in this family depend on the complex structure of C. The exceptions are J, which can be characterized as the complex structure in which $\mathcal{A} = A + i\phi$ is holomorphic, and $-J$, in which $\overline{\mathcal{A}} = A - i\phi$ is holomorphic. The other complex structures on $\mathcal{W}$ treat differently the z and $\overline{z}$ components of A and ϕ, a distinction that depends on the complex structure of C.

Does the symmetry group $\mathcal{U}_1$ come from a symmetry of four-dimensional Yang-Mills theory? The answer to this question is a little subtle. The twisted $\mathcal{N} = 4$ super Yang-Mills theory compactified on C does have the symmetry group $\mathcal{U}_1$. It acts by rotation of the $\phi_2 - \phi_3$ plane; this is a

subgroup of $SU(4)_{\mathcal{R}}$ that is unbroken in this particular compactification. However, this does not give a symmetry of the underlying family of four-dimensional TQFT's, because it does not preserve the family of supersymmetry generators $\epsilon = u\epsilon_\ell + v\epsilon_r$ (it maps this family to a larger family, whose interpretation we discuss in section 5.1, of supersymmetries that lack the full $Spin'(4)$ symmetry).

But the element $-1 \in \mathcal{U}_1$ does arise from a symmetry of the four-dimensional TQFT. This element acts on $\mathcal{M}_H$ just like the central element $\mathcal{I} \in SU(4)_{\mathcal{R}}$, which we introduced in section 2.2. Indeed, $\mathcal{I}$ acts as i in the $\mathbf{4}$ of $SU(4)_{\mathcal{R}}$, so it acts as -1 on $\mathbf{6} = \wedge^2\mathbf{4}$, which is the representation that contains ϕ. (Note that $\mathcal{I}$ also reverses the sign of untwisted scalars.) The transformation $\phi \to -\phi$ is holomorphic in complex structure I, but antiholomorphic in structures J and K. The obvious component of the fixed point set of this transformation is defined by $\phi = 0$ (giving, as we discuss later, an embedding of $\mathcal{M}$, the moduli space of stable bundles, in $\mathcal{M}_H$). There also are [20] other components of the fixed point set in which ϕ is gauge-equivalent to $-\phi$.

Symplectic Structures The three symplectic structures associated with the hyper-Kahler structure on $\mathcal{W}$ can be defined as $\omega_I = I^t \otimes 1(ds^2)$, and similarly $\omega_J = J^t \otimes 1(ds^2)$ and $\omega_K = K^t \otimes 1(ds^2)$. We get

$$\begin{aligned}
\omega_I &= -\frac{i}{2\pi}\int_C |d^2z|\,\mathrm{Tr}\,(\delta A_{\overline{z}} \wedge \delta A_z - \delta\phi_{\overline{z}} \wedge \delta\phi_z) \\
&= -\frac{1}{4\pi}\int_C \mathrm{Tr}\,(\delta A \wedge \delta A - \delta\phi \wedge \delta\phi) \\
\omega_J &= \frac{1}{2\pi}\int_C |d^2z|\,\mathrm{Tr}\,(\delta\phi_{\overline{z}} \wedge \delta A_z + \delta\phi_z \wedge \delta A_{\overline{z}}) \\
\omega_K &= \frac{i}{2\pi}\int_C |d^2z|\,\mathrm{Tr}\,(\delta\phi_{\overline{z}} \wedge \delta A_z - \delta\phi_z \wedge \delta A_{\overline{z}}) \\
&= \frac{1}{2\pi}\int_C \mathrm{Tr}\,\delta\phi \wedge \delta A.
\end{aligned} \tag{4.8}$$

Here, ω_J depends on the complex structure of C, while ω_I and ω_K do not.

Moreover, ω_I is of type $(1,1)$ with respect to complex structure I, and similarly for ω_J and ω_K. We also define $\Omega_I = \omega_J + i\omega_K$, along with cyclic

permutations $\Omega_J = \omega_K + i\omega_I$, $\Omega_K = \omega_I + i\omega_J$:

(4.9)
$$\Omega_I = \frac{1}{\pi}\int_C |d^2z|\,\mathrm{Tr}\,\delta\phi_z \wedge \delta A_{\overline{z}}$$
$$\Omega_J = -\frac{i}{4\pi}\int_C \mathrm{Tr}\,\delta\mathcal{A}\wedge\delta\mathcal{A}$$
$$\Omega_K = -\frac{i}{2\pi}\int_C |d^2z|\,\mathrm{Tr}\,\left(\delta A_{\overline{z}}\wedge\delta A_z - \delta\phi_{\overline{z}}\wedge\delta\phi_z - \delta\phi_{\overline{z}}\wedge\delta A_z - \delta\phi_z\wedge\delta A_{\overline{z}}\right).$$

Ω_I is of type $(2,0)$ in complex structure I, and similarly Ω_J and Ω_K are of type $(2,0)$ in their respective complex structures. Ω_I and Ω_K depend on the complex structure of C, but Ω_J does not.

The closed two-forms ω_J and ω_K are actually exact. Indeed, we can write very explicitly $\omega_J = \delta\lambda_J$ with $\lambda_J = \frac{1}{2\pi}\int_C |d^2z|\,\mathrm{Tr}\,(\phi_{\overline{z}}\delta A_z + \phi_z\delta A_{\overline{z}})$. λ_J is gauge-invariant, because ϕ and δA (unlike A itself) both transform in the adjoint representation of G. Similarly $\omega_K = \delta\lambda_K$ with $\lambda_K = \frac{1}{2\pi}\int_C \mathrm{Tr}\,\phi\wedge\delta A$. Among other things, this means that a topologically trivial line bundle over $\mathcal{M}_H$ can be endowed with a connection whose curvature is any desired linear combination of ω_J and ω_K. One simply picks the connection to be a suitable linear combination of λ_J and λ_K.

On the other hand, ω_I is topologically non-trivial. To write $\omega_I = \delta\lambda_I$ along the above lines, one would need in λ_I a term proportional to $\int \mathrm{Tr}\,A\wedge\delta A$, but this is not gauge-invariant because of the inhomogeneous transformation of A under gauge transformations. However, we can write $\omega_I = \omega_I' + \delta\lambda_I$, where $\lambda_I = (1/4\pi)\int_C \mathrm{Tr}\,\phi\wedge\delta\phi$ and

$$\omega_I' = -\frac{1}{4\pi}\int_C \mathrm{Tr}\,\delta A\wedge\delta A. \tag{4.10}$$

So ω_I and ω_I' are cohomologous, and in particular the cohomology class of ω_I is a pullback from the subspace with $\phi = 0$. For simply-connected G, the Picard group of line bundles over $\mathcal{M}$ is $\mathbb{Z}$, generated by a line bundle $\mathfrak{L}$ whose first Chern class is represented in de Rham cohomology by $\omega_I'/2\pi$. We will loosely call $\mathfrak{L}$ the determinant line bundle, since for classical G it can be defined [60] as the determinant of a $\overline{\partial}$ operator.[14]

[14] For $G = SU(N)$, if $E \to C$ is a holomorphic vector bundle of rank N, the fiber $\mathfrak{L}_E$ at E of the determinant line bundle $\mathfrak{L}$ is $\mathfrak{L}_E = \det H^0(C,E)^{-1}\otimes\det H^1(C,E)$,

Hitchin Moduli Space as a Hyper-Kahler Quotient We now consider the group of gauge transformations acting on $\mathcal{W}$ in an obvious way. It manifestly preserves the metric (4.2) and the three symplectic structures (4.8). A short computation shows that the Hitchin equations (4.1) are equivalent to the vanishing of the three moment maps corresponding to the symplectic structures (4.8). Explicitly, let ϵ generate an infinitesimal gauge transformation (thus, ϵ is a zero-form with values in $\mathrm{ad}(E)$), and let $V(\epsilon)$ be the corresponding vector field on $\mathcal{W}$, which acts by $\delta A = -D\epsilon$, $\delta\phi = [\epsilon, \phi]$. Then the moment map μ for symplectic structure ω is found by solving the condition $\iota_{V(\epsilon)}\omega = \delta\mu$ (here $\iota_{V(\epsilon)}$ is the operator of contraction with $V(\epsilon)$). Taking ω to be ω_I, ω_J, or ω_K, the corresponding moment maps turn out to be

$$\begin{aligned} \mu_I &= -\frac{1}{2\pi}\int_C \mathrm{Tr}\,\epsilon(F - \phi\wedge\phi), \\ \mu_J &= -\frac{1}{2\pi}\int_C |d^2z|\,\mathrm{Tr}\,\epsilon\left(D_z\phi_{\overline{z}} + D_{\overline{z}}\phi_z\right), \\ \mu_K &= -\frac{i}{2\pi}\int_C |d^2z|\,\mathrm{Tr}\,\epsilon\left(D_z\phi_{\overline{z}} - D_{\overline{z}}\phi_z\right). \end{aligned} \tag{4.11}$$

In general, if X is a hyper-Kahler manifold on which a Lie group H acts with moment maps $\vec{\mu} = (\mu_I, \mu_J, \mu_K)$, then the quotient by H of the solution set of the equations $\vec{\mu} = 0$ is a hyper-Kahler manifold [61], called the hyper-Kahler quotient $X///H$ of X by H. So in particular $\mathcal{M}_H$ is the hyper-Kahler quotient of the infinite-dimensional space $\mathcal{W}$ by the group of gauge transformations, and is a hyper-Kahler manifold.

It is possible to give a more "physical" explanation of why the target of the sigma-model is the hyper-Kahler quotient of $\mathcal{W}$. In general, in a two-dimensional gauge theory with $(4,4)$ supersymmetry and gauge group H, the hypermultiplets parametrize a hyper-Kahler manifold X, and the low energy physics (assuming for simplicity that H acts freely on X) is a sigma-model with target the hyper-Kahler quotient $X///H$. (This is the context in which the hyper-Kahler quotient construction was originally discovered [61].) In the present example, compactification of the twisted topological gauge theory in four dimensions gives in two dimensions a theory with $(4,4)$ supersymmetry. (The twisting and compactification preserve eight of the original sixteen supersymmetries, of which four have one two-dimensional

---

where $\det V$ denotes the highest exterior power of a vector space V. The line bundle $\mathfrak{L} \to \mathcal{M}$ so defined is ample. Some authors write $\mathfrak{L}^{-1}$ for what we call $\mathfrak{L}$.

chirality and four have the other.) The gauge group of this two-dimensional theory is the group $\mathcal{G}$ of maps of C to G, and the hypermultiplets parametrize $\mathcal{W}$. This is precisely the setup that leads to $\mathcal{M}_H$ as a hyper-Kahler quotient. So the low energy physics is a sigma-model with target the hyper-Kahler quotient of $\mathcal{W}$.

More Precise Account Of Dimensional Reduction We can now be more precise about the dimensional reduction from twisted four-dimensional super Yang-Mills theory to the sigma-model with target $\mathcal{M}_H$. Such a sigma-model has a metric and a B-field. We want to determine what they are.

We consider four-dimensional super Yang-Mills on $M = \Sigma \times C$, where the radius of Σ is much greater than the radius of C. The kinetic energy of the gauge fields is

$$\boldsymbol{I}_{kin} = -\frac{1}{2e^2}\int_{\Sigma\times C} |d^2y|\,|d^2z|\,\mathrm{Tr}\,F_{\mu\nu}F^{\mu\nu}, \tag{4.12}$$

where $|d^2y|$ and $|d^2z|$ are the Riemannian measures on Σ and C, respectively. This has contributions of type (p,q), $p+q=2$, where p is the number of indices μ, ν that are tangent to Σ and q is the number tangent to C. It is useful to write the gauge field on $\Sigma\times C$ as $A = A_\Sigma + A_C$, where A_Σ is tangent to Σ and A_C is tangent to C.

The terms of type $(0,2)$ are part of the potential energy for A_C that causes the theory to be localized on slowly varying maps $\Phi : \Sigma \to \mathcal{M}_H$. The terms of type $(2,0)$ give a kinetic energy for A_Σ, which away from singularities of $\mathcal{M}_H$ is massive, because the gauge group is completely broken, and can be integrated out in the infrared (as we describe explicitly in section 8.1).

The kinetic energy for the sigma-model comes from the terms of type $(1,1)$. These terms are

$$-\frac{1}{e^2}\int |d^2y|\,|d^2z| \sum_{\mu=0,1}\sum_{\nu=2,3} \mathrm{Tr}\,F_{\mu\nu}F^{\mu\nu}. \tag{4.13}$$

We want to compare this to a general sigma-model of maps $\Phi : \Sigma \to X$, where the target space X has metric $ds_X^2 = g_{IJ}(X)\,dX^I\,dX^J$ in terms of local coordinates X^I. The kinetic energy of the sigma-model is then

$$\int |d^2y|\,g_{IJ}\sum_{\mu=0,1}\partial_\mu X^I \partial^\mu X^J. \tag{4.14}$$

In comparing (4.13) and (4.14), the role of the X^I is played by A_C, and F_{ij} corresponds to $\partial_i X^I$. If we compare (4.13) and (4.14) to our original

definition (4.2) of the hyper-Kahler metric ds^2 on $\mathcal{W}$ (which induces on the hyper-Kahler quotient $\mathcal{M}_H$ a metric that we also call ds^2), we see that the metric ds_g^2 that is induced by the four-dimensional gauge theory is related to ds^2 by

$$ds_g^2 = \frac{4\pi}{e^2}\, ds^2 = (\mathrm{Im}\,\tau)\, ds^2. \tag{4.15}$$

We similarly want to reduce the θ term of the gauge theory, which in Euclidean signature is

$$\boldsymbol{I}_\theta = \frac{i\theta}{8\pi^2}\int_{\Sigma\times C} \mathrm{Tr}\, F\wedge F. \tag{4.16}$$

Again, we evaluate this coupling for a gauge field that represents a slowly varying map to $\mathcal{M}_H$. Because of the topological invariance of $\boldsymbol{I}_\theta$, in evaluating (4.16) we can deform to a situation in which the restriction of F to C vanishes (that is, $F_{23}=0$). The deformation to $F|_C = 0$ can be achieved via a complex gauge transformation or using Hitchin's second fibration $\mathcal{M}_H\to\mathcal{M}$ (these are both notions that we explain later). Given this, in evaluating $\boldsymbol{I}_\theta$, we can replace $F = \frac{1}{2}\sum_{\mu,\nu=0}^3 dx^\mu\wedge dx^\nu\, F_{\mu\nu}$ by $F^{1,1} = \sum_{\mu=0,1}\sum_{\nu=2,3} dx^\mu\wedge dx^\nu\, F_{\mu\nu}$. In the sigma-model limit, $F^{1,1}$ corresponds to $\delta A_C = \sum_{\mu=0,1} dx^\mu\, D_\mu A_C$, and $\boldsymbol{I}_\theta$ reduces to[15]

$$-\frac{i\theta}{8\pi^2}\int_\Sigma\int_C \mathrm{Tr}\,\delta A_C\wedge\delta A_C. \tag{4.17}$$

Comparing to the definition (4.10) of ω_I', we see that this is equivalent to

$$\frac{i\theta}{2\pi}\int_\Sigma \Phi^*(\omega_I') = \frac{i\theta}{2\pi}\int_\Sigma \Phi^*(\omega_I), \tag{4.18}$$

where the last step uses the fact that $\omega_I-\omega_I'$ is exact. In the general sigma-model, the B-field is a closed two-form on the target space X; it appears in the action via a term $-i\int_\Sigma \Phi^*(B)$. So the B-field of the effective low energy sigma-model is

$$B = -\frac{\theta}{2\pi}\omega_I = -(\mathrm{Re}\,\tau)\,\omega_I. \tag{4.19}$$

[15] A minus sign enters because in defining ω_I', δ, the exterior derivative on the space of connections A_C, is taken to commute with a one-form on C such as dx^j, $j=2,3$, which does not depend on A_C. For differential forms on $\Sigma\times C$, in terms of which $\boldsymbol{I}_\theta$ is defined, they would anticommute.

We can describe this more conceptually. For simply-connected G, the instanton number $(1/8\pi^2)\int_M \mathrm{Tr}\, F\wedge F$ is integer-valued for any four-manifold M. In addition, for simply-connected G, $\mathcal{M}_H$ is simply-connected, and hence $H^2(\mathcal{M}_H,\mathbb{Z}) = \pi_2(\mathcal{M}_H)$. The form ω_I has been normalized so that $\omega_I/2\pi$ is the image in de Rham cohomology of a generator of $H^2(\mathcal{M}_H,\mathbb{Z})$. These facts imply that for any Σ and C, there is a map $\Phi:\Sigma\to\mathcal{M}_H$ with $\int_\Sigma \Phi^*(\omega_I/2\pi) = 1$ (and this is the smallest achievable value). Such a map corresponds to a gauge field on $M=\Sigma\times C$ that has instanton number ±1 depending on one's conventions.[16] More briefly, the theta angles are the same in two and four dimensions because the instantons are the same. If G is not simply-connected, things are a little more complicated; the instanton number is not always an integer, and there are corresponding complications on the sigma-model side.

4.2. Complex Structures Of $\mathcal{M}_H$

Our next task is to understand more concretely the complex structures on $\mathcal{M}_H$.

Description By Holomorphic Data Let us look at the hyper-Kahler quotient of $\mathcal{W}$ from the standpoint of complex structure I. The combination $\nu_I = \mu_J + i\mu_K$ is holomorphic in this complex structure:

$$\nu_I = -\frac{1}{\pi}\int_C |d^2z|\,\mathrm{Tr}\,\epsilon\, D_{\overline{z}}\phi_z. \tag{4.20}$$

ν_I is holomorphic in complex structure I because it is the moment map derived from Ω_I, which is of type $(2,0)$ in complex structure I (explicitly ν_I is holomorphic because it depends only on $A_{\overline{z}}$ and ϕ_z). In such problems, it is often convenient [61] to consider separately the vanishing of the holomorphic moment map ν_I and the real moment map μ_I.

Any connection A on a smooth G-bundle E over a Riemann surface C automatically turns E into a holomorphic G-bundle (which we will denote by the same name, hoping this causes no confusion). One simply defines the $\overline{\partial}$

[16] In the physics literature, instantons in four dimensions are generally taken to be selfdual gauge fields, obeying $F^- = 0$, and in the math literature, they are generally taken to be anti-selfdual, obeying $F^+ = 0$. If an instanton in the sigma model is understood as a holomorphic map to $\mathcal{M}_H$, and the usual physics convention is followed in four dimensions, then a sigma model field of instanton number 1 comes from a gauge theory instanton of instanton number -1. The usual math convention avoids this minus sign.

operator as $\overline{D} = d\overline{z}\, D_{\overline{z}}$, which in complex dimension one trivially obeys $\overline{D}^2 = 0$. The vanishing of the holomorphic moment map ν_I tells us that $\overline{D}\phi = 0$; in other words, $\varphi = \phi_z\, dz$ is a holomorphic section of $K_C \otimes \mathrm{ad}(E)$, where K_C is the canonical line bundle of C. Differently put, φ is a holomorphic one-form valued in the adjoint representation, that is in the adjoint bundle of E. We will call a pair (E, φ) consisting of a G-bundle E and a holomorphic section φ of $K_C \otimes \mathrm{ad}(E)$ a Hitchin pair. φ is often called the Higgs field, and the bundle E endowed with the choice of φ is also called a Higgs bundle.

To obtain the moduli space $\mathcal{M}_H$, we must also set to zero the real moment map μ_I and divide by the group of G-valued gauge transformations. However, as exploited in [20] (and as is often the case in moduli problems) there is a simpler way to understand these combined steps. The space $\mathcal{W}_0$ of zeroes of the holomorphic moment map ν_I admits the action not just of ordinary G-valued gauge transformations, but of gauge transformations valued in the complexification $G_{\mathbb{C}}$ of G. This is manifest from the holomorphy of ν_I. Thus we can perform on $\mathcal{W}_0$ either of the following two operations:

(I) Restrict to the subspace of $\mathcal{W}_0$ with $\mu_I = 0$ and divide by G-valued gauge transformations.

(II) Divide $\mathcal{W}_0$ by $G_{\mathbb{C}}$-valued gauge transformations.

Operation *(I)* gives the desired moduli space $\mathcal{M}_H$, but operation *(II)* is much easier to understand and nearly coincides with it. (The reason they nearly coincide is that almost every orbit of complex gauge transformations contains a unique orbit of ordinary gauge transformations on which $\mu_I = 0$.) The easiest way to understand operation *(I)* is often to first understand operation *(II)* and then understand its relation to operation *(I)*.

In the present case, the result of operation *(II)* is easy to describe. It means that we do not care about the particular choice of $A_{\overline{z}}$, ϕ_z, but only about the holomorphic G-bundle E that is determined by $A_{\overline{z}}$, and the associated holomorphic section φ of $K_C \otimes \mathrm{ad}(E)$. Thus, in the present example, operation *(II)* gives us the set of equivalence classes of Hitchin pairs (E, φ).

This set is for many purposes a very good approximation to $\mathcal{M}_H$; for example, they differ in rather high codimension if C has genus at least 2. To be more precise (most of the present paper does not depend on the details), we need the notion of stability. For $G = SU(2)$, we interpret E as a rank two holomorphic vector bundle over C, and φ as a holomorphic map $\varphi : E \to E \otimes K_C$. A line bundle $\mathcal{L} \subset E$ is called φ-invariant if $\varphi(\mathcal{L}) \subset \mathcal{L} \otimes K_C$. A Hitchin pair (E, φ) is called stable if every φ-invariant line bundle $\mathcal{L} \subset E$ has negative first Chern class. It is called semistable if each such $\mathcal{L}$ has non-positive first Chern class. For general G, one must consider φ-invariant

reductions of the structure group of E to a maximal parabolic subgroup P of $G_{\mathbb{C}}$. The bundle E with such a reduction has a natural first Chern class (because P has a distinguished $U(1)$), and the pair (E,φ) is called stable (or semistable) if for every such reduction the first Chern class is negative (or nonpositive). A pair that is semistable but not stable is called strictly semistable.

Stability is a mild condition in the sense that, for example, if a pair (E,φ) is stable, then so is every nearby pair. Moreover, the pairs that are not stable have high codimension (if the genus of C is greater than 1).

Now we can go back to the question of comparing operations *(I)* and *(II)*, or equivalently, describing in holomorphic terms the moduli space $\mathcal{M}_H$ of solutions of Hitchin's equations. The result proved in [20] is that $\mathcal{M}_H$ is the same as the "moduli space of stable pairs (E,φ)," i.e., stable Hitchin pairs. We get this moduli space by throwing away unstable Hitchin pairs, imposing a certain equivalence relation on the semistable ones, and then dividing by the complex-valued gauge transformations. This slightly modified version of operation *(II)* – call it operation *(II)′* – agrees precisely with operation *(I)*.

In sum, $\mathcal{M}_H$ parametrizes the Hitchin pairs (E,φ) that are stable, as well as certain equivalence classes of strictly semistable Hitchin pairs. The stable pairs correspond to smooth points in $\mathcal{M}_H$ as well as (for some G) certain orbifold singularities. The strictly semistable pairs generally lead to singularities in $\mathcal{M}_H$ that are worse than orbifold singularities.

Analog For Complex Structure J All this has an analog in complex structure J. The holomorphic moment map in complex structure J is $\nu_J = \mu_K + i\mu_I$:

$$\nu_J = -\frac{i}{2\pi}\int_C \mathrm{Tr}\,\epsilon\mathcal{F} \tag{4.21}$$

To construct $\mathcal{M}_H$, we first impose the vanishing of ν_J. A zero of ν_J is simply a pair A,ϕ such that the curvature $\mathcal{F}$ of the $G_{\mathbb{C}}$-valued connection $\mathcal{A} = A + i\phi$ is equal to zero. If therefore we were to divide the zero set of ν_J by the group of $G_{\mathbb{C}}$-valued gauge transformations, we would simply get the set $\mathcal{Y}_0$ of all homomorphisms $\vartheta : \pi_1(C) \to G_{\mathbb{C}}$, up to conjugation. This is operation *(II)*.

Instead, to get $\mathcal{M}_H$, we are supposed to carry out operation *(I)*. In other words, we are supposed to set $\mu_J = 0$, i.e., to impose the condition $D^*\phi = 0$, and divide only by the G-valued gauge transformations. As in complex structure I, in comparing these two operations, a notion of stability intrudes.

For $G = SU(2)$, a homomorphism $\vartheta : \pi_1(C) \to G_{\mathbb{C}}$ is considered stable if the monodromies cannot be simultaneously reduced to the triangular form

$$\begin{pmatrix} \alpha & \beta \\ 0 & \alpha^{-1} \end{pmatrix}. \tag{4.22}$$

If ϑ has such a reduction, we call it strictly semistable. (In complex structure J, there is no notion of an unstable representation.) We consider two strictly semistable representations to be equivalent if they have the same diagonal monodromy elements, that is, the same α's. Note that each such equivalence class has a distinguished representative with $\beta = 0$.

For any G, the analog of putting the monodromies in triangular form is to conjugate them into a parabolic subgroup P of G. A representation is stable if it cannot be so conjugated, and otherwise strictly semistable. The equivalence relation on strictly semistable representations has a natural analog for any G (two strictly semistable representations that both reduce to P are equivalent if the two flat P bundles become equivalent when projected to the maximal reductive subgroup of P).

A theorem of Corlette [58] and Donaldson (see the appendix to [20]) is that $\mathcal{M}_H$, in complex structure J, is the moduli space $\mathcal{Y}$ that parametrizes stable homomorphisms $\vartheta : \pi_1(C) \to G_{\mathbb{C}}$, as well as the equivalence classes of semistable ones.

Analog For I_w What we have just said has a direct analog in each of the complex structure I_w, $w \neq 0, \infty$. The holomorphic variables in complex structure I_w are $A_{\overline{z}} - w\phi_{\overline{z}}$, $A_z + w^{-1}\phi_z$. Two of Hitchin's equations combine to the holomorphic condition $[D_{\overline{z}} - w\phi_{\overline{z}}, D_z + w^{-1}\phi_z] = 0$, and the third is a moment map condition. The holomorphic condition says that the complex-valued connection with components $(A_{\overline{z}} - w\phi_{\overline{z}}, A_z + w^{-1}\phi_z)$ is flat. Setting to zero the moment map and dividing by gauge transformations gives the same moduli space $\mathcal{Y}$ that we found in complex structure J. Thus all the complex structures I_w, $w \neq 0, \infty$ are equivalent.

In fact, in complex structure $I = I_{w=0}$, we can define a symmetry group $\mathcal{U} \cong \mathbb{C}^*$ that acts on a Higgs bundle by $(E, \varphi) \to (E, \lambda\varphi)$, $\lambda \in \mathbb{C}^*$. (The stability condition is invariant under this transformation.) For $|\lambda| = 1$, this reduces to the group $\mathcal{U}_1$, described in section 4.0, which is a symmetry group of the hyper-Kahler metric of $\mathcal{M}_H$. For $|\lambda| \neq 1$, we do not get a symmetry of the hyper-Kahler metric of $\mathcal{M}_H$, but as is shown on pp. 107-8 of [20], we get a group of diffeomorphisms of $\mathcal{M}_H$ that preserves complex structure I and

transforms the other complex structures by

$$I_w \longrightarrow I_{\lambda^{-1}w}, \tag{4.23}$$

generalizing eqn. (4.7) to $|\lambda| \neq 1$.

4.3. Hitchin's Fibrations

Now we will discuss in more detail the geometry of $\mathcal{M}_H$.

If E is a stable G-bundle, then the pair $(E, 0)$ is a stable Hitchin pair. So we get a natural embedding of $\mathcal{M}$, the moduli space of stable G-bundles on C, into the Hitchin moduli space $\mathcal{M}_H$. $\mathcal{M}$ is a holomorphic submanifold of $\mathcal{M}_H$ in complex structure I, since it is defined by the equations $\phi_z = 0$, which are holomorphic in complex structure I. In complex structures J and K, $\mathcal{M}$ is not holomorphic. Instead, it is Lagrangian, since $\phi = 0$ implies $\delta\phi = 0$ and hence $\mu_J = \mu_K = 0$.

If E is stable, then the Hitchin pair (E, φ) is stable for every $\varphi \in H^0(C, K_C \otimes \mathrm{ad(E)})$. Moreover, the tangent bundle to $\mathcal{M}$ at the point E is $H^1(C, \mathrm{ad(E)})$ (which parametrizes first order deformations of E), so by Serre duality, the cotangent space to $\mathcal{M}$ at the point E is $H^0(C, K_C \otimes \mathrm{ad(E)})$. So φ takes values in this cotangent space, and the space of all (E, φ) with stable E is the cotangent bundle $T^*\mathcal{M}$. We thus actually get an embedding of $T^*\mathcal{M}$ in $\mathcal{M}_H$. The holomorphic symplectic form Ω_I of $\mathcal{M}_H$ in complex structure I restricts on $T^*\mathcal{M}$ to its natural symplectic structure as a holomorphic cotangent bundle (that is, for any local holomorphic coordinates q^α on $\mathcal{M}$, we have $\Omega_I = \sum_\alpha dp_\alpha \wedge dq^\alpha$ for some holomorphic functions p_α that are linear on the fibers of $T^*\mathcal{M} \to \mathcal{M}$).[17]

The image of $T^*\mathcal{M}$ in $\mathcal{M}_H$ is not all of $\mathcal{M}_H$ because a Hitchin pair (E, φ) may be stable even if E is unstable. However, the stable Hitchin pairs (E, φ) for which E is unstable are a set of very high codimension. Upon throwing away this set, $\mathcal{M}_H$ becomes isomorphic to $T^*\mathcal{M}$, and has a natural map to $\mathcal{M}$ by forgetting φ. Even though it is only defined away from a set of very high

[17] This statement is true because the analogous statement is actually true on the space $\mathcal{W}$ of all pairs (A, ϕ) before taking the hyper-Kahler quotient. $\mathcal{W}$ can be understood as the cotangent space of the space $\mathcal{A}$ of connections, with A parameterizing $\mathcal{A}$ and ϕ parameterizing the cotangent space to $\mathcal{A}$; Ω_I is the natural holomorphic symplectic form of $\mathcal{W}$ regarded as $T^*\mathcal{A}$. In general, this type of structure persists in taking a complex symplectic quotient (in the present case, setting $\nu_I = 0$ and dividing by $G_{\mathbb{C}}$-valued gauge transformations).

codimension, this map is extremely useful. We will call it Hitchin's second fibration (the first one being another map that we introduce presently).

The Hitchin Fibration Another natural operation in complex structure I, apart from mapping φ to zero, is to calculate the gauge-invariant polynomials in φ. For $G = SU(2)$, this simply means that we consider the quadratic Casimir operator $w = \mathrm{Tr}\,\varphi^2$. Since $\overline{D}\varphi = 0$, we have $\overline{\partial} w = 0$, so w is a holomorphic quadratic differential, taking values in $\boldsymbol{B} = H^0(C, K_C^2) \cong \mathbb{C}^{3g-3}$. The Hitchin fibration, as it is most commonly called, is the map $\pi : \mathcal{M}_H \to \boldsymbol{B}$ obtained by mapping the pair (E, φ) to $w = \mathrm{Tr}\,\varphi^2$.

For any G, the Hitchin fibration is defined similarly, except that one characterizes φ by all of its independent Casimirs (that is, all of the independent G-invariant polynomials in φ), not just the quadratic one. For example, for $G = SU(N)$, we define $w_n = \mathrm{Tr}\,\varphi^n$, $n = 2, \dots, N$, and let $\boldsymbol{B} = \oplus_{n=2}^N H^0(C, K_C^n)$. The Hitchin fibration is then the map that takes (E, φ) to the point $(w_2, w_3, \dots, w_n) \in \boldsymbol{B}$. For any G, one considers instead of $\mathrm{Tr}\,\varphi^n$ the appropriate independent homogeneous G-invariant polynomials $\mathcal{P}_i$. The number of these polynomials equals the rank r of G, and their degrees d_i obey the identity

$$\sum_{i=1}^{r} (2d_i - 1) = \dim(G). \tag{4.24}$$

For example, for $G = SU(N)$, the d_i are $2, 3, \dots, N$, whence $\sum_i (2d_i - 1) = N^2 - 1 = \dim(G)$. And finally in general we define $\boldsymbol{B} = \oplus_i H^0(C, K_C^{d_i})$.

Since $\dim H^0(C, K_C^n) = (2n-1)(g-1)$, it follows from (4.24) that the complex dimension of $\boldsymbol{B}$ is $(g-1)\dim(G)$, which equals the dimension of $\mathcal{M}$, and one half of the dimension of $\mathcal{M}_H$. The Hitchin fibration $\pi : \mathcal{M}_H \to \boldsymbol{B}$ is surjective, as we will discuss momentarily. As the base $\boldsymbol{B}$ of the Hitchin fibration $\pi : \mathcal{M}_H \to \boldsymbol{B}$ has one half the dimension of $\mathcal{M}_H$, it follows that the dimension of a typical fiber $\boldsymbol{F}$ of π is also half the dimension of $\mathcal{M}_H$ and so equal to the dimension of $\boldsymbol{B}$:

$$\dim\, \boldsymbol{F} = \dim\, \boldsymbol{B} = \frac{1}{2} \dim\, \mathcal{M}_H = (g-1) \dim\, G. \tag{4.25}$$

Let us explain qualitatively why the Hitchin fibration is surjective. For example, take $G = SU(2)$. Pick a stable $SU(2)$ bundle E. Consider the equations $\mathrm{Tr}\,\varphi^2 = w$, where φ varies in the $(3g-3)$-dimensional space $H^0(C, K_C \otimes \mathrm{ad(E)})$ and w is a fixed element of the $(3g-3)$-dimensional space $\boldsymbol{B} = H^0(C, K_C^2)$. This is a system of $3g-3$ quadratic equations for

$3g-3$ complex variables. The number of solutions is generically 2^{3g-3}. A similar counting can be made for other G.

Complete Integrability We now want to explain one of Hitchin's main results [62]: $\mathcal{M}_H$ is a completely integrable Hamiltonian system in the complex structure I. We can find $\frac{1}{2}\dim\mathcal{M}_H$ functions H_a on $\mathcal{M}_H$ that are holomorphic in complex structure I, are algebraically independent, and are Poisson-commuting using the Poisson brackets obtained from the holomorphic symplectic form Ω_I.[18]

In fact, we can take the H_a to be linear functions on $\boldsymbol{B}$, since the dimension of $\boldsymbol{B}$ is the same as the desired number of functions. We will explain the construction first for $G = SU(2)$. We begin by picking a basis α_a, $a = 1, \dots, 3g-3$ of the $(3g-3)$-dimensional space $H^1(C, T_C)$, which is dual to $H^0(C, K_C^2) \cong \boldsymbol{B}$. (Here T_C is the holomorphic tangent bundle to C.) We represent α_a by $(0,1)$-forms valued in T_C, which we call by the same name, and we define

$$H_a = \int_C \alpha_a \wedge \mathrm{Tr}\,\varphi^2. \tag{4.26}$$

We claim that these functions are Poisson-commuting with respect to the holomorphic symplectic form Ω_I.

A natural proof uses the fact that the definition of the H_a makes sense on the infinite-dimensional space $\mathcal{W}$, before taking the hyper-Kahler quotient. Using the symplectic structure Ω_I on $\mathcal{W}$ to define Poisson brackets, the H_a are obviously Poisson-commuting. For in these Poisson brackets, given the form (4.9) of Ω_I, φ_z is conjugate to $A_{\overline{z}}$ and commutes with itself. But the H_a are functions of φ_z only, not $A_{\overline{z}}$.

The functions H_a can be restricted to the locus with $\nu_I = 0$, and then, because they are invariant under the $G_{\mathbb{C}}$-valued gauge transformations, they descend to holomorphic functions on $\mathcal{M}_H$. A general property of symplectic

[18] The reader may be unaccustomed to completely integrable systems in this holomorphic sense. From such systems, one can extract completely integrable Hamiltonian systems in the ordinary real sense (and moreover, interesting and significant constructions arise in this way; see [63,64] for some examples). We pick C to admit a real structure – that is, an involution that reverses its complex structure. This induces real structures on $\mathcal{M}$ and $\mathcal{M}_H$. By specializing to a real slice in $\mathcal{M}_H$, one gets then completely integrable Hamiltonian systems in which the phase space coordinates, the symplectic structure, and the commuting Hamiltonians are all real.

reduction (in which one sets to zero a moment map, in this case ν_I, and then divides by the corresponding group, in this case the group of $G_{\mathbb{C}}$-valued gauge transformations) is that it maps Poisson-commuting functions to Poisson-commuting functions. So the H_a are Poisson-commuting as functions on $\mathcal{M}_H$. There are enough of them to establish the complete integrability of $\mathcal{M}_H$.

One point to note in this construction is that the H_a, as functions on $\mathcal{W}$, depend on the particular choice of T-valued $(0,1)$-forms α_a that represent the cohomology classes. Any choice will do, but we have to make a choice. But after restricting and descending to $\mathcal{M}_H$, the functions we get on $\mathcal{M}_H$ depend only on the cohomology classes of the α_a. In fact, once we have $\overline{D}\varphi = 0$ and hence $\overline{\partial}\mathrm{Tr}\,\varphi^2 = 0$, the H_a are clearly invariant under $\alpha_a \to \alpha_a + \overline{\partial}\epsilon_a$.

The Poisson-commuting functions H_a generate commuting flows on $\mathcal{M}_H$ that are holomorphic in complex structure I. Complex tori admit commuting flows, and one might surmise that the orbits generated by the H_a are complex tori at least generically. This follows from general results about holomorphic symplectic structures and compactness of the fibers of the Hitchin fibration and can also be demonstrated more directly, using the technique of the spectral cover [20,62]. This technique has many applications in further development of this subject, as will be explained elsewhere [26]. In this paper, we will get as far as we can without using the spectral cover construction.

The analog of the above construction for any G is to replace $\mathrm{Tr}\,\varphi^2$ by a general gauge-invariant polynomial $\mathcal{P}_i$ of degree d_i. The associated commuting Hamiltonians take the form $H_{\alpha,i} = \int_C \alpha\,\mathcal{P}_i(\varphi)$, for $\alpha \in H^1(C, K_C^{1-d_i})$. A simple dimension counting, using the dimension formula (4.24), shows that these Hamiltonians are precisely sufficient in number to establish the complete integrability of $\mathcal{M}_H$. By using the fact that the fibers of the Hitchin fibration are compact, so that a holomorphic function must be a pullback from the base of this fibration, one can show that the commuting Hamiltonians generate the ring of holomorphic functions on $\mathcal{M}_H$.

One easy and important consequence of complete integrability is that the fibers of the Hitchin fibration are Lagrangian submanifolds in the holomorphic symplectic structure Ω_I or equivalently in the real symplectic structures ω_J and ω_K. Indeed, a fiber of this fibration is defined by equations $H_k - h_k = 0$, where H_k are the commuting Hamiltonians and h_k are complex constants. In general, the zero set of a middle-dimensional collection of Poisson-commuting functions, such as $H_k - h_k$ in the present case, is Lagrangian.

5. Topological Field Theory In Two Dimensions

In section 3, we analyzed a twisted version of $\mathcal{N} = 4$ super Yang-Mills theory on a general four-manifold M. The twisting preserved those supersymmetries that are invariant under a subgroup $Spin'(4) \subset Spin(4) \times Spin(6)$. The condition for a supersymmetry generator to be invariant under $Spin'(4)$ is

$$(\Gamma_{\mu\nu} + \Gamma_{4+\mu,4+\nu})\,\epsilon = 0, \quad \mu,\nu = 0,\dots,3. \tag{5.1}$$

In addition, all 16 supersymmetry generators obey a ten-dimensional chirality condition,

$$\Gamma_0\Gamma_1\dots\Gamma_9\epsilon = \begin{cases} \epsilon & \text{in Lorentz signature} \\ -i\epsilon & \text{in Euclidean signature.} \end{cases} \tag{5.2}$$

In the twisted theory, four of the six scalar fields $\phi_0,\dots,\phi_5$ of $\mathcal{N} = 4$ super Yang-Mills theory are reinterpreted as a one-form $\phi = \sum_{\mu=0}^{3}\phi_\mu\,dx^\mu$. Two scalar fields, ϕ_4 and ϕ_5, are untwisted, and are rotated by a group $SO(2)_{\mathcal{R}}$ (whose generator we called $\mathcal{K}$ in section 3.0). Allowing for the fermions, the symmetry group of the twisted theory is $Spin(2)_{\mathcal{R}}$, the double cover of $SO(2)_{\mathcal{R}}$.

The twisted theory is a formal construction in the sense that twisting violates unitarity and only works in Euclidean signature.[19] Suppose, however, that we split off the time direction and take $M = \mathbb{R}\times W$, where $\mathbb{R}$ parametrizes the time, W is a three-manifold, and the metric on M is a product. Then time is not involved in the twisting, and the twisted theory makes sense with Lorentz or Euclidean signature and is unitary and physically sensible. In this case, as the holonomy of M reduces to $SO(3)$, unbroken supersymmetries need only be invariant under a subgroup $Spin'(3) \subset Spin'(4)$; equivalently, we need only impose (5.1) for $\mu,\nu = 1,\dots,3$. There are four unbroken supersymmetry generators, which obey (5.1) for $\mu,\nu = 1,\dots 3$. Three scalars are untwisted, namely ϕ_0, ϕ_4 and ϕ_5, and are rotated by a $Spin(3)_{\mathcal{R}}$ symmetry that extends the $Spin(2)_{\mathcal{R}}$ symmetry that is present on any M.

[19] To construct a twisted theory in Lorentz signature, we would have needed a suitable homomorphism $Spin(1,3) \to Spin(6)$. Because of the compactness of $Spin(6)$, a non-trivial homomorphism does not exist. If one replaces $Spin(6)$ by $Spin(1,5)$ to work around this, the couplings to fermions will no longer be hermitian and the energy is no longer bounded from below.

Let us specialize this further to the case $M = \mathbb{R}^{1,1} \times C$ (or $\mathbb{R}^2 \times C$), for C a Riemann surface. The holonomy group of M reduces to $SO(2)$, and (5.1) collapses to the single condition

$$\Gamma_{2367}\epsilon = \epsilon, \tag{5.3}$$

leaving an eight-dimensional space of unbroken supersymmetries. Four scalars ϕ_0, ϕ_1, ϕ_4, and ϕ_5 are untwisted, and the symmetry group that rotates the untwisted scalars is now extended to $Spin(4)_{\mathcal{R}}$.

In fact, the structure in two dimensions is precisely that of $(4,4)$ supersymmetry. If we combine (5.3) and (5.2), we find (in Lorentz signature)

$$\Gamma_{01}\epsilon = \Gamma_{4589}\epsilon. \tag{5.4}$$

Here Γ_{01} is the operator that distinguishes the two spinor representations of the Lorentz group $SO(1,1)$ of $\mathbb{R}^{1,1}$. Similarly, Γ_{4589} distinguishes the two spin representations of $Spin(4)_{\mathcal{R}} = SU(2)' \times SU(2)''$. These two spin representations transform under $SU(2)' \times SU(2)''$ as $(\mathbf{2},\mathbf{1})$ and $(\mathbf{1},\mathbf{2})$, respectively. Eqn. (5.4) says that the eigenvalues of Γ_{01} and Γ_{4589} are equal for all two-dimensional supersymmetries.

So with a suitable choice of orientation of $\mathbb{R}^{1,1}$, the supersymmetries transform as $(\mathbf{2},\mathbf{1})^- \oplus (\mathbf{1},\mathbf{2})^+$, where the superscript labels the eigenvalue of Γ_{01}. But this is the usual structure of $(4,4)$ supersymmetry in two dimensions: one $SU(2)$ group of R-symmetries – namely $SU(2)'$ – acts on the supersymmetries of negative $Spin(1,1)$ chirality, while a second such group – $SU(2)''$ – acts on supersymmetries of positive chirality. In fact, in section 3, we have already identified the effective low energy theory as a sigma-model in which the target space is the hyper-Kahler manifold $\mathcal{M}_H$. In general, sigma-models with hyper-Kahler targets have $(4,4)$ supersymmetry [65].

In addition to the $Spin(4)_{\mathcal{R}}$ global symmetry, the twisted gauge theory on $\mathbb{R}^{1,1} \times C$ has an additional symmetry that appears because the holonomy group of C, by which we twist, is abelian. Even after twisting ϕ_2 and ϕ_3, it is still possible to make a rotation of the $\phi_2 - \phi_3$ plane. This is the symmetry group, called $\mathcal{U}_1$ in section 4.0, that acts on a Hitchin pair (E,φ) by rotating the phase of φ.

5.1. Twisted Topological Field Theories

We now want to study the sigma-model of target $\mathcal{M}_H$ from the viewpoint of topological field theory, so we replace $\mathbb{R}^{1,1}$ by a general two-manifold Σ that

we eventually will take to have Euclidean signature. We want to discuss what twisted topological field theories on Σ can be constructed from the sigma-model of target $\mathcal{M}_H$, and which of these actually arise by compactifying on the Riemann surface C the four-dimensional topological field theories that we constructed in section 3.

To construct a topological field theory from a sigma-model with a target X, one picks a *pair* of complex structures (J_+, J_-) on X that obey certain conditions of $(2,2)$ supersymmetry [66].[20] These conditions have recently been reinterpreted [67] in terms of generalized complex geometry [69]. Once one has $(2,2)$ supersymmetry, there is a standard recipe [70], via the twisting procedure that we reviewed in section 3.0, for constructing a topological field theory.

If X is hyper-Kahler, things simplify. The sigma-model with target X has $(4,4)$ supersymmetry, and a structure of $(2,2)$ supersymmetry can be obtained by picking a $(2,2)$ subalgebra of the $(4,4)$ supersymmetry algebra.[21] In terms of choosing complex structures, this amounts to the following. X has a family of complex structures parametrized by a copy of $\mathbb{CP}^1$ that we call $\mathbb{CP}^1_h$, and a $(2,2)$ structure can be defined by picking a pair of points $(J_+, J_-) \in \mathbb{CP}^1_h$. The conditions in [66] are automatically obeyed for such a pair. In the case of $\mathcal{M}_H$, this gives the family of topological field theories that is important for the geometric Langlands program.

We can conveniently characterize as follows the topological field theory associated with the pair (J_+, J_-). Let $\Phi : \Sigma \to \mathcal{M}_H$ be the bosonic field of the sigma-model. Then the equations of unbroken supersymmetry read

$$(1 - iJ_+)\overline{\partial}\Phi = 0, \tag{5.5}$$

and

$$(1 - iJ_-)\partial\Phi = 0. \tag{5.6}$$

Here ∂ and $\overline{\partial}$ are the usual $(1,0)$ and $(0,1)$ parts of the exterior derivative on Σ; and $\partial\Phi$ and $\overline{\partial}\Phi$ are understood as one-forms on Σ with values in the

[20] Generalized complex geometry leads to an extension of this construction, which we will not need here, in which only one of the two generalized complex structures that can be formed [67] from the pair (J_+, J_-) is integrable. One then aims as in [68] to define a topological field theory using this integrable generalized complex structure.

[21] Two $(2,2)$ subalgebras that differ by conjugation by $Spin(4)_{\mathcal{R}}$ are considered equivalent.

pullback of the tangent bundle TX (on which J_+ and J_- act). The first condition says that the map $\Phi : \Sigma \to X$ is holomorphic in complex structure J_+; the second says that it is antiholomorphic in complex structure J_-. Analogously to how we found the conditions of unbroken supersymmetry in section 3.1, these equations arise because there are fermi fields χ_+, χ_- with $\{Q, \chi_+\} = (1 - iJ_+)\overline{\partial}\Phi$, $\{Q, \chi_-\} = (1 - iJ_-)\partial\Phi$.

There is much more information in the equations (5.5) and (5.6) than in their solutions. To discuss the solutions of the equations, one usually specializes to Σ of Euclidean signature (the most natural case for topological field theory). Then for generic J_+, J_-, they imply that the map $\Phi : \Sigma \to X$ must be constant, a result that does not depend on J_+ and J_-. The equations themselves, understood algebraically, and without fixing a real structure on the two-manifold Σ, depend on J_+ and J_-.

An important example is the case $J_+ = J_- = \widehat{J}$ (where, in our applications, $\widehat{J}$ will be one of the complex structures I_w of $\mathcal{M}_H$, as described in eqn. (4.6)). The equations combine to $(1 - i\widehat{J})d\Phi = 0$, which imply that $d\Phi = 0$. The model is called the B-model in complex structure $\widehat{J}$. If instead $J_+ = -J_- = \widehat{J}$, we get the A-model in complex structure $\widehat{J}$. (This model really depends on the symplectic form $\omega_{\widehat{J}}$ associated with complex structure $\widehat{J}$ rather than the complex structure *per se*. But calling it the A-model in complex structure $\widehat{J}$ is sometimes a convenient shorthand.) For Σ of Euclidean signature, the equations of the A-model are redundant, as eqn. (5.6) is the complex conjugate of (5.5). Either one of them asserts that the map $\Phi : \Sigma \to \mathcal{M}_H$ is holomorphic in complex structure $\widehat{J}$. If $J_- \neq \pm J_+$, we get a model that is neither an A-model nor a B-model, but can be reduced to an A-model using generalized complex geometry, as we discuss in section 5.2.

We recall from eqn. (4.6) the explicit form of the family of complex structures on $\mathcal{M}_H$ parametrized by $\mathbb{CP}^1_h$. In terms of an affine parameter w for $\mathbb{CP}^1_h$, we defined a complex structure I_w in which the holomorphic coordinates are $A_{\overline{z}} - w\phi_{\overline{z}}$ and $A_z + w^{-1}\phi_z$. Explicitly, we found

$$I_w = \frac{1 - \overline{w}w}{1 + \overline{w}w} I + \frac{i(w - \overline{w})}{1 + \overline{w}w} J + \frac{w + \overline{w}}{1 + \overline{w}w} K. \tag{5.7}$$

The pair (J_+, J_-) corresponds in this parametrization to a pair (w_+, w_-).

Reduction From Four Dimensions Now let us compare this to what we get by reduction from four dimensions. We learned in section 3 that gauge theory leads to a family of topological quantum field theories (TQFT's) defined on any four-manifold M and parametrized by a copy of $\mathbb{CP}^1$ that we

will call $\mathbb{CP}^1_g$. This family certainly reduces for $M = \Sigma \times C$ to a subfamily of the family $\mathbb{CP}^1_h \times \mathbb{CP}^1_h$ that is natural from the two-dimensional point of view. Both families come from a twisting procedure applied to $\mathcal{N} = 4$ super Yang-Mills theory on a four-manifold M. In this procedure, a topological field theory is obtained from the cohomology of a suitable linear combination Q of the supersymmetries. To get the family $\mathbb{CP}^1_g$, we require Q to be $Spin'(4)$-invariant so that the construction works for all M. But to get the sigma-model of target $\mathcal{M}_H$, we specialize to the manifold $M = \Sigma \times C$ of reduced holonomy, and then, as we discuss more explicitly later, a weaker condition on Q suffices, leading to the larger family $\mathbb{CP}^1_h \times \mathbb{CP}^1_h$.

We aim to identify the four-dimensional topological field theory family $\mathbb{CP}^1_g$ as a subfamily of the sigma-model family $\mathbb{CP}^1_h \times \mathbb{CP}^1_h$. The embedding of $\mathbb{CP}^1_g$ in $\mathbb{CP}^1_h \times \mathbb{CP}^1_h$ can be described by functions $w_+(t)$, $w_-(t)$. We can compute these functions by specializing to a convenient configuration or physical state.

A very convenient way to proceed is to simply abelianize the problem, working in a vacuum in which G is broken to its maximal torus by the expectation values of some of the untwisted scalar fields. In the abelian case, from (3.29), the equations of unbroken supersymmetry in four dimensions are

$$\begin{aligned} (F + t\, d\phi)^+ &= 0 \\ (F - t^{-1} d\phi)^- &= 0. \end{aligned} \tag{5.8}$$

Here, the one-form ϕ is a section of $T^*M = T^*\Sigma \oplus T^*C$, and the gauge field A is locally a one-form or section of T^*M. To compare to a sigma-model on Σ with target $\mathcal{M}_H(G, C)$, we must take Σ to have radius much greater than that of C. In that limit, the parts of A and ϕ that take values in $T^*\Sigma$ are "massive" and can be dropped; so A and ϕ can be interpreted as sections of T^*C, slowly varying on Σ.

We write y for a local complex coordinate on Σ and z for a local complex coordinate on C. The two-form $d\overline{y} \wedge d\overline{z}$ is selfdual, while $dy \wedge d\overline{z}$ is anti-selfdual. So the first equation in (5.8) leads to

$$\partial_{\overline{y}} \left(A_{\overline{z}} + t\phi_{\overline{z}}\right) = 0, \tag{5.9}$$

and the second leads to

$$\partial_y \left(A_{\overline{z}} - t^{-1}\phi_{\overline{z}}\right) = 0. \tag{5.10}$$

By comparing these results to eqns. (5.5) and (5.6), we can read off the functions $w_+(t)$ and $w_-(t)$. Eqn. (5.9) asserts that the map $\Sigma \to \mathcal{M}_H(G,C)$ is holomorphic if we take on $\mathcal{M}_H(G,C)$ the complex structure in which $A_{\overline{z}} + t\phi_{\overline{z}}$ is holomorphic. This agrees with eqn. (5.5), which asserts such holomorphy in complex structure I_{w_+}, if and only if $w_+ = -t$. Similarly, eqn. (5.10) asserts that the map $\Sigma \to \mathcal{M}_H(G,C)$ is antiholomorphic if we take on $\mathcal{M}_H(G,C)$ the complex structure in which $A_{\overline{z}} - t^{-1}\phi_{\overline{z}}$ is holomorphic. This agrees with eqn. (5.6), which asserts antiholomorphy in complex structure I_{w_-}, if and only if $w_- = t^{-1}$.

So the embedding of $\mathbb{CP}^1_g$ in $\mathbb{CP}^1_h \times \mathbb{CP}^1_h$ is defined by

$$\begin{aligned} w_+ &= -t \\ w_- &= t^{-1}. \end{aligned} \tag{5.11}$$

Some Simple Considerations Let us now make a few simple simple observations about this result.

First of all, when do we get a B-model? For a B-model, we want $w_+ = w_-$. So the condition is $t^{-1} = -t$, which occurs precisely for $t = \pm i$. Since the complex structures I_w for $w = \mp i$ coincide with J and $-J$, we get this way the B-model in complex structure J or $-J$.

When do we get an A-model? For an A-model, the complex structures I_{w_+} and I_{w_-} should be opposite. The map $w \to -1/\overline{w}$ maps I_w to its opposite (this is clear from (5.7)), so we get an A-model in two dimensions if $w_- = -1/\overline{w}_+$, which works out to $t = \overline{t}$. In other words, precisely for real t, we get an A-model. For example, for $t = 1$ or -1, we get the A-model in complex structure K or $-K$. For $t = 0$ or ∞, we get the A-model in complex structure I or $-I$.

Now, let us compare this family to the more complete family $\mathbb{CP}^1_h \times \mathbb{CP}^1_h$ of the sigma-model. At the end of section 4.2, we described the group $\mathcal{U} \cong \mathbb{C}^*$ of diffeomorphisms of $\mathcal{M}_H(G,C)$; it acts on the family of complex structures I_w by $w \to \lambda^{-1}w$, $\lambda \in \mathbb{C}^*$. The topological field theory determined by a pair[22] (w_+, w_-) is the same as that determined by $(\lambda^{-1}w_+, \lambda^{-1}w_-)$. How many really inequivalent TQFT's can we define from the two-dimensional point of view? The only invariant we can form from the pair (w_+, w_-) is the

[22] This topological field theory depends also on the hyper-Kahler metric and B-field of $\mathcal{M}_H$. The full dependence on all variables is determined in section 5.2 using generalized complex geometry.

ratio $q = w_+/w_-$. We see that

$$q = -t^2, \tag{5.12}$$

so all values of q can be achieved, but not quite uniquely. The points t and $-t$ on $\mathbb{CP}^1_g$ lead to equivalent theories in two dimensions. This equivalence reflects the action of the center of $SU(4)_{\mathcal{R}}$ (recall eqn. (3.14)).

Although all values of the invariant q do come from four-dimensional TQFT's, it is not quite true that all $\mathbb{C}^*$ orbits on $\mathbb{CP}^1_h \times \mathbb{CP}^1_h$ have representatives with such an origin. The missing orbits are the $\mathbb{C}^*$-invariant points $(0,0)$ and (∞,∞), and also the orbits in which w_+ or w_-, but not both, is 0 or ∞. Particularly notable is the fact that the points $(0,0)$ and (∞,∞) are not equivalent to theories that originate in four-dimensional TQFT's. These points correspond to the B-models in complex structures I and $-I$.

The B-model in complex structure I has been the starting point in mathematical efforts – briefly surveyed in the introduction – to interpret the geometric Langlands program in terms of the geometry of $\mathcal{M}_H$. Because the Hitchin fibration is holomorphic in complex structure I, the T-duality on the fibers of the Hitchin fibration (whose relation to four-dimensional S-duality we review soon) maps the B-model of complex structure I to itself, acting on D-branes via the Fourier-Mukai transform that is the starting point in the mathematical description. Although the point $(0,0)$ corresponding to this B-model is not in the family $\mathbb{CP}^1_g$ that comes from four-dimensional TQFT's, it is interesting that it can be infinitesimally perturbed to give points on the $\mathbb{C}^*$ orbits corresponding to almost any point on $\mathbb{CP}^1_g$. We simply perturb $(0,0)$ to (α,β) for infinitesimal α, β; the invariant $q = w_+/w_-$ is then $q = \alpha/\beta$, and can take any value for arbitrarily small α,β. Perhaps this fact will lead eventually to an understanding of the geometric Langlands program based on perturbing from the B-model in complex structure I. Our approach, however, will rely on the family $\mathbb{CP}^1_g$ that comes directly from four dimensions.

Although certain $\mathbb{C}^*$ orbits, such as the point $(0,0)$, do not arise by specializing a four-dimensional TQFT to $M = \Sigma\times C$, this does not mean that they cannot be described in four-dimensional gauge theory. In section 3.1, we obtained the family $\mathbb{CP}^1_g$ using the supersymmetry generator

$$\epsilon = u\epsilon_\ell + v\epsilon_r \tag{5.13}$$

that was constrained to be $Spin'(4)$-invariant. If we specialize to $M = \Sigma\times C$, there is no need to ask for $Spin'(4)$ invariance. The holonomy group of $\Sigma\times C$

is $SO(2) \times SO(2) \subset SO(4)$, and $Spin'(2) \times Spin'(2)$ invariance is enough. This means that we can generalize (5.13) to

$$\widehat{\epsilon} = (u + \widetilde{u}\,\Gamma^*)\,\epsilon_\ell + (v + \widetilde{v}\,\Gamma^*)\,\epsilon_r, \tag{5.14}$$

where

$$\Gamma^* = \begin{cases} \Gamma_{01} & \text{in Lorentz signature} \\ i\Gamma_{01} & \text{in Euclidean signature} \end{cases} \tag{5.15}$$

is the operator that distinguishes the two chiralities of two-dimensional spinors. We will adopt Euclidean signature here, as this is more natural for topological field theory.

The supersymmetry generators in (5.14) are precisely the ones that obey the following conditions:

$$\begin{aligned} \Gamma_{2367}\epsilon &= \epsilon \\ \Gamma_{0145}\epsilon &= \epsilon. \end{aligned} \tag{5.16}$$

The first is eqn. (5.3), which says that the supersymmetry generated by ϵ is unbroken by the curvature of C; the second says that it is similarly unbroken by the curvature of Σ. With this more general starting point, it is possible to get the whole family $\mathbb{CP}^1_h \times \mathbb{CP}^1_h$ of two-dimensional TQFT's from four-dimensional gauge theory, though not from a four-dimensional TQFT.

It actually is convenient to rewrite (5.14) in an eigenbasis of Γ^*:

$$\widehat{\epsilon} = \frac{1}{2}\,(1 - \Gamma^*)\,(u'\epsilon_\ell + v'\epsilon_r) + \frac{1}{2}\,(1 + \Gamma^*)\,(u''\epsilon_\ell + v''\epsilon_r). \tag{5.17}$$

Here (u', v') and (u'', v'') are, respectively, homogeneous coordinates for the two factors of $\mathbb{CP}^1_h \times \mathbb{CP}^1_h$. Let $\widehat{\delta}_T$ be the extended topological symmetry generated by $\widehat{\epsilon}$. To determine the two complex structures (J_+, J_-), or equivalently, to determine the pair (w_+, w_-), we just need to compute $\widehat{\delta}_T \chi^+_{\overline{y}\,\overline{z}}$ and $\widehat{\delta}_T \chi^-_{y\overline{z}}$. Setting these to zero will give the generalization of (5.8). To determine the generalization of the usual formulas from eqn. (3.27), all we need to know is that if $\widetilde{\chi}^+ = \Gamma^{\overline{y}\,\overline{z}} \chi^+_{\overline{y}\,\overline{z}}$ and $\widetilde{\chi}^- = \Gamma^{y\overline{z}} \chi^-_{y\overline{z}}$, then $\widetilde{\chi}^+ \Gamma^* = \widetilde{\chi}^+$ and $\widetilde{\chi}^- \Gamma^* = -\widetilde{\chi}^-$. Using this, we get

$$\begin{aligned} \widehat{\delta}_T \chi^+_{\overline{y}\,\overline{z}} &= u'(F - \phi \wedge \phi)_{\overline{y}\,\overline{z}} + v'(D\phi)_{\overline{y}\,\overline{z}} \\ \widehat{\delta}_T \chi^-_{y\overline{z}} &= v''(F - \phi \wedge \phi)_{yz} - u''(D\phi)_{yz}. \end{aligned} \tag{5.18}$$

If, therefore, $t' = v'/u'$ and $t'' = v''/u''$, then the generalizations of eqns. (5.9) and (5.10) are

$$\begin{aligned} \partial_{\overline{y}}(A_{\overline{z}} + t'\phi_{\overline{z}}) &= 0 \\ \partial_y(A_{\overline{z}} - (t'')^{-1}\phi_{\overline{z}}) &= 0. \end{aligned} \tag{5.19}$$

This determines the two complex structures:

$$(w_+, w_-) = (-t', (t'')^{-1}). \tag{5.20}$$

Now we can determine how S-duality acts on the full family $\mathbb{CP}^1_h \times \mathbb{CP}^1_h$. We learned in eqn. (3.13) that a duality transformation $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z})$ (or its analog for a gauge group that is not simply-laced) acts on t by

$$t \longrightarrow t\frac{c\tau + d}{|c\tau + d|}. \tag{5.21}$$

The same reasoning shows that t' and t'' transform in precisely the same way. From this and (5.20), it follows that the action of S-duality on $\mathbb{CP}^1_h \times \mathbb{CP}^1_h$ is

$$\begin{aligned} w_+ &\longrightarrow w_+\frac{c\tau + d}{|c\tau + d|} \\ w_- &\longrightarrow w_-\frac{|c\tau + d|}{c\tau + d}. \end{aligned} \tag{5.22}$$

An important special case is that the B-model in complex structure I, which corresponds to $(w_+, w_-) = (0, 0)$, is completely invariant under duality transformations. The A-model in complex structure I, which corresponds to $(w_+, w_-) = (0, \infty)$, is likewise invariant. This and other statements about two-dimensional TQFT's that are not on the distinguished family $\mathbb{CP}^1_g$ will be useful auxiliary tools, but as will become clear, the geometric Langlands program is really a statement or collection of statements about the distinguished family.

Dependence On The Metric In general, a two-dimensional TQFT with target $\mathcal{M}_H(G, C)$ will depend on the complex structure of C, because this influences the hyper-Kahler structure of $\mathcal{M}_H(G, C)$. However, if a two-dimensional TQFT, such as those parametrized by $\mathbb{CP}^1_g$, descends from a four-dimensional one, it must be independent of the complex structure on C. After all, the four-dimensional TQFT on $M = \Sigma \times C$ is independent of the metric on M and in particular on C.

In some cases, we can see directly that these models are independent of the complex structure of C. At $t = \pm i$, we get the B-model in complex structure $\pm J$; this complex structure is independent of the metric of C, as we observed in section 4.1. For real t, we get the A-model in a complex structure that is a linear combination of complex structures I and K. The A-model is determined by the corresponding symplectic structure, which is a linear combination of symplectic structures ω_I and ω_K; we found these in section 4.1 to be independent of the metric of C.

To go farther, we will use generalized complex geometry.

5.2. The Role Of Generalized Complex Geometry

Generalized complex geometry is a natural framework for describing the topological field theories that can be constructed by twisting two-dimensional sigma-models, such as the sigma-model with target $\mathcal{M}_H$. Our goal here is to use generalized complex geometry to clarify a few questions from a two-dimensional point of view. What is the family $\mathbb{CP}^1_g$ and why is it independent of the complex structure of C? How, from a two-dimensional point of view, can we understand the canonical parameter Ψ introduced in section 3.5? And what geometry of $\mathcal{M}_H$ is really needed in these constructions?

Let TX and T^*X be the tangent and cotangent bundles of a manifold X. We let $\widehat{T} = TX \oplus T^*X$ and we write a section of $\widehat{T}$ as $\begin{pmatrix} v \\ \xi \end{pmatrix}$, where v is a section of TX and ξ is a section of T^*X. $\widehat{T}$ has a natural indefinite metric, in which TX and T^*X are both null and the inner product of TX and T^*X is the natural one, in which the inner product of $\begin{pmatrix} v \\ \xi \end{pmatrix}$ and $\begin{pmatrix} v' \\ \xi' \end{pmatrix}$ is $v^i\xi'_i + v'^i\xi_i$.

A generalized almost complex structure on a manifold X is a linear transformation $\mathcal{I}$ of $\widehat{T} = TX \oplus T^*X$ that preserves the metric and obeys $\mathcal{I}^2 = -1$. If a certain integrability condition is obeyed, it is called a generalized complex structure [69]. (For more detail, see Gualtieri's thesis [67], as well as [71–74,68] for applications to sigma-models and [75–77] for applications to supergravity.) One basic example of a generalized complex structure is

$$\mathcal{I}_{\widehat{J}} = \begin{pmatrix} \widehat{J} & 0 \\ 0 & -\widehat{J}^t \end{pmatrix}, \tag{5.23}$$

where $\widehat{J} : TX \to TX$ is an ordinary complex structure, and $\widehat{J}^t : T^*X \to T^*X$ is its transpose. If ω is a symplectic structure, then a second basic

example of generalized complex structure is given by

$$\mathcal{I}_\omega = \begin{pmatrix} 0 & -\omega^{-1} \\ \omega & 0 \end{pmatrix}. \tag{5.24}$$

Here we regard ω as a linear map from TX to T^*X; ω^{-1} is the inverse map from T^*X to TX. The topological field theory associated with $\mathcal{I}_{\widehat{J}}$ is the B-model in complex structure $\widehat{J}$, and the one associated with $\mathcal{I}_\omega$ is the A-model with symplectic structure ω.[23]

In each of these cases, the B-field vanishes. A B-field can be turned on as follows. For any closed two-form B_0, let

$$\mathcal{M}(B_0) = \begin{pmatrix} 1 & 0 \\ B_0 & 1 \end{pmatrix}. \tag{5.25}$$

Then if $\mathcal{I}$ is an integrable generalized complex structure, so is

$$\mathcal{I}(B_0) = \mathcal{M}(B_0)\mathcal{I}\mathcal{M}(B_0)^{-1}. \tag{5.26}$$

The transformation $\mathcal{I} \to \mathcal{I}(B_0)$ has the effect of shifting the B-field by B_0. So in particular the topological field theory derived from $\mathcal{I}_{\widehat{J}}(B_0)$ is the B-model with complex structure $\widehat{J}$ and B-field B_0, and $\mathcal{I}_\omega(B_0)$ is similarly related to the A-model with symplectic structure ω and B-field B_0. Conjugation by $\mathcal{M}(B_0)$ is called the B-field transform.

It is shown in chapter 6 of [67] that the conditions [66] for a sigma model to have $(2,2)$ supersymmetry are equivalent to the existence of a pair of generalized complex structures obeying a certain compatibility condition. If the $(2,2)$ model has anomaly-free R-symmetries, then it can be twisted in two ways to make a topological field theory. It is believed that, in general, each of these topological field theories is determined by one of the two generalized complex structures and independent of the second. (For example, if the target is a Kahler manifold and the B-field is flat, then the two generalized complex structures are $\mathcal{I}_{\widehat{J}}$ and $\mathcal{I}_\omega$, as described above; the two topological field theories are the B-model, which is determined by the first, and the A-model, which is determined by the second.) Arguments supporting this claim can be found in [72,68], though a complete proof is not yet known.

[23] The A-model is most commonly considered on a Kahler manifold, and then ω is the Kahler form. Because the A-model makes sense on symplectic manifolds more generally, we simply refer to ω as the symplectic form.

In the case relevant to us that the $(2,2)$ model comes from a hyper-Kahler metric g with a pair of points $J_+, J_- \in \mathbb{CP}^1_h$, there is a slight simplification in the formulas of [67], because the B-field can be taken to vanish (and restored later by a B-field transform). We let $\omega_\pm$ be the two symplectic structures $\omega_\pm = gJ_\pm$. One of the two generalized complex structures determined by the pair J_+, J_- with the hyper-Kahler metric g is then, according to eqn. 6.3 of [67],

$$\mathcal{J} = \frac{1}{2}\begin{pmatrix} J_+ + J_- & -(\omega_+^{-1} - \omega_-^{-1}) \\ \omega_+ - \omega_- & -(J_+^t + J_-^t) \end{pmatrix}. \tag{5.27}$$

The second one, not relevant for us, is obtained by reversing the sign of J_- and ω_-.

Using (5.7) and $(w_+, w_-) = (-t, t^{-1})$, we have

$$\begin{aligned} J_+ &= \frac{1-\overline{t}t}{1+\overline{t}t} I - \frac{i(t-\overline{t})}{1+\overline{t}t} J - \frac{t+\overline{t}}{1+\overline{t}t} K \\ J_- &= -\frac{1-\overline{t}t}{1+\overline{t}t} I - \frac{i(t-\overline{t})}{1+\overline{t}t} J + \frac{t+\overline{t}}{1+\overline{t}t} K. \end{aligned} \tag{5.28}$$

The associated symplectic structures are

$$\begin{aligned} \omega_+ &= \operatorname{Im}\tau \left(\frac{1-\overline{t}t}{1+\overline{t}t}\omega_I - \frac{i(t-\overline{t})}{1+\overline{t}t}\omega_J - \frac{t+\overline{t}}{1+\overline{t}t}\omega_K \right) \\ \omega_- &= \operatorname{Im}\tau \left(-\frac{1-\overline{t}t}{1+\overline{t}t}\omega_I - \frac{i(t-\overline{t})}{1+\overline{t}t}\omega_J + \frac{t+\overline{t}}{1+\overline{t}t}\omega_K \right). \end{aligned} \tag{5.29}$$

(The factor of $\operatorname{Im}\tau$ was obtained in (4.15).) So we compute that the generalized complex structure determined by this data is

$$\mathcal{I}_t = \frac{1}{1+\overline{t}t} \begin{pmatrix} -i(t-\overline{t})J & -(\operatorname{Im}\tau)^{-1}((1-\overline{t}t)\omega_I^{-1} - (t+\overline{t})\omega_K^{-1}) \\ \operatorname{Im}\tau((1-\overline{t}t)\omega_I - (t+\overline{t})\omega_K) & i(t-\overline{t})J^t \end{pmatrix}. \tag{5.30}$$

(To get this formula, it helps to know that on a hyper-Kahler manifold, if $a^2+b^2+c^2=1$, then $(a\omega_I + b\omega_J + c\omega_K)^{-1} = a\omega_I^{-1} + b\omega_J^{-1} + c\omega_K^{-1}$.)

The first important observation is that I, K, and ω_J have disappeared. $\mathcal{I}_t$ depends only on J, ω_K, and ω_I, or equivalently on J and the holomorphic two-form $\Omega_J = \omega_K + i\omega_I$ (and its complex conjugate). But this data, as we noted in section 4.1, is independent of the metric of C. This explains,

from the two-dimensional point of view, why the family of topological field theories that we get by dimensional reduction from four dimensions does not depend on the metric.

An analogous family of generalized complex structures can be constructed on any hyper-Kahler manifold X and is described in section 4.6 of [67]. As is explained there (Proposition 4.34), the generic element in such a family is a B-field transform of a generalized complex structure derived from a symplectic structure. Indeed, a small calculation shows that

$$\mathcal{I}_t = \mathcal{M}(B_0) \begin{pmatrix} 0 & -\omega_0^{-1} \\ \omega_0 & 0 \end{pmatrix} \mathcal{M}(B_0)^{-1}, \tag{5.31}$$

where

$$\begin{aligned} \omega_0 &= \mathrm{Im}\,\tau \frac{1-\overline{t}^2 t^2}{(1+t^2)(1+\overline{t}^2)} \left(\omega_I - \frac{t+\overline{t}}{1-\overline{t}t}\omega_K\right) \\ B_0 &= -\mathrm{Im}\,\tau\, \frac{i(t^2-\overline{t}^2)}{(1+t^2)(1+\overline{t}^2)} \left(\omega_I + \frac{1-\overline{t}t}{t+\overline{t}}\omega_K\right). \end{aligned} \tag{5.32}$$

Therefore, for $t \neq \pm i$ the TQFT derived from the generalized complex structure $\mathcal{I}_t$ is equivalent to an A-model with symplectic structure ω_0 and B-field B_0. This fact will enable us to understand from a two-dimensional point of view the canonical parameter Ψ introduced in section 3.5.

In general, the A-model with symplectic form ω_0 and B-field B_0 depends only on the cohomology class $[B_0 + i\omega_0]$. (On a Kahler manifold, this is called the complexified Kahler class.) In the present problem, since the cohomology class of ω_K vanishes (as we explained in section 4.1), we have

$$[B_0 + i\omega_0] = -i\,\mathrm{Im}\,\tau \left(\frac{t-t^{-1}}{t+t^{-1}}\right)[\omega_I]. \tag{5.33}$$

Thus, when the four-dimensional θ angle vanishes (we took it to vanish by starting with $\mathcal{I}_t$ rather than a B-field transform thereof), the model depends on $\mathrm{Im}\,\tau$ and t only in the combination that appears in (5.33).

The four-dimensional θ angle induces, as we explained at the end of section 4.0, an additional contribution $B' = -(\theta/2\pi)\omega_I = -(\mathrm{Re}\,\tau)\,\omega_I$ to the two-dimensional B-field. This can be incorporated in the generalized complex structure simply by conjugating with $\mathcal{M}(B')$. The resulting model depends on the cohomology class $[B_0 + B' + i\omega_0]$, which is $-[\omega_I]$ times

$$\Psi = \mathrm{Re}\,\tau + i\,\mathrm{Im}\,\tau \left(\frac{t-t^{-1}}{t+t^{-1}}\right). \tag{5.34}$$

This gives a two-dimensional interpretation of why the model depends on t and τ only in the combination Ψ.

A Few Loose Ends Finally, let us wrap up a few loose ends.

First of all, if t is real, the TQFT determined by complex structures $J_+(t), J_-(t)$ is an A-model to begin with, and the above argument showing that Ψ is the only relevant parameter did not really require generalized complex geometry. We could have based our reasoning on the real case together with holomorphy in t. We chose not to do this since we think that generalized complex geometry is a natural framework for understanding this problem, even though there are some technical gaps in the current understanding of it.

Second, in the above analysis, we used the fact that the A-model only depends on the cohomology class of $B + i\omega$. This is proved by writing the action as $\{Q, V\} + \int_\Sigma \Phi^*(\omega - iB)$, where $\Phi : \Sigma \to X$ is the sigma-model map.

In the above derivation, the B-field is not just of type $(1,1)$ (in a complex structure in which the symplectic form is Kahler); it also has components of type $(2,0) \oplus (0,2)$. In fact, the A-model is sensitive to all components of the B-field, including the part of type $(2,0) \oplus (0,2)$, but this point may require some clarification.

If $\Phi : \Sigma \to X$ is a holomorphic map, and B is a form on X of type $(2,0) \oplus (0,2)$, then $\Phi^*(B) = 0$. This makes the $(2,0) \oplus (0,2)$ part of B appear irrelevant, if one interprets the A-model purely as a mechanism for computing correlation functions by summing over holomorphic maps. But if one considers branes (as we most definitely will do to understand the geometric Langlands program), one immediately sees that all parts of the Hodge decomposition of B are relevant. A Lagrangian submanifold $N \subset X$ endowed with a Chan-Paton line bundle $\mathcal{L}$ of curvature F gives an A-brane if $F + B|_N = 0$. This condition is certainly sensitive to the $(2,0) \oplus (0,2)$ part of B.

Finally, our analysis showing that Ψ is the only relevant parameter is really not valid at $t = \pm i$, because of poles in the formulas. At these values of t, the model is actually not the B-field transform of an A-model; it is a B-model in complex structure $\pm J$. To complete our analysis for these values of t, we will argue directly using the B-model.

At $t = \pm i$, we have $\Psi = \infty$, independent of τ, so to complete the demonstration that Ψ is the only relevant parameter, we must show that τ is completely irrelevant at $t = \pm i$. $\mathrm{Im}\,\tau$ controls the Kahler class of $\mathcal{M}_H$, and this is certainly irrelevant in the B-model. Varying $\mathrm{Re}\,\tau$ adds to the B-field a multiple of ω_I. To show that this term is irrelevant in the B-model with complex structure J, we write $\omega_I = -i\Omega_J + i\omega_K$. The contribution

from ω_K is irrelevant because the form ω_K is exact. The contribution from Ω_J is irrelevant because Ω_J is a form of type $(2,0)$. But in the B-model, the B-field contribution $-i\int_\Sigma \Phi^*(B)$ can be written as $\{Q,\dots\}$ if B is of type $(1,1)$ or $(2,0)$. (By contrast, a $(0,2)$ component of the B-field does affect the category of B-branes [71].) Of course, in complex structure $-J$, we make the same argument, starting with $\omega_I = i\overline{\Omega}_J - i\omega_K$.

5.3. Some Specializations

In the rest of this paper except section 11.3, we focus primarily on the distinguished values $\Psi = 0$ and $\Psi = \infty$, and the duality

$$S = \begin{pmatrix} 0 & 1/\sqrt{n_{\mathfrak{g}}} \\ -\sqrt{n_{\mathfrak{g}}} & 0 \end{pmatrix} \tag{5.35}$$

that maps $\Psi \to -1/n_{\mathfrak{g}}\Psi$ and hence exchanges $\Psi = 0$ and $\Psi = \infty$. We recall that $n_{\mathfrak{g}} = 1$ for simply-laced G. When we speak of S-duality, we will be a little imprecise about whether we mean the full duality group Γ generated by S and by

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \tag{5.36}$$

or just the subgroup generated by S. In fact, a complete story requires including also T, but S is the main actor in the geometric Langlands program.

To get $\Psi = \infty$, we take $t = i$ and thus $(w_+, w_-) = (-i, -i)$. The resulting model is the B-model in complex structure J. This statement is unaffected by the choice of τ, but it is convenient to take τ imaginary.

Applying the transformation S, we find that τ remains imaginary, and t is mapped to $t' = -t(\tau/|\tau|) = -it = 1$. This leads to $(w_+, w_-) = (-1, 1)$, and hence to the A-model in complex structure K. Thus S-duality implies that the B-model for $\mathcal{M}_H(G,C)$ in complex structure J is equivalent to the A-model for $\mathcal{M}_H({}^LG,C)$ in complex structure K. If we combine S with a transformation $\mathcal{I} : \varphi \to i\varphi$ that rotates complex structure K back to complex structure J, we get a mirror symmetry that maps the B-model in complex structure J to the A-model in complex structure J. So the complex manifolds $\mathcal{M}_H(G,C)$ and $\mathcal{M}_H({}^LG,C)$ are a mirror pair in complex structure J.

The automorphism $\mathcal{I}$ that we used here is not quite natural, however, in the sense that it moves us out of the four-dimensional family of TQFT's parametrized by $\mathbb{CP}^1_g$. It has to move us out of this family, since the A-model in complex structure J is not a member of this family!

Table 2: A model and its S-dual. I_B, for example, represents the B-model in complex structure I, which is its own S-dual.

Model	S-Dual
I_B	I_B
I_A	I_A
J_B	K_A
J_A	K_B
K_B	J_A
K_A	J_B

Additionally, we already noted in discussing (5.22) that S (and the whole duality group) leaves invariant the B-model in complex structure I, and likewise the A-model in that complex structure. And finally, since S exchanges the B-model in complex structure J with the A-model in complex structure K, we may guess that it likewise exchanges the A-model in complex structure J with the B-model in complex structure K. In fact, the automorphism $\mathfrak{I}$ relates these two statements, but instead of arguing this way, let us use eqn. (5.22). The A-model in complex structure J corresponds to $(w_+, w_-) = (-i, i)$. This is mapped by S to $(w_+, w_-) = (1, 1)$, which indeed corresponds to the B-model in complex structure K. We summarize these statements in Table 2.

5.4. *S*-Duality Of The Hitchin Fibration

Our aim here and in section 5.5 is to show that S-duality acts classically on the base of the Hitchin fibration $\pi : \mathcal{M}_H \to \boldsymbol{B}$, while acting by T-duality on the fibers.

We recall that the base of the Hitchin fibration is a complex vector space $\boldsymbol{B}$. As described in section 4.2, the linear functions on $\boldsymbol{B}$ are the commuting Hamiltonians of the integrable system $\mathcal{M}_H$. They are of the form

$$H_{\mathcal{P},\alpha} = \int_C \alpha \mathcal{P}(\varphi), \tag{5.37}$$

where $\mathcal{P}$ is one of the fundamental homogeneous invariant polynomials on the Lie algebra $\mathfrak{g}$ of G. If $\mathcal{P}$ of degree n, then α is an element of $H^1(C, K_C^{1-n})$. The group $\mathcal{U} \cong \mathbb{C}^*$, which acts on φ by $\varphi \to \lambda\varphi$, acts as $\mathcal{P} \to \lambda^n \mathcal{P}$. This group action endows the holomorphic polynomial functions on $\boldsymbol{B}$ with the structure of a graded ring.

Roughly speaking, we want to show that S-duality preserves not only the Hitchin fibration but also this graded ring of functions on the base. This statement is trivial for T (which acts trivially on everything in sight) so it is really a statement about the duality transformation S. S maps G to ${}^LG$, and the claim is that it maps the sigma-model of $\mathcal{M}_H(G,C)$ to that of $\mathcal{M}_H({}^LG,C)$, mapping one Hitchin fibration to the other, and preserving the grading.

However, we need to explain exactly how to interpret these notions quantum-mechanically. After compactifying the topological gauge theory on $\Sigma\times C$, we pick a point $z\in\Sigma$ and then, if $\widehat{P}$ is any polynomial in the $H_{\mathcal{P},\alpha}$'s, we evaluate $\widehat{P}$ at z to get a local operator $\mathcal{O}_{\widehat{P}}(z)$ in the effective two-dimensional sigma-model on Σ. The operators of this type, for any z, form a graded ring R; holomorphy in φ ensures that there are no singularities in products of these operators. The claim we wish to justify is that S establishes an isomorphism between the graded rings $\mathsf{R}(G)$ and $\mathsf{R}({}^LG)$ for the two dual groups. (There are no local operators analogous to the $\mathcal{O}_{\widehat{P}}$'s that can be similarly used to measure the fibers of the Hitchin fibration, so there is no analogous way to show that S-duality acts classically on the fibers. It hardly can, given its relation to mirror symmetry!)

The claim can be established directly by considering the sigma-model with target $\mathcal{M}_H$. We simply note that in the B-model of complex structure I, the ring we have just defined is the same as the subring of ghost number zero (or cohomological degree zero) of what is customarily called the chiral ring [78]. In general, in the B-model with target X, the chiral ring is the bi-graded ring $H^q(X,\wedge^pTX)$ (TX is the tangent bundle of X). In particular, the subring of the chiral ring with $p=q=0$ is just the ring of holomorphic functions on X. But in complex structure I, a holomorphic function on $\mathcal{M}_H$ must be constant on the fibers of the Hitchin fibration (which are compact complex submanifolds) and hence must come from a holomorphic function on the base. So the holomorphic functions on $\mathcal{M}_H$, in complex structure I, are precisely the holomorphic functions on $\boldsymbol{B}$. Now the duality transformation S preserves the B-model in complex structure I, as we have learned above, or more precisely it maps this B-model for $\mathcal{M}_H(G,C)$ to the corresponding model for $\mathcal{M}_H({}^LG,C)$. So S maps the chiral ring of $\mathcal{M}_H(G,C)$ to that of $\mathcal{M}_H({}^LG,C)$. Equivalently, the base $\boldsymbol{B}$ of the Hitchin fibration for G maps to the analogous base ${}^L\boldsymbol{B}$ for ${}^LG$. Moreover, S commutes with the R-symmetry group $\mathcal{U}_1$ (the unitary subgroup of $\mathcal{U}\cong\mathbb{C}^*$). This plus holomorphy ensures that the action of S is compatible with the $\mathbb{C}^*$ action on the two sides.

What we have just seen is a typical example of exploiting the fact that $\mathcal{M}_H$ is a hyper-Kahler manifold. We made the argument in the B-model of

complex structure I even though we will apply the results to the B-model in complex structure J and the A-model in complex structure K.

One can argue the same result more explicitly starting from four-dimensional gauge theory. The four-dimensional argument gives more information; it will tell us precisely how the duality acts on the base of the Hitchin fibration.

We begin with $\mathcal{N} = 4$ super Yang-Mills theory on $\mathbb{R}^4$. For any positive integer n, let $\mathcal{D}_n$ (or $\mathcal{D}_n(G)$ if we wish to specify the gauge group) be the space of local operators of the following form. An element of $\mathcal{D}_n$ is a G-invariant polynomial function $\widehat{Q}(\phi_0, \dots, \phi_5)$ in the ϕ_i, homogeneous of degree n, and moreover constrained as follows: under $Spin(6)_{\mathcal{R}} = SU(4)_{\mathcal{R}}$, $\widehat{Q}$ transforms in the representation $\mathrm{Sym}^n_0\mathbf{6}$ (the representation consisting of traceless n^{th} order polynomials in the $\mathbf{6}$ or in other words the irreducible representation whose highest weight is n times the highest weight of the $\mathbf{6}$).

We interpret an element of $\mathcal{D}_n$ as a local operator that can be evaluated at an arbitrary point $x \in \mathbb{R}^4$. These operators are precisely the "half-BPS" operators of dimension n in $\mathcal{N} = 4$ super-Yang-Mills theory, that is, the operators that are annihilated by one-half of the superconformal symmetries that leave fixed the point x.

Because the space $\mathcal{D}_n$ is defined by an intrinsic criterion in terms of the action of supersymmetry, S-duality must establish an isomorphism between $\mathcal{D}_n(G)$ and $\mathcal{D}_n({}^LG)$. In particular, if the Lie algebra of G is selfdual ($\mathcal{N} = 4$ super Yang-Mills theory on $\mathbb{R}^4$ depends only on the Lie algebra of G), one could hope that S-duality acts trivially on $\mathcal{D}_n$. This is indeed true for simply-laced G, if the scalars are normalized to have canonical kinetic energy, as can be seen from string-theoretic derivations of S-duality. For G_2 and F_4, the duality transformation S acts on $\mathcal{D}_n$ as a nontrivial involution [47].

If the graded sum $\mathcal{D} = \oplus_n \mathcal{D}_n$ were a graded ring under operator products, we would use this ring to draw the conclusions we want about the Hitchin fibration. This actually is not true, mainly because of the condition on how $\widehat{Q}$ transforms under $Spin(6)_{\mathcal{R}}$.

We do, however, get a graded ring if we fix a Weyl chamber of $Spin(6)_{\mathcal{R}}$ and restrict to those operators in $\mathcal{D}_n$ that transform as highest weight vectors. Indeed, consider the subgroup $Spin(2) \times Spin(4) \subset Spin(6)_{\mathcal{R}}$, where, in our construction of the twisted TQFT, $Spin(2)$ rotates the $\phi_2 - \phi_3$ plane and $Spin(4)$ rotates the untwisted scalars. (Thus, $Spin(2)$ is precisely the group $\mathcal{U}_1$ that acts by phase rotations of φ.) Then in the representation of $Spin(6)$ spanned by the scalars $\phi_0, \dots, \phi_5$, the field φ can be interpreted as a highest weight vector, for some choice of Weyl chamber. And the subspace

$\mathcal{D}'_n \subset \mathcal{D}_n$ of highest weight vectors consists of the gauge-invariant polynomials $P(\varphi)$. Alternatively, these are the $Spin(4)$-invariants with the largest possible eigenvalue of $Spin(2)$. Since S-duality commutes with $Spin(6)_{\mathcal{R}}$ and hence with $Spin(2)\times Spin(4)$, these conditions are invariant under S-duality. Hence, S-duality maps $\mathcal{D}'_n$ to itself.

Now $\mathcal{D}' = \oplus_{n\geq 0}\mathcal{D}'_n$ *does* form a graded ring under operator products. It is part of what is usually called the chiral ring of the gauge theory (the operators $P(\varphi)$ are chiral superfields with respect to one of the supersymmetries, and their products are also chiral superfields). The action of S on $\mathcal{D}'$ preserves its structure as a graded ring, since this structure is part of the operator product expansion of the theory.

The transformation of gauge-invariant polynomials in φ under S-duality is, for dimensional and symmetry reasons, not affected by twisting and compactification. Moreover, a commuting Hamiltonian $H_{\mathcal{P},\alpha}$ of degree n in φ is obtained by integrating an element $\mathcal{P}(\varphi) \in \mathcal{D}'_n$ over C (more exactly, over $z \times C \subset \Sigma \times C = M$, for some point $z \in \Sigma$) with some weight $\alpha \in H^1(C, K_C^{1-n})$. This process of integration over C does not affect the transformation under S-duality, so the action of S-duality on the graded ring of holomorphic functions on the base $\boldsymbol{B}$ of the Hitchin fibration is simply determined by the action on the four-dimensional graded ring $\mathcal{D}'$. In particular, therefore, the transformation S maps the base of the Hitchin fibration of G to the base of the Hitchin fibration for ${}^LG$, intertwining between the two $\mathbb{C}^*$ actions.

5.5. Two-Dimensional Interpretation Of *S*-Duality

Now we can obtain a useful characterization of how the operation S acts on the sigma-model of target $\mathcal{M}_H$. Because the S-transformation reduces essentially to mirror symmetry, we will follow the insight of Strominger, Yau, and Zaslow [79]: the key point is to understand the action of S on zerobranes. In this analysis, and in much of our later discussion of branes, it will suffice to consider branes in the effective two-dimensional sigma-model of target $\mathcal{M}_H$. How to interpret these branes in the underlying gauge theory is briefly discussed in section 12 and will be further described in [26]. The picture we describe has been argued [18,19] previously and has been exploited mathematically [21].

We start with a zerobrane $\mathcal{B}_p$ supported at a point $p \in \mathcal{M}_H(G, C)$. Like any brane, it corresponds to a boundary condition in the sigma-model. The specific meaning of a zerobrane supported at p is as follows: if this boundary

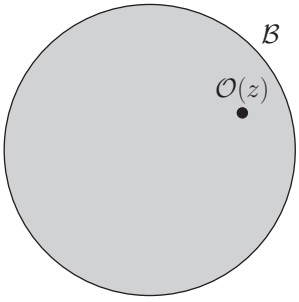

Figure 3: A local operator $\mathcal{O}(z)$ inserted at a point z in the interior of a disc (shaded region) whose boundary is labeled by a brane $\mathcal{B}$. As z approaches the boundary of the disc, $\mathcal{O}(z)$ may approach a complex constant. If this holds for suitable operators $\mathcal{O}_{H_\alpha}$, we say that $\mathcal{B}$ is supported on the fiber $\boldsymbol{F}$ of the Hitchin fibration.

condition is imposed on a component γ of the boundary of Σ (in which case we say that γ is *labeled* by $\mathcal{B}_p$), this means that the sigma-model map $\Phi : \Sigma \to \mathcal{M}_H(G, C)$ maps γ to the point p.

A point is a complex submanifold in every complex structure. So a zero-brane is what we will call a brane of type (B, B, B); that is, it is a B-brane in each of the three complex structures I, J, K.

Because the zerobrane is supported at a single point, it lies on a single fiber $\boldsymbol{F}_p$ of the Hitchin fibration. This fiber is characterized by equations $H_\alpha = h_\alpha$, where H_α are the commuting Hamiltonians and h_α are complex constants. Quantum mechanically, we say that a brane $\mathcal{B}$ is supported on a fiber $\boldsymbol{F}_p$ if the operators $\mathcal{O}_{H_\alpha}(z)$, as z approaches a boundary labeled by the brane $\mathcal{B}$, approach the c-numbers h_α (Figure 3). The brane $\mathcal{B}_p$ certainly has this property.

Now we consider the duality operation S. It replaces G with the Langlands dual group ${}^L G$, and so turns the zerobrane $\mathcal{B}_p$ into a dual brane $\widetilde{\mathcal{B}}_p$ in the sigma-model whose target is $\mathcal{M}_H({}^L G, C)$. Looking back at Table 2, we see that, since $\mathcal{B}_p$ is a brane of type (B, B, B), $\widetilde{\mathcal{B}}_p$ will be a brane of type (B, A, A), that is, a B-brane in complex structure I and an A-brane in complex structures J and K.

The transformation S gives a map $\Xi : \boldsymbol{B} \to {}^L\boldsymbol{B}$, where $\boldsymbol{B}$ is the base of the Hitchin fibration for G and ${}^L\boldsymbol{B}$ is its analog for ${}^L G$. Concretely, if a point $v \in \boldsymbol{B}$ is defined by equations $H_\alpha - h_\alpha = 0$, and S maps H_α to ${}^L H_\alpha$, then $\Xi(v)$ is defined by equations ${}^L H_\alpha - h_\alpha = 0$. As explained above, the map Ξ is the identity map for simply-laced G.

Applying S to the situation of Figure 3, we observe that if for a brane $\mathcal{B}$, the operator $\mathcal{O}_{H_\alpha}(z)$ approaches h_α as z approaches the boundary, then for the dual brane $\widetilde{\mathcal{B}}$, the dual operator $\mathcal{O}_{\widetilde{H}_\alpha}(z)$ likewise approaches h_α in the same limit. Hence, if $\mathcal{B}$ is supported on a fiber $\boldsymbol{F}$ of the the Hitchin fibration for G, then its S-dual is a brane $\widetilde{\mathcal{B}}$ supported on the corresponding fiber $\widetilde{\boldsymbol{F}} = \Xi(\boldsymbol{F})$ of the Hitchin fibration for ${}^L G$.

Therefore, $\widetilde{\mathcal{B}}_p$ is a brane of type (B, A, A) that is supported on a fiber $\widetilde{\boldsymbol{F}}$ of the dual Hitchin fibration. Let us focus on complex structures J and K in which $\widetilde{\mathcal{B}}_p$ is an A-brane. In general, the support of an A-brane in a space X is at least middle-dimensional. The most familiar A-branes are supported on middle-dimensional Lagrangian submanifolds; there are also more exotic coisotropic A-branes, with support above the middle dimension [80]. But the middle dimension is the lower bound on the dimension of the support of an A-brane.

However, an A-brane such as $\widetilde{\mathcal{B}}_p$ that is supported on the fiber $\widetilde{\boldsymbol{F}}$ has support that is *at most* middle-dimensional. To reconcile these constraints, $\widetilde{\mathcal{B}}_p$ must have support that is precisely $\widetilde{\boldsymbol{F}}$. Indeed, as explained in section 4.3, $\widetilde{\boldsymbol{F}}$ is holomorphic in complex structure I and Lagrangian with respect to symplectic structures ω_J and ω_K. So a brane wrapped on $\widetilde{\boldsymbol{F}}$ and endowed with a flat unitary Chan-Paton bundle $\mathcal{L}$ is indeed a brane of type (B, A, A).

Let us next compare the moduli on the two sides, restricting ourselves to branes supported on $\boldsymbol{F}$ on one side or on $\widetilde{\boldsymbol{F}}$ on the other. With this restriction, the moduli space of $\mathcal{B}_p$ is just a copy of $\boldsymbol{F}$: p can be any point in $\boldsymbol{F}$.

Meanwhile, the moduli of the dual brane $\widetilde{\mathcal{B}}_p$ are a complex torus $\mathcal{J}_{\widetilde{\boldsymbol{F}}}$ that parametrizes flat Chan-Paton bundles on $\widetilde{\boldsymbol{F}}$ of rank 1. (The rank is 1, as otherwise the moduli space of $\widetilde{\mathcal{B}}_p$ would have dimension greater than that of $\mathcal{B}_p$, and S-duality could not hold.) $\mathcal{J}_{\widetilde{\boldsymbol{F}}}$ has the same dimension as $\boldsymbol{F}$ or $\widetilde{\boldsymbol{F}}$. Clearly, S-duality establishes an isomorphism between $\boldsymbol{F}$ and $\mathcal{J}_{\widetilde{\boldsymbol{F}}}$.

Of course, we could run this backwards. A zerobrane in $\mathcal{M}_H({}^L G, C)$ at a point in $\widetilde{\boldsymbol{F}}$ is similarly transformed by S to a brane in $\mathcal{M}_H(G, C)$ that is wrapped on $\boldsymbol{F}$, and endowed with a flat unitary Chan-Paton bundle. So S-duality likewise establishes an isomorphism between $\widetilde{\boldsymbol{F}}$ and $\mathcal{J}_{\boldsymbol{F}}$, the moduli space of flat Chan-Paton bundles on $\boldsymbol{F}$ of rank 1.

This picture is SYZ duality between torus fibrations. $\boldsymbol{F}$ parametrizes flat Chan-Paton bundles of rank one on $\widetilde{\boldsymbol{F}}$, and vice-versa. Informally, we can describe this by saying that the operation S acts by a map $\Xi : \boldsymbol{B} \to {}^L\boldsymbol{B}$ from

the base of one Hitchin fibration to the base of the dual fibration, together with a T-duality on the corresponding torus fibers. This is the usual SYZ picture of mirror symmetry, except that the map Ξ is usually assumed to be trivial.

That the corresponding fibers of the Hitchin fibrations for G and ${}^LG$ are dual complex tori has been shown for unitary groups by Hausel and Thaddeus [21]. For the exceptional Lie group G_2, this duality has been established recently by Hitchin [81]. The question has also been analyzed by Donagi and Pantev for any semi-simple Lie group using an abstract approach to spectral covers [82]. See also the work of Faltings on G-bundles [83].

A number of important subtleties about this duality have been omitted in this explanation. Some questions involve the center and fundamental group of G and are discussed briefly in section 7. Other questions involve the role of a spin structure on C and will not be analyzed in this paper. A discussion of the role of spin structures and more detail on the role of the center and fundamental group and the duality for unitary groups will appear elsewhere [26].

5.6. Branes On $\mathcal{M}_H$

We have met several interesting examples of branes on $\mathcal{M}_H$ – the zerobrane and the brane wrapped on a fiber. We will encounter others. As in our initial examples, most of the important branes will have supersymmetric properties with respect to all three complex structures I, J, and K, and this is very useful in understanding their behavior. The branes of most interest preserve half the supersymmetry of the effective two-dimensional sigma-model, which in turn has half the supersymmetry of the underlying gauge theory.

Examples include (B,B,B)-branes, which are B-branes in each complex structure I,J,K, as well as branes of type (B,A,A), (A,B,A), and (A,A,B), which are of B-type in one complex structure and of A-type in the two others. We have already seen examples of S exchanging a brane of type (B,B,B) with one of type (B,A,A). From Table 2, we further see that S maps a brane of type (A,B,A) to another brane of the same type, and likewise for a brane of type (A,A,B).

It is not difficult to give examples of branes of each of these four types; they will be the important branes in our study of the geometric Langlands program. Of course, we can more generally have a brane that is of B type

for any linear combination of the complex structures I, J, and K, and of A-type for the two orthogonal complex structures.[24]

The zerobranes that we have already studied are the most obvious branes of type (B, B, B). Other examples of (B, B, B)-branes are space-filling branes, whose target is all of $\mathcal{M}_H$, endowed with a Chan-Paton vector bundle that is holomorphic in each of the three complex structures. Bundles of that sort include the trivial bundle as well as other examples whose role in the geometric Langlands program will be discussed elsewhere [26].

A large class of examples of (B, A, A)-branes is as follows. Let N be a complex submanifold of $\mathcal{M}_H$ in complex structure I which is a Lagrangian submanifold with respect to the holomorphic symplectic structure Ω_I, and hence with respect to ω_J and ω_K. Then a brane supported on N with trivial Chan-Paton bundle is a (B, A, A)-brane. We have already encountered one important example: N might be a fiber of the Hitchin fibration. For another example, we can take N to be, from the standpoint of complex structure I, the space of all pairs (E, φ) where E is held fixed and φ is allowed to vary. From the point of view of gauge theory, this means that $A_{\overline{z}}$ is held fixed up to a gauge transformation, which ensures the vanishing of $\Omega_I = (1/\pi)\int_C |d^2z|\,\mathrm{Tr}\,\delta\phi_z \wedge \delta A_{\overline{z}}$. Hence it gives us a brane of type (B, A, A).

Similarly, examples of (A, B, A) and (A, A, B)-branes come from complex Lagrangian manifolds with respect to the complex structure J or K. A simple example of a complex Lagrangian manifold in complex structure J is obtained simply by keeping $\mathcal{A}_{\overline{z}} = A_{\overline{z}} + i\phi_{\overline{z}}$ fixed and letting $\mathcal{A}_{\overline{z}} = A_{\overline{z}} + i\phi_{\overline{z}}$ vary. This gives us the family of all flat $G_{\mathbb{C}}$ bundles that have a specified holomorphic structure. Similarly, we get a complex Lagrangian manifold in complex structure K by specifying $A_{\overline{z}} - \phi_{\overline{z}}$ and letting $A_z + \phi_z$ vary. In section 11, we will construct more sophisticated examples of branes of type (A, B, A) or (A, A, B) as "coisotropic branes."

6. Loop and Line Operators

What is really unusual about the four-dimensional TQFT's introduced in section 3.1, compared to other theories with a superficially similar origin, is that they admit operators that are associated to oriented one-manifolds

[24] However, there are no branes of type (B, B, A). A subvariety that is holomorphic in complex structures I and J is automatically holomorphic in complex structure K. And there likewise are no branes of type (A, A, A).

$\mathcal{S} \subset M$, and depend on the homotopy class, not just the homology class, of $\mathcal{S}$. The point is not that we want to study the fundamental group of M, but that these one-manifold operators prove to have very interesting properties in relation to branes.

If M is compact and without boundary, then $\mathcal{S}$ is required to be a closed loop. Reflecting this case, one-manifold operators are commonly called loop operators. More generally, however, on a noncompact four-manifold, we allow line operators that go off to infinity. An important example is a static line operator, whose trajectory spans all time at a given point in space. Such a line operator must be included in quantizing the theory to construct a Hilbert space of physical states. This will actually be the most important case in our applications.[25]

In section 6.1, we discuss the most elementary loop or line operators, usually called Wilson operators. Wilson operators in TQFT's similar to the ones we consider here, with the connection modified by a scalar field to ensure the topological symmetry, were first constructed by Blau and Thompson [84]. Similar gauge-invariant operators were also used in studying five-dimensional supersymmetric Yang-Mills theory [85]. Such operators are important in contemporary developments in string theory (along with their magnetic duals, which we introduce in section 6.2). They were first introduced in the half-BPS case in [86,87]. The topological line operators that we consider are most similar to the 1/16 BPS line operators defined in [88] for any loop $\mathcal{S} \subset \mathbb{R}^4$.

In section 6.2, we describe the dual 't Hooft operators. In section 6.3, we discuss the interpretation of such operators after compactification to two dimensions. Finally, in section 6.4, we study line operators as operators on branes.

6.1. Topological Wilson Operators

In gauge theory, the most elementary loop operator is the Wilson loop operator. We consider gauge theory with gauge group G and connection A on a G-bundle $E \to M$. We let $\mathcal{S}$ be an oriented loop in M, oriented say by the choice of a one-form ds, where s is a parameter along $\mathcal{S}$. We pick an

[25] Thus, other cases such as line operators that end on branes, though they make sense in general, will not be important in the present paper.

irreducible representation R of G, and let[26]

$$W_0(R,\mathcal{S}) = \mathrm{Tr}_R\, P\, \exp\left(-\oint_{\mathcal{S}} A\right). \tag{6.1}$$

In other words, $W_0(R,\mathcal{S})$ is the trace, in the representation R, of the holonomy of A around $\mathcal{S}$.

Now let us look back to eqn. (3.23) from section 3.1, in which we determined the transformation law of A under the topological symmetry:

$$\delta_T A_\mu = iu\psi_\mu + iv\widetilde{\psi}_\mu. \tag{6.2}$$

Clearly, for any non-zero (u,v), the Wilson loop operator is not invariant. However, twisted $\mathcal{N}=4$ Yang-Mills theory also has an adjoint-valued one-form $\phi = \sum_{\mu=0}^{3} \phi_\mu\, dx^\mu$, transforming as

$$\delta_T \phi_\mu = iv\psi_\mu - iu\widetilde{\psi}_\mu. \tag{6.3}$$

It is possible to modify the connection A by adding to it a multiple of ϕ.

Precisely if $t = v/u$ is equal to $\pm i$, a linear combination of A and ϕ is invariant. For $t=i$, the invariant combination is $\mathcal{A} = A + i\phi$; at $t=-i$, it is $\overline{\mathcal{A}} = A - i\phi$. We then define the supersymmetric Wilson loop operator as the holonomy of $\mathcal{A}$ or $\overline{\mathcal{A}}$ around $\mathcal{S}$. Thus

$$W(R,\mathcal{S}) = \mathrm{Tr}_R\, P \exp\left(-\oint_{\mathcal{S}} \mathcal{A}\right) = \mathrm{Tr}_R\, P \exp\left(-\oint_{\mathcal{S}} (A + i\phi)\right), \quad t = i \tag{6.4}$$

and

$$W(R,\mathcal{S}) = \mathrm{Tr}_R\, P \exp\left(-\oint_{\mathcal{S}} \overline{\mathcal{A}}\right) = \mathrm{Tr}_R\, P \exp\left(-\oint_{\mathcal{S}} (A - i\phi)\right), \quad t = -i. \tag{6.5}$$

The central generator $\mathcal{I}$ of $SU(4)_{\mathcal{R}}$, which acts by $t \to -t$ and $\phi \to -\phi$, transforms one formula into the other.

[26] Since our covariant derivative is $D = d + A$, the holonomy operator has a minus sign in the exponent. In terms of a hermitian gauge field $A' = iA$ (recall eqn. (2.2)), the holonomy operator is $\mathrm{Tr}_R\, P \exp\left(i \oint_{\mathcal{S}} A'\right)$.

The existence of Wilson loop operators possessing the topological symmetry is extremely natural at $t = \pm i$. At these values of t, the topological equations (3.29) assert the flatness of $\mathcal{A}$. So the holonomies of $\mathcal{A}$ or $\overline{\mathcal{A}}$ around closed loops are natural invariants. Equally well, such holonomies are not natural invariants if $t \neq \pm i$, since the topological equations do not necessarily imply flatness. (The vanishing theorems of section 3.3, which relate supersymmetric configurations to complex flat connections for generic t, require some global input and do not hold in all situations, for instance involving branes, to which one might apply the TQFT.)

Though we formulated these definitions for $\mathcal{S}$ a closed loop, they have a good analog for the open case. If $\mathcal{S}$ has endpoints p and q, we define $W(R, \mathcal{S})$ not as a trace but as the matrix of parallel transport (of the connection $\mathcal{A}$ or $\overline{\mathcal{A}}$) from the fiber of E at p, taken in the representation R, to the fiber at q. $W(R, \mathcal{S})$ is thus a map from a vector space at p to a vector space at q. It is then incorporated as part of a larger quantum construction involving initial and final quantum states, as mentioned in our introductory remarks. This extended version of the Wilson operator will be essential when we get to the geometric Langlands program in earnest.

When the context is clear, we abbreviate $W(R, \mathcal{S})$ as $W(R)$ or simply W.

6.2. Topological 't Hooft Operators

We have found Wilson operators in the topological field theory at $t = \pm i$, that is at $\Psi = \infty$. By the S-duality $S : \Psi \to -1/n_{\mathfrak{g}}\Psi$, there must be dual operators at $\Psi = 0$. The topological Wilson operators were classified by the choice of a representation R of G. So, as S exchanges G and ${}^LG$, the dual operators that appear in the TQFT at $\Psi = 0$ must be classified by a representation ${}^LR$ of ${}^LG$.

The Wilson operator is an example of what is often called an "order" operator in statistical mechanics. An order operator is built from a classical expression (such as the holonomy of the connection around a loop) that is then interpreted as a quantum operator and included as a factor in a path integral.

The dual of an order operator is frequently a "disorder" operator. A disorder operator is defined by specifying a singularity that the classical fields are supposed to have, and performing a path integral (or quantizing) in the presence of the singularity. Classically, the order and disorder operators seem like completely different kinds of things. But quantum mechanically, they turn out to have, in suitable cases, the same formal properties, and are frequently exchanged by duality. In our problem, the dual of a

Wilson operator is an 't Hooft operator, which is an example of a disorder operator.[27]

To define a Wilson loop operator associated with a loop $\mathcal{S} \subset M$, we required an orientation of $\mathcal{S}$. The 't Hooft loop operator instead requires an orientation of the normal bundle to $\mathcal{S}$. In fact, it is convenient to identify a small neighborhood of $\mathcal{S}$ with $\mathcal{S} \times \mathbb{R}^3$ and to write ϵ_3 for a volume-form on $\mathbb{R}^3$. The definition of the 't Hooft operator will depend only on the orientation of the normal bundle, not the details of these choices. If r denotes the distance from $\mathcal{S}$ (in some metric), then we can also write

$$\epsilon_3 = dr\, d\mathrm{Vol}, \tag{6.6}$$

where dVol is a volume-form on the two-sphere at fixed r.

Throughout this paper, when we write explicit formulas for 't Hooft operators in $U(N)$ or $SU(N)$ gauge theory, we consider the gauge fields as connections on a rank N vector bundle. This means that A and ϕ are are one-forms valued in $N \times N$ skew-hermitian matrices; for $N = 1$, they are simply imaginary one-forms.

Let us first describe the most basic 't Hooft operator in $U(1)$ gauge theory. On $\mathbb{R}^4 = \mathbb{R} \times \mathbb{R}^3$, we pick coordinates $(s, \vec{x})$. The singular Dirac monopole is a classical solution of Maxwell's equations that is invariant under translations of s and rotations of $\vec{x}$, and has a singularity at the line L defined by $\vec{x} = 0$. If $F = dA$ is the curvature of a $U(1)$ connection A, then Maxwell's equations read $dF = d\star F = 0$, and can be solved on $\mathbb{R}^3 \backslash \{0\}$ (that is, on the complement of the point $\vec{x} = 0$ in $\mathbb{R}^3$) by

$$F = \frac{i}{2} \star d\left(\frac{1}{|\vec{x}|}\right). \tag{6.7}$$

The use of the Hodge $\star$ operator shows explicitly the dependence on the orientation of the normal bundle to L. To get a solution on $\mathbb{R}^4 \backslash L$, we simply use the projection $\mathbb{R}^4 \backslash L \to \mathbb{R}^3 \backslash \{0\}$, and pull F back to a solution of Maxwell's equations on $\mathbb{R}^4 \backslash L$. F has been normalized so that if Y is a two-sphere enclosing the singularity at $\vec{x} = 0$, then $\int_Y iF/2\pi = 1$. Hence there exists a unitary line bundle $\mathcal{L} \to \mathbb{R}^4 \backslash L$ that admits a connection of curvature F. This is the singular Dirac monopole. It is characterized by $\int_Y c_1(\mathcal{L}) = 1$.

The basic 't Hooft operator of abelian gauge theory is defined by saying that the path integral is to be performed over connections on a line bundle

[27] For an explicit demonstration that the dual of a Wilson operator is an 't Hooft operator in the case of $G = U(1)$, see [89].

over $\mathbb{R}^4\backslash L$ that agree with the singular Dirac monopole near L. Here there is no need to consider only straight lines. L can be replaced by any closed curve $\mathcal{S}$ in a general four-manifold M, or (if M is geodesically complete) any curve that goes off to infinity. The path integral with an insertion of the 't Hooft loop operator $T_{(0)}$ on the curve $\mathcal{S}$ is obtained by performing a path integral for abelian gauge fields[28] on $M\backslash\mathcal{S}$ such that the curvature near $\mathcal{S}$ diverges as $F\sim\frac{i}{2}\star_3 d(1/r)$, where $r=|\vec{x}|$ is now the distance from $\mathcal{S}$ and $\star_3$ is an operation on forms that near $\mathcal{S}$ looks like the $\star$ operator on planes normal to $\mathcal{S}$. More generally, picking an integer m and requiring that the singular part of F is $F\sim(im/2)\star_3 d(1/r)$, we get the 't Hooft loop operator $T_{(0)}(m)$ of charge m.

Now let us consider $\mathcal{N}=4$ super Yang-Mills theory with gauge group $U(1)$. We can define an 't Hooft operator by asking that the gauge field have the Dirac monopole singularity while other fields are non-singular. But is such an operator compatible with the topological symmetry at $\Psi=0$?

We can decide this question as follows. To get to $\Psi=0$, we take $t=1$ and τ imaginary. In order for the 't Hooft operator to preserve the topological symmetry, the singular behavior of the fields near $\mathcal{S}$ must be compatible with that symmetry. Since the question concerns the local behavior near $\mathcal{S}$, it is convenient to temporarily revert to the case that $\mathcal{S}$ is a line $L\subset\mathbb{R}^4$.

The conditions for a set of fields (A,ϕ) to preserve the topological symmetry were found in eqns. (3.29). If those equations are compatible with the type of singularity required by the 't Hooft operator, then the 't Hooft operator preserves supersymmetry. Otherwise, it is impossible for any field configuration in the presence of the 't Hooft operator to be supersymmetric, and we say that the 't Hooft operator violates supersymmetry.

At $t=1$ and for abelian gauge theory, the requisite conditions read

$$(F+d\phi)^+=(F-d\phi)^-=0. \tag{6.8}$$

These equations are not satisfied if F has the Dirac monopole singularity and ϕ is non-singular. To get an 't Hooft operator that preserves the topological symmetry, just as in the Wilson case, we have to include ϕ in a suitable fashion.

[28] If the homology class of $\mathcal{S}$ in $H_1(M,\mathbb{Z})$ is nonzero, there are no such abelian gauge fields and we declare that this path integral vanishes. The dual of this is that if the homology class of $\mathcal{S}$ is nonzero, the Wilson operator $W(R,\mathcal{S})$ has a vanishing expectation value after integrating over all connections; this vanishing follows from the behavior in twisting by a flat line bundle.

What sort of singularity of ϕ must accompany the Dirac monopole singularity of F in order for the equations (6.8) to be satisfied? This question has a simple answer: we should take $\phi = (i/2|\vec{x}|)ds$, where s is the "time" coordinate along L. Thus, the equations (6.8) are obeyed if we take

$$\begin{aligned} F &= \frac{i}{2} \star_3 d\frac{1}{|\vec{x}|} \\ \phi &= \frac{i}{2|\vec{x}|} ds, \end{aligned} \tag{6.9}$$

where $\star_3$ is the three-dimensional Hodge $\star$ operator in the directions normal to L.

This tells us how to define, for a general closed one-manifold $\mathcal{S} \subset M$, the basic 't Hooft operator T that preserves the topological symmetry at $\Psi = 0$: to evaluate a matrix element of this operator, we perform a path integral over fields that possess near $\mathcal{S}$ the singularity described in (6.9). Similarly, for any integer m, to get a charge m topological 't Hooft operator $T(m)$, we simply require that the fields have a singularity along $\mathcal{S}$ that is m times what is described in (6.9).

Generalization To Any G And Langlands Duality The next step is to generalize this construction to arbitrary G. This is done in a simple fashion [90]. We simply pick an arbitrary homomorphism $\rho : U(1) \to G$. Then we define a topological 't Hooft operator by asking that the fields should have a singularity along $\mathcal{S}$ that looks like the image under ρ of the singularity in (6.9).

The operator defined this way depends on ρ only up to conjugacy. But conjugacy classes of homomorphisms $\rho : U(1) \to G$ are classified by highest weights of the dual group ${}^LG$. In other words, they are classified by irreducible representations of ${}^LG$.

So for every irreducible representation ${}^LR$ of ${}^LG$, and every one-manifold $\mathcal{S}$, we have obtained a topological 't Hooft operator $T({}^LR, \mathcal{S})$. As in the Wilson case, we refer to this operator as just $T({}^LR)$ or T if the context is clear. The fact that 't Hooft operators of G are classified by representations of ${}^LG$ is the classic GNO duality [10], as recently reinterpreted in terms of operators rather than states [90].

The classification of 't Hooft operators is also intimately related to Grothendieck's classification of G-bundles on $\mathbb{CP}^1$. According to this classification, a G-bundle on $\mathbb{CP}^1$ is associated to the fundamental line bundle $\mathcal{O}(1) \to \mathbb{CP}^1$ via a homomorphism $\rho : U(1) \to G$ (or its complexification $\rho : \mathbb{C}^* \to G_{\mathbb{C}}$). Let us identify with $\mathbb{CP}^1$ a small two-sphere surrounding the

singularity that defines an 't Hooft operator. Every G-bundle on a Riemann surface, even if unstable, admits a connection obeying the Yang-Mills equations [91], and on $\mathbb{CP}^1$ every solution of the Yang-Mills equations is equivalent to one of the abelian solutions used to construct 't Hooft operators. So the classification of 't Hooft operators is the same as the classification of G-bundles on $\mathbb{CP}^1$.

Topologically Non-Trivial 't Hooft Operators A holomorphic G-bundle over $S^2 = \mathbb{CP}^1$ may be topologically non-trivial, though the topological classification of G-bundles is much less fine than the holomorphic classification.

A G-bundle over S^2 is classified by a characteristic class $\xi \in H^2(S^2, \pi_1(G)) \cong \pi_1(G)$. But $\pi_1(G)$ is naturally isomorphic to the center of ${}^LG$. The bundle over S^2 that is used in constructing the 't Hooft operator $T({}^LR)$ is topologically non-trivial if and only if the center of ${}^LG$ acts nontrivially on ${}^LR$.

Combined Wilson-'t Hooft Operators: Abelian Case We have so far found Wilson line operators at $\Psi = \infty$ and 't Hooft line operators at $\Psi = 0$. By applying duality transformations, we find other special values of Ψ at which line operators exist that preserve the topological symmetry. In fact, all rational values of Ψ, and only those, can be obtained from $\Psi = 0$ (or $\Psi = \infty$) by a duality transformation. This is true regardless of whether the duality group Γ is $SL(2,\mathbb{Z})$, which it is for G simply-laced, or one of the Hecke groups that arise when G is not simply-laced.

To get to rational values of Ψ, we will take $t = 1$ with $\theta = 2\pi\Psi$. Thus, to the action of the gauge field, we add a term

$$\boldsymbol{I}_\theta = i\frac{\Psi}{4\pi}\int \mathrm{Tr}\, F\wedge F. \tag{6.10}$$

At this stage, it is necessary to assume that M is oriented, say by the choice of a volume-form ϵ_4. $\mathcal{N} = 4$ super Yang-Mills theory makes sense on an unorientable four-manifold M. But the interaction (6.10), whose role we now wish to consider, can only be introduced if M is oriented, since this is required in order to define the integral $\int_M \mathrm{Tr}\, F\wedge F$. (More generally, Ψ is odd under orientation-reversal, as we noted at the end of section 3.5, so to get to nonzero rational Ψ, we require an orientation on M.) So we assume that M is oriented, and express the orientation by the choice of a volume-form ϵ_4.

The definition of a Wilson loop operator supported on $\mathcal{S}$ required an orientation of $\mathcal{S}$ by a one-form ds. The definition of an 't Hooft loop operator

supported on $\mathcal{S}$ required an orientation ϵ_3 of the normal bundle to $\mathcal{S}$. Once M is oriented, we can ask for the orientations of $\mathcal{S}$ and its normal bundle to be compatible in the sense that

$$\epsilon_4 = ds \wedge \epsilon_3 \tag{6.11}$$

along $\mathcal{S}$. Identifying a small neighborhood of $\mathcal{S}$ in M with $\mathcal{S} \times \mathbb{R}^3$, we can extend ds over this neighborhood so that (6.11) remains true throughout $\mathcal{S} \times \mathbb{R}^3$. Then writing $\epsilon_3 = dr \wedge d\mathrm{Vol}$ as in (6.6), we have

$$\epsilon_4 = ds \wedge dr \wedge d\mathrm{Vol} = -dr \wedge ds \wedge d\mathrm{Vol}. \tag{6.12}$$

The line operators that preserve the topological symmetry at rational values of Ψ are called mixed Wilson-'t Hooft operators. Because these operators will combine the properties of 't Hooft and Wilson operators, to define them it is important to choose compatible orientations on $\mathcal{S}$ and its normal bundle, as we did in the last paragraph. To construct these operators, we first consider the case $G = U(1)$.

At first sight it may appear that the 't Hooft operators as we have defined them already possess the topological symmetry at any Ψ. This is in fact not true, because the singularity of the 't Hooft operator causes $\boldsymbol{I}_\theta$ not to be invariant under the topological symmetry.

This requires some explanation. We let A_0 be a fixed connection whose curvature has the Dirac monopole singularity along a curve $\mathcal{S}$. A path integral in the presence of a charge m operator $T_m(\mathcal{S})$ is evaluated as follows. One takes the gauge field to be $A = mA_0 + \widehat{A}$, where $\widehat{A}$ is smooth near $\mathcal{S}$. The path integral is then evaluated by integrating over $\widehat{A}$, with A_0 fixed. The precise choice of A_0 does not matter (as long as it has the right singularity), since we are going to integrate over all $\widehat{A}$.

As a step in this direction, let us evaluate $\boldsymbol{I}_\theta$ as a function of A. We write $F = mF_0 + \widehat{F}$, where F_0 and $\widehat{F}$ are, respectively, the curvatures of A_0 and $\widehat{A}$. Since we are going to encounter some singularities, we introduce a cutoff. Letting V_ϵ be a tube of radius ϵ centered on $\mathcal{S}$, we will evaluate $\int_{M-V_\epsilon} F \wedge F$ and consider the limit as $\epsilon \to 0$.

One term comes from $\int_{M-V_\epsilon} m^2 F_0 \wedge F_0$. This term is not very interesting, because it is independent of the integration variable $\widehat{A}$.[29] Another contribution that is not interesting for our present purposes is $\int_{M-V_\epsilon} \widehat{F} \wedge \widehat{F}$. Since $\widehat{A}$ is

[29] This contribution to $\boldsymbol{I}_\theta$ actually converges as $\epsilon \to 0$, because the singular part of F_0 is of rank two.

smooth, this has a limit as $\epsilon \to 0$, namely $\int_M \widehat{F} \wedge \widehat{F}$. The limiting functional is a topological invariant, and so is certainly invariant under the change in $\widehat{A}$ generated by the topological symmetry.

The interesting contribution is the cross term $\widetilde{\boldsymbol{I}}_\theta = (im\Psi/2\pi) \int_{M-V_\epsilon} F_0 \wedge \widehat{F}$. We want to determine if this expression is annihilated by the generator δ_T of the topological symmetry. We define the action of δ_T on A_0 and $\widehat{A}$ by $\delta_T A_0 = 0$ (since A_0 is a fixed gauge field) and $\delta_T \widehat{A} = \delta_T A$. We have $\delta_T \widehat{F} = d(\delta_T \widehat{A})$, so

$$\delta_T \widetilde{\boldsymbol{I}}_\theta = \frac{im\Psi}{2\pi} \int_{M-V_\epsilon} d\left(F_0 \wedge \delta_T \widehat{A}\right) = \frac{im\Psi}{2\pi} \int_{\partial V_\epsilon} F_0 \wedge \delta_T \widehat{A}. \tag{6.13}$$

To justify the integration by parts here, we must orient ∂V_ϵ by contracting the outward normal vector to $M - V_\epsilon$ with the orientation form of M. As the outward normal to $M - V_\epsilon$ is $-d/dr$, we see from (6.12) that ∂V_ϵ must be oriented by $ds \wedge d\mathrm{Vol}$.

Now the boundary of V_ϵ is a two-sphere bundle over the curve $\mathcal{S}$. The integral over the fibers of $V_\epsilon \to \mathcal{S}$ can be performed using the fact that the integral of F_0 over a fiber is $-2\pi i$. So we get

$$\delta_T \boldsymbol{I}_\theta = m\Psi \int_{\mathcal{S}} \delta_T \widehat{A}. \tag{6.14}$$

So $\boldsymbol{I}_\theta$ breaks the topological symmetry in the presence of the 't Hooft operator, or equivalently, the 't Hooft operator spoils the topological symmetry when $\theta \neq 0$. How can we restore the symmetry? Let us remember what one does with the action $\boldsymbol{I}$ in quantum mechanics: one integrates over all fields with a factor of $\exp(-\boldsymbol{I})$.[30] So it is useful to re-express (6.14) as a formula for the variation of $\exp(-\boldsymbol{I}_\theta)$:

$$\delta_T(\exp(-\boldsymbol{I}_\theta)) = -m\Psi \exp(-\boldsymbol{I}_\theta) \int_{\mathcal{S}} \delta_T \widehat{A}. \tag{6.15}$$

From here, it is a short step to see that

$$\delta_T \left(\exp\left(m\Psi \int_{\mathcal{S}} \widehat{A} \right) \exp\left(-\boldsymbol{I}_\theta\right) \right) = 0. \tag{6.16}$$

[30] In Lorentz signature, this factor would be $\exp(i\boldsymbol{I})$, and $\boldsymbol{I}_\theta$ would have an additional factor of i, leading to the same result.

So we can restore the topological symmetry if we include a Wilson operator $\exp(m\Psi\int_{\mathcal{S}}\widehat{A})$ as an additional factor in the path integral. (Notice that we have to use here a naive Wilson operator, not a supersymmetric one with a contribution from ϕ.)

However, the expression $\exp(-n\int_{\mathcal{S}}\widehat{A})$ is not gauge-invariant for generic real n. It is gauge-invariant if and only if n is an integer. So $n=-m\Psi$ must be an integer if we are to restore the topological symmetry. To put it differently, Ψ must be a rational number $-n/m$. For rational $\Psi=-n/m$, we have learned that the topological symmetry is preserved by the product of a charge n Wilson operator $\exp(-n\int_S\widehat{A})$ and a charge m 't Hooft operator. The product is called a Wilson-'t Hooft operator. We will denote it as $WT_{n,m}(\mathcal{S})$.

For rational Ψ, we write $\Psi=-n/m$ with coprime integers n and m. Under a duality transformation $\Psi\to(a\Psi+b)/(c\Psi+d)$, n and m transform by

$$(m\quad n)\longrightarrow(m\quad n)\begin{pmatrix}d&-b\\-c&a\end{pmatrix}. \tag{6.17}$$

As in eqn. (2.23), this is the standard transformation of the magnetic and electric charges $\mathfrak{m}$ and $\mathfrak{n}$ under duality. The reason for this is that the state obtained by quantizing the abelian gauge theory in the presence of a static Wilson-'t Hooft operator $WT_{n,m}$ (and no other charges or sources) has $(\mathfrak{m}\ \mathfrak{n})=(m\ n)$. For the magnetic charges, this is fairly clear from the fact that m and $\mathfrak{m}$ have both been defined in terms of the first Chern class (measured either at infinity or near the singularity). For the electric charges, it can be demonstrated by solving Maxwell's equations to determine the asymptotic behavior of the electric field in the presence of the Wilson operator. The computation we have just performed shows that $n=-\Psi m$ transforms under $\Psi\to\Psi+1$ as claimed in (6.17).

Combined Wilson-'t Hooft Operators: Nonabelian Case Just as for 't Hooft operators, the generalization of Wilson-'t Hooft operators to a general simple gauge group G is made by embedding the abelian construction in G, using a homomorphism $\rho:U(1)\to G$. This, however, requires some explanation.

We write $A=\rho(A_0)+\widehat{A}$, where A_0 is the singular $U(1)$ gauge field with the Dirac singularity along $\mathcal{S}$, and $\widehat{A}$ is smooth. (For $G=U(1)$, the homomorphism ρ is at the Lie algebra level multiplication by m for some integer m, so we wrote $A=mA_0+\widehat{A}$.) Now, however, we must clarify what

it means for $\widehat{A}$ to be smooth. Because of the nonlinear term in the curvature $F = dA + A \wedge A$, we must require

$$[\rho(A_0), \widehat{A}] = 0 \tag{6.18}$$

along $\mathcal{S}$, or else the singular part of the curvature along $\mathcal{S}$ will depend on $\widehat{A}$ and will not coincide with the Dirac singularity.

For the condition (6.18) to make sense, we must also restrict the gauge transformations along $\mathcal{S}$ to the subgroup $H \subset G$ that commutes with $\rho(U(1))$. So, when restricted to $\mathcal{S}$, $\widehat{A}$ has structure group H.

The Lie algebra of H has an H-invariant projection π_ρ to $\rho(\mathfrak{u}(1))$, the Lie algebra of $\rho(U(1))$. Along $\mathcal{S}$, we can define the projection $\pi_\rho(\widehat{A})$, a $\rho(\mathfrak{u}(1))$-valued gauge field.

The same calculation as in the abelian case shows that the 't Hooft operator that corresponds to ρ is no longer invariant under the topological symmetry if $\Psi \neq 0$. Instead,

$$\delta_T \boldsymbol{I}_\theta = \Psi \int_{\mathcal{S}} \mathrm{Tr}\, \rho(1)\, \delta_T \widehat{A}. \tag{6.19}$$

Here $\rho(1)$ is simply the image of $1 \in \mathfrak{u}(1)$ under the Lie algebra homomorphism $\rho : \mathfrak{u}(1) \to \mathfrak{g}$. The expression $\mathrm{Tr}\, \rho(1)\, \delta_T \widehat{A}$ is a multiple of $\pi_\rho(\delta_T \widehat{A})$.

Just as in the abelian case, topological invariance can be restored if we multiply the 't Hooft operator by

$$\exp\left(\Psi \int_{\mathcal{S}} \mathrm{Tr}\, \rho(1) \widehat{A}\right). \tag{6.20}$$

For this to be gauge-invariant places a condition on Ψ, which informally is that ρ must be divisible by the denominator of Ψ. For any rational Ψ, this condition is obeyed for suitable ρ.

The operators obtained this way are called Wilson-'t Hooft operators. The underlying supersymmetric gauge theory has more general Wilson-'t Hooft operators with linearly independent electric and magnetic weights [90]. The Wilson-'t Hooft operators in the topological field theory all arise by duality from Wilson operators at $\Psi = \infty$, so their electric and magnetic weights are proportional.

See also [92] for a recent related discussion of Wilson-'t Hooft operators.

6.3. Compactification To Two Dimensions

We now want to consider the interpretation of loop or line operators after compactification to two dimensions. We set $M = \Sigma \times C$, with Σ and C being Riemann surfaces. Thinking of Σ as being much bigger than C, we want to ask what loop operators look like in an effective two-dimensional theory on Σ. To begin with, we assume that Σ has no boundary.

Several cases can be distinguished. One case (Figure 4(a)) is a loop of the form $\mathcal{S} = p \times \mathcal{S}'$, where p is a point in Σ and $\mathcal{S}'$ is a loop in C. Clearly, an operator supported on a loop of this kind simply looks like a pointlike local operator in the effective theory on Σ. All ordinary quantum field theories and many TQFT's have such local operators, so in this example, the loop operator, after compactification, turns into something fairly ordinary in the effective theory on Σ.

The opposite case (fig. 4(b)) is $\mathcal{S} = \mathcal{S}'' \times q$, with $\mathcal{S}''$ a loop in Σ and q a point in C. Here the loop operator remains as a loop operator in the effective two-dimensional theory.

The general case, of course, is a curve $\mathcal{S}$ that propagates non-trivially in both Σ and C. Any such curve is homotopic to a curve that propagates

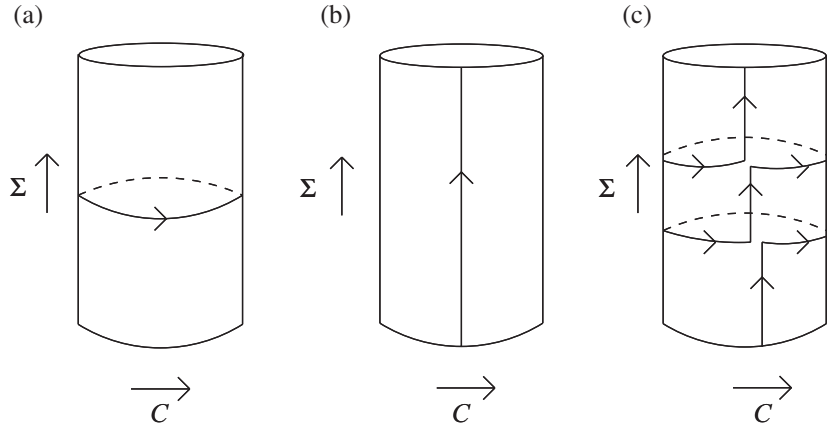

Figure 4: A schematic depiction of the four-manifold $\Sigma \times C$, with Σ running vertically and C horizontally. (a) A Wilson line that propagates in the C direction. (b) A Wilson line that propagates in the Σ direction. (c) A Wilson line that propagates in Σ, except at two moments at which it loops around one-cycles in C.

first on Σ, then on C, then on Σ, and so on, as indicated in Figure 4(c). Let $\overline{\mathcal{S}}$ be the projection of $\mathcal{S}$ to Σ. A loop operator on $\mathcal{S}$ reduces in the effective theory on Σ to a loop operator on $\overline{\mathcal{S}}$ with local operators inserted at distinguished points on $\overline{\mathcal{S}}$, namely the points where $\mathcal{S}$ propagates around C. We will momentarily describe a different interpretation of such a loop operator.

The essential point is clearly to understand the meaning of a loop or line operator in two dimensions. Here we should note that a two-manifold is locally divided by a one-manifold into two disjoint regions. Hence, a loop or line operator might produce a long-range effect; the couplings in the two-dimensional effective theory might be different on the two sides. (How Wilson and 't Hooft line operators can have such an effect is discussed in section 8.1.) This possibility is incorporated in Figure 5(a), where we sketch a two-dimensional line operator that divides the plane into two regions labeled by two distinct theories, Y and Z.

To put this situation in a more familiar framework, we can use a "folding" trick (see for example [93-95]). At the cost of reversing the orientation of region Z to get what we will call theory $\overline{Z}$, we can "fold" Figure 5(a) to get a similar figure in which regions Y and Z are on the same side and end at the location of the line operator (Figure 5(b)).

What we have now from an abstract point of view is a boundary condition in the tensor product theory $Y\otimes\overline{Z}$. For purposes of this paper, by a "brane," in general, we simply mean a local boundary condition in a quantum field theory. (In other words, our branes are all D-branes.) We will refer to branes

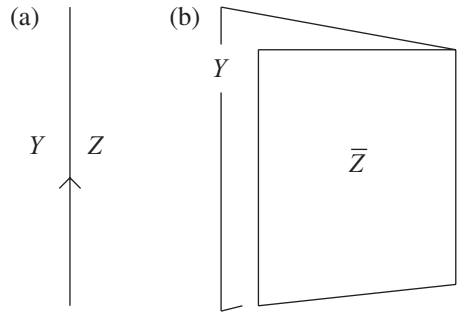

Figure 5: (a) A line operator dividing the plane into two regions, labeled by theories Y and Z. (b) A folded version of the same picture, interpreted in terms of theory $Y\times\overline{Z}$ on a half-plane; the boundary is labeled by a (Y,Z)-brane.

of the product theory $Y \otimes \overline{Z}$ as (Y, Z)-branes. So the line operator reduces in two dimensions to a (Y, Z)-brane.

Now we can also understand Figure 4(c) a little better. We previously interpreted this configuration in terms of a line operator with local operators inserted on it. After folding, the effective two-dimensional theory is formulated on a Riemann surface Σ with boundary and with local operators inserted on the boundary. Their existence is characteristic of brane physics; in general, for any brane, there is a certain space of local operators that can be inserted on a boundary component of Σ that is labeled by that brane.

Let us recall how to characterize such local operators. This will also help us recall a few basic facts about branes. For every two-dimensional quantum field theory X and pair of branes $\mathcal{B}_1$ and $\mathcal{B}_2$, one defines a vector space $\mathcal{H}_{\mathcal{B}_1,\mathcal{B}_2}$ of $(\mathcal{B}_1, \mathcal{B}_2)$ strings. In unitary quantum field theory, these spaces are Hilbert spaces; even without unitarity, $\mathcal{H}_{\mathcal{B}_1,\mathcal{B}_2}$ is dual to $\mathcal{H}_{\mathcal{B}_2,\mathcal{B}_1}$. One defines $\mathcal{H}_{\mathcal{B}_1,\mathcal{B}_2}$ by quantizing the theory X on $\Sigma = \mathbb{R} \times I$, with I a unit interval whose ends are labeled respectively by $\mathcal{B}_1$ and $\mathcal{B}_2$ (Figure 6(a)). $\mathcal{H}_{\mathcal{B}_1,\mathcal{B}_2}$ is also called the space of physical states of the theory with the given boundary conditions at the two ends of I.

Now in topological field theory, there are for any branes $\mathcal{B}_1, \mathcal{B}_2$, and $\mathcal{B}_3$ natural maps $\mathcal{H}_{\mathcal{B}_1,\mathcal{B}_2} \otimes \mathcal{H}_{\mathcal{B}_2,\mathcal{B}_3} \to \mathcal{H}_{\mathcal{B}_1,\mathcal{B}_3}$ defined by joining Riemann surfaces together (Figure 6(b)). (In two-dimensional quantum field theory without topological invariance, one must take into account the metrics or conformal structures of the surfaces, as a result of which an analogous discussion leads to the operator product expansion.) These maps obey the obvious associativity relation, which says that when three strings are joined (Figure 6(c)), one does not have to say which two joined first. In particular, setting all the branes $\mathcal{B}_i$ equal to $\mathcal{B}$, we find that $\mathcal{H}_{\mathcal{B},\mathcal{B}}$ always has the natural structure of an associative algebra.

In fact, in TQFT, the space of local operators $\mathcal{O}$ that can be inserted on a boundary labeled by a brane $\mathcal{B}$ is the same as $\mathcal{H}_{\mathcal{B},\mathcal{B}}$, since (Figure 6(d)) a "cutting" operation applied near such an $\mathcal{O}$ reveals an element of $\mathcal{H}_{\mathcal{B},\mathcal{B}}$. By similar reasoning (Figure 6(e)), we find that, for any $\mathcal{B}$ and $\widetilde{\mathcal{B}}$, $\mathcal{H}_{\mathcal{B},\widetilde{\mathcal{B}}}$ always has a natural structure of left $\mathcal{H}_{\mathcal{B},\mathcal{B}}$-module. Similarly, it has a natural structure of right $\mathcal{H}_{\widetilde{\mathcal{B}},\widetilde{\mathcal{B}}}$ module. Indeed, $\mathcal{H}_{\mathcal{B},\widetilde{\mathcal{B}}}$ is a $(\mathcal{H}_{\mathcal{B},\mathcal{B}}, \mathcal{H}_{\widetilde{\mathcal{B}},\widetilde{\mathcal{B}}})$ bimodule. This means simply that one can act with $\mathcal{H}_{\mathcal{B},\mathcal{B}}$ by attaching a string on the left, or by $\mathcal{H}_{\widetilde{\mathcal{B}},\widetilde{\mathcal{B}}}$ by attaching a string on the right, and moreover these two actions commute.

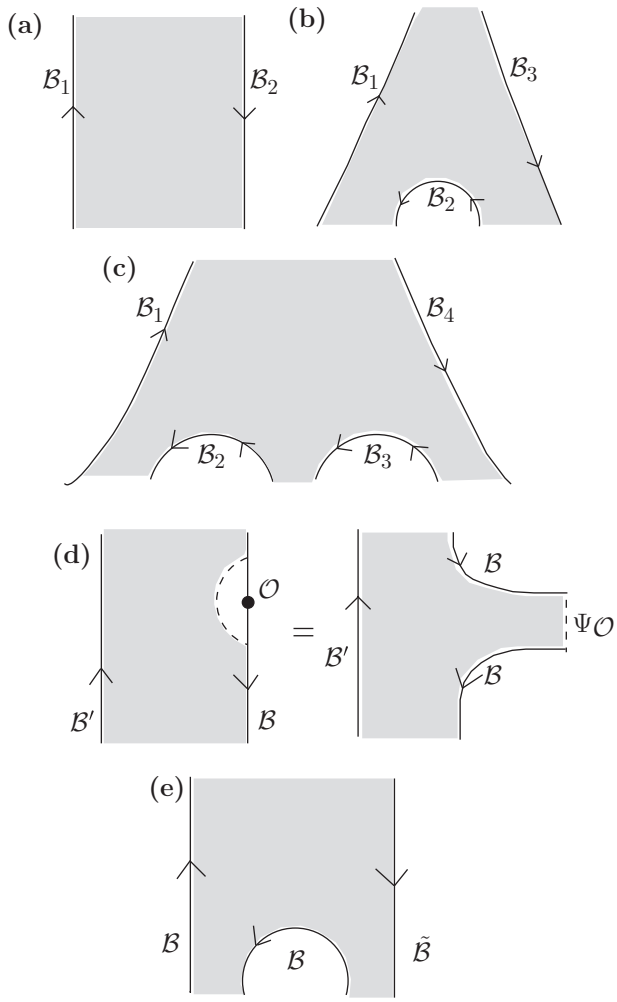

Figure 6: (a) The space of $(\mathcal{B}_1, \mathcal{B}_2)$ strings. (b) Joining of strings to make a product $\mathcal{H}_{\mathcal{B}_1,\mathcal{B}_2} \otimes \mathcal{H}_{\mathcal{B}_2,\mathcal{B}_3} \to \mathcal{H}_{\mathcal{B}_1,\mathcal{B}_3}$. (c) The associativity relation that comes by joining three strings. (d) The space of local operators $\mathcal{O}$ that can be inserted on a boundary labeled by a brane $\mathcal{B}$ is the same as the space $\mathcal{H}_{\mathcal{B},\mathcal{B}}$. To see this, we perform the path integral in a small region around the insertion point of $\mathcal{O}$ (the unshaded region on the left) to get a physical state $\Psi_{\mathcal{O}}$ that can be inserted on the dotted line to reproduce the effects of $\mathcal{O}$. The resulting picture can be put in the form shown on the right. (e) For every $\mathcal{B}$ and $\widetilde{\mathcal{B}}$, the space $\mathcal{H}_{\mathcal{B},\widetilde{\mathcal{B}}}$ is a module for the algebra $\mathcal{H}_{\mathcal{B},\mathcal{B}}$.

6.4. Line Operator Near A Boundary

The geometric Langlands program revolves around the case that Σ has a boundary and the line operator runs parallel to the boundary.

Defining a quantum field theory on the Riemann surface Σ with boundary requires a choice of boundary condition – that is a choice of a brane, which we will call $\mathcal{B}$. In Figure 7, we sketch a Riemann surface Σ with a boundary and a line L parallel to the boundary. The boundary has been labeled by a brane $\mathcal{B}$, which defines the boundary conditions at that boundary. A line operator $\mathcal{X}$ is supported on L. $\mathcal{X}$ may be a Wilson, 't Hooft, or mixed Wilson-'t Hooft operator.

From a macroscopic point of view, $\mathcal{X}$ and $\mathcal{B}$ simply combine together into an effective boundary condition $\mathcal{B}'$. One can characterize $\mathcal{B}'$ as the effective boundary condition that one sees at a big distance due to the combined effects of $\mathcal{B}$ and $\mathcal{X}$. Or one can simply take the limit that L approaches the boundary and let $\mathcal{B}'$ be the limiting boundary condition that results.

Either way, we write $\mathcal{B}' = \mathcal{X}\mathcal{B}$ for the new brane that is produced by acting with $\mathcal{X}$ on the old brane. In this way, we regard line operators as acting on boundary conditions to produce new boundary conditions.

This is a more abstract form of action than the usual action of an operator on a vector space. As we saw above, to get a vector space $\mathcal{H}_{\mathcal{B},\widetilde{\mathcal{B}}}$ associated with $\mathcal{B}$ on which an operator might act, we need a pair of branes $\mathcal{B}$ and $\widetilde{\mathcal{B}}$. Of course, we can just take $\widetilde{\mathcal{B}} = \mathcal{B}$. But even once we construct a space $\mathcal{H}_{\mathcal{B},\widetilde{\mathcal{B}}}$ of physical states, it is really a *loop* operator, localized in time, that acts on this space, by quantizing the picture of Figure 8. An open line operator that runs in the time direction, as in Figure 7, is instead part of the data that must be incorporated in quantization to get the right space of physical states.

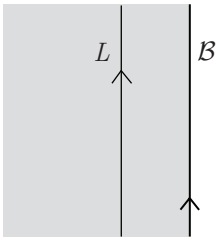

Figure 7: A Riemann surface with a boundary, labeled by a brane $\mathcal{B}$, and a line operator L parallel to the boundary.

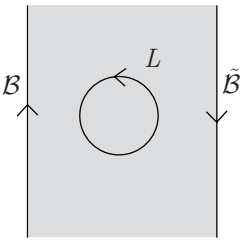

Figure 8: Action of a loop operator as an ordinary operator on the space $\mathcal{H}_{\mathcal{B},\widetilde{\mathcal{B}}}$.

If we *do* have a vector space $\mathcal{H}$ and an operator $\mathcal{O}\colon \mathcal{H} \to \mathcal{H}$, we can look for a vector $\psi \in \mathcal{H}$ that is an eigenvector of $\mathcal{O}$:

$$\mathcal{O}\psi = \lambda\psi. \tag{6.21}$$

Here λ is an ordinary complex number.

In our problem, instead of an operator that acts on vectors in a Hilbert space, we have a more exotic operator that acts on theories, via $\mathcal{B} \to \mathcal{B}' = \mathcal{X}\mathcal{B}$. An ordinary complex eigenvalue does not make sense in this situation. Rather, what plays the role of an "eigenvalue" is a vector space V.

For every brane $\mathcal{B}$ and vector space V, one defines a new brane $\mathcal{B}\otimes V$. An abstract way to characterize $\mathcal{B}\otimes V$ is to say that for any brane $\widetilde{\mathcal{B}}$, the Hilbert space $\mathcal{H}_{\widetilde{\mathcal{B}},\mathcal{B}\otimes V}$ is equal to $\mathcal{H}_{\widetilde{\mathcal{B}},\mathcal{B}}\otimes V$.

We can be more concrete in the case of branes that are constructed from geometry. In the case that theory X is a two-dimensional sigma-model with a target space $\widehat{X}$, a brane can be constructed from a vector bundle $\mathcal{W} \to \widehat{X}$ (or from a more general sheaf, possibly suported on a submanifold). $\mathcal{W}$ is known as the Chan-Paton bundle. Tensoring such a brane with a fixed vector space V just means replacing $\mathcal{W}$ by $\mathcal{W}\otimes V$. This operation generalizes without any problem when $\mathcal{W}$ is replaced by a more general sheaf.

If we identify V as $\mathbb{C}^k$ for some k, then tensoring the Chan-Paton bundle with V amounts to taking k copies of the original brane. So for $\mathcal{B}$ to be an eigenbrane for $\mathcal{X}$ means that $\mathcal{X}\mathcal{B}$ is isomorphic to the sum of k copies of $\mathcal{B}$. However, it is much better to think of the "eigenvalue" as a vector space V, not just its integer dimension k, since V may in fact vary with some additional parameters as the fiber of a vector bundle over some parameter space.

Having established the notion of tensoring a brane by a vector space, we now can define the notion of eigenbrane. A brane $\mathcal{B}$ is an eigenbrane for a

line operator $\mathcal{X}$ if

$$\mathcal{X}\mathcal{B} = \mathcal{B} \otimes V, \tag{6.22}$$

for some vector space V. This concept corresponds to the concept of an eigensheaf as defined by Beilinson and Drinfeld in the geometric Langlands program [5].

If $\mathcal{X}_1$ and $\mathcal{X}_2$ are two line operators, supported on parallel lines L_1 and L_2, we say that they commute if for all $\mathcal{B}$ we have

$$\mathcal{X}_1\mathcal{X}_2\mathcal{B} = \mathcal{X}_2\mathcal{X}_1\mathcal{B}. \tag{6.23}$$

This statement can be illustrated in a convenient picture: it means that the lines L_1 and L_2 can be passed through each other without any discontinuity (Figure 9). Just as for ordinary operators on a Hilbert space, if two line operators commute in this sense, it is possible for them to have a simultaneous eigenbrane, that is, a brane $\mathcal{B}$ obeying:

$$\begin{aligned} \mathcal{X}_1\mathcal{B} &= \mathcal{B} \otimes V_1 \\ \mathcal{X}_2\mathcal{B} &= \mathcal{B} \otimes V_2. \end{aligned} \tag{6.24}$$

More generally, any collection of commuting line operators can have a simultaneous eigenbrane.

Line Operators That Descend From Four Dimensions Now in our problem, we are interested in a two-dimensional TQFT that arises by

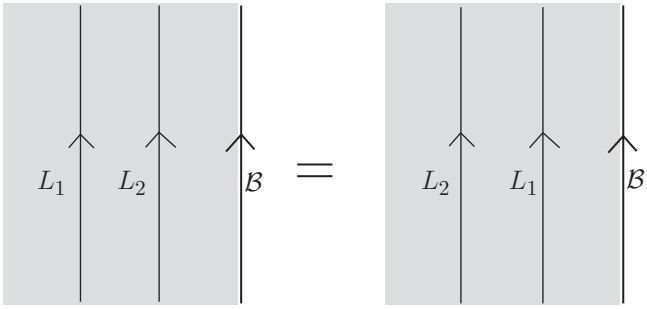

Figure 9: Line operators L_1 and L_2 are said to commute if they can be passed through each other without singularity. In topological field theory, this simply means that the two figures pictured here can be considered equivalent.

reducing a four-dimensional TQFT on the Riemann surface C. Moreover, our line operators all descend from line operators in four dimensions. This is enough to ensure some special behavior.

First of all, each line operator $\mathcal{X}$ can be defined for any point $p \in C$. To make this explicit, we write the line operator as $\mathcal{X}_p$. If a brane $\mathcal{B}$ that respects the topological symmetry is an eigenbrane of $\mathcal{X}_p$ for one p, it must be an eigenbrane for every p, since p is irrelevant anyway. So

$$\mathcal{X}_p\mathcal{B} = \mathcal{B} \otimes V_p \tag{6.25}$$

for some V_p. Here, locally, the vector space V_p must be independent of p, since p is irrelevant. But topological invariance allows the possibility that V_p might have global monodromies as p varies on the Riemann surface C. So the V_p are fibers of a flat vector bundle over C.

To be more explicit about this, moving p is equivalent to a special case of changing the metric on C, and this changes the underlying action by a BRST-trivial term $\{Q, \cdot\}$. This fact enables one to find a flat connection on the family of vector spaces V_p for $p \in C$ and thus to prove locally that V_p is independent of p. But this flat connection may have nontrivial global monodromies.

Suppose that we are given several loop operators $\mathcal{X}_i$ that all preserve the same topological symmetry. For example, at $\Psi = \infty$, they may be Wilson operators associated with representations R_i of G; at $\Psi = 0$, they may be 't Hooft operators associated with representations ${}^LR_i$ of ${}^LG$. In any event, the $\mathcal{X}_i$ all depend on points $p_i \in C$. As we are free to take the points p_i to be distinct, we can perform the interchange of Figure 9, which looks potentially singular from a two-dimensional point of view, without meeting any singularity. (The line operators do not meet in four dimensions, so there is no possibility of a singularity. Moreover, as C is two-dimensional, there is enough room in moving the line operators that we do not encounter monodromies that are local along C.[31]) Hence, topological line operators with a four-dimensional origin automatically commute.

[31] If L and L' are two line operators supported at the same point $p \in C$, to move them past each other in $\Sigma \times C$ without a singularity, we separate them to two distinct points $p, p' \in C$, with p' near p. This can be done in a small open ball $U \subset C$. For $\Sigma = \mathbb{R} \times I$, static line operators are supported at points in the three-manifold $W = I \times C$ and the operation of moving L past L' can be carried out in the subspace $W' = I \times U$. The space of pairs of distinct points in W' is simply-connected, so there are no monodromies in this local model and no ambiguity about how to move L past L'. This argument would not work if C were one-dimensional.

Electric And Magnetic Eigenbranes As a special case, the topological Wilson operators at $\Psi = \infty$ commute. It is therefore possible for them to have simultaneous eigenbranes. A joint eigenbrane of the Wilson line operators will be called an electric eigenbrane.

Similarly, the topological 't Hooft operators at $\Psi = 0$ commute and can likewise have simultaneous eigenbranes. A joint eigenbrane of the 't Hooft operators will be called a magnetic eigenbrane.

The S-transformation $S : \Psi \to -1/n_{\mathfrak{g}}\Psi$ maps topological Wilson operators to topological 't Hooft operators and will hence map electric eigenbranes to magnetic eigenbranes.

Like eigenbranes of any family of line operators that descend from four dimensions, electric and magnetic eigenbranes are automatically associated with flat vector bundles over C. This is an important statement in the geometric Langlands program. As we have seen, it reflects the existence of an underlying TQFT above two dimensions.

Algebras Of Commuting Line Operators The fact that the line operators that descend from four dimensions commute raises another question. What is the commutative algebra that they generate?

In the case of Wilson line operators, it is simply the tensor algebra of representations of G. In other words, let R and $\widetilde{R}$ be two representations of G, and suppose that the decomposition of $R \otimes \widetilde{R}$ in irreducible representations is

$$R \otimes \widetilde{R} = \oplus_\alpha n_\alpha R_\alpha. \tag{6.26}$$

Here we take the R_α to be *distinct* irreducible representations; the integer n_α is the multiplicity with which R_α appears in the decomposition of $R \otimes \widetilde{R}$. (The relation (6.26) and others below can be more precisely stated in terms of vector spaces N_α of dimension n_α, but we postpone this refinement to section 10.4.)

We consider parallel lines L and L' and consider the limit as L approaches L'. We claim that

$$\lim_{L \to L'} W(R, L) \cdot W(\widetilde{R}, L') = \sum_\alpha n_\alpha W(R_\alpha, L'). \tag{6.27}$$

In fact, because of supersymmetry, the limit can be evaluated classically. Had we considered ordinary Wilson operators rather than supersymmetric ones, there would be various quantum effects in the limit $L \to L'$. In the supersymmetric case, we can just set $L = L'$ and calculate classically.

Given this, the calculation is clear sailing. In the case of a closed loop, $W(R)$ and $W(\widetilde{R})$ are holonomies in representations R and $\widetilde{R}$. Their product is a holonomy in the representation $R\otimes\widetilde{R}$, and in view of (6.26), this can be expanded as a sum of holonomies in the representations R_α, leading to (6.27). In the case of a line operator, $W(R)$ or $W(\widetilde{R})$ acts by parallel transport on initial states in the R or $\widetilde{R}$ representation, whose tensor product has a direct sum decomposition as in (6.27).

More informally, the Wilson operators $W(R)$ and $W(\widetilde{R})$ describe external charges in representations R and $\widetilde{R}$. Their product describes a pair of charges in the representation $R\otimes\widetilde{R}$, whose decomposition in irreducibles gives the OPE of supersymmetric Wilson operators.

Dually, we expect that if ${}^LR$ and ${}^L\widetilde{R}$ are representations of ${}^LG$, with

$$ {}^LR\otimes{}^L\widetilde{R}=\oplus_\alpha n_\alpha\,{}^LR_\alpha, \tag{6.28}$$

then

$$ \lim_{L\to L'}T({}^LR,L)T({}^L\widetilde{R},L')=\sum_\alpha n_\alpha T({}^LR_\alpha,L'). \tag{6.29}$$

For some groups and representations, for example in the cases studied in [86,87], duality between gauge theory and strings makes this result clear. But we do not know of any attempts in the physics literature to compute the operator product expansion of 't Hooft operators directly in gauge theory. In section 9, we will establish a relation between 't Hooft operators and Hecke transformations, and then we will see in section 10.4 how the statement (6.29) is related to what is known in the mathematical literature [96-98].

We can get some information about the coefficients in (6.29) by studying the operator product expansion of 't Hooft operators on the Coulomb branch, where the gauge group is broken down to an abelian subgroup. Given two 't Hooft operators $T({}^LR,L)$ and $T({}^LR',L')$, let ${}^Lw$ and ${}^Lw'$ be the highest weights of the two representations ${}^LR$ and ${}^LR'$. On the Coulomb branch, these 't Hooft operators act via singular monopole solutions with magnetic charges that are weights ${}^L\widetilde{w}$ and ${}^L\widetilde{w}'$ of ${}^LG$; here ${}^L\widetilde{w}$ and ${}^L\widetilde{w}'$ can be any Weyl transforms of ${}^Lw$ and ${}^Lw'$. The product of the two 't Hooft operators is represented by a singular abelian monopole with magnetic charge ${}^L\widetilde{w}+{}^L\widetilde{w}'$. In order for a representation ${}^LR_i$ to appear in the operator product expansion (6.29), its highest weight ${}^Lw_i$ must equal ${}^L\widetilde{w}+{}^L\widetilde{w}'$, for some choice of ${}^L\widetilde{w}$ and ${}^L\widetilde{w}'$. Whichever choice we make for ${}^L\widetilde{w}$ and ${}^L\widetilde{w}'$, their sum ${}^L\widehat{w}={}^L\widetilde{w}+{}^L\widetilde{w}'$ is dominated by ${}^L\overline{w}={}^Lw+{}^Lw'$ (meaning that ${}^L\overline{w}-{}^L\widehat{w}$ is a

dominant weight, possibly zero). One possibility for ${}^L R_i$ is that it may be the representation ${}^L\overline{R}$ whose highest weight is exactly ${}^L\overline{w}$. This representation appears in the expansion (6.29) with a multiplicity that is precisely $n = 1$, since there is precisely one choice of ${}^L\widetilde{w}$ and ${}^L\widetilde{w}'$ that add up to ${}^L\overline{w}$. The other representations that can appear are the representations whose highest weight is dominated by ${}^L\overline{w}$. We will call these the associated representations of ${}^L\overline{R}$. There is no straightforward way to deduce the coefficient in the operator product expansion for these subdominant or associated representations. Ultimately, we see in section 10.4 that they arise from singularities of the space of Hecke modifications, accounting for why it is subtle to determine them.

Functors Acting On The Category Of Branes What we have said so far does not fully capture the formal properties of line operators in two dimensions.

We have explained so far that a line operator $\mathcal{X}$ that respects the appropriate topological symmetry has a natural action on branes. If $\mathcal{B}$ and $\mathcal{B}'$ are any two branes, we can act on them with $\mathcal{X}$ to get new branes $\mathcal{X}\mathcal{B}$ and $\mathcal{X}\mathcal{B}'$. Associated with the pair of branes $\mathcal{B}$ and $\mathcal{B}'$ is a space of physical states $\mathcal{H}_{\mathcal{B},\mathcal{B}'}$ obtained by quantizing the $(\mathcal{B},\mathcal{B}')$ open strings. Likewise, associated to the pair of branes $\mathcal{X}\mathcal{B}$ and $\mathcal{X}\mathcal{B}'$ is a space $\mathcal{H}_{\mathcal{X}\mathcal{B},\mathcal{X}\mathcal{B}'}$ of physical $(\mathcal{X}\mathcal{B},\mathcal{X}\mathcal{B}')$ strings.

Associated with the line operator $\mathcal{X}$ is a natural map $\widehat{\mathcal{X}} : \mathcal{H}_{\mathcal{B},\mathcal{B}'} \to \mathcal{H}_{\mathcal{X}\mathcal{B},\mathcal{X}\mathcal{B}'}$. The definition of $\widehat{\mathcal{X}}$ should be clear from Figure 10(a). In this picture, from a microscopic point of view, the operator $\mathcal{O}$ determines a state in $\mathcal{H}_{\mathcal{B},\mathcal{B}'}$, but from a macroscopic point of view, as L approaches the boundary, what we see is a state in $\mathcal{H}_{\mathcal{X}\mathcal{B},\mathcal{X}\mathcal{B}'}$.

Precisely if $\mathcal{O}$ is a multiple of the identity operator, there is no distinguished point on the boundary of Σ at which $\mathcal{O}$ is inserted, and this is true before or after the line operator $\mathcal{X}$ approaches the boundary. So $\widehat{\mathcal{X}}$ maps the identity in $\mathcal{H}_{\mathcal{B},\mathcal{B}}$ (that is, the open string state that corresponds to the identity operator $\mathcal{O} = 1$) to the identity in $\mathcal{H}_{\mathcal{X}\mathcal{B},\mathcal{X}\mathcal{B}}$. It also preserves the multiplication of open strings. Indeed, given three branes $\mathcal{B}$, $\mathcal{B}'$, and $\mathcal{B}''$, the associative multiplication $\mathcal{H}_{\mathcal{B},\mathcal{B}'} \otimes \mathcal{H}_{\mathcal{B}',\mathcal{B}''} \to \mathcal{H}_{\mathcal{B},\mathcal{B}''}$ is defined by joining open strings as in Figure 6(b), or equivalently by an operator product expansion of operators inserted on the boundary, as in Figure 10(b). This obviously commutes with bringing a line operator up to the boundary. The fancy way to summarize these statements is to say that the line operator $\mathcal{X}$ determines a functor mapping the category of branes to itself.

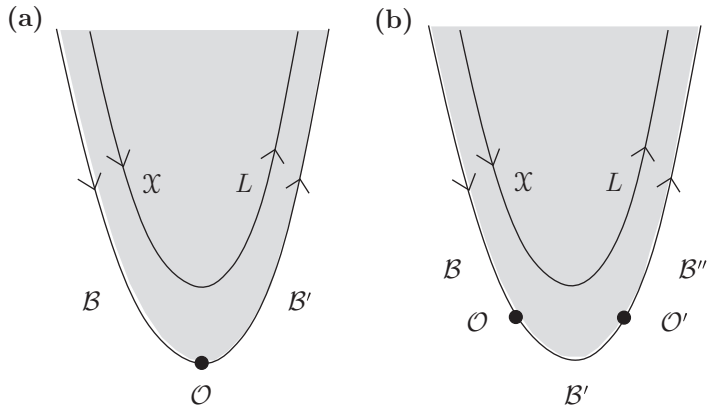

Figure 10: (a) A line operator $\mathcal{X}$ supported on a line L maps branes $\mathcal{B}$ and $\mathcal{B}'$ to branes $\mathcal{X}\mathcal{B}$ and $\mathcal{X}\mathcal{B}'$, and likewise maps $(\mathcal{B},\mathcal{B}')$ strings to $(\mathcal{X}\mathcal{B},\mathcal{X}\mathcal{B}')$ strings. In the picture, the operator $\mathcal{O}$ determines a $(\mathcal{B},\mathcal{B}')$ string, which is mapped by the line operator to an $(\mathcal{X}\mathcal{B},\mathcal{X}\mathcal{B}')$ string. (b) These maps preserve the associativity of open string multiplication.

Line operators as symmetries of the category of branes have been discussed independently and from a different point of view in [99]. Their relation to duality transformations will be discussed elsewhere [26]. For an example of the use of Wilson-like operators to describe boundary perturbations in two-dimensional conformal field theory, see section 5.5 of [100].

Line Operators and Enhanced Supersymmetry Special Wilson or 't Hooft operators can have more supersymmetry than what we have claimed so far.

Let us go back to the definition (6.4) of the supersymmetric Wilson operator. It is a function of $\mathcal{A}_\mu = A_\mu + i\phi_\mu$, so it is invariant under any supersymmetry that leaves $A_\mu + i\phi_\mu$ invariant. According to the microscopic formula (2.5) for the transformation of the fields under supersymmetry, the variation of $A_\mu + i\phi_\mu$ under a supersymmetry generated by an infinitesimal parameter ϵ is proportional to $(\Gamma_\mu + i\Gamma_{\mu+4})\epsilon$. The condition for ϵ to generate a supersymmetry of an arbitrary Wilson line operator is that this vanishes for $\mu = 0,\dots,3$:

$$\Gamma_\mu\Gamma_{\mu+4}\epsilon = i\epsilon,\ \ \mu = 0,\dots,3. \tag{6.30}$$

This condition leaves only one unbroken supersymmetry, which is the one we have exploited so far.

Suppose, however, that $M = \Sigma \times C$ with a product metric. In this case, the twisting preserves the four supersymmetries that were characterized in (5.16). Consider also a Wilson operator defined on a one-manifold $\mathcal{S}$ of the form $\gamma \times p$ where γ is a curve in Σ and p is a point in C. For an operator of this type, which we will call a special Wilson operator, the conditions (6.30) collapse to

$$\Gamma_{04}\epsilon = \Gamma_{15}\epsilon = i\epsilon. \tag{6.31}$$

These imply the second condition in (5.16), and the combined set of conditions leaves two unbroken supersymmetries.

We can easily see what the extra supersymmetry means. Special Wilson operators do not depend on φ, so they are invariant under the $\mathcal{U}_1$ symmetry introduced in section 4.1, which acts on the Higgs field by $\varphi \to \lambda\varphi$, $|\lambda| = 1$. This symmetry rotates complex structure J to complex structure K, so it maps the B-model in complex structure J to the B-model in complex structure K. Hence the two supersymmetries are the topological supercharge of the B-models in complex structures J and K. The topological supercharge of the B-model in any of the complex structures I_w is a linear combination of these. So it follows that a special Wilson operator is what we might call a line operator of type (B, B, B), that is, it preserves the topological supersymmetry of the B-model in every complex structure of $\mathcal{M}_H$.

We can reason in the same way for special 't Hooft operators, that is 't Hooft operators supported on a curve $\mathcal{S}$ of the same type. Special 't Hooft operators are again independent of φ and so invariant under the group $\mathcal{U}_1$. This group rotates the A-model in complex structure K into the A-model in complex structure J. General 't Hooft operators preserve the A-type supersymmetry in complex structure K, so special ones do so also in complex structure J. The topological supercharge of the B-model in complex structure I is a linear combination of those for the A-models in complex structures J and K. So actually, the special 't Hooft operators are of type (B, A, A), that is, they preserve the indicated topological supersymmetries for complex structures I, J, and K.

This is completely in accord with expectations. The S-transformation maps Wilson operators to 't Hooft operators, and as one can see from Table 2 of section 5.3, maps supersymmetry of type (B, B, B) to supersymmetry of type (B, A, A).

One can further specialize this situation by taking the metric on Σ to be flat and $\mathcal{S}$ to be a straight line. In this case, the line operator describes a static external electric or magnetic charge. A Wilson or 't Hooft operator of this form actually preserves four supersymmetries. For example, for a Wilson operator supported on a straight line that runs in the time direction, the conditions for an unbroken supersymmetry reduce to $\Gamma_{04}\epsilon = i\epsilon$ and $\Gamma_{2367}\epsilon = \epsilon$. This system is physically sensible (if we interpret it with Lorentz signature); it is described by a real, positive-definite Hilbert space with a positive-definite, hermitian Hamiltonian H. The four unbroken supersymmetries are the two topological supercharges Q_i, $i = 1, 2$ described earlier, and their hermitian adjoints $Q_i^\dagger$. They obey a physical supersymmetry algebra $\{Q_i, Q_j^\dagger\} = 2\delta_{ij}H$, $\{Q_i, Q_j\} = 0 = \{Q_i^\dagger, Q_j^\dagger\}$, for $i, j = 1, 2$.

7. Fluxes and *S*-Duality

In this section, we will examine phenomena involving the center of the gauge group that are relevant to understanding S-duality of Hitchin moduli space and the geometric Langlands program.

We begin section 7.1 with a review of discrete electric and magnetic fluxes and the action of S-duality on them. More detail can be found in [45], section 3, as well as in [26]. For discrete electric and magnetic flux in gauge theory, the original reference is [101]. We also explain some mathematical notions such as the concept of the universal bundle. Then we develop our applications in section 7.2.

7.1. Review

In this section, G denotes a compact semi-simple Lie group; $\overline{G}$ is its universal cover; and $G_{\rm ad}$ is the corresponding adjoint group. The center of $\overline{G}$, which is also the fundamental group of $G_{\rm ad}$, is a finite abelian group $\mathcal{Z}$. For simple G, $\mathcal{Z}$ is a cyclic group in all cases except for $\overline{G} = Spin(4n)$, as is summarized in Table 3.

One happy fact that we notice from the table is that some of the subtleties with S-duality for non-simply-laced groups are irrelevant, since G_2 and F_4 have trivial centers.

An important property is that the group $\mathcal{Z}$ is naturally selfdual. The dual of an abelian group B is $B^\vee = \mathrm{Hom}(B, U(1))$. A finite abelian group is always isomorphic to its own dual, but not naturally. The center $\mathcal{Z}$ of $\overline{G}$

Table 3: Simple and simply-connected Lie groups with non-trivial centers. F_2, G_4, and E_8, which are omitted, have trivial center.

Group	Center
$SU(N)$	$\mathbb{Z}_N$
$Spin(2n+1)$	$\mathbb{Z}_2$
$Spin(4n+2)$	$\mathbb{Z}_4$
$Spin(4n)$	$\mathbb{Z}_2 \times \mathbb{Z}_2$
$Sp(n)$	$\mathbb{Z}_2$
E_6	$\mathbb{Z}_3$
E_7	$\mathbb{Z}_2$

always has, however, a natural selfduality. This duality is best expressed as a homomorphism

$$\Upsilon : \mathcal{Z} \times \mathcal{Z} \longrightarrow U(1), \tag{7.1}$$

which is symmetric, $\Upsilon(a,b) = \Upsilon(b,a)$, and such that any homomorphism of $\mathcal{Z}$ to $U(1)$ is $a \to \Upsilon(a,b)$ for a unique b. Such a Υ is called a "perfect pairing." Υ is constructed as follows. If G is simply-laced, then $\mathcal{Z} = \Lambda^\vee/\Lambda$, where Λ is the root lattice and $\Lambda^\vee$ is the weight lattice of $\overline{G}$. The pairing Υ is then defined as

$$\Upsilon(a,b) = \exp(2\pi i \langle a, b\rangle), \tag{7.2}$$

where $\langle\ ,\ \rangle$ is the usual quadratic form on $\Lambda^\vee$ (which is even and integral when restricted to Λ and takes rational values on $\Lambda^\vee$). If G is not simply-laced, its center is trivial or $\mathbb{Z}_2$, and so admits precisely one selfduality, which we call Υ.

A $G_{\rm ad}$-bundle E on M has a characteristic class $\xi(E)$ that takes values in $H^2(M,\mathcal{Z})$. For example, if $G_{\rm ad} = SO(3)$, $\xi(E)$ is the second Stieffel-Whitney class $w_2(E)$. One can define a partition function Z_ξ for every choice of ξ by restricting the path integral to bundles with given ξ. The Z_ξ transform in a unitary representation of the S-duality group [45,26], but this is something that we will not explain here.

Instead, we concentrate on the Hamiltonian description, and specialize to $M = S^1 \times W$ for a three-manifold W. In this case, we have $H^2(M,\mathcal{Z}) = H^1(W,\mathcal{Z}) \oplus H^2(W,\mathcal{Z})$, so $\xi(E)$ can be decomposed as $\xi = \mathbf{a} \oplus \mathbf{m}$, with $\mathbf{a} \in H^1(W,\mathcal{Z})$, $\mathbf{m} \in H^2(W,\mathcal{Z})$.

Consider the pairing

$$\widehat{\Upsilon} : H^2(M,\mathcal{Z}) \times H^2(M,\mathcal{Z}) \longrightarrow H^4(M,U(1)) = U(1) \tag{7.3}$$

obtained by composing the cup product $H^2(M,\mathcal{Z})\times H^2(M,\mathcal{Z}) \to H^4(M,\mathcal{Z}\times\mathcal{Z})$ with the map $\Upsilon : \mathcal{Z}\times\mathcal{Z}\to U(1)$. According to Poincaré duality, $\widehat{\Upsilon}$ is a perfect pairing, making $H^2(M,\mathcal{Z})$ a selfdual abelian group. More generally, for a closed oriented manifold Y of dimension n, Poincaré duality together with selfduality of $\mathcal{Z}$ gives perfect pairings

$$H^d(Y,\mathcal{Z}) \times H^{n-d}(Y,\mathcal{Z}) \longrightarrow U(1), \quad 0 \leq d \leq n. \tag{7.4}$$

In particular, we have the perfect pairing

$$\overline{\Upsilon} : H^1(W,\mathcal{Z}) \times H^2(W,\mathcal{Z}) \longrightarrow U(1). \tag{7.5}$$

So these are dual abelian groups, and in particular

$$H^2(W,\mathcal{Z}) = \mathrm{Hom}(H^1(W,\mathcal{Z}),U(1)). \tag{7.6}$$

Let us now recall how S-duality is implemented in a Hamiltonian framework. The partition function of the gauge theory with $\mathbf{m}$ specified has a natural Hilbert space interpretation, because $\mathbf{m}$ is part of the data at an initial time and is independent of time. However, because $\mathbf{a}$ cannot be expressed in terms of the data at a fixed time, the partition function with fixed $\mathbf{a}$ cannot be given a Hamiltonian interpretation. Instead one must introduce a character $\mathbf{e}$ of the finite abelian group $H^1(M,\mathcal{Z})$, that is, a homomorphism

$$\mathbf{e} : H^1(W,\mathcal{Z}) \longrightarrow U(1). \tag{7.7}$$

For each choice of $\mathbf{e}$ and $\mathbf{m}$, the sum $\sum_{\mathbf{a}} \mathbf{e}(\mathbf{a}) Z_{\mathbf{a},\mathbf{m}}$ can be interpreted in terms of a trace in a Hilbert space $\mathcal{H}_{\mathbf{e},\mathbf{m}}$. This sum has a Hilbert space interpretation because summing over $\mathbf{a}$ with the weight factor $\mathbf{e}(\mathbf{a})$ is compatible with cutting and pasting (or what in quantum field theory is usually called cluster decomposition) in the S^1 direction. $\mathbf{e}$ and $\mathbf{m}$ are called respectively the discrete electric and magnetic flux [101]. Because of (7.6), $\mathbf{e}$ can be alternatively viewed as an element of $H^2(W,\mathcal{Z})$, just like $\mathbf{m}$.

So we get a Hilbert space $\mathcal{H}_{\mathbf{m},\mathbf{e}}$ for each choice of

$$\begin{aligned} \mathbf{e} &\in \mathbf{E} = \mathrm{Hom}(H^1(W,\mathcal{Z}),U(1)) \\ \mathbf{m} &\in \mathbf{M} = H^2(W,\mathcal{Z}). \end{aligned} \tag{7.8}$$

Happily, the finite groups $\mathbf{E}$ and $\mathbf{M}$ are isomorphic according to (7.6). This makes it possible to have S-duality symmetry. The duality transformation S exchanges $\mathbf{E}$ and $\mathbf{M}$ and maps $(\mathbf{e},\mathbf{m}) \to (-\mathbf{m},\mathbf{e})$:

$$S : \mathcal{H}_{\mathbf{e},\mathbf{m}} \longrightarrow \mathcal{H}_{-\mathbf{m},\mathbf{e}}. \tag{7.9}$$

It also, of course, maps $\tau \to -1/\tau$. (Eqn. (7.9) is a discrete analog of (2.23), with $(\vec{\mathfrak{m}},\vec{\mathfrak{n}})$ replaced by their discrete counterparts $(\mathbf{m},\mathbf{e})$.)

While it is possible to define all the Hilbert spaces $\mathcal{H}_{\mathbf{e},\mathbf{m}}$, a given gauge theory construction may not use all of them. If we do gauge theory on $M = S^1 \times W$ with simply-connected gauge group $\overline{G}$, then we must set to zero the characteristic classes $\mathbf{a}$ and $\mathbf{m}$ that enter in the partition function $Z_{\mathbf{a},\mathbf{m}}(M)$. For the Hilbert space, this means that $\mathbf{e}$, which arises by a Fourier transform with respect to $\mathbf{a}$, is arbitrary, while $\mathbf{m} = 0$. Hence

$$\mathcal{H}(\overline{G},W) = \oplus_{\mathbf{e}}\mathcal{H}_{\mathbf{e},0}. \tag{7.10}$$

Taking into account the assumed transformation law for $\mathbf{e},\mathbf{m}$ under S-duality, we see that the Hilbert space of the dual theory must be

$$\oplus_{\mathbf{m}}\mathcal{H}_{0,\mathbf{m}}. \tag{7.11}$$

How is this Hilbert space related to a path integral on $M = S^1 \times W$? Setting $\mathbf{e}$ to zero means summing over $\mathbf{a}$. So in terms of the variables $\mathbf{a},\mathbf{m}$, (7.11) means one must sum over all $\mathbf{a},\mathbf{m}$ with equal weight. In other words, in the dual theory we sum over ${}^LG_{\rm ad}$-bundles with all possible ξ. Thus the dual theory has gauge group ${}^LG_{\rm ad}$, as expected.

The Universal Bundle We need another piece of background, which is the concept of the universal bundle.

Consider first the moduli space $\mathcal{M}(G,C)$ of G-bundles over a Riemann surface C. For every point $p \in \mathcal{M}(G,C)$, there is a corresponding G-bundle $E_p \to C$ parametrized by p; it is determined up to isomorphism. A universal bundle, if it exists, is a G-bundle $\mathcal{E} \to \mathcal{M} \times C$ such that for any $p \in \mathcal{M}$, $\mathcal{E}$ restricted to $p \times C$ is isomorphic to E_p. Gauge theory gives [91] a very direct method to analyze the universal bundle, though we will not explain this here.

Naively, we just define $\mathcal{E}$ so that its restriction to $p \times C$ is "the" bundle over C that is determined by p. The reason that a problem arises is that, to begin with, p only determines an isomorphism class of G-bundle over C, not

an actual bundle. This causes no difficulty locally[32]; we just pick a bundle in the right isomorphism class at a particular p and then deform it in a small neighborhood. So for each small open set $U_i \subset \mathcal{M}$, we can pick a universal bundle $\mathcal{E}_i \to U_i \times C$, and moreover $\mathcal{E}_i$ is unique up to isomorphism.

The next step is to glue together the $\mathcal{E}_i$ over $(U_i \cap U_j) \times C$. The bundles $\mathcal{E}_i$ and $\mathcal{E}_j$ are isomorphic over $(U_i \cap U_j) \times C$, so we pick an isomorphism $\Theta_{ij} : \mathcal{E}_i \cong \mathcal{E}_j$ over this intersection, with $\Theta_{ji} = \Theta_{ij}^{-1}$.

Now if on triple intersections $(U_i \cap U_j \cap U_k) \times C$, the composition $\Theta_{ki}\Theta_{jk}\Theta_{ij}$ is equal to 1, then we can consistently glue together the $\mathcal{E}_i$ to get the desired universal bundle $\mathcal{E}$.

If the gauge group is the adjoint group $G_{\rm ad}$, there is no problem in the gluing. A generic $G_{\rm ad}$ bundle has no automorphisms (exceptions occur at singularities of $\mathcal{M}$). So $\Theta_{ki}\Theta_{jk}\Theta_{ij}$, just because it is an isomorphism of $\mathcal{E}_i$ (restricted to $(U_i \cap U_j \cap U_k) \times C$) to itself, is the identity at a generic point of the triple intersection and hence everywhere.

Thus, the universal $G_{\rm ad}$-bundle $\mathcal{E}_{\rm ad}$ does exist. Now let us repeat this discussion, taking the gauge group to be the simply-connected cover $\overline{G}$ (or more generally, any cover of $G_{\rm ad}$ with a non-trivial center). We start in the same way. For each small open set $U_i \in \mathcal{M}$, we pick a universal bundle $\overline{\mathcal{E}}_i$ over $U_i \times C$, and try to glue to make a universal $\overline{G}$-bundle $\overline{\mathcal{E}}$.

The reason that the result is different is that a generic $\overline{G}$-bundle does have a non-trivial group of automorphisms; the center $\mathcal{Z}$ is a group of automorphisms of any $\overline{G}$-bundle E.

Hence, in the above argument, the composite map $\Theta_{ki}\Theta_{jk}\Theta_{ij}$ over a triple intersection is not necessarily the identity; instead

$$\Theta_{ki}\Theta_{jk}\Theta_{ij} = f_{ijk}, \tag{7.12}$$

with $f_{ijk} \in \mathcal{Z}$. The f_{ijk} combine to a two-cycle defining an element $\zeta \in H^2(\mathcal{M}, \mathcal{Z})$. The element ζ is known to be non-trivial; it is the obstruction to the existence of the universal $\overline{G}$-bundle $\overline{\mathcal{E}}$.

There is another way to explain the existence of ζ. Let $\mathcal{M}(G_{\rm ad}, C)_0$ be the component of $\mathcal{M}(G_{\rm ad}, C)$ that parametrizes topologically trivial bundles – the ones that when restricted to $p \times C$, for $p \in \mathcal{M}(G_{\rm ad}, C)$, can be lifted to $\overline{G}$-bundles. Consider the universal bundle $\mathcal{E}_{\rm ad}$ over $\mathcal{M}(G_{\rm ad}, C)_0 \times C$. If we could lift $\mathcal{E}$ to a $\overline{G}$-bundle $\overline{\mathcal{E}}$, this (after being pulled back to $\mathcal{M}(\overline{G}, C) \times C$,

[32] This statement holds near a smooth point in $\mathcal{M}$. A more precise analysis than we will give shows that the class $\zeta(\mathcal{E}_{\rm ad})$ that obstructs the universal bundle can have a local contribution at singularities.

which is a finite cover of $\mathcal{M}(G_{\rm ad},C)_0\times C)$ would be the universal $\overline{G}$-bundle. In general, the obstruction to lifting a $G_{\rm ad}$-bundle $\mathcal{E}_{\rm ad}$ to a $\overline{G}$-bundle $\overline{\mathcal{E}}$ is the characteristic class $\xi(\mathcal{E}_{\rm ad})$. So the obstruction to existence of the universal $\overline{G}$-bundle is $\zeta=\xi(\mathcal{E}_{\rm ad})$.

There is a similar story for Higgs bundles. The universal Higgs bundle is a pair $(\mathcal{E},\widehat{\varphi})$, with $\mathcal{E}$ being a G-bundle over $\mathcal{M}_H(G,C)\times C$, and $\widehat{\varphi}\in H^0(\mathcal{M}_H(G,C)\times C,{\rm ad}(\mathcal{E})\otimes K_C)$, obeying the following condition. For each $p\in\mathcal{M}_H(G,C)$, the restriction of $(\mathcal{E},\widehat{\varphi})$ to $p\times C$ should be isomorphic to the Hitchin pair (E_p,φ_p) parametrized by p.

The arguments that we have already given can be carried over with no essential change to show that for gauge group $G_{\rm ad}$, the universal Higgs bundle $(\mathcal{E}_{\rm ad},\widehat{\varphi})$ does exist. But for the simply-connected gauge group $\overline{G}$ (or any nontrivial cover of $G_{\rm ad}$), the universal Higgs bundle does not exist. It is obstructed by the fact that a generic Hitchin pair (E,φ), with E a $\overline{G}$-bundle, has the group $\mathcal{Z}$ of automorphisms. The obstruction is just the obstruction to lifting the "bundle" part of the universal $G_{\rm ad}$ Higgs bundle $(\mathcal{E}_{\rm ad},\widehat{\varphi})$ to a $\overline{G}$-bundle. It is therefore

$$\zeta=\xi(\mathcal{E}_{\rm ad}). \tag{7.13}$$

The Universal Bundle As A Twisted Vector Bundle Although for G not of adjoint type, the universal G-bundle does not exist as a vector bundle or as a principal bundle, it does exist as a twisted vector bundle. We will describe this concept in a very pedestrian way.

One way to construct an ordinary vector bundle V of rank N over a manifold X is to cover X with small open sets U_i, on each of which we pick a rank N trivial bundle $V_i\cong U_i\times\mathbb{C}^N$. Then we glue V_i to V_j on the intersection $U_i\cap U_j$ via a gluing map $v_{ij}:V_i\cap V_j\to U(N)$, with $v_{ji}=v_{ij}^{-1}$. If on triple intersections we have

$$v_{ki}v_{jk}v_{ij}=1, \tag{7.14}$$

then the V_i can be glued together consistently to make a rank N vector bundle $V\to X$.

Now suppose that we are given $b\in H^2(X,U(1))$. Then b can be represented explicitly by a $U(1)$-valued cocycle b_{ijk} defined on triple intersections $U_i\cap U_j\cap U_k$. A twisted vector bundle is defined by the same sort of data $v_{ij}:V_i\cap V_j\to U(N)$. But now, instead of (7.14), we ask for

$$v_{ki}v_{jk}v_{ij}=b_{ijk}. \tag{7.15}$$

Thus, a b-twisted vector bundle V is not a vector bundle in the usual sense. However, the associated adjoint bundle $\mathrm{ad}(V)$ is an ordinary vector bundle, since the phase b_{ijk} in (7.15) disappears if we pass to the adjoint representation.

Now let us return to the problem of constructing a universal bundle. In this paper, we are generally a little imprecise about whether by a G-bundle, we mean a principal G-bundle, whose fiber is a copy of G, or an associated vector bundle in some faithful representation of G. Principal bundles make possible a uniform analysis good for any G, but for a group like $U(N)$ that has a convenient faithful representation (the N-dimensional representation) it is useful to think in terms of vector bundles.

In discussing the universal bundle, it is helpful to be more precise. We interpret the transition functions Θ_{ij} of our discussion above as $\overline{G}$-valued functions, with no particular choice of representation. Their projection to G_{ad} gives us transition functions for a universal principal G_{ad}-bundle $\mathcal{E}_{\mathrm{ad}}$, but we cannot lift this to a principal $\overline{G}$-bundle, because of the relation

$$\Theta_{ki}\Theta_{jk}\Theta_{ij} = f_{ijk}, \tag{7.16}$$

where the f_{ijk} define the class $\zeta = \xi(\mathcal{E}_{ad}) \in H^2(\mathcal{M}_H(G_{\mathrm{ad}}, C), \mathcal{Z})$. Now let us pick an irreducible representation ϱ of $\overline{G}$. We write Θ^{ϱ}_{ij} for the transition functions Θ_{ij} evaluated in the representation ϱ. Likewise we write $\varrho(f)$ for f evaluated in the representation ϱ. In the irreducible representation ϱ, the center of $\overline{G}$ acts by scalar multiplication, so $\varrho(f)$ takes values in $U(1)$. We have

$$\Theta^{\varrho}_{ki}\Theta^{\varrho}_{jk}\Theta^{\varrho}_{ij} = \varrho(f_{ijk}). \tag{7.17}$$

The quantities $\varrho(f_{ijk})$ are a cocycle which defines the element $\varrho(\zeta) \in H^2(\mathcal{M}_H(G_{\mathrm{ad}}, C), U(1))$.

If $\varrho(f) = 1$, the objects Θ^{ϱ} are transition functions that define a vector bundle $\mathcal{E}_{\varrho} \to \mathcal{M}_H(\overline{G}, C) \times C$ that we may call the universal bundle in the representation ϱ. In general this is not the case. However, comparing eqns. (7.16) and (7.15), we see that while $\mathcal{E}_{\varrho}$ may not exist as an ordinary vector bundle, it does always exist as a twisted vector bundle, twisted by $\varrho(\zeta) \in H^2(\mathcal{M}_H(G_{\mathrm{ad}}, C), U(1))$.

7.2. Compactification To Two Dimensions

Now we compactify to two dimensions. Consider $\mathcal{N} = 4$ super Yang-Mills theory on $M = \Sigma \times C$ with gauge group G_{ad}. The data in this theory are a

$G_{\rm ad}$-bundle $E \to \Sigma \times C$, a connnection A on E, and various other fields. Any choice of E determines a class $\xi(E) \in H^2(\Sigma \times C, \mathcal{Z})$. Since $H^2(\Sigma \times C, \mathcal{Z}) = H^2(\Sigma, \mathcal{Z}) \oplus H^1(\Sigma, H^1(C, \mathcal{Z})) \oplus H^2(C, \mathcal{Z})$, we have

$$\xi(E) \in H^2(\Sigma, \mathcal{Z}) \oplus H^1(\Sigma, H^1(C, \mathcal{Z})) \oplus H^2(C, \mathcal{Z}). \tag{7.18}$$

Relative to this decomposition we write $\xi(E) = \xi(E)^{2,0} + \xi(E)^{1,1} + \xi(E)^{0,2}$.

We can describe the same thing in the low energy effective sigma-model. Here we consider low energy data on $\Sigma \times C$, which when restricted to $p \times C$ for a point $p \in \Sigma$ determine a point in $\mathcal{M}_H(G_{\rm ad})$. As p varies, we get a map $\Phi : \Sigma \to \mathcal{M}_H(G_{\rm ad})$. Composing Φ with the identity map on C, we get a map $\Phi \times 1 : \Sigma \times C \to \mathcal{M}_H(G_{\rm ad}) \times C$. Now recall the universal bundle $\mathcal{E}_{\rm ad}$ and the class $\zeta = \xi(\mathcal{E}_{\rm ad}) \in H^2(\mathcal{M}_H(G_{\rm ad}) \times C, \mathcal{Z})$. Pulling it back to $\Sigma \times C$, we get a class $(\Phi \times 1)^*(\zeta) \in H^2(\Sigma \times C, \mathcal{Z})$. By chasing through the definitions, one can see that

$$\xi(E) = (\Phi \times 1)^*(\zeta). \tag{7.19}$$

In fact, the universal bundle $\mathcal{E}_{ad}$ is defined so that $(\Phi \times 1)^*(\mathcal{E}_{\rm ad}) = E$, so $(\Phi \times 1)^*(\zeta) = (\Phi \times 1)^*(\xi(\mathcal{E}_{\rm ad})) = \xi((\Phi \times 1)^*(\mathcal{E}_{\rm ad})) = \xi(E)$.

Let us further specialize to $\Sigma = S^1 \times \widetilde{S}^1$, where the two factors are circles. We write $W = \widetilde{S}^1 \times C$. So $\Sigma \times C = S^1 \times W$. $\xi(E)$ has the familiar expansion $\xi(E) = \mathbf{a} + \mathbf{m}$ with $\mathbf{a} \in H^1(W, \mathcal{Z})$, $\mathbf{m} \in H^2(W, \mathcal{Z})$. Now, however, we can further expand $H^1(W, \mathcal{Z}) = H^1(\widetilde{S}^1, \mathcal{Z}) \oplus H^1(C, \mathcal{Z})$, so we have $\mathbf{a} = \mathbf{a}_0 + \mathbf{a}_1$ with

$$\begin{aligned} &\mathbf{a}_0 \in H^1(\widetilde{S}^1, \mathcal{Z}) = \mathcal{Z}, \\ &\mathbf{a}_1 \in H^1(C, \mathcal{Z}). \end{aligned} \tag{7.20}$$

And similarly, we can expand $H^2(W, \mathcal{Z}) = H^1(\widetilde{S}^1, H^1(C, \mathcal{Z})) \oplus H^2(C, \mathcal{Z})$, so we have $\mathbf{m} = \mathbf{m}_0 + \mathbf{m}_1$ with

$$\begin{aligned} &\mathbf{m}_0 \in H^2(C, \mathcal{Z}) = \mathcal{Z} \\ &\mathbf{m}_1 \in H^1(\widetilde{S}^1, H^1(C, \mathcal{Z})) = H^1(C, \mathcal{Z}). \end{aligned} \tag{7.21}$$

Comparing back to the expansion (7.18), we see that

$$\xi(E)^{2,0} = \mathbf{a}_0, \ \xi(E)^{0,2} = \mathbf{m}_0. \tag{7.22}$$

We can similarly decompose the character $\mathbf{e}$ that is dual to $\mathbf{a}$. Clearly, we have $\mathbf{e} = \mathbf{e}_0 + \mathbf{e}_1$, where $\mathbf{e}_0$ is a character of $\mathbf{a}_0 \in \mathcal{Z}$ and $\mathbf{e}_1$ is a character

of $\mathbf{a}_1 \in H^1(C, \mathcal{Z})$. Hence, using the selfduality of $\mathcal{Z}$ and Poincaré duality,

$$\begin{aligned} &\mathbf{e}_0 \in \mathcal{Z}^\vee = \mathcal{Z} \\ &\mathbf{e}_1 \in H^1(C, \mathcal{Z})^\vee = H^1(C, \mathcal{Z}). \end{aligned} \tag{7.23}$$

In view of (7.9), the transformation under S-duality is

$$\begin{aligned} &(\mathbf{e}_0, \mathbf{m}_0) \longrightarrow (-\mathbf{m}_0, \mathbf{e}_0) \\ &(\mathbf{e}_1, \mathbf{m}_1) \longrightarrow (-\mathbf{m}_1, \mathbf{e}_1). \end{aligned} \tag{7.24}$$

Our next goal is to interpret $\mathbf{e}_0$, $\mathbf{e}_1$, $\mathbf{m}_0$, and $\mathbf{m}_1$ in the effective two-dimensional sigma-model with target $\mathcal{M}_H(G_{\rm ad})$. We will consider both the case of a closed Riemann surface Σ, and the case when Σ has a boundary. The latter case will enable us to understand the implications of (7.24) for the geometric Langlands program.

Interpretation Of $\mathbf{m_0}$ The easiest to interpret is $\mathbf{m}_0 = \xi(E)^{0,2}$. The target space $\mathcal{M}_H(G_{\rm ad}, C)$ of our sigma-model is not connected. Its components are labeled by the topological type of the $G_{\rm ad}$-bundle $E \to C$. But this is exactly what is measured by $\mathbf{m}_0$.

Interpretation Of $\mathbf{e_0}$ Consider in general a sigma-model of maps $\Phi\colon \Sigma \to X$, for some X. A flat B-field is an element $b \in H^2(X, U(1))$. A flat B-field is incorporated in the sigma-model path integral as follows, for the case that Σ has no boundary. Given a map $\Phi : \Sigma \to X$, one pulls back b to $\Phi^*(b) \in H^2(\Sigma, U(1)) = U(1)$, and then one includes in the path integral a factor of $\Phi^*(b)$. Thus, in this situation, incorporating the flat B-field has the effect of weighting by phases the different components of maps of Σ to X.

Now $\mathbf{e}_0$ determines a flat B-field in the sigma-model of maps $\Phi : \Sigma \to \mathcal{M}_H(G_{\rm ad}, C)$. Indeed, there is as we have explained a natural class $\zeta \in H^2(\mathcal{M}_H(G_{\rm ad}, C), \mathcal{Z})$, which expresses the obstruction to lifting the universal $G_{\rm ad}$ bundle $\mathcal{E}_{\rm ad} \to \mathcal{M}_H(G_{\rm ad}, C) \times C$ to a $\overline{G}$-bundle. By composing ζ with $\mathbf{e}_0 : \mathcal{Z} \to U(1)$, we get a flat B-field $b_{\mathbf{e}_0} = \mathbf{e}_0(\zeta) \in H^2(\mathcal{M}_H(G_{\rm ad}, C), U(1))$. We claim that the role of $\mathbf{e}_0$ in the effective sigma-model of maps $\Sigma \to \mathcal{M}_H(G_{\rm ad})$ is precisely to weight every map in the way that one would expect for a sigma-model with the flat B-field $b_{\mathbf{e}_0}$.

In fact, the definition of $\mathbf{e}_0$ is that the path integral includes a phase factor $\mathbf{e}_0(\mathbf{a}_0) = \mathbf{e}_0(\xi(E)^{2,0})$. But $\xi(E)^{2,0} = \Phi^*(\zeta)$; this follows upon restricting (7.19) to a point in C. So the phase factor induced in the path integral by $\mathbf{e}_0$ is $\mathbf{e}_0(\Phi^*(\zeta)) = \Phi^*(\mathbf{e}_0(\zeta)) = \Phi^*(b_{\mathbf{e}_0})$. This justifies our claim that the effect of $\mathbf{e}_0$ in the sigma-model is to generate a flat B-field $b_{\mathbf{e}_0}$.

(7.24) therefore means that S-duality exchanges the topological class $\mathbf{m}_0$ of a flat $G_{\rm ad}$-bundle with the flat B-field determined by $\mathbf{e}_0$.

Incorporation Of Branes For applications to the geometric Langlands program, the real payoff is to understand the implications of all this for branes.

So we take Σ to be a Riemann surface with boundary. $M = \Sigma \times C$ is therefore a four-manifold with boundary $\partial M = \partial\Sigma \times C$. On ∂M, we place some supersymmetric boundary condition.

The effective two-dimensional description is a sigma-model of maps $\Phi : \Sigma \to \mathcal{M}_H(G_{\rm ad})$, with boundary condition corresponding to some brane. We would like to understand the role of $\mathbf{e}_0$ and $\mathbf{m}_0$ in this description.

There is little new to say about $\mathbf{m}_0$. It labels the components of $\mathcal{M}_H(G_{\rm ad})$, whether Σ has a boundary or not.

The role of $\mathbf{e}_0$ is more subtle. As we have seen, the sigma-model with target $\mathcal{M}_H(G_{\rm ad})$ is endowed with a flat B-field $\mathbf{e}_0(\zeta)$. A flat B-field has a very interesting effect on branes [102]. In the absence of a B-field, a brane on $\mathcal{M}_H(G_{\rm ad})$ has a Chan-Paton bundle, which is a vector bundle over $\mathcal{M}_H(G_{\rm ad})$ (or more generally a sheaf, perhaps supported on a submanifold, that defines a K-theory class of $\mathcal{M}_H(G_{\rm ad})$). However, in the presence of a flat B-field, associated with an element $b \in H^2(\mathcal{M}_H(G_{\rm ad}), U(1))$, the Chan-Paton bundle becomes a twisted vector bundle (or more generally a twisted sheaf related to an element of the twisted K-theory of $\mathcal{M}_H(G_{\rm ad})$), twisted by b in the sense of eqn. (7.15). It is because of this that we introduced the concept of a twisted vector bundle.

So in short, for $\mathbf{e}_0 \neq 0$, the Chan-Paton bundle of a brane is a twisted vector bundle, twisted by $b_{\mathbf{e}_0} = \mathbf{e}_0(\zeta)$. Luckily, from the analysis at the end of section 7.1, we have a plentiful supply of such twisted vector bundles. If ϱ is any irreducible representation of $\overline{G}$ such that the character of the center of $\overline{G}$ defined by ϱ is equal to $\mathbf{e}_0$, then the universal bundle $\mathcal{E}_\varrho$ in the representation ϱ is an example of a twisted vector bundle for the flat B-field $b_{\mathbf{e}_0}$.

Other important examples of twisted branes can be constructed by picking a submanifold $Y \subset \mathcal{M}_H(G_{\rm ad})$ such that $b_{\mathbf{e}_0}$ is trivial when restricted to Y. In this case, the definition of a brane supported on Y is independent of $\mathbf{e}_0$ (up to a not quite canonical isomorphism). For example, Y could be a point in $\mathcal{M}_H(G_{\rm ad})$. Certainly $b_{\mathbf{e}_0}$ is trivial when restricted to a point, so zerobranes exist for any $\mathbf{e}_0$. From such zerobranes, we can form electric eigenbranes, as we explain in section 8. A slightly more subtle example is a magnetic eigenbrane, a brane of rank one supported on a fiber $\boldsymbol{F}$ of the Hitchin fibration.

The Chan-Paton bundle of such a brane should be a flat[33] twisted line bundle. In [21], it is shown that the flat b-field $b_{\mathbf{e}_0}$ is trivial when restricted to $\boldsymbol{F}$, as a result of which the space of flat twisted line bundles is isomorphic, but not canonically isomorphic, to the space of ordinary flat line bundles on $\boldsymbol{F}$.

We can now deduce from S-duality a statement about branes. The duality transformation S maps $(\mathbf{e}_0, \mathbf{m}_0) \to (-\mathbf{m}_0, \mathbf{e}_0)$, so it exchanges the topology of the component of $\mathcal{M}_H$ on which a brane is supported with the flat B-field $b_{\mathbf{e}_0}$ by which its Chan-Paton bundle is twisted. It also, of course, exchanges $G_{\rm ad}$ with[34] ${}^LG_{\rm ad}$, and (as we explain in the concluding remark of this section), exchanges $\mathcal{M}_H(G_{\rm ad})$ with $\widehat{\mathcal{M}}_H({}^LG_{\rm ad})$, which we define to be the universal cover of $\mathcal{M}_H({}^LG_{\rm ad})$.

As a special case of this duality, a point on one side, contained in a fiber $\boldsymbol{F}$ of the Hitchin fibration, is mapped on the other side to a brane of rank one supported on the corresponding fiber ${}^L\boldsymbol{F} = \Xi(\boldsymbol{F})$ of the Hitchin fibration, and endowed with a flat twisted line bundle. $\boldsymbol{F}$ is a union of complex tori, labeled by the characteristic class $\mathbf{m}_0 = \xi(E)$ of the Higgs bundle. The choice of a component of $\boldsymbol{F}$ on one side determines on the other side the discrete electric field $\mathbf{e}_0$ and hence the twist. The choice of a point on $\boldsymbol{F}$ determines a flat twisted line bundle on ${}^L\boldsymbol{F}$ (to which it maps under the duality transformation S). Of course, this relationship between $\boldsymbol{F}$ and ${}^L\boldsymbol{F}$ is reciprocal. This twisted duality between $\boldsymbol{F}$ and ${}^L\boldsymbol{F}$ is in fact one of the main results of Hausel and Thaddeus [21].

Interpretation Of $\mathbf{e}_1$ And $\mathbf{m}_1$ Finally, let us discuss the interpretation of $\mathbf{e}_1$ and $\mathbf{m}_1$ in two-dimensional terms. It will be helpful to begin by comparing the theories with gauge groups $\overline{G}$ and $G_{\rm ad}$.

$\mathcal{M}_H(\overline{G})$ is simply-connected. It has a natural group of symmetries $\mathbf{E}_C = H^1(C, \mathcal{Z})$. Indeed, $\mathbf{E}_C$ parametrizes $\mathcal{Z}$-bundles over C. A $\overline{G}$ Higgs bundle (E, φ) can be tensored with a $\mathcal{Z}$-bundle to make a new $\overline{G}$ Higgs bundle. (Concretely, this operation multiplies the holonomies of E around noncontractible loops in C by elements of the center of $\overline{G}$.) So this gives a

[33] One can here interpret flatness to mean, just as for ordinary line bundles, that the transition functions are constants.

[34] By ${}^LG_{\rm ad}$, we mean the adjoint form of the group ${}^LG$. The statement we are describing here is best expressed in terms of adjoint bundles on both sides to allow all possible topologies. Momentarily we indicate explicitly whether we want a given component of $\mathcal{M}_H$ or its universal cover.

group $\mathbf{E}_C$ of symmetries of the sigma-model with target $\mathcal{M}_H(\overline{G})$. The Hilbert space of this sigma-model can be decomposed in characters of $\mathbf{E}_C$.

On the other hand, $\mathcal{M}_H(G_{\rm ad})$ has no such geometrical symmetries. But it has a fundamental group $\mathbf{M}_C = H^1(C,\mathcal{Z})$, which is isomorphic to $\mathbf{E}_C$. To understand where this fundamental group comes from, a shortcut is to note that one component $\mathcal{M}_H(G_{\rm ad})_0$ of $\mathcal{M}_H(G_{\rm ad})$, namely the component that parametrizes Higgs bundles that can be lifted to $\overline{G}$, is simply $\mathcal{M}_H(G_{\rm ad})_0 = \mathcal{M}_H(\overline{G})/\mathbf{E}_C$. Dividing by $\mathbf{E}_C$ eliminates the geometrical symmetries of $\mathcal{M}_H(\overline{G})$, but of course it creates a fundamental group.[35] So $\pi_1(\mathcal{M}_H(G_{\rm ad})_0) \cong \mathbf{E}_C$.

Actually, the fundamental group is the same for any component of $\mathcal{M}_H(G_{\rm ad})$. $\mathcal{M}_H(G_{\rm ad})$ is defined by dividing the space of all solutions of Hitchin's equations, for gauge group $G_{\rm ad}$, by the group $\mathcal{G}_{\rm ad}(C)$ of all $G_{\rm ad}$-valued gauge transformations on C. If one were to divide only by the connected component of $\mathcal{G}_{\rm ad}(C)$, one would get the universal cover $\widehat{\mathcal{M}}_H(G_{\rm ad},C)$. The fundamental group of $\mathcal{M}_H(G_{\rm ad})$ is therefore the group of components of $\mathcal{G}_{\rm ad}(C)$, and this is $\mathbf{E}_C = H^1(C,\mathcal{Z})$, for any component of $\mathcal{M}_H(G_{\rm ad})$.

In sum, strings moving on $\mathcal{M}_H(G_{\rm ad})$ have a discrete group of conserved winding numbers $\mathbf{M}_C = \pi_1(\mathcal{M}_H(G_{\rm ad})) = H^1(C,\mathcal{Z})$. Likewise strings moving on $\mathcal{M}_H(\overline{G})$ have a discrete group of conserved momenta $\mathbf{E}_C = H^1(C,\mathcal{Z})$.

Let us compare the symmetries $\mathbf{E}_C$ and $\mathbf{M}_C$ to what we can see in the underlying gauge theory. In $G_{\rm ad}$ gauge theory on $M = \Sigma\times C = S^1\times\widetilde{S}^1\times C$, the bundle E and other data determine a map $\Phi : S^1\times\widetilde{S}^1\to\mathcal{M}_H(G_{\rm ad})$. In analyzing the topology of this situation, we expanded the characteristic class $\xi(E)$ as $\xi(E) = \mathbf{a}\oplus\mathbf{m}$, where $\mathbf{m}$ is the restriction of $\xi(E)$ to $\widetilde{S}^1\times C$ (more precisely, to $p\times\widetilde{S}^1\times C$, for a point $p\in S^1$). In the low energy sigma-model, $\mathbf{a}$ and $\mathbf{m}$ are invariants of $\Phi:\Sigma\to\mathcal{M}_H(G_{\rm ad})$. In particular, $\mathbf{m}$ only depends on the restriction $\Phi|_{\widetilde{S}^1}$ of Φ to $\widetilde{S}^1$.

In (7.21), we further expanded $\mathbf{m} = \mathbf{m}_0\oplus\mathbf{m}_1$ where $\mathbf{m}_1\in H^1(\widetilde{S}^1, H^1(C,\mathcal{Z}))$. $\mathbf{m}_1$ is a topological invariant of $\Phi|_{\widetilde{S}^1}$ that vanishes for constant maps of $\widetilde{S}^1$ to $\mathcal{M}_H(G_{\rm ad})$. So it measures the homotopy class of the map $\Phi|_{\widetilde{S}^1}$ in $\pi_1(\mathcal{M}_H(G_{\rm ad})) = H^1(C,\mathcal{Z})$.

Similarly, we could exchange the role of S^1 and $\widetilde{S}^1$. The restriction of $\xi(E)$ to $S^1\times C$ (that is, to $S^1\times q\times C$, for a point $q\in\widetilde{S}^1$), is in the

[35] $\mathbf{E}_C$ does not act freely, so $\mathcal{M}_H(G_{ad})$ has orbifold singularities. (It also has more severe singularities from reducible Higgs bundles.) As is familiar in sigma models, the fundamental group of $\mathcal{M}_H(G_{ad})$ must be understood in an orbifold sense. Alternatively, one can rely on the four-dimensional gauge theory instead of reducing to the sigma model with its singularities.

above notation $\mathbf{m}_0 \oplus \mathbf{a}_1$. In particular, $\mathbf{a}_1 \in H^1(S^1, H^1(C, \mathcal{Z}))$ measures the winding of Φ in the S^1 or "time" direction. The character $\mathbf{e}_1$ which is dual to $\mathbf{a}_1$ therefore measures the conserved momentum of the strings.

An Example For an important illustration of all this, consider the Langlands dual pair $G_{\rm ad}$ and ${}^L\overline{G}$. In $G_{\rm ad}$ gauge theory, we set $\mathbf{e} = 0$ and consider a sigma-model with target $\mathcal{M}_H(G_{\rm ad})$. In this model, there are no conserved momenta, but there is a symmetry $\mathbf{M}_C$ of string windings.

In the dual picture, the gauge group is ${}^L\overline{G}$, we set $\mathbf{m} = 0$, and the sigma-model with target $\mathcal{M}_H(\overline{G})$ has no conserved string windings, but a symmetry group $\mathbf{E}_C$ of discrete conserved momenta.

S-duality or Montonen-Olive or Langlands duality exchanges the two pictures, exchanging $\mathbf{E}_C({}^L\overline{G})$ with $\mathbf{M}_C(G_{\rm ad})$. The fact that the duality exchanges the discrete conserved momenta and windings of strings is an aspect of its relation to T-duality in two dimensions.

A Concluding Comment More generally, we can simply specify $\mathbf{m}_0$ and $\mathbf{e}_0$ as we please. Then we consider branes on a component of $\mathcal{M}_H(G_{\rm ad})$ labeled topologically by $\mathbf{m}_0$; the Chan-Paton bundles of the branes are twisted by $\mathbf{e}_0$.

We still have two ways to proceed with the quantization. If we divide by all $G_{\rm ad}$-valued gauge transformations, then the target space is the component of $\mathcal{M}_H(G_{\rm ad}, C)$ labeled by $\mathbf{m}_0$. In this case, there is a finite group $\mathbf{M}_1$ that classifies the string winding numbers, but there is no group $\mathbf{E}_1$ of geometrical symmetries. If we divide by only the connected gauge transformations, then the target is the universal cover $\widehat{\mathcal{M}}_H(G_{\rm ad}, C)$. In this case, there is a group $\mathbf{E}_1$ of geometrical symmetries, but no group $\mathbf{M}_1$ of string winding numbers. S-duality exchanges $\mathbf{E}_1$ and $\mathbf{M}_1$, so (in addition to exchanging G and ${}^LG$) it exchanges the two methods of quantization.

In particular, when we apply S-duality to branes of specified $\mathbf{e}_0$ and $\mathbf{m}_0$, we must exchange the two methods of quantization in addition to exchanging the two adjoint groups $G_{\rm ad}$ and ${}^LG_{\rm ad}$.

8. Electric Eigenbranes

In section 6, we introduced the topological Wilson and 't Hooft operators, their action on branes, and the concept of an electric or magnetic eigenbrane. Our goal in section 8.1 will be to understand explicitly how Wilson operators act on branes. Then in section 8.2 we will use this information to identify zerobranes as electric eigenbranes.

8.1. How Wilson Operators Act On Branes

As always in this paper, the theory of interest is twisted $\mathcal{N} = 4$ super Yang-Mills theory, with gauge group G, on a four-manifold M. The fields include a connection A on a G-bundle $E \to M$. We assume M to have a boundary that is labeled by a brane $\mathcal{B}$ that respects the topological symmetry, a condition that of course depends on the parameter Ψ.

Topological Wilson line operators exist if $\Psi = \infty$, which we achieve by setting $t = i$ with arbitrary gauge coupling τ. Let R be a representation of G. A Wilson operator in the representation R and supported on a curve $\mathcal{S}$ contributes in the path integral a factor

$$W_{\mathcal{S}}(R) = P \exp\left(- \int_{\mathcal{S}} (A + i\phi) \right). \tag{8.1}$$

More explicitly, write $E(R)$ for the bundle associated to E in the representation R of G. The Wilson line operator is the matrix of parallel transport along $\mathcal{S}$, in the bundle $E(R)$, with the connection $\mathcal{A} = A + i\phi$. The effect of including this operator is to add an external charge, in the representation R, whose trajectory in spacetime is $\mathcal{S}$. If $\mathcal{S}$ is a closed loop, we take the trace of $W_{\mathcal{S}}(R)$, and if instead $\mathcal{S}$ goes off to infinity, we combine $W_{\mathcal{S}}(R)$ with initial and final states at the ends of $\mathcal{S}$ so as to make a gauge-invariant expression.

Now we specialize to $M = \Sigma \times C$, so as to be able to reduce the discussion to a two-dimensional sigma-model with target $\mathcal{M}_H(G, C)$. In this sigma-model, $\mathcal{B}$ is represented by a Chan-Paton bundle U (or possibly a more general sheaf or complex of sheaves [103,104]) over the target space $\mathcal{M}_H$. U is endowed with a connection α. Since the brane $\mathcal{B}$ is defined by its Chan-Paton bundle (or sheaf), a Wilson line operator will have to act on branes by acting on this bundle or sheaf in some way.

Let us recall how Chan-Paton bundles enter sigma-models. We consider a sigma-model of maps $\Phi : \Sigma \to \mathcal{M}_H$, and a brane $\mathcal{B}$ that is endowed with a Chan-Paton bundle $U \to \mathcal{M}_H$ with connection α. The quantum theory is defined by an integral over the possible maps Φ, along with certain fermionic variables. One factor in the path integral comes from the bulk action $\boldsymbol{I}$ and takes the form $\exp\left(-\int_\Sigma \boldsymbol{I}\right)$. There also is a boundary factor that involves parallel transport in the Chan-Paton bundle. Let Q be the part of the boundary of Σ that is labeled by the brane $\mathcal{B}$, and write Φ_{Q} for the restriction of Φ to Q. The boundary factor in the path integral involving Q is given by

the parallel transport or holonomy along Q of the bundle $\Phi_{\mathrm{Q}}^*(U)$:

$$P\exp\left(-\int_{\mathrm{Q}}\left(\Phi_{\mathrm{Q}}^*(\alpha)+\ldots\right)\right). \tag{8.2}$$

If Q is a closed circle, we take a trace of this holonomy, and otherwise this factor combines at the endpoints of Q with other factors, depending on the precise calculation that one chooses to perform, to make a gauge-invariant expression. The ellipses in (8.2) are fermionic corrections to the connection $\Phi_{\mathrm{Q}}^*(\alpha)$ on $\Phi_{\mathrm{Q}}^*(U)$. They are required by supersymmetry, rather as the shift $A\to\mathcal{A}=A+i\phi$ was needed in section 6.1 to define supersymmetric Wilson operators.

There is an obvious analogy between the factor (8.2) by which Chan-Paton bundles enter in sigma-models and the factor (8.1) by which a Wilson line operator influences the underlying four-dimensional gauge theory. The analogy is even closer because to define the action of the Wilson operator on the brane $\mathcal{B}$, we must take the limit as $\mathcal{S}$ (or rather its projection from $\Sigma\times C$ to Σ) approaches Q.

To get something precise from this analogy, we begin with the following observation. When gauge theory on a G-bundle $E\to M=\Sigma\times C$ is described in terms of a map $\Phi:\Sigma\to\mathcal{M}_H$, the G-bundle E can be identified as $(\Phi\times 1)^*(\mathcal{E})$, where $\mathcal{E}$ is the "bundle" part of the universal Higgs bundle $(\mathcal{E},\widehat{\varphi})$ over $\mathcal{M}_H$, and $(\Phi\times 1)^*(\mathcal{E})$ is its pullback via the map $\Phi\times 1:\Sigma\times C\to\mathcal{M}_H\times C$. This statement just means that, to the extent that the sigma-model is a good description, for each point $q\in\Sigma$, the bosonic fields A,ϕ of the gauge theory, when restricted to $q\times C$, are given by the solution of Hitchin's equations corresponding to the point $\Phi(q)\in\mathcal{M}_H$. This solution is simply, up to a gauge transformation, the restriction of the universal Higgs bundle $(\mathcal{E},\widehat{\varphi})$ to $\Phi(q)\times C$.

To interpret the connection $\mathcal{A}=A+i\phi$ in (8.1) in terms of the sigma-model, we note that in general, this connection involves both $\mathcal{A}_\Sigma$, the part of the connection tangent to Σ, and $\mathcal{A}_C$, the part tangent to C. Here in the low energy theory, we can assume that $\mathcal{A}_C$ obeys Hitchin's equations, and as long as we avoid singularities in $\mathcal{M}_H$, the fields $\mathcal{A}_\Sigma$ are massive in the sigma-model. They can therefore be integrated out in favor of the sigma-model fields $\mathcal{A}_C$. For very large $\mathrm{Im}\,\tau$, it is sufficient to integrate out $\mathcal{A}_\Sigma$ at the classical level. The part of the gauge theory action which depends on only A and ϕ was written in (3.47). Assuming that A_C and ϕ_C satisfy Hitchin equations and dropping the terms which vanish as the volume of C

goes to zero, we find a quadratic action for $\mathcal{A}_\Sigma$. The corresponding equations of motion read

$$\mathcal{D}_C{}^\dagger \mathcal{D}_C \mathcal{A}_\Sigma = \mathcal{D}_C{}^\dagger d_\Sigma \mathcal{A}_C + \dots, \tag{8.3}$$

where $\mathcal{D}_C$ is the covariant differential with respect to the connection $\mathcal{A}_C$. The ellipses refer to terms involving zero modes of the fermions ψ, $\widetilde{\psi}$, etc., of the four-dimensional gauge theory; we will not write these terms explicitly. A map $\Phi : \Sigma \to \mathcal{M}_H$ determines $\mathcal{A}_C$ and hence also $d_\Sigma \mathcal{A}_C$, and then, assuming we keep away from singularities of $\mathcal{M}_H$, the equation (8.3) has a unique solution for $\mathcal{A}_\Sigma$.

So once $\Phi : \Sigma \to \mathcal{M}_H$ is given (and assuming that we keep away from singularities of $\mathcal{M}_H$), the connection $\mathcal{A} = (\mathcal{A}_\Sigma, \mathcal{A}_C)$ is determined. $\mathcal{A}$ is, of course, a connection on the bundle $E = (\Phi \times 1)^*(\mathcal{E})$. The connection $\mathcal{A}$ is actually the pullback by $\Phi \times 1$ of a connection $\widehat{\mathcal{A}}$ on $\mathcal{E} \to \mathcal{M}_H \times C$. In fact, to define $\widehat{\mathcal{A}}$, we must specify its components $\mathcal{A}_{\mathcal{M}_H}, \mathcal{A}_C$ tangent to $\mathcal{M}_H$ and C. $\mathcal{A}_C$ is the appropriate solution of Hitchin's equations, and $\mathcal{A}_{\mathcal{M}_H}$ is defined by generalizing (8.3) in an obvious way:[36]

$$\mathcal{D}_C{}^\dagger \mathcal{D}_C \mathcal{A}_{\mathcal{M}_H} = \mathcal{D}_C{}^\dagger d_{\mathcal{M}_H} \mathcal{A}_C + \dots. \tag{8.4}$$

Now let us specialize to the case that $\mathcal{S} = \gamma \times p$, with γ a curve in Σ and p a point in C. We write $\mathcal{E}_p(R)$ for the restriction of $\mathcal{E}(R)$ to $\mathcal{M}_H \times p$. We also write Φ_p for the restriction to $\Sigma \times \{p\} \subset \Sigma \times C$ of the map $\Phi \times 1 : \Sigma \times C \to \mathcal{M}_H \times C$. We can replace the connection $\mathcal{A} = A + i\phi$ in (8.1) by $\Phi_p^*(\widehat{\mathcal{A}})$. Hence the factor in the path integral that comes from the inclusion of a Wilson operator on the contour $\mathcal{S}$ in the representation R can be written as

$$W_R(\mathcal{S}) = P \exp\left(- \int_\gamma \Phi_p^*(\widehat{\mathcal{A}}) \right). \tag{8.5}$$

In the limit that γ approaches the boundary Q of Σ, this has the same form as the term that comes anyway from the Chan-Paton bundle U of the

[36] The ellipses in (8.3) involve fermionic zero modes, which represent tangent vectors to $\mathcal{M}_H$ and so have analogs in the case of $\mathcal{M}_H \times C$.

original brane $\mathcal{B}$:[37]

$$P\exp\left(-\int_{\mathrm{Q}}(\Phi^*_{\mathrm{Q}}(\alpha)+\ldots)\right). \tag{8.6}$$

So we learn how Wilson lines act on branes. A Wilson line in the representation R and supported at a point $p\in C$ transforms the Chan-Paton bundle U of a brane $\mathcal{B}$ by

$$U\longrightarrow U\otimes\mathcal{E}_p(R). \tag{8.7}$$

Transformation Of The *B*-Field As we have discussed in section 7.0, $\mathcal{E}(R)$, and hence also $\mathcal{E}_p(R)$, in general does not exist as a vector bundle. But it always exists as a twisted vector bundle, twisted by the flat B-field $\theta_R(\zeta)$, where θ_R is the character of the center of the gauge group determined by R.

The category of branes depends on a choice of B-field, a fact that we exploited in section 7.2. For a given discrete electric field $\mathbf{e}_0$, the background B-field in the sigma-model on $\mathcal{M}_H$ is $b_{\mathbf{e}_0}=\mathbf{e}_0(\zeta)$, where $\zeta=\xi(\mathcal{E}_{\mathrm{ad}})$ is the obstruction to existence of a universal $\overline{G}$ Higgs bundle. Tensoring with a twisted bundle that is twisted by $\theta_R(\zeta)$ maps a brane that is twisted by a flat B-field b to a brane that is twisted by $b+\theta_R(\zeta)$. Therefore, the action of a Wilson line operator on branes changes the B-field, by $b\to b+\theta_R(\zeta)$. In other words, it changes the discrete electric field studied in section 7 by $\mathbf{e}_0\to\mathbf{e}_0+\theta_R$.

We want to understand what this result means for S-duality. So we write it as a statement about the dual gauge theory, with gauge group ${}^LG$, and a Wilson line $W({}^LR)$ determined by a representation ${}^LR$. This Wilson line transforms the discrete electric field by

$$\mathbf{e}_0\longrightarrow\mathbf{e}_0+\theta({}^LR), \tag{8.8}$$

where, for convenience, we write $\theta({}^LR)$ rather than $\theta_{{}^LR}$.

[37] The ellipses in (8.6) represent fermionic terms whose analog in (8.5) arises from the ellipses in (8.3), which reflect fermionic contributions to $\mathcal{A}_\Sigma$. All these terms are uniquely determined by the topological symmetry, so we do not need to worry about comparing them.

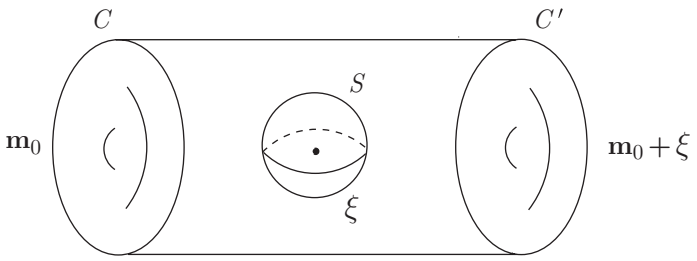

Figure 11: Insertion of an 't Hooft operator changes the topology of a G-bundle, as shown here. Sketched is a three manifold with boundary components consisting of two Riemann surfaces C and C' and a small two-sphere S enclosing a point at which an 't Hooft operator is inserted. Cobordism invariance of the characteristic class implies that if $\mathbf{m_0}$ is the characteristic class of the G-bundle $E \to C$, then the characteristic class of $E \to C'$ must be $\mathbf{m}_0 + \xi$, where $\xi = \xi({}^L R)$ is the characteristic class associated with the 't Hooft operator.

Under S-duality, the fact that a Wilson line operator can change the discrete electric field $\mathbf{e}_0$ maps to the fact that an 't Hooft line operator can change the characteristic class $\mathbf{m}_0 \in H^2(C, \pi_1(G))$ which classifies the topology of the G-bundle $E \to C$. Indeed, as we explained in section 6.1, an 't Hooft operator $T({}^L R)$ is constructed from a G-bundle, which we may call $E({}^L R)$, over $S^2 \cong \mathbb{CP}^1$. This G-bundle has a characteristic class $\xi({}^L R) = \xi(E({}^L R))$. The action of the 't Hooft operator on $\mathbf{m}_0$ is

$$\mathbf{m}_0 \longrightarrow \mathbf{m}_0 + \xi({}^L R). \tag{8.9}$$

This statement, which is the S-dual of (8.8), just comes from the behavior of the characteristic class under cobordism, as in Figure 11.

Wilson Operators And Supersymmetry Let us see what kind of supersymmetry the operation (8.7) preserves. For a generic choice of curve $\mathcal{S} \subset \Sigma \times C$, the supersymmetric Wilson line $W(R, \mathcal{S})$ preserves B-type supersymmetry in the complex structure J of $\mathcal{M}_H$ and nothing else. However, if we take $\mathcal{S} = \gamma \times p$, for γ a curve in Σ and p a point in C, then the Wilson operator preserves more supersymetry. In fact, as we showed at the end of section 6.4, it preserves supersymmetry of type (B, B, B), that is, it preserves B-type supersymmetry in each complex structure.

We can now see explicitly how this is reflected in the action of the Wilson operator on branes. For a point $p \in C$, the bundle $\mathcal{E}_p(R)$ is holomorphic

in each of the complex structures on $\mathcal{M}_H$. (This can be naturally proved using the hyper-Kahler quotient construction of $\mathcal{M}_H$.) So the operation (8.7) preserves B-type supersymmetry for each complex structure.

8.2. Zerobranes As Electric Eigenbranes

Now we can look for electric eigenbranes. Consider a Wilson line $W_q(R)$ in a representation R and at a point $q \in C$. It maps a brane $\mathcal{B}$ with Chan-Paton sheaf U to a brane with Chan-Paton sheaf $U \otimes \mathcal{E}_q(R)$. By definition, $\mathcal{B}$ is an electric eigenbrane if $U \otimes \mathcal{E}_q(R)$ is isomorphic to $U \otimes V$ for some fixed vector space V. The condition, therefore, is that $\mathcal{E}_q(R)$ equipped with the connection $\Phi_q^*(\widehat{\mathcal{A}})$ must be holomorphically trivial – isomorphic to a vector bundle with constant fiber V – when restricted to the support of U.

Let x be a point in $\mathcal{M}_H$, and let $\mathcal{B}_x$ be a zerobrane supported at x. In other words, $\mathcal{B}_x$ is a brane whose Chan-Paton sheaf is a skyscraper sheaf U_x supported at x. We will suppose that x is a smooth point in $\mathcal{M}_H$, so that one can integrate out the fields A_Σ, ϕ_Σ on the classical level.

Any vector bundle is trivial when restricted to a point, and in particular the bundle $\mathcal{E}_q(R)$ is trivial when restricted to $x \in \mathcal{M}_H$. Its restriction to x, which we denote as $\mathcal{E}_q(R)|_x$, is just a fixed vector space. So

$$U_x \otimes \mathcal{E}_q(R) = U_x \otimes \mathcal{E}_q(R)|_x. \tag{8.10}$$

Therefore, $\mathcal{B}_x$ is an electric eigenbrane, the "eigenvalue" of the Wilson loop operator $W_q(R)$ being the vector space $\mathcal{E}_q(R)|_x$.

Equivalently, the boundary conditions on the sigma-model fields corresponding to the zerobrane say that $\mathcal{A}_C$ is constant up to a gauge transformation on the boundary of Σ. By virtue of (8.3), this implies that $\mathcal{A}_\Sigma$ is trivial on the boundary of Σ.

We argued in section 6.4, based on the underlying four-dimensional topological field theory, that the "eigenvalue" of an electric or magnetic eigenbrane must vary with $q \in C$ as the fiber of a flat bundle over C. Indeed, as $q \in C$ varies, the space $\mathcal{E}_q(R)|_x$ does vary as the fiber of a flat vector bundle. In fact, taking complex structure J, the restriction $(\mathcal{E}, \widehat{\varphi})|_x$ of the universal bundle to $x \times C$ gives a flat $G_{\mathbb{C}}$ bundle $\mathcal{E}|_x \to C$ (since a solution of Hitchin's equations corresponds in complex structure J to a flat $G_{\mathbb{C}}$-bundle). By taking this bundle in the representation R, we get the desired flat bundle whose fiber at $q \in C$ is $\mathcal{E}_q(R)|_x$. We picked here complex structure J because it is singled out by the underlying topological supersymmetry.

In describing the zerobrane as an electric eigenbrane, we have been slightly informal on one key point: the role of the discrete B-field $\mathbf{e}_0(\zeta)$. A B-field is trivial when restricted to a point, so as long as one is considering zerobranes, one can informally ignore the B-field, as we have just done. However, a more precise way to describe things is as follows. For every $\mathbf{e}_0$, the category of branes contains an object that is a zerobrane supported at the (smooth) point $x \in \mathcal{M}_H$. Because the action of a Wilson line operator can change $\mathbf{e}_0$, the electric eigenbrane is really a sum of twisted zerobranes with different $\mathbf{e}_0$ (all of which are isomorphic, as the twisting is trivial when restricted to a point). Dually, since an 't Hooft operator can change $\mathbf{m}_0$, the magnetic eigenbranes will be sums of branes supported on different components of $\mathcal{M}_H$, for all possible $\mathbf{m}_0$. In standard approaches to the geometric Langlands program, one says that the Hecke eigensheaves are supported on the union of all topological components of $\mathcal{M}(G,C)$.

Relation To The Geometric Langlands Program Now we can get at least a glimmering of how all this is related to the geometric Langlands program.

In the geometric Langlands program, one begins with the group ${}^LG$ and a homomorphism $\vartheta : \pi_1(C) \to {}^LG_{\mathbb{C}}$. (One requires that this homomorphism be semistable in a sense explained in section 4.2.) The space of such homomorphisms is the Hitchin moduli space $\mathcal{M}_H({}^LG, C)$. So ϑ defines a point $x(\vartheta)$ in $\mathcal{M}_H({}^LG, C)$.

If ϑ is irreducible, then $x(\vartheta)$ is a smooth point and a zerobrane $\mathcal{B}_{x(\vartheta)}$ supported at $x(\vartheta)$ is an electric eigenbrane. S-duality applied to this electric eigenbrane will give a magnetic eigenbrane in the sigma-model of target $\mathcal{M}_H(G,C)$. From sections 5.4 and 5.5, we know that the S-dual of a zerobrane of $\mathcal{M}_H({}^LG, C)$ is a brane in $\mathcal{M}_H(G,C)$ whose support is a fiber $\boldsymbol{F}$ of the Hitchin fibration, endowed with a flat Chan-Paton bundle of rank 1. We refer to such a brane as a brane of type $\boldsymbol{F}$.

The main claim of the geometric Langlands program is that a homomorphism $\vartheta : \pi_1(C) \to {}^LG_{\mathbb{C}}$ is associated in a natural way to a sheaf on $\mathcal{M}(G,C)$ that is a Hecke eigensheaf and also a holonomic $\mathcal{D}$-module. Sections 9-11 of this paper will be devoted to explaining why a brane of type $\boldsymbol{F}$ on $\mathcal{M}_H(G,C)$ has the right properties. In sections 9 and 10, we relate the 't Hooft operators of quantum gauge theory to the Hecke operators of the geometric Langlands program, showing how our notion of a magnetic eigenbrane is related to the mathematical concept of a Hecke eigensheaf. In section 11, we argue that by virtue of the existence of a certain coisotropic A-brane on $\mathcal{M}_H(G,C)$,

the brane of type $\boldsymbol{F}$ is naturally associated to a module for the differential operators on $\mathcal{M}(G,C)$.

9. 't Hooft And Hecke Operators

The goal of the present section is to begin the study of the supersymmetric 't Hooft operators that appear at $\Psi = 0$, a value we reach by setting $t = 1$ and $\theta = 0$. The main result will be to show that 't Hooft operators correspond to the Hecke operators of the geometric Langlands program.

We will mainly consider static 't Hooft operators. So our four-manifold will be $M = \mathbb{R} \times W$, for some three-manifold W, and our 't Hooft line operators will be supported on one-manifolds of the form $\mathbb{R} \times p$, for some point $p \in W$. As we have discussed in section 6, line "operators" of this kind are not operators in the usual sense, acting on a pre-existing vector space; rather, they must be incorporated in the definition of the space of physical, supersymmetric states.

Our four-dimensional TQFT at $\Psi = 0$ reduces in two dimensions to the A-model with target $\mathcal{M}_H$ in complex structure K. So let us first recall some facts about A-models. In general, in a two-dimensional A-model with target X, a supersymmetric classical field configuration is a holomorphic map $\Phi : \Sigma \to X$. Moreover, the first approximation to the space of supersymmetric states is the cohomology of the space of time-independent supersymmetric classical fields on $\Sigma = \mathbb{R} \times I$, with I being an interval. Supersymmetric fields are holomorphic maps $\Phi : \Sigma \to X$. But a holomorphic map that is also time-independent is necessarily a constant. So in the A-model, the classical approximation to the space of physical states is simply the cohomology of the space of constant maps to X that obey the boundary conditions. For example, in Floer theory, one takes boundary conditions such that the two boundary components of $\mathbb{R} \times I$ are mapped to two specified Lagrangian submanifolds $N, N' \subset X$. A constant map with these boundary conditions must map Σ to the intersection $N \cap N'$, so the classical approximation to the space of supersymmetric states is simply the cohomology of $N \cap N'$. (One must shift the dimensions or ghost numbers of the cohomology classes in order to account for the number of filled fermion states in the vacuum. Mathematically, the shift involves the Maslov index.)

In general, there are instanton corrections to this classical intersection; their study is the main content of Floer theory. But we will see that for the branes we consider, there are no instanton corrections to the A-model at $\Psi = 0$; the supersymmetric configurations are all time-independent.

Reduction To Two Dimensions Aiming to reduce to a two-dimensional model on $\mathbb{R} \times I$, we specialize our four-manifold, which so far has been $M = \mathbb{R} \times W$, by taking $W = I \times C$, where C is a Riemann surface and I an interval at whose ends we take boundary conditions defined by choices of brane. So altogether, we work on $M = \mathbb{R} \times I \times C$. The conditions for supersymmetry are familiar from section 3.1:

$$\begin{aligned} (F - \phi \wedge \phi + tD\phi)^+ &= 0 \\ (F - \phi \wedge \phi - t^{-1}D\phi)^- &= 0 \\ D^*\phi &= 0. \end{aligned} \tag{9.1}$$

At $t = 1$, which we generally assume in our study of 't Hooft operators, the first two equations in (9.1) can be written $F - \phi \wedge \phi + \star D\phi = 0$. In terms of the complex connection $\mathcal{A} = A + i\phi$ and curvature $\mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A}$, this is equivalent to

$$\mathcal{F} + i \star \overline{\mathcal{F}} = 0. \tag{9.2}$$

In the absence of 't Hooft loops, and with suitable boundary conditions, vanishing theorems similar to those of section 3.3 show that solutions of these equations on $M = \mathbb{R} \times I \times C$ are pulled back from C and come from a constant map of $\Sigma = \mathbb{R} \times I$ to $\mathcal{M}_H$. This is part of the reduction of the four-dimensional TQFT to a two-dimensional sigma-model.

With 't Hooft lines included, things are more complicated (Figure 12). As we recall from section 6.2, 't Hooft operators create specified singularities in classical field configurations. For an 't Hooft operator supported on the line $\mathbb{R} \times p$, with some $p \in I \times C$, the singularity will prevent us from proving

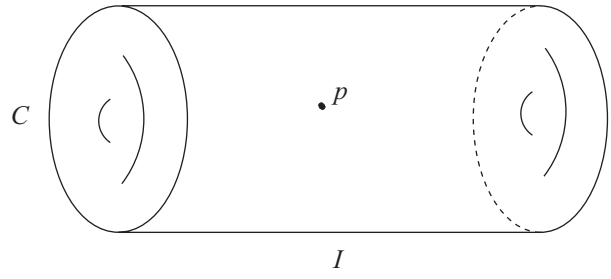

Figure 12: A static 't Hooft line inserted at a point $p = y \times p_0$ in $I \times C$. Near this point, the solution of the Bogomolny equation has a prescribed singularity.

that the solution is a pullback from C. But for certain boundary conditions, a weaker vanishing result still applies: the relevant solutions of (9.1) are time-independent, that is, invariant under translations of $\mathbb{R}$.

To explain this, recall that from the point of view of geometric Langlands duality, the most important A-branes are the branes of type $\boldsymbol{F}$, supported on a fiber of the Hitchin fibration with a flat Chan-Paton bundle of rank 1. They are S-duals of zerobranes and hence are expected to be magnetic eigenbranes. As explained in section 5.5, they are branes of type (B, A, A), i.e. the corresponding boundary conditions are compatible with two linearly independent topological supercharges, suitable linear combinations of which generate the topological supersymmetry of the B-model in complex structure I and the A-models in complex structures J and K.

When we consider BRST-invariant field configurations with (B, A, A) boundary conditions and an insertion of the 't Hooft operator, we can distinguish between configurations which preserve both fermionic symmetries and those which preserve only the A-type supersymmetry in complex structure K. In the latter case, the broken supersymmetry generates an extra fermionic zero mode that is not lifted by quantum corrections, since it is generated by a symmetry, and will prevent such instantons from contributing to any topological observables. (There may also be other fermionic zero modes whose existence does not follow from the fermionic symmetry.) Hence, contributions of interest come only from instantons that preserve both supersymmetries.

In fact, instantons with the requisite properties are time-independent. One would guess this from the sigma-model, since the B-model in complex structure I has no non-trivial instantons. The result is also true in the context of the full twisted supersymmetric gauge theory on $\mathbb{R} \times I \times C$, even in the presence of static 't Hooft operators. The most general proof of this involves a moment map argument and will appear elsewhere [26]. Here we sketch another argument that is adequate for the most important (B, A, A)-branes, such as branes of type $\boldsymbol{F}$.

First we need a vanishing theorem for ϕ_1, the component of ϕ in the I direction of $M = \mathbb{R} \times I \times C$. In any classical instanton solution, ϕ_1 will obey its second order Euler-Lagrange equations, which can be deduced from the classical action (3.46) and read

$$-D_\mu D^\mu \phi_1 + [\phi_\mu, [\phi_1, \phi^\mu]] = 0. \tag{9.3}$$

We suppose that ϕ_1 obeys Dirichlet or Neumann boundary conditions on each brane; this is true for the most important branes. Multiplying the

equation by ϕ_1, taking the trace and integrating over space-time, we get

$$\int d^4x \ \mathrm{Tr}\,(D_\mu\phi_1 D^\mu\phi_1 + [\phi_1,\phi_\mu][\phi_1,\phi^\mu]) = 0. \tag{9.4}$$

Integration by parts produces no boundary terms because of the assumed boundary conditions, and there is no problem with the singularities associated with the 't Hooft operators because ϕ_1 is nonsingular near a static 't Hooft operator. Since the integrand is negative-definite, the above identity implies that $D_\mu\phi_1 = [\phi_\mu,\phi_1] = 0$, so that $\phi_1 = 0$ in the case of an irreducible solution.

In this argument, we tacitly assumed that the integration by parts gives no boundary term in the far past or future. We do not know *a priori* that a supersymmetric solution of the equations is time-independent. However, we at least assume that the desired solutions approach time-independent "vacuum" solutions in the far past or the far future. By first running the argument of the last paragraph for the time-independent case, we learn that $\phi_1 = 0$ in each vacuum. It then follows, since W is compact, that $\phi \to 0$ exponentially fast in the far past or the far future. Hence, there is no boundary term in the integration by parts used in the last paragraph to show that $\phi_1 = 0$ everywhere.

Now we want to show that a field that possesses the topological supersymmetry of the A-model in both complex structures J and K and also has $\phi_1 = 0$ is actually time-independent. The supersymmetric equations of the A-model in complex structure K are the equations (9.1) at $t = 1$; those in the A-model in complex structure J are obtained from these by making a $\mathcal{U}_1$ transformation $\varphi \to i\varphi$. It is a short exercise to write out the combined system of equations on $M = \mathbb{R}\times I\times C$ and show that, if in addition $\phi_1 = 0$, they imply that the fields are time-independent up to a gauge transformation.

The time-independent supersymmetric fields can usefully be analyzed in the more general context of $M = \mathbb{R}\times W$ with static Wilson lines (without specializing to $W = I\times C$), even though our argument that all relevant supersymmetric fields are time-independent does not hold in this generality. When we specialize to $M = \mathbb{R}\times W$, there is a further symmetry $\mathcal{T}$ of time-reversal invariance, acting as $s\to -s$ where s is a "time" coordinate on $\mathbb{R}$; $\mathcal{T}$ acts trivially on W and maps $(A,\phi)\to(\mathcal{T}^*A, -\mathcal{T}^*\phi)$. (This is an example of the orientation-reversing symmetries discussed at the end of section 3.4.) The sign change of ϕ is needed because $\mathcal{T}$ reverses the sign of the Hodge $\star$ operator in (9.2).

Therefore, we can consistently restrict ourselves to solutions that are both time-independent and $\mathcal{T}$-invariant. In such a solution, the G-bundle E and connection A are pullbacks from W. Also, ϕ takes the form $\phi_0\, ds$, where ϕ_0 is an $\mathrm{ad}(E)$-valued zero-form on W. The equations reduce to

$$F = \star D\phi_0. \tag{9.5}$$

Our orientation conventions are such that in four dimensions, $\star(dx^0 \wedge dx^1) = dx^2 \wedge dx^3$, and in three dimensions, $\star(dx^1) = dx^2 \wedge dx^3$. The equations (9.5) are the standard Bogomolny equations [105] for supersymmetric or BPS monopoles[38] and have been widely studied, for instance in [107–109]. A more familiar way to obtain the Bogomolny equations is to look for solutions of the Yang-Mills instanton equation on $\mathbb{R}^4$ that are invariant under translations in one direction.

In this section, we will first analyze static 't Hooft operators in the context of the standard Bogomolny equations (9.5). We will show that they implement Hecke transformations of G-bundles, as usually defined in the geometric Langlands program. This will involve studying BPS monopole solutions with point singularities due to the 't Hooft operators. Such singular solutions of the Bogomolny equations have been studied first in unpublished work by Kronheimer [110]. They can arise as limits of smooth monopole solutions for larger gauge groups [111–114], and from certain string theory brane configurations [115], which have motivated more recent study in [119,120]. Then we will provide an elementary introduction to Hecke modifications of G-bundles – claiming neither novelty nor completeness from a mathematical point of view!

In section 10.5, we will generalize the analysis to relax the assumption of time-reversal symmetry. As we will see, in this case the supersymmetric equations give a sort of complexified or extended analog of the Bogomolny equations, which does not appear to have been studied before. In the context of these extended equations, static 't Hooft operators act by Hecke transformations of Higgs bundles or in other words of Hitchin pairs (E, φ). Hecke operators in this extended sense are natural mathematically and have been considered in [119,120].

Applications to the geometric Langlands program really depend upon the extended Bogomolny equations and Hecke operators. We can reduce to ordinary Bogomolny equations and Hecke operators only when the boundary

[38] The name reflects the fact that the basic supersymmetric monopole solution was also found in [106] as a solution of the second order Yang-Mills-Higgs equations.

conditions (the choices of branes) ensure that the Higgs field φ vanishes. This is not the case for typical applications. However, we begin with ordinary Bogomolny equations and Hecke operators and devote most of our attention to them because this case is simpler and exhibits the main ideas.

9.1. 't Hooft Operators And Hecke Modifications

To begin with, we consider the Bogomolny equations on $W = \mathbb{R}^3$, which we regard as $\mathbb{R} \times \mathbb{C}$, where $\mathbb{C}$ is parametrized by $z = x^2 + ix^3$, and $\mathbb{R}$ by $y = x^1$. We have

$$\begin{aligned} \star dy &= \frac{i}{2} dz \wedge d\overline{z} \\ \star dz &= -i dz \wedge dy \\ \star d\overline{z} &= i d\overline{z} \wedge dy. \end{aligned} \tag{9.6}$$

Upon expanding $F = dz \wedge d\overline{z}\, F_{z\overline{z}} + dy \wedge dz F_{yz} + dy \wedge d\overline{z} F_{y\overline{z}}$, and similarly $D\phi_0 = dy D_y \phi_0 + dz D_z \phi_0 + d\overline{z} D_{\overline{z}} \phi_0$, the Bogomolny equations become

$$\begin{aligned} F_{z\overline{z}} &= \frac{i}{2} D_y \phi_0 \\ F_{yz} &= i D_z \phi_0 \\ F_{y\overline{z}} &= -i D_{\overline{z}} \phi_0. \end{aligned} \tag{9.7}$$

More generally, take $W = \mathbb{R} \times C$, with a Riemann surface C. We write z for a local complex coordinate on C, endow C with a Kahler metric $h(z, \overline{z})|dz|^2$, for some positive function h, and take the metric on W to be $h(z, \overline{z})|dz|^2 + dy^2$. The first equations in (9.6) and (9.7) become respectively

$$\begin{aligned} \star dy &= \frac{ih}{2}\, dz \wedge d\overline{z} \\ F_{z\overline{z}} &= \frac{ih}{2} D_y \phi_0. \end{aligned} \tag{9.8}$$

The other equations are unchanged.

There is no integrability condition for $\overline{\partial}$ operators in complex dimension 1. So any connection A on a G-bundle $E \to C$ endows E with a holomorphic structure. We simply define the $\overline{\partial}$ operator as $\overline{D} = d\overline{z} D_{\overline{z}} = d\overline{z}(\partial_{\overline{z}} + A_{\overline{z}})$.

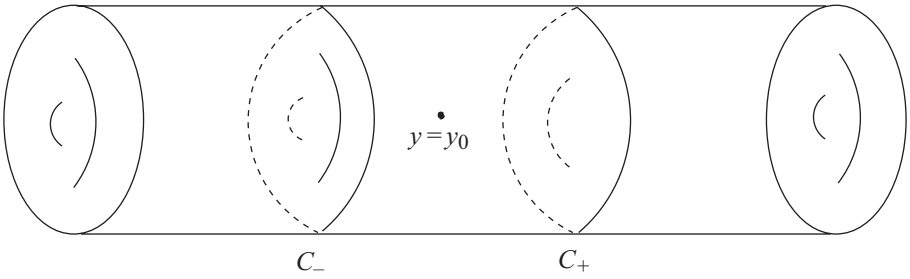

Figure 13: In a solution of the Bogomolny equations on a G-bundle E, the holomorphic type of E_y – the restriction of E to $\{y\} \times C$ – is constant except when one crosses the position of an 't Hooft operator. Sketched are an 't Hooft operator at $y = y_0$ and copies of $C_y = \{y\} \times C$ to the left and right of y_0. They are denoted C_- and C_+.

Thus, as y varies in $\mathbb{R}$, any connection A on a G-bundle $E \to \mathbb{R} \times C$ defines a family of holomorphic G-bundles $E_y \to C$. E_y is simply the restriction of E to $C_y = \{y\} \times C$ (Figure 13). Now suppose that A obeys the Bogomolny equations. Let us work in the gauge $A_y = 0$. Then $F_{y\overline{z}} = \partial_y A_{\overline{z}}$. Hence the last equation in (9.7) tells us that

$$\frac{\partial}{\partial y} A_{\overline{z}} = -iD_{\overline{z}}\phi_0. \tag{9.9}$$

But the right hand side is the change in $A_{\overline{z}}$ under a gauge transformation generated by $-i\phi_0$. Hence the holomorphic type of E_y is independent of y. (Another way to state this argument is to note that $\partial_y \overline{D} = -i[\overline{D}, \phi_0]$, showing that $\overline{D}$ is independent of y up to conjugation.)

Now let us see what happens if we incorporate an 't Hooft operator at a point $p = y_0 \times p_0 \in \mathbb{R} \times C$. This means that we consider a bundle E and connection A that are defined on the complement of the point p, and which have a certain type of singularity near p, as described in section 6.2.

Away from $y = y_0$, the above argument applies and the holomorphic type of E_y is independent of y. At $y = y_0$, because of the singularity, the bundle E_y is not defined. In crossing $y = y_0$, the holomorphic type of the bundle may jump. However, it can only jump in a very special way. If we restrict E to $C \backslash p_0$ (that is, the complement of p_0 in C), then the above argument applies and the restricted bundle E_y has a holomorphic type that is constant even at $y = y_0$. So the holomorphic type of E_y may jump at $y = y_0$, but only in a way that is trivial if we omit the point p_0 from C.

We will now try to describe precisely how E_y can jump at $y = y_0$. We consider first the abelian case, and then the nonabelian case.

The Abelian Case We begin with $G = U(1)$. In this case, E is a complex line bundle $\mathcal{L}$. We write $\mathcal{L}_-$ for $\mathcal{L}_y$ with $y < y_0$ and $\mathcal{L}_+$ for $\mathcal{L}_y$ with $y > y_0$. The line bundles $\mathcal{L}_-$ and $\mathcal{L}_+$ are isomorphic on the complement of the point $p_0 \in C$.

This means that the line bundle $\mathcal{L}_+ \otimes \mathcal{L}_-^{-1}$ is trivial away from p_0. It hence has a section s that has neither a zero nor a pole away from p_0; let q be the order of its zero at p_0. Then $\mathcal{L}_+ \otimes \mathcal{L}_-^{-1} \cong \mathcal{O}(p_0)^q$. Here as usual $\mathcal{O}(p_0)$ is the line bundle whose holomorphic sections are functions holomorphic away from p_0 with a possible single pole at p_0. So the relation between $\mathcal{L}_+$ and $\mathcal{L}_-$ is $\mathcal{L}_+ \cong \mathcal{L}_- \otimes \mathcal{O}(p_0)^q$, for some integer q.

We recall that an 't Hooft operator $T(m)$ for the group $G = U(1)$ is classified by the choice of an integer m. (We write the operator as $T(m; p_0)$ if we want to specify the point p_0.) This suggests an obvious hypothesis – that m may coincide with the integer q of the last paragraph.

The operator $T(m)$ is defined by saying that near $p = p_0 \times y_0$, the curvature has the singular behavior

$$F \sim \star d\left(\frac{im}{2}\frac{1}{|\vec{x} - p|}\right), \tag{9.10}$$

where $|\vec{x} - p|$ is the distance from p to a nearby point $\vec{x} \in \mathbb{R} \times C$. This implies that if S is a small sphere enclosing the point p, as in the figure, then $\int_S c_1(\mathcal{L}) = m$. The 't Hooft operators $T(m)$ are thus all topologically distinct from one another. This statement, which is far from being valid if G is nonabelian, will enable us in the abelian case to determine the action of $T(m)$ just on topological grounds.

Now pick $y_\pm$ with $y_- < y_0 < y_+$, and write $C_\pm$ for $y_\pm \times C$, so that $\mathcal{L}_\pm$ is the restriction of $\mathcal{L}$ to $C_\pm$. Let $q_\pm = \int_{C_\pm} c_1(\mathcal{L})$. With an obvious choice of orientations, as in Figure 14, the homology cycle $D = C_+ - C_- - S$ is a boundary in $(\mathbb{R} \times C)\backslash p$. Hence $0 = \int_D c_1(\mathcal{L}) = \int_{C_+} c_1(\mathcal{L}) - \int_{C_-} c_1(\mathcal{L}) - \int_S c_1(\mathcal{L}) = q_+ - m - q_-$.

So $q_+ = m + q_-$. Consequently, $\mathcal{L}_+$ and $\mathcal{L}_-$, which are isomorphic away from p_0, differ in first Chern class by m. Hence $\mathcal{L}_+ \cong \mathcal{L}_- \otimes \mathcal{O}(p_0)^m$.

Thus, we have determined the action of the 't Hooft operators for $G = U(1)$. The 't Hooft operator $T(m; p_0)$, that is the operator $T(m)$ inserted at the point $p_0 \in C$, acts by twisting with $\mathcal{O}(p_0)^m$. This result agrees with the standard definition of the Hecke operators for $U(1)$.

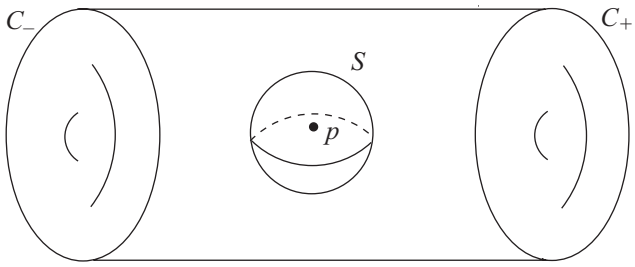

Figure 14: A cobordism between the two-cycles $C_- + S$ and C_+, showing that the homology cycle $D = C_+ - C_- - S$ is a boundary.

We see that, in accord with general arguments in section 6.4, 't Hooft operators inserted at distinct points in C commute. Moreover, the 't Hooft operators at a given point $p_0 \in C$ form a commutative group, with $T(m; p_0) = T(1; p_0)^m$.

The Nonabelian Case: $U(N)$ Now let us discuss what 't Hooft operators do in the nonabelian case. We first consider the group $U(N)$. There are two things that make this group simple to analyze. $U(N)$ has a convenient representation, the N-dimensional representation that we will call V. And $U(N)$ is its own Langlands dual.

The definition of an 't Hooft operator depends on a choice of homomorphism $\rho : U(1) \to U(N)$, up to conjugation. The most general such homomorphism maps $\exp(i\alpha) \in U(1)$ to the diagonal matrix $\mathrm{diag}(\exp(im_1\alpha), \exp(im_2\alpha), \dots, \exp(im_N\alpha))$, with an N-plet of integers ${}^L w = (m_1, m_2, \dots, m_N)$ that are unique up to permutation. (We call this quantity ${}^L w$ because in the general case, it will be a weight of the dual group ${}^L G$.) We will usually order the m_i so $m_1 \geq m_2 \geq \dots \geq m_N$. At the Lie algebra level, ρ maps $1 \in \mathfrak{u}(1)$ to the diagonal matrix

$$\begin{pmatrix} m_1 & & & \dots & \\ & m_2 & & \dots & \\ & & m_3 & \dots & \\ & & \vdots & \dots & \\ & & & & m_N \end{pmatrix}. \tag{9.11}$$

The corresponding 't Hooft operator $T({}^L w)$ is defined by saying that near the point $p \in \mathbb{R} \times C$, the gauge field has a singularity obtained by embedding the basic $U(1)$ singularity in $U(N)$, via this embedding. For example, the

singular part of the curvature near p is diagonal

$$F \sim \star d\left(\frac{i}{2}\frac{1}{|\vec{x}-p|}\right)\begin{pmatrix} m_1 & & & \dots & \\ & m_2 & & \dots & \\ & & m_3 & \dots & \\ & & \vdots & \dots & \\ & & & & m_N \end{pmatrix}. \tag{9.12}$$

Near p, the Bogomolny equations reduce to equations in some maximal torus $\mathcal{T} = U(1)^N$ of $U(N)$. (From a holomorphic point of view, it is more natural to consider the corresponding reduction to a complexified maximal torus in the complexification of the gauge group, namely $U(N)_{\mathbb{C}} = GL(N,\mathbb{C})$.) Corresponding to the reduction of the equations to a maximal torus, the bundle E_y splits up, near $p_0 \times y \in C_y$, as a sum $\mathcal{L}_1 \oplus \mathcal{L}_2 \oplus \dots \oplus \mathcal{L}_N$ of line bundles (each of which, of course, is trivial near p). The effect of the 't Hooft operator on $\mathcal{L}_i$ is precisely what it was in the abelian case, namely $\mathcal{L}_i \to \mathcal{L}_i \otimes \mathcal{O}(p_0)^{m_i}$.

Thus, we have arrived at a description of how the 't Hooft operator $T({}^L w)$ acts on a bundle E. Relative to some decomposition of E as $\oplus_{i=1}^N \mathcal{L}_i$ near p, it acts by $\mathcal{L}_i \to \mathcal{L}_i \otimes \mathcal{O}(p_0)^{m_i}$. This coincides with the usual mathematical description of the action of the Hecke operators for $U(N)$. This is the basic link between 't Hooft and Hecke operators.

In sections 9.2 and 9.3, we will address the question of what sort of data is contained in the local decomposition of E as $\oplus_{i=1}^N \mathcal{L}_i$, and how much of this data is relevant to the action of the 't Hooft operators. The general answer to this question involves something called the affine Grassmannian, which we briefly describe in section 9.3.

Some Examples In general, as we know from section 6.1, representations of the Langlands dual group ${}^L G$ correspond naturally to 't Hooft operators of G. A representation $R({}^L w)$ with highest weight ${}^L w$ corresponds to an 't Hooft operator that we write as $T(R({}^L w))$ or simply as $T({}^L w)$. In the present discussion, ${}^L G = G = U(N)$.

For example, the fundamental N-dimensional representation V of $U(N)$ has, with our convention $m_1 \geq m_2 \geq \dots \geq m_N$, the highest weight ${}^L w(1) = (1,0,0,\dots,0)$. It corresponds to an 't Hooft operator $T({}^L w(1))$ or simply $T_{(1)}$. The k^{th} antisymmetric tensor representation $\wedge^k \mathsf{V}$, for $k = 1,\dots,N$, has highest weight

$${}^L w(k) = (\overbrace{1,1,\dots,1}^{k \text{ times}},0,\dots,0) \tag{9.13}$$

and corresponds to an 't Hooft operator $T_{(k)}$.

The representation $\wedge^N \mathbf{V}$ is one-dimensional; it is the representation in which $g \in U(N)$ acts by multiplication by $\det(g)$. It can be raised to any integer power u, positive or negative, to get a representation in which g acts by $(\det(g))^u$. This representation, which we denote as $(\wedge^N \mathbf{V})^u$, corresponds to the weight ${}^L w = (u, u, \dots, u)$. The 't Hooft/Hecke operator $T_{(N)}$ corresponding to $\wedge^N \mathbf{V}$ simply acts by $E \to E \otimes \mathcal{O}(p_0)$, as in the abelian case, and is obviously invertible. Its u^{th} power corresponds to the representation $(\wedge^N \mathbf{V})^u$. The $T_{(k)}$ with $k < N$ are not invertible.

The central element $\exp(i\alpha)$ of $U(N)$ acts on the representation $\wedge^k \mathbf{V}$ as multiplication by $\exp(ik\alpha)$. More generally, it acts on any representation of highest weight ${}^L w = (m_1, m_2, \dots, m_N)$ as multiplication by $\exp(i\alpha \sum_{i=1}^N m_i)$. A special case of this statement is that the center of $U(N)$ acts trivially if and only if

$$\sum_i m_i = 0. \tag{9.14}$$

S-duality, as we discussed in section 8, maps the action of the center of ${}^L G$ on a representation ${}^L R$ to the action of the corresponding 't Hooft operator $T({}^L R)$ on the topology of a G-bundle. Indeed, since $T({}^L w)$ locally maps $\oplus_i \mathcal{L}_i$ to $\oplus_i (\mathcal{L}_i \otimes \mathcal{O}(p_0)^{m_i})$, it changes $c_1(E)$ by

$$c_1(E) \longrightarrow c_1(E) + \sum_i m_i, \tag{9.15}$$

and thus the topology of E is unchanged if and only if $\sum_i m_i = 0$.

From S-duality, we expect 't Hooft operators of G to form a commutative algebra isomorphic to the representation ring of ${}^L G$. (In fact, as we discussed in section 6.4, commutativity of 't Hooft operators follows from more general considerations of four-dimensional topological field theory, without resort to S-duality.) We will examine this in section 10.3 from the standpoint of the Bogomolny equations. The representation ring of $U(N)$ is freely generated by the representations $\wedge^k \mathbf{V}$, $k = 1, \dots, N$ along with $(\wedge^N \mathbf{V})^{-1}$. Accordingly the commutative algebra of 't Hooft operators for $U(N)$ is freely generated by the $T_{(k)}$ together with $T_{(N)}^{-1}$. In the mathematical literature on the geometric Langlands program, these generators are usually taken as the basic Hecke operations of $U(N)$.

't Hooft/Hecke Operators For Other Groups The generalization of this to any compact gauge group G is rather direct. To keep things relatively simple, we will pick a representation R of G and and consider a G-bundle

E_R which we take in the representation R. (Thus, if E denotes a principal G-bundle, then $E_R = E \times_G R$.) Picking a maximal torus $\mathcal{T}$ of G, let w denote a weight of G, that is a character $w : \mathcal{T} \to U(1)$. Let $R = \oplus_w R_w$ be the decomposition of R in weight spaces R_w (all but finitely many of which vanish).

An 't Hooft operator of G is classified by a dominant coweight ${}^L w : U(1) \to \mathcal{T}$, which is the highest weight of a representation ${}^L R$ of ${}^L G$. The composition of $w : \mathcal{T} \to U(1)$ and ${}^L w : U(1) \to \mathcal{T}$ is a homomorphism $w \circ {}^L w : U(1) \to U(1)$ which takes the form $\exp(i\alpha) \to \exp(i\langle {}^L w, w\rangle \alpha)$ for some integer $\langle {}^L w, w\rangle$.

For any choice of a complexified maximal torus $\mathcal{T}_{\mathbb{C}}$ of $G_{\mathbb{C}}$, the fiber of E_R at $p_0 \in C$ has a decomposition $E_R|_{p_0} = \oplus_w E_{R,w}|_{p_0}$ in weight spaces. If we extend $\mathcal{T}$ to a holomorphically varying family of maximal tori near p_0, then we get a corresponding local decomposition of E_R in subbundles of definite weight: $E_R = \oplus_w E_{R,w}$. By solving the Bogomolny equations near p_0, and taking account of the abelian nature of the singularity, we find that relative to some choice of such a decomposition, $T({}^L w)$ acts by $E_{R,w} \to E_{R,w} \otimes \mathcal{O}(p_0)^{\langle {}^L w, w\rangle}$.

We will make this more concrete for the case $G = SU(N)$, ${}^L G = PSU(N) = SU(N)/\mathbb{Z}_N$, and the reciprocal case $G = PSU(N)$, ${}^L G = SU(N)$.

A homomorphism $\rho : U(1) \to G = SU(N)$ takes the same form as for $U(N)$, namely $\exp(i\alpha) \to \mathrm{diag}(\exp(im_1\alpha), \exp(im_2\alpha), \ldots, \exp(im_N\alpha))$, except that now we must require $\sum_i m_i = 0$, as in (9.14). The action of the corresponding 't Hooft operator $T({}^L w)$ on an $SU(N)$ bundle E is as before: relative to some decomposition $E = \oplus_{i=1}^N \mathcal{L}_i$ near p, we have $\mathcal{L}_i \to \mathcal{L}_i \otimes \mathcal{O}(p_0)^{m_i}$.

With our usual ordering $m_1 \geq m_2 \geq \ldots \geq m_N$, the N-plet ${}^L w = (m_1, m_2, \ldots, m_N)$ is a positive integer combination of the simple roots $e_1 = (1, -1, 0, \ldots, 0)$, $e_2 = (0, 1, -1, 0, \ldots, 0), \ldots, e_{N-1} = (0, 0, \ldots, 0, 1, -1)$ of $SU(N)$. Hence, it is the highest weight of a representation of the adjoint group $PSU(N)$, which of course coincides with ${}^L G$ in this case.

Now let us discuss the case that $G = PSU(N)$, and therefore ${}^L G = SU(N)$. A homomorphism $\rho : U(1) \to PSU(N)$ can be lifted to a homomorphism from an N-fold cover of $U(1)$ to $SU(N)$. Such a homomorphism takes the form $\exp(i\alpha) \to \mathrm{diag}(\exp(im_1\alpha), \exp(im_2\alpha), \ldots, \exp(im_N\alpha))$, but the m_i are not integers; rather, they take values in $\frac{1}{N}\mathbb{Z}$, with the differences $m_i - m_j$ being integers. Such an N-plet ${}^L w = (m_1, m_2, \ldots, m_N)$, ordered so that $m_1 \geq m_2 \geq \ldots \geq m_N$, is a dominant weight of ${}^L G = SU(N)$. It corresponds to a representation ${}^L R$ of $SU(N)$ and an 't Hooft operator $T({}^L R)$

of $G = PSU(N)$. The representations ${}^LR$ on which the center of $SU(N)$ acts nontrivially, and hence those for which $T({}^LR)$ changes the topology of a $PSU(N)$-bundle, are precisely those for which the m_i are non-integral.

Let E be a principal $PSU(N)$ bundle. As $PSU(N)$ does not have an N-dimensional representation corresponding to the N-dimensional representation V of $SU(N)$, we cannot derive from E a rank N complex vector bundle E_{V}. However, E_{V} does exist as a twisted vector bundle, a notion that we explained briefly in section 7.2. The obstruction to lifting E_{V} to an ordinary vector bundle is the characteristic class $\xi(E)$ introduced in section 7.1. After picking a family of maximal tori, E_{V} has a local decomposition $E_{\mathsf{V}} = \oplus_{i=1}^N \mathcal{L}_i$ in terms of twisted line bundles $\mathcal{L}_i$ (all of which are twisted in the same way, since the obstruction to lifting E_{V} to a vector bundle is central).

The 't Hooft/Hecke operators act, relative to some such decomposition, in the familiar way $\mathcal{L}_i \to \mathcal{L}_i \otimes \mathcal{O}(p_0)^{m_i}$. Now, however, the m_i may not be integers. We define $\mathcal{O}(p_0)^{1/N}$ as a twisted line bundle, and we define $\mathcal{O}(p_0)^{m/N}$, for any integer m, as the m^{th} power of $\mathcal{O}(p_0)^{1/N}$. In particular, the N^{th} power of $\mathcal{O}(p_0)^{1/N}$ is the ordinary line bundle $\mathcal{O}(p_0)$. Since the differences $m_i - m_j$ are integers, the twisted line bundles $\mathcal{O}(p_o)^{m_i}$ are all twisted in the same way, and the action of the 't Hooft/Hecke operators by $\mathcal{L}_i \to \mathcal{L}_i \otimes \mathcal{O}(p_0)^{m_i}$ preserves the fact that the $\mathcal{L}_i$ all have the same twist.

Of course, we can avoid talking about twisted bundles by picking a representation R of $PSU(N)$, and applying the 't Hooft operators to the ordinary vector bundle $E_R = E \times_{PSU(N)} R$. For example, let $N = 2$, so that $PSU(2)$ becomes $SO(3)$. The rank two twisted bundle E_{V} has the local decomposition $\mathcal{L} \oplus \mathcal{L}^{-1}$ in terms of a twisted line bundle $\mathcal{L}$. If we take R to be the adjoint representation of $SO(3)$, then the corresponding adjoint bundle E_R is $\mathrm{Sym}^2(E_{\mathsf{V}})$ (the symmetric part of $E_{\mathsf{V}} \otimes E_{\mathsf{V}}$) or $\mathcal{L}^2 \oplus \mathcal{O} \oplus \mathcal{L}^{-2}$, where $\mathcal{L}^2$ is an ordinary line bundle. The transformation $\mathcal{L} \to \mathcal{L} \otimes \mathcal{O}(p_0)^{1/2}$ acts on E_R by $E_R \to (\mathcal{L}^2 \otimes \mathcal{O}(p_0)) \oplus \mathcal{O} \oplus (\mathcal{L}^{-2} \otimes \mathcal{O}(p_0)^{-1})$. This transformation involves no fractional exponents.

9.2. The Space Of Hecke Modifications

The rest of this section is devoted to a closer look at the holomorphic data in this problem.

We will first give a direct description of the holomorphic data, emphasizing simple examples, and then explain in section 9.3 the more powerful framework in which the space of Hecke modifications is usually placed in the mathematical literature.

The considerations will be local, so to keep things simple we choose $C = \mathbb{CP}^1$, which we think of as the complex z-plane plus a point at infinity. We take the point p_0 to be $z = 0$. To start with, we take the gauge group to be $G = U(N)$. We let E_- be a trivial bundle of rank n. (Since the considerations will be local, it does not matter what E_- we start with.) We want to discuss how an 't Hooft/Hecke operator can modify E_- to make a new bundle E_+. We write a local holomorphic section of E_- (that is, a section holomorphic near p_0), as $s = (f_1, f_2, \dots, f_N)$, where the f_i are local holomorphic functions.

If s is any holomorphic section of E_- near p_0, we can define a line bundle $\mathcal{L}$ near p_0, as follows. $\mathcal{L}$ is a subbundle of E_-, defined, over a neighborhood of p_0, by saying that sections of $\mathcal{L}$ are sections of E_- of the form gs for an arbitrary holomorphic function g. We call $\mathcal{L}$ the line bundle generated by s. Now, suppose that $s_1, \dots, s_N$ are N local holomorphic sections of E_- whose restrictions to $z = 0$ are linearly independent. (If we write $s_i(z) = (f_{1i}(z), f_{2i}(z), \dots, f_{Ni}(z))$, the condition is that the $N \times N$ matrix $f_{ji}(0)$ has nonzero determinant.) For each i, we let $\mathcal{L}_i$ be the line bundle generated by s_i. The linear independence ensures that near $z = 0$ we have a decomposition $E_- = \oplus_{i=1}^N \mathcal{L}_i$, which is equivalent to saying that any local holomorphic section s of E_- is of the form $s = \sum_i g_i s_i$ with some local holomorphic functions g_i. In this situation, we say that the sections s_i generate E_-.

Conversely, any such local decomposition of E_- takes this form. If $E_- = \oplus_{i=1}^N \mathcal{L}_i$, we let s_i be any section of $\mathcal{L}_i$ that is nonzero at $z = 0$. Then near $z = 0$, $\mathcal{L}_i$ is the line bundle generated by s_i, and so the decomposition of E_- takes the form described in the last paragraph.

Now we consider the action of the 't Hooft/Hecke operator $T({}^L w)$ with ${}^L w = (m_1, m_2, \dots, m_N)$. With our chosen decomposition of E, it maps E_- to the bundle that near $z = 0$ is $E_+ = \oplus_i (\mathcal{L}_i \otimes \mathcal{O}(p_0)^{m_i})$. Away from $z = 0$, E_+ is just the same as E_-.

But $\mathcal{L}_i \otimes \mathcal{O}(p_0)^{m_i}$ is the line bundle whose general section, near $z = 0$, is s_i/z^{m_i}. Thus, near $z = 0$ a general section of E_+ takes the form $\sum_{i=1}^N g_i(z) s_i(z) z^{-m_i}$, where the functions g_i are holomorphic near $z = 0$. Away from $z = 0$, a holomorphic section of E_+ is the same as a holomorphic section of E_-: it takes the form $(f_1, \dots, f_N)$ with any holomorphic functions $f_1, \dots, f_N$.

Different decompositions may lead to the same E_+. This is precisely what we want to investigate; we want to investigate the space of all bundles E_+ to which E_- can be mapped by a given 't Hooft operator. In formulating this question, we must remember that any Hecke modification E_+ of E_- is naturally isomorphic to E_- away from $z = 0$, so if E_+ and $\widetilde{E}_+$ are two

Hecke modifications of E_-, they are naturally isomorphic to each other away from $z = 0$. We consider E_+ and $\widetilde{E}_+$ equivalent if and only if the natural isomorphism away from $z = 0$ can be extended over $z = 0$. We want to classify the possible E_+'s up to this equivalence. What this means in practice will become clear as we consider examples.

We begin with a few simple examples with $N = 2$. First we take ${}^Lw = (1, 0)$. We suppose that $s_1(z) = (e_1(z), f_1(z))$ and $s_2(z) = (e_2(z), f_2(z))$, with $(e_1(0), f_1(0))$ and $(e_2(0), f_2(0))$ linearly independent. E_+ is generated by $s_1' = z^{-1}s_1$ together with s_2. In particular, $s_1 = zs_1'$ is a section of E_+, as of course is s_2. Since s_1 and s_2 generate E_-, this means that any section of E_- is a section of E_+. But a section of E_+ may have a polar part proportional to $z^{-1}s_1(0)$.

In sum, in this example, any section of E_+ takes the form

$$s(z) = \frac{c}{z}s_1(0) + (h_1(z), h_2(z)) \tag{9.16}$$

near $z = 0$, where c is a complex constant and h_1, h_2 are local holomorphic functions. Clearly, E_+ depends on the choice of $s_1(0) = (e_1(0), f_1(0))$ only up to complex scaling. So we should think of $(e_1(0), f_1(0))$ as defining a point in $\mathbb{CP}^1$, and this $\mathbb{CP}^1$ parametrizes all bundles E_+ that can be constructed from E_- by acting with $T({}^Lw)$ for ${}^Lw = (1, 0)$.

We will say that $\mathbb{CP}^1$ is the space of all Hecke modifications of type ${}^Lw = (1, 0)$. We denote this space as $\mathcal{Y}({}^Lw)$, or $\mathcal{Y}({}^Lw, p_0)$ if we want to specify the point, in this case p_0, at which the Hecke modification is made. We also denote the same space as $\mathcal{Y}({}^LR)$, where ${}^LR$ is the representation of ${}^LG$ with highest weight ${}^Lw$, and refer to it alternatively as the space of Hecke modifications of type ${}^LR$.

In this particular example, the space of Hecke modifications is compact and smooth. But that is not typical. For another example, take ${}^Lw = (2, 0)$. The Hecke operator $T({}^Lw)$, acting on E_- with the same decomposition as before, now produces a bundle E_+ that is generated by $s_1' = z^{-2}s_1$ and s_2. Since $s_1 = z^2s_1'$ is a section of E_+, along with s_2, it is still true that any section (h_1, h_2) of E_- defines a section of E_+. In addition, E_+ has a two-dimensional space of polar sections, generated by $z^{-2}s_1$ and $z^{-1}s_1$. Let us expand $e_1(z) = u + zu' + \mathcal{O}(z^2)$, $f_1(z) = v + zv' + \mathcal{O}(z^2)$, with $u, u', v, v' \in \mathbb{C}$. Then the general local holomorphic section of E_+ takes the form

$$s(z) = \frac{c}{z^2}(u + zu', v + zv') + \frac{d}{z}(u, v) + (h_1(z), h_2(z)), \tag{9.17}$$

where c and d are complex constants, and $h_1(z), h_2(z)$ are holomorphic functions.

How can we describe the space of Hecke modifications in this case? Obviously, an overall scaling of the four variables

$$(u, u', v, v') \longrightarrow \lambda(u, u', v, v') \tag{9.18}$$

is inessential; it can be absorbed in scaling c and d. But there is another equivalence relation that we should take into account. If we transform

$$(u', v') \longrightarrow (u', v') + w(u, v), \tag{9.19}$$

for $w \in \mathbb{C}$, then this can be absorbed in $d \to d - wc$. So we should regard this transformation as another equivalence relation on the parameters u, u', v, v'. The two symmetries both arise from the fact that E_+ is invariant under $s_1 \to g s_1$, where g is a holomorphic function that is invertible at $z = 0$.

So the space of Hecke modifications is parametrized by (u, u', v, v'), with u and v not both zero (since $s_1(0)$ must be nonzero), and subject to the equivalences (9.18) and (9.19). The invariants under the transformation (9.19) are u, v, and $b = vu' - uv'$. So we can take this scaling into account by simply using u, v, and b to parametrize the Hecke modifications. b scales under (9.18) with weight 2, $b \to \lambda^2 b$. So any triplet (u, v, b), not identically zero, defines, modulo the scaling, a point in a weighted projective space $\mathbb{WCP}^2(1, 1, 2)$. The space of Hecke modifications as we have defined it so far is not all of the weighted projective space; we are missing the single point $(0, 0, 1)$, since we do not allow $u = v = 0$.

Although it is possible to work with a noncompact space of Hecke modifications, it is inconvenient. The spaces of physical states in the A-model are the cohomology groups of various moduli spaces. Compactness makes the definition of the appropriate cohomology groups much more straightforward; without it, one needs a detailed discussion of how the wavefunctions are supposed to behave near infinity (i.e., near $(0, 0, 1)$ in the present case).

The space of Hecke modifications $\mathcal{Y}({}^L w)$ as we have defined it so far does have a natural compactification, coming from the fact that the bundle E_+ has a natural limit as $u, v \to 0$. Introduce a small parameter ϵ and let $u = \epsilon\overline{u}$, $v = \epsilon\overline{v}$, and $d = d'/\epsilon$. A general section of E_+ takes the form

$$s(z) = \frac{c}{z^2}(\epsilon\overline{u} + zu', \epsilon\overline{v} + zv') + \frac{d'}{z}(\overline{u}, \overline{v}) + (h_1(z), h_2(z)). \tag{9.20}$$

The limit of this as $\epsilon \to 0$, keeping the other variables fixed, is just

$$s(z) = \frac{c}{z}(u', v') + \frac{d'}{z}(\overline{u}, \overline{v}) + (h_1(z), h_2(z)). \tag{9.21}$$

The bundle whose general section takes this form can be given a more simple interpretation. The point $(0, 0, 1)$ on the weighted projective space has $b \neq 0$, which is equivalent to linear independence of the vectors (u, v) and (u', v') or equivalently $(\overline{u}, \overline{v})$ and (u', v'). So when we approach $(0, 0, 1)$ by taking a limit as $\epsilon \to 0$, we should assume this linear independence. That being so, (9.21) can be described more simply by saying that $s(z)$ can have an arbitrary simple pole at $z = 0$. Consequently, the bundle in this limit is simply $E_+ = E_- \otimes \mathcal{O}(p_0)$.

But this is what we get if we make a Hecke modification of the bundle E_- with ${}^Lw = (1, 1)$. The Hecke operation $T({}^Lw)$ for this ${}^Lw$ is simply the operation of tensoring with $\mathcal{O}(p_0)$, as we have explained in section 9.1. The conclusion, then, is that the space of Hecke modifications with ${}^Lw = (2, 0)$ has a natural compactification in which one also allows a Hecke modification with ${}^Lw = (1, 1)$. We will write $\overline{\mathcal{Y}}({}^Lw)$ for this sort of compactification of the space of Hecke modifications of type $\mathcal{Y}({}^Lw)$ (we write more explicitly $\overline{\mathcal{Y}}({}^Lw, p_0)$ if we wish to specify the point at which the Hecke modification is made).

We recall that in general an 't Hooft operator is specified by the choice of a G-bundle over $\mathbb{CP}^1$; once such a choice is given, one then solves the Bogomolny equations requiring a local singularity determined by that choice of G-bundle. In particular, for $G = U(2)$, the 't Hooft operator with ${}^Lw = (m_1, m_2)$ is defined using the $U(2)$ bundle $\mathcal{O}(m_1) \oplus \mathcal{O}(m_2)$ over $\mathbb{CP}^1$. The only topological invariant of such a bundle is its first Chern class, which is $m_1 + m_2$. The bundle $\mathcal{O}(2) \oplus \mathcal{O}$, which corresponds to ${}^Lw = (2, 0)$, is an "unstable bundle"; it can be infinitesimally perturbed to change its holomorphic type to $\mathcal{O}(1) \oplus \mathcal{O}(1)$, which corresponds to ${}^Lw = (1, 1)$. The analysis that we have just described shows that the Hecke modification corresponding to ${}^Lw = (1, 1)$ must be included in order to compactify the space of Hecke modifications with ${}^Lw = (2, 0)$.

The other example that we considered is quite different. The bundle $\mathcal{O}(1) \oplus \mathcal{O}(0)$ cannot be perturbed to any other holomorphic type. So the space of Hecke modifications with ${}^Lw = (1, 0)$ is smooth and compact.

The Singularity of The Weighted Projective Space To understand more deeply the example with ${}^Lw = (2, 0)$, let us reexamine the weighted projective space $\mathbb{WCP}^2(1, 1, 2)$, which has homogeneous coordinates (u, v, b) of weights $(1, 1, 2)$. At the point $(0, 0, 1)$, this space has a $\mathbb{Z}_2$ orbifold singularity.

The local structure is therefore that of the A_1 singularity $\mathbb{C}^2/\mathbb{Z}_2$. $\mathbb{C}^2/\mathbb{Z}_2$ is a hyper-Kahler orbifold. Unlike a typical complex singularity, the A_1 singularity can be resolved and deformed in a compatible fashion. This is most naturally understood [121] by exhibiting $\mathbb{C}^2/\mathbb{Z}_2$ as a hyper-Kahler quotient of $\mathbb{H}^2 \cong \mathbb{C}^4$ by $U(1)$. Instead of following that route, we will here just describe by hand the deformation and resolution of the singularity of the weighted projective space $\mathbb{WCP}^2(1,1,2)$.

To describe the deformation, we set $a_1 = u^2$, $a_2 = uv$, $a_3 = v^2$. The variables (a_1, a_2, a_3, b) all have weight two, and modulo scaling they parametrize an ordinary projective space $\mathbb{CP}^3$. However, they obey $a_1 a_3 - a_2^2 = 0$, so $\mathbb{WCP}^2(1,1,2)$ is the subvariety of $\mathbb{CP}^3$ defined by this equation. The equation $a_1 a_3 - a_2^2 = 0$ can be deformed to $a_1 a_3 - a_2^2 = \epsilon b^2$ with a small parameter ϵ. This last equation defines a smooth quadric in $\mathbb{CP}^3$, which is isomorphic to[39] $\mathbb{CP}^1 \times \mathbb{CP}^1$, so $\mathbb{WCP}^2(1,1,2)$ can be deformed to $\mathbb{CP}^1 \times \mathbb{CP}^1$.

To describe the resolution, we begin with $\mathbb{C}^4$ with coordinates u, v, b, and b'. We consider the action of $U(1) \times U(1)$ by $(u, v, b, b') \to (\lambda u, \lambda v, \lambda^2 \widetilde{\lambda} b, \widetilde{\lambda} b')$, where $|\lambda| = |\widetilde{\lambda}| = 1$. We define the moment maps $\mu = |u|^2 + |v|^2 + 2|b|^2$, $\widetilde{\mu} = |b|^2 + |b'|^2$, and we take a symplectic quotient of $\mathbb{C}^4$ by $U(1)^2$. To do this, impose the moment map equations

$$\begin{aligned} |u|^2 + |v|^2 + 2|b|^2 &= 1 \\ |b|^2 + |b'|^2 &= d \end{aligned} \tag{9.22}$$

for some constant d, and divide by $U(1) \times U(1)$. For small positive d, the equations do not permit $u = v = 0$, so the pair (u, v) can be taken as homogeneous coordinates for a copy of $B = \mathbb{CP}^1$. The fiber over a given point in $\mathbb{CP}^1$ is, for small d, another copy of $\mathbb{CP}^1$, parametrized by b and b'. Since (b, b') transform to $(\lambda^2 b, b')$ under $(u, v) \to (\lambda u, \lambda v)$, b and b' take values in the line bundles $\mathcal{O}(2)$ and $\mathcal{O}$, respectively, over B. So the symplectic quotient for small positive d is fibered over $B = \mathbb{CP}^1$ with the fiber being $Y = \mathbb{P}(\mathcal{O}(2) \oplus \mathcal{O})$, that is, the projectivization of the rank two vector bundle $\mathcal{O}(2) \oplus \mathcal{O}$. In particular, the symplectic quotient for small positive d is a $\mathbb{CP}^1$ bundle over $\mathbb{CP}^1$, and thus a Hirzebruch surface.

If we increase d, nothing happens until we reach $d = 1/2$. At that point, the Hirzebruch surface develops a singularity and reduces to the weighted

[39] By a linear change of variables, one can arrange a_1, a_2, a_3 and b as the matrix elements of a 2×2 matrix M such that the equation becomes $\det(M) = 0$. This equation says that M is of rank 1, and so takes the form $M_{ij} = x_i y_j$, where the x_i and y_j are homogeneous coordinates for, respectively, the two factors in $\mathbb{CP}^1 \times \mathbb{CP}^1$.

projective space $\mathbb{WCP}^2(1,1,2)$. At $d = 1/2$, we simply write the second moment map equation as $|b'|^2 = 1/2 - |b|^2$, and observe that, by virtue of the first equation, the right hand side is nonnegative. So a solution for b' always exists and is unique up to the action[40] of $\widetilde{\lambda}$. The symplectic quotient at $d = 1/2$ is thus parametrized by u, v, and b, subject to $|u|^2+|v|^2+2|b|^2 = 1$ and the equivalence $(u, v, b) \cong (\lambda u, \lambda v, \lambda^2 b)$ for $|\lambda| = 1$. The combined operation gives the weighted projective space $\mathbb{WCP}^2(1,1,2)$.

Running this in reverse, reducing d from $1/2$ to a smaller value resolves the singularity of $\mathbb{WCP}^2(1,1,2)$ to give the Hirzebruch surface. Moreover, this particular Hirzebruch surface can be deformed to give a simple product $\mathbb{CP}^1 \times \mathbb{CP}^1$. For this, we just observe that the bundle $\mathcal{O}(2) \oplus \mathcal{O}$ is unstable and can be deformed to $\mathcal{O}(1) \oplus \mathcal{O}(1)$. But the projectivization of $\mathcal{O}(1) \oplus \mathcal{O}(1)$ is the same as the projectivization of $\mathcal{O} \oplus \mathcal{O}$ and so is a trivial $\mathbb{CP}^1$ bundle over $\mathbb{CP}^1$, that is a product $\mathbb{CP}^1 \times \mathbb{CP}^1$.

The conclusion is that, in contrast to a generic singular algebraic variety, the singularity of $\mathbb{WCP}^2(1,1,2)$ can be deformed or resolved to give the same result topologically, and moreover the deformation and resolution can be carried out simultaneously. Moreover, this is also related to the fact that the local structure near the singularity is the hyper-Kahler A_1 singularity. At the end of section 10.2 and in section 10.3, we will explain this behavior from the point of view of the Bogomolny equations. We also in section 10.4 will use the structure near this singularity to illustrate the computation of the operator product expansion of 't Hooft operators.

General Case For $U(2)$ The space of Hecke modifications of type ${}^L w = (m_1, m_2)$ can be analyzed similarly. First of all, the space of Hecke modifications of type $(m_1 + c, m_2 + c)$ is the same as that of type (m_1, m_2), since adding (c, c) just has the effect of tensoring the output of the Hecke transformation with $\mathcal{O}(p_0)^c$. So it is enough to take ${}^L w = (a, 0)$ with some integer a, which we can assume nonnegative. (Otherwise we replace $(a, 0)$ by $(0, a)$ and then, adding $-a$ to each weight, by $(-a, 0)$.) In describing Hecke modifications of type $(a, 0)$, it is convenient, given a function g with a pole of order n at $z = 0$, to write $[g]_-$ for the polar part of g: $[g]_- = g_{-n} z^{-n} + g_{-n+1} z^{-n+1} + \ldots + g_{-1} z^{-1}$.

Applying a Hecke transform with ${}^L w = (a, 0)$ relative to the decomposition $E_- = \mathcal{L}_1 \oplus \mathcal{L}_2$ introduced above, we get a bundle E_+ whose general

[40] $\widetilde{\lambda}$ acts freely on b' except at the point $b' = u = v = 0$, where the action of $\widetilde{\lambda}$ generates the same orbit as λ. So at $d = 1/2$, upon fixing b' to be a nonnegative real number, we can omit $\widetilde{\lambda}$ from the description.

section takes the form

$$(9.23)\quad s = [c_0 s_1]_- + [c_1 z s_1]_- + [c_2 z^2 s_1]_- + \ldots + [c_{a-1} z^{a-1} s_1]_- + (h_1(z), h_2(z)),$$

with complex coefficients $c_0, c_1, \ldots, c_{a-1}$.

The bundle E_+ therefore depends on s_1 modulo terms of order z^a. This leaves a total of $2a$ complex parameters. However, the bundle E_+ is unaffected if we replace s_1 by gs_1, where g is a holomorphic function that is invertible at $z = 0$. By choice of g, we can eliminate a of the complex parameters in s_1, generalizing what we did for $a = 1, 2$. So the space of Hecke modifications of type $(a, 0)$ has complex dimension a. To compactify this space, one has to include Hecke modifications of types $(a-1, 1)$, $(a-2, 2), \ldots, (a-[a/2], [a/2])$, which appear when successive terms in the Taylor series expansion of s_1 vanish.

Some of what we have explained will be more transparent to most physicists if we shift from ${}^L w = (a, 0)$ to ${}^L w = (a/2, -a/2)$. (The space of Hecke modifications is of course unchanged in this shift.) As we have explained in section 9.0, weights $(a/2, -a/2)$ make sense for gauge group $G = PSU(2) = SO(3)$, ${}^L G = SU(2)$, and moreover the 't Hooft operator with these weights is S-dual to a Wilson loop of $SU(2)$ in the representation of spin a.

In studying confinement in $SU(2)$ gauge theory, it is redundant to consider Wilson loops based on an arbitrary representation R. Two representations on which the center of $SU(2)$ act in the same way are equivalent for that purpose, so the only non-trivial representation of $SU(2)$ that one must consider is the representation of spin $1/2$. Dually, any 't Hooft operator of $SO(3)$ is topologically equivalent to either the trivial one or the operator with weights $(1/2, -1/2)$. Precisely because the higher 't Hooft operators are topologically unstable, they are considered inessential in applications of 't Hooft operators (such as 't Hooft's original work) that aim at understanding the phases of gauge theories.

From our point of view, the unstable 't Hooft operators are important, but can "mix" with lower operators of the same topological type. In section 10.2, we will explain what this mixing means in terms of the Bogomolny equations.

The general nature of the mixing is most easily stated in case the Lie groups G and ${}^L G$ are simple. Let ${}^L R$ be a representation of ${}^L G$. Let us say that a representation ${}^L R'$ of ${}^L G$ is associated to ${}^L R$ if the highest weight of ${}^L R'$ is a weight of ${}^L R$ but not the highest weight. Writing ${}^L w$ and ${}^L w'$ for the highest weights of ${}^L R$ and ${}^L R'$, we also say in this situation that ${}^L w'$

is associated to ${}^L w$. The condition is that ${}^L w - {}^L w'$ is a nonzero dominant weight of the adjoint form of ${}^L G$.

Then in general, the natural compactification of the space $\mathcal{Y}({}^L R)$ of Hecke modifications of type ${}^L R$ includes the Hecke modifications of type ${}^L R'$ for all representations ${}^L R'$ associated to ${}^L R$. For $G = U(N)$ (or $SU(N)$ or $PSU(N)$), this statement just means that for a weight ${}^L w = (m_1, m_2, \dots, m_N)$, if $m_i \geq m_j + 2$ for some i, j, the natural compactification includes Hecke modifications with m_i replaced by $m_i - 1$ and m_j replaced by $m_j + 1$, as well as all other weights that can be obtained by repeating this move. That this move is needed for the compactification can be seen simply by embedding in $U(N)$ a family of $U(2)$ Hecke modifications whereby weights (m_i, m_j) degenerate to $(m_i - 1, m_j + 1)$. For more general groups, one can argue similarly, using moves generated by various $SU(2)$ subgroups of ${}^L G$. The role of these moves in the compactification can also be seen in terms of the affine Grassmannian, a notion that we briefly introduce in section 9.3, or from the Bogomolny equations, as in section 10.2.

Minuscule Representations Generalizing the weight $(1, 0)$ of $U(2)$, we will describe a few more examples where the space of Hecke modifications is smooth and compact. For $G = U(N)$, let us consider as in (9.13) the weight ${}^L w(k) = (1, 1, \dots, 1, 0, 0, \dots, 0)$, with the number of 1's equal to k. Picking a decomposition $E_- = \oplus_{i=1}^N \mathcal{L}_i$, where $\mathcal{L}_i$ is generated by a section s_i, the Hecke operator $T({}^L w(k))$ maps E_- to $E_+ = \oplus_{i \leq k} (\mathcal{L}_i \otimes \mathcal{O}(p_0)) \oplus (\oplus_{i>k} \mathcal{L}_i)$. The general section of E_+ takes the form

$$s = \frac{1}{z}(a_1, \dots, a_N) + (h_1, \dots, h_N) \tag{9.24}$$

near p_0, where $\vec{a} = (a_1, \dots, a_N) \in \mathbb{C}^N$ is a linear combination of $s_1(0), \dots, s_k(0)$. E_+ therefore depends on the choice of s_i precisely through the subspace $W \subset \mathbb{C}^N$ generated by $s_1(0), \dots, s_k(0)$. W can be any k-dimensional subspace of $\mathbb{C}^N$, so the space $\mathcal{Y}({}^L w(k))$ of Hecke modifications of type ${}^L w(k)$ is the Grassmannian $\mathrm{Gr}(k, N)$ of complex k-planes in $\mathbb{C}^N$.

Clearly, then, the space of Hecke modifications of type ${}^L w(k)$ is smooth and compact. This generalizes the $U(2)$ example with ${}^L w = (1, 0)$. As we noted in relation to eqn. (9.13), ${}^L w(k)$ is the highest weight of the representation $\wedge^k \mathbf{V}$ of ${}^L G = U(N)$. This representation is "minuscule," which means that its weights form a single Weyl orbit. In general, minuscule representations are associated to stable 't Hooft operators with smooth, compact spaces of Hecke modifications. Minuscule representations are precisely those that

have no associated representations, since the highest weight of a minuscule representation is its only dominant weight.

The highest weight ${}^L w'(k)$ of the analogous representation $\wedge^k \mathbf{V}$ of $SU(N)$ is obtained by subtracting the constant $-k/N$ from each weight of ${}^L w(k)$. The space of Hecke modifications is unchanged. The representations $\wedge^k \mathbf{V}$ are minuscule and generate the representation ring of $SU(N)$. However, $SU(N)$ is the only simple and simply-connected Lie group whose representation ring is generated by minuscule representations. For general simple G, there is precisely one minuscule representation for each character of the center of G; groups other than $SU(N)$ have relatively small centers and few minuscule representations.

The Dimension Of The Space Of Hecke Modifications For $U(N)$ (or $SU(N)$, or $PU(N)$), let us determine the dimension of the space of Hecke modifications of an arbitrary type. For weight ${}^L w = (a_1, a_2, \dots, a_N)$, with $a_1 \geq a_2 \geq \dots \geq a_N$, the dimension is

$$\Delta_{{}^L w} = \sum_{i<j} (a_i - a_j). \tag{9.25}$$

To justify this statement, we consider N sections s_i that generate the trivial rank N bundle $E_- = \mathcal{O} \oplus \mathcal{O} \oplus \dots \oplus \mathcal{O}$ near $z = 0$. We recall that this means that the $s_i(0)$ are linearly independent. Then we consider the bundle E_+ whose general section is

$$q = \sum_{i=1}^{N} z^{-a_i} g_i s_i, \tag{9.26}$$

where the functions $g_i(z)$ are holomorphic at $z = 0$. The bundle E_+ is invariant under

$$s_i \to s_i + \sum_{j \leq i} h_j s_j + \sum_{j>i} z^{a_i - a_j} h_j s_j, \tag{9.27}$$

for generic functions h_j that are holomorphic at $z = 0$. Such a transformation of the s_i can be absorbed in redefining the functions g_i that appear in (9.26). Expanding s_i in a Taylor series near $z = 0$, the number of coefficients that cannot be eliminated by the transformation (9.27) is $\sum_{j=i+1}^{N} (a_i - a_j)$. After summing over i, we arrive at the formula (9.25) for the dimension of the space of Hecke modifications.

9.3. The Affine Grassmannian

The space of Hecke modifications associated with an 't Hooft operator can be described as a finite-dimensional subvariety in a certain infinite-dimensional variety called the affine Grassmannian. This description enables one to construct a natural compactification of the space of Hecke modifications. We will not really use this framework in the present paper, but we explain it here because it gives a powerful way of understanding some things and is the framework in which the subject is usually placed mathematically. The affine Grassmannian is familiar to physicists primarily for its applications to two-dimensional conformal field theory [122].

For the sake of clarity, let us consider the case $G = U(N)$. In this case, an 't Hooft operator is specified by a set of integers $m_1, \dots, m_N$ modulo permutations. Since the problem is local, we can fix the curve C to be $\mathbb{CP}^1$ and take E_- to be trivial. A Hecke modification E of E_- at a point $p_0 \in C$ is a bundle that is endowed with a chosen isomorphism $\sigma : E \cong E_-$ outside of p_0; σ may not extend over p_0.

It is natural to regard the space of Hecke modifications for fixed $m_1, \dots, m_N$ as a subspace of the space of pairs (E, σ), where E is a holomorphic vector bundle on C and σ is its trivialization outside p_0. Let us recall an explicit description of the space of such pairs (E, σ), following chapter 8 of [122].

It is convenient to think of $C = \mathbb{CP}^1$ as a one-point compactification of the complex z-plane $\mathbb{C}$ and to set p_0 to be the point $z = 0$. We also let

$$U_\infty = \mathbb{CP}^1 - \{0\}, \quad U_0 = \mathbb{CP}^1 - \{\infty\}. \tag{9.28}$$

We are given a trivialization σ of E over U_∞. Although σ may not extend over U_0, we can pick a trivialization σ' of E over U_0. Over $U_0 \bigcap U_\infty \simeq \mathbb{C}^*$, σ and σ' are related by a $GL(N, \mathbb{C})$ gauge transformation $g(z)$. This is a $GL(N, \mathbb{C})$-valued function whose entries are holomorphic on $\mathbb{C}^*$ but may have poles at 0 and ∞. (σ' can always be chosen so that the singularities of $g(z)$ are poles.) Let us denote by $\mathcal{X}$ the ring of functions holomorphic on $\mathbb{C}^*$ and having a meromorphic extension to $\mathbb{CP}^1$; then the group of holomorphic gauge transformations on $\mathbb{C}^*$ can be identified with $GL(N, \mathcal{X})$, and $g(z)$ is an element of this group.

If we change σ' by a gauge transformation $h_0(z)$ which is holomorphic throughout U_0, then $g(z)$ is replaced by $g(z)h_0^{-1}(z)$. $h_0(z)$ takes values in $GL(N, \mathcal{O})$, the group of invertible matrices whose entries are holomorphic functions on $U_0 \simeq \mathbb{C}$. The set of holomorphic vector bundles of rank N

over $\mathbb{CP}^1$ equipped with a fixed holomorphic trivialization σ outside p_0 is isomorphic to the quotient $GL(N,\mathcal{X})/GL(N,\mathcal{O})$. The latter space is known as the affine Grassmannian Gr_N for the group $GL(N)$. Another name for it is the loop Grassmannian, because it is isomorphic [122] to the space of based loops in $U(N)$.

The definition of the affine Grassmannian admits slight variations. For example, since any function in $\mathcal{X}$ can be expanded in a Laurent series, one can embed $\mathcal{X}$ into the ring $\mathbb{C}((z))$ of formal Laurent series. Similarly, one can embed the ring of holomorphic functions $\mathcal{O}$ into the ring $\mathbb{C}[[z]]$ of formal power series. One can show that the quotient $GL(N,\mathbb{C}((z)))/GL(N,\mathbb{C}[[z]])$ is isomorphic to the affine Grassmannian Gr_N. Intuitively, this happens because given any two elements of $GL(N,\mathcal{X})$, one can tell whether they are in the same $GL(N,\mathcal{O})$ orbit by studying a finite number of terms in their Laurent expansions, and therefore it is immaterial whether the Laurent series have a nonzero region of convergence.

Now we have to identify those points in Gr_N which can be obtained from the trivial vector bundle by a Hecke modification of weight ${}^L w = (m_1,\dots,m_N)$. A Hecke modification of the trivial vector bundle with this weight is obtained by choosing a trivialization of the trivial vector bundle over U_0 by N linearly independent holomorphic sections $f_1, f_2, \dots, f_N$, and declaring that E_+ is generated over U_0 by sections $s_1,\dots,s_N$ which upon restriction to $U_\infty \bigcap U_0$ are given by

$$z^{-m_j} f_j, \quad j = 1,\dots,N. \tag{9.29}$$

Here we made use of the fact that on U_∞ we are given an isomorphism between E_+ and the trivial vector bundle. Thus if $f_j = (f_j^1, f_j^2, \dots, f_j^N)$, then the matrix $g(z)$ corresponding to E_+ is given by

$$g_j^i(z) = z^{-m_j} f_j^i(z). \tag{9.30}$$

The simplest choice is $f_j^i(z) = \delta_j^i$; all other choices can be obtained from this one by acting on $g(z)$ from the left by an element of $GL(N,\mathcal{O})$. We conclude that the space of Hecke modifications $\mathcal{Y}({}^L w; p_0)$ is the orbit of the point

$$g_j^i(z) = z^{-m_j}\delta_j^i \tag{9.31}$$

in the affine Grassmannian under the left action of the group $GL(N,\mathcal{O})$. One can replace $\mathcal{X}$ with $\mathbb{C}((z))$ and $\mathcal{O}$ with $\mathbb{C}[[z]]$ throughout and get an equivalent result. Note that $g(z)$ given by (9.31) is a homomorphism of $\mathbb{C}^*$

to $GL(N,\mathbb{C})$. In fact, it is the complexification of the homomorphism

$$\rho : U(1) \longrightarrow U(N), \quad \rho : e^{i\alpha} \longrightarrow \begin{pmatrix} e^{im_1\alpha} & & \cdots & \\ & e^{im_2\alpha} & \cdots & \\ & \vdots & \cdots & \\ & & & e^{im_N\alpha} \end{pmatrix} \tag{9.32}$$

which enters into the definition of the 't Hooft operator $T(m_1,\dots,m_N)$.

We have associated to each set of integers $m_1,\dots,m_N$ an orbit of $GL(N,\mathcal{O})$ in Gr_N. It turns out that all $GL(N,\mathcal{O})$ orbits in Gr_N are obtained in this way [122]. Thus Gr_N is stratified by spaces $\mathcal{Y}(m_1,\dots,m_N)$, which are called Schubert cells. Equivalently, by applying Hecke modifications to the trivial vector bundle, one can obtain an arbitrary holomorphic vector bundle on $\mathbb{CP}^1$ with an arbitrary trivialization on U_∞.

The examples of spaces of Hecke modifications discussed in section 9.1 suggest that these spaces are always finite-dimensional. This can be shown in general as follows. Let $\mathrm{Gr}_N(i)$ be the subset of Gr_N defined by the condition that $g(z)$ has at $z=0$ a pole of order not higher than i. The variety $\mathrm{Gr}_N(i)$ is finite-dimensional, $GL(N,\mathcal{O})$-invariant, and compact, and any point in Gr_N belongs to $\mathrm{Gr}_N(i)$ for some i. This implies that any orbit of $GL(N,\mathcal{O})$ belongs to some $\mathrm{Gr}_N(i)$ and therefore is finite-dimensional.

It may be helpful to mention at this point that the infinite-dimensional space Gr_N has another stratification with strata labelled by sets of integers $k_1,\dots,k_N$. To define this stratification, we recall that by a theorem of Grothendieck any holomorphic vector bundle of rank N on $\mathbb{CP}^1$ is isomorphic to

$$\oplus_{i=1}^N \mathcal{O}(p_0)^{k_i} \tag{9.33}$$

for some integers $k_1,\dots,k_N$. We define the stratum in Gr_N corresponding to $k_1,\dots,k_N$ as the set of those pairs (E,σ) where E is isomorphic to (9.33). In other words, a stratum is obtained by fixing E and varying the trivialization σ. Obviously, the strata are infinite-dimensional in this case, in contrast with the stratification given by $\mathcal{Y}(m_1,\dots,m_N)$. The relation between the two stratifications of Gr_N is studied in [122].

As we have seen, spaces of Hecke modifications or Schubert cells are noncompact, in general. A natural way to compactify them is to consider the closure of the corresponding orbits in $\mathrm{Gr}_N(i)$ for sufficiently large i. For a given $\rho : U(1) \to U(N)$, the closure of the orbit C_ρ is in general a singular variety which is a union of C_ρ and a finite number of orbits of lower dimension. It is called a Schubert cycle. As discussed in section 9.2, the

structure of the closure of C_ρ reflects the "mixing" between 't Hooft operators with different ρ but the same topological type. We also will describe this process in terms of monopole bubbling in section 10.2.

The construction of Gr and the description of the spaces of Hecke modifications as subvarieties in Gr can be generalized to other gauge groups. For any simple compact Lie group G, with complexification $G_{\mathbb{C}}$, one defines Gr_G as the quotient

$$\mathrm{Gr}_G = G_{\mathbb{C}}(\mathcal{X})/G_{\mathbb{C}}(\mathcal{O}). \tag{9.34}$$

An 't Hooft operator is parametrized by the conjugacy class of a homomorphism $\rho : U(1) \to G$, which can be analytically continued to a homomorphism $\rho_{\mathbb{C}} : \mathbb{C}^* \to G_{\mathbb{C}}$. Obviously, $\rho_{\mathbb{C}}$ defines a point on Gr_G, and the orbit of this point under the left action of $G_{\mathbb{C}}(\mathcal{O})$ depends only on the conjugacy class of ρ. This orbit is the Schubert cell or space of Hecke modifications $\mathcal{Y}(\rho)$.

10. The Bogomolny Equations And The Space Of Hecke Modifications

In this section, we return to the Bogomolny equations with singularities and study them more closely. The first goal is to exhibit the space of Hecke modifications of a given type as a moduli space of solutions of the Bogomolny equations, with suitable boundary conditions. This will enable us to get a clear picture of many relatively subtle aspects of the spaces of Hecke modifications. Then we relax the assumption of time-reversal invariance and consider the complexified or extended Bogomolny equations. These extended equations enable one to define the Hecke transform of an arbitrary A-brane on $\mathcal{M}_H(G, C)$ and therefore are the important ones for geometric Langlands duality.

10.1. Boundary Conditions

We will study the Bogomolny equations on a G-bundle E over a three-manifold $W = I \times C$ (as usual, I is an interval and C a Riemann surface, and there are prescribed singularities at points $p_i \in W$ where 't Hooft operators are inserted). We write C_- and C_+ for the two ends of W, E_- and E_+ for the restriction of E to the two ends, and A_- and A_+ for the restrictions of the connection A.

There are two closely related problems that we might consider. One natural question is to describe the action of the 't Hooft/Hecke operators on bundles, and the second is to describe the Hecke correspondence.

For the first problem (Figure 15(a)), we specify a bundle E_- at the left end of W and then ask what bundle E_+ can appear at the other end as a result of solving the Bogomolny equations with singularities. In this way we should recover the space of Hecke modifications, as described in section 9.0. An alternative problem (Figure 15(b)) is to try to describe the Hecke correspondence, which parametrizes pairs of bundles with a suitable Hecke relation. We will describe boundary conditions appropriate for either of these two problems.

To specify the bundle E_-, we specify the gauge field A_- on C_-. We want to specify an actual bundle E_-, not an isomorphism class of bundles, so we specify an actual connection A_- on C_-. So we will only divide by gauge

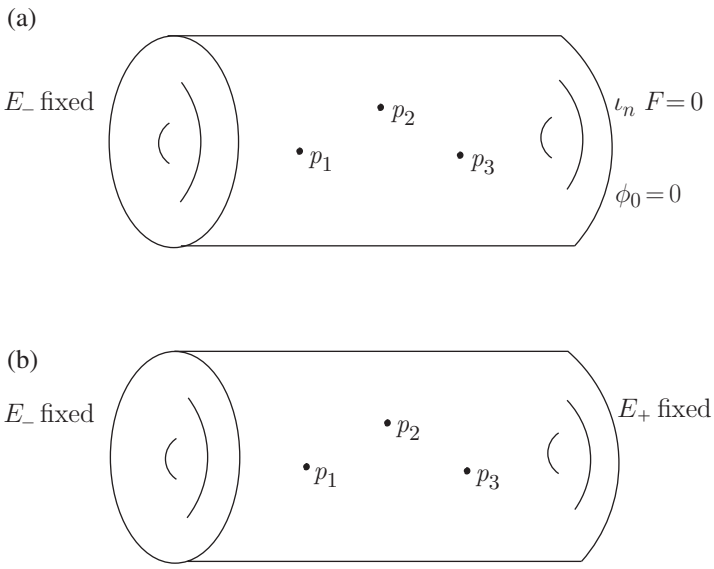

Figure 15: (a) On the three-manifold $W = I \times C$ with 't Hooft operator insertions, a bundle E_- is specified on the left, and one wishes to determine possible Hecke modifications E_+ that may appear on the right. On the left, the connection A_- obeys Dirichlet boundary conditions and ϕ_0 is undetermined; on the right, their roles are reversed. (b) To describe a fiber of the Hecke correspondence, one specifies both E_- and E_+ and leaves ϕ_0 undetermined at each end.

transformations that are trivial on C_-. It is convenient for some purposes to assume that the bundle E_- is stable, in which case we can assume that the connection A_- is flat. We will not assume that the output bundle E_+ that results from solving the Bogomolny equations is stable.

We want the boundary conditions on C_- to be elliptic (modulo the gauge transformations). The simplest elliptic boundary conditions in the Bogomolny equations are obtained by either imposing Dirichlet boundary conditions on A – that is, specifying its boundary values – and leaving the boundary values of ϕ_0 unspecified, or vice-versa. If the boundary values of A are specified, then the Bogomolny equations determine the normal derivative of ϕ_0 at the boundary, and vice-versa. At E_-, since we want to specify A, we leave ϕ_0 unspecified. If A_- is flat, then the Bogomolny equations require the normal derivative of ϕ_0 to vanish at C_-.

At C_+, on the other hand, we want to allow all possible Hecke modifications E_+ that may arise from E_- by solving the Bogomolny equations. So the connection A is unspecified on C_+, and instead we set $\phi_0 = 0$ on C_+. The Bogomolny equations then require that A obeys covariant Neumann boundary conditions at E_+; that is, its curvature $F(A)$ vanishes when contracted with the normal vector to the boundary. We are only interested in the output bundle E_+ up to gauge transformations, so we allow the gauge transformations to be non-trivial on C_+. These boundary conditions are summarized in Figure 15(a). In section 10.3, we will understand more deeply the naturalness of these boundary conditions.

The Hecke Correspondence What we have just described is the boundary condition that we will generally use in employing the Bogomolny equations to study Hecke operators. But we pause to discuss the differential geometric analog of another formulation that is standard in algebraic geometry. (The details are not needed for the rest of the paper.) Instead of thinking of E_- as input and E_+ as output, as we have done above, one can treat the two symmetrically by describing the "Hecke correspondence." In the usual formulation in algebraic geometry, we let Bun_G be the "stack" of all G-bundles over C, not necessarily stable. The Hecke correspondence of type ${}^L w, z$ (for a weight ${}^L w$ and a point $z \in C$) is then a variety $\mathcal{Q}({}^L w, z)$ that maps to $\mathrm{Bun}_G \times \mathrm{Bun}_G$. The fiber of $\mathcal{Q}({}^L w, z)$ over $E_- \times E_+ \in \mathrm{Bun}_G \times \mathrm{Bun}_G$ parametrizes ways of obtaining E_+ from E_- by a Hecke transformation of type ${}^L w$ at z.

The relationship between E_- and E_+ can be stated more symmetrically, as C_- and C_+ are exchanged by an orientation-preserving automorphism $\mathcal{PT}$

that acts trivially on C, exchanges the two ends of I, and also reverses the time coordinate. Such an automorphism maps ϕ to $\mathcal{PT}^*(\phi)$, and maps an 't Hooft operator of weight ${}^Lw$ to one of weight ${}^Lw'$, where ${}^Lw'$ is a dominant weight that is Weyl conjugate to $-{}^Lw$. So E_+ is produced from E_- by a Hecke transformation of type ${}^Lw$ at z, or equivalently E_- is produced from E_+ by a Hecke transformation of type ${}^Lw'$ at z.

The fiber of the map $\mathcal{Q}({}^Lw, z)$ to $\mathrm{Bun}_G \times \mathrm{Bun}_G$ is generically empty if the weight ${}^Lw$ is small and the genus of C is large. On the other hand, for any C, this map is generically a fibration if ${}^Lw$ is large enough. One can extend the definition of $\mathcal{Q}$ to allow several Hecke operators with specified positions and weights, in which case the nature of the map $\mathcal{Q} \to \mathrm{Bun}_G \times \mathrm{Bun}_G$ depends on the number and weights of the 't Hooft operators.

In differential geometry, Bun_G corresponds to the space $\mathcal{A}$ of all connections. $\mathcal{A}$ is of course infinite-dimensional, and Bun_G is an algebro-geometric analog of an infinite-dimensional manifold (it includes bundles that are arbitrarily unstable with no bound on the dimensions of spaces of deformations and automorphism groups). The Hecke correspondence $\mathcal{Q}$ likewise is infinite-dimensional and will not come from elliptic boundary conditions. A differential-geometric version of the Hecke correspondence is to simply define $\mathcal{Q}$ as the space of all solutions of the Bogomolny equations on $W = I \times C$, with specified singularities, modulo gauge transformations that are trivial at each end. The map of $\mathcal{Q}$ to $\mathcal{A} \times \mathcal{A}$ is obtained by restricting a solution to the two ends of W.

$\mathcal{Q}$ is infinite-dimensional, as in defining it we have specified neither A nor ϕ_0 at either end; these boundary conditions are not elliptic. In differential geometry, it seems more natural to define the fiber of the Hecke correspondence over a pair of connections $A_- \times A_+ \subset \mathcal{A} \times \mathcal{A}$. This is the gauge theory analog of studying in algebraic geometry the fiber of the map $\mathcal{Q} \to \mathrm{Bun}_G \times \mathrm{Bun}_G$. For this, we fix A_- and A_+ at each end, and divide by gauge transformations that are trivial at each end (Figure 15(b)). The space of all solutions of the Bogomolny equations with these boundary conditions is the fiber of the Hecke correspondence over the pair $A_- \times A_+$. The boundary conditions defining this fiber are elliptic and correspond to simply using at both ends of W the boundary conditions that in Figure 15(a) are imposed only on C_-.

The virtual complex dimension of the fiber of the Hecke correspondence can be computed via index theory, by modifying the discussion presented below. It equals $-(g-1)\dim(G)$ plus the dimension Δ (also calculated below) of the space of Hecke modifications of type ${}^Lw$. If the virtual dimension is a negative number $-d$, it means that generically $\mathcal{Q}$ is a subvariety of $\mathcal{A} \times \mathcal{A}$

of codimension d. If it is a positive integer d', it means that generically $\mathcal{Q}$ fibers over $\mathcal{A} \times \mathcal{A}$ with fibers of dimension d'.

For the rest of our analysis, however, we concentrate on describing the space of possible Hecke modifications of a given bundle E_-, and hence we use the boundary conditions of Figure 15(a). Before investigating more detailed properties, first we will see what can be learned from index theory.

Index Theory And The Virtual Dimension We let $\mathcal{Z}$ be the moduli space of solutions of the Bogomolny equations with the elliptic boundary conditions of Figure 15(a). From our discussion of the relationship between 't Hooft and Hecke operators in section 9, we anticipate that $\mathcal{Z}$ is a complex manifold. We demonstrate this in section 10.3. Linearization of the Bogomolny equations leads to an elliptic complex studied in [107,108]:

$$0 \to \Omega^0(\mathrm{ad}(E)) \xrightarrow{d_1} \Omega^1(\mathrm{ad}(E)) \oplus \Omega^0(\mathrm{ad}(E)) \xrightarrow{d_2} \Omega^2(\mathrm{ad}(E)) \to 0. \tag{10.1}$$

(Here $\Omega^i(\mathrm{ad}(E))$ is the space of $\mathrm{ad}(E)$-valued i-forms. d_1 is the map from an element of $\Omega^0(\mathrm{ad}(E))$, the Lie algebra of gauge transformations, to the linearization of A and ϕ, which take values in $\Omega^1(\mathrm{ad}(E))$ and $\Omega^0(\mathrm{ad}(E))$, respectively. d_2 is the linearization of the Bogomolny equations.) The tangent space to $\mathcal{Z}$ is the first cohomology group of this complex, namely $H^1 = \ker(d_2)/\mathrm{im}(d_1)$. The index of this elliptic complex is called the virtual dimension of $\mathcal{Z}$, and coincides with the actual dimension if the other cohomology groups $H^0 = \ker(d_1)$ and $H^2 = \mathrm{coker}(d_2)$ vanish. H^0 consists of covariantly constant sections of $\Omega^0(\mathrm{ad}(E))$, which generate unbroken gauge symmetries. With our boundary conditions, there are none, since we require the gauge transformations to be trivial on one end of W. As we will see, with the boundary conditions of Figure 15(a), the virtual dimension is nonnegative. This being so, H^2 vanishes away from singularities of $\mathcal{Z}$, and the smooth part of $\mathcal{Z}$ is a manifold whose dimension equals the virtual dimension or the index of the complex.

We first analyze the Bogomolny equations in the absence of any 't Hooft operators. We assume for simplicity that E_- is stable and represented by a flat connection with $F = 0$. (We can always deform to this situation without changing the index.) In this case, a standard sort of argument shows that a solution of the Bogomolny equations with any of the boundary conditions above is a pullback from C, so in particular E_+ and E_- are isomorphic. This is actually a special case of the vanishing theorems of section 3.3, but the argument is so simple that we present it separately here.

If the Bogomolny equations $F - \star D\phi_0 = 0$ are obeyed, then $0 = \int_W \mathrm{Tr}\,(F - \star D\phi_0) \wedge \star(F - \star D\phi_0)$. Expanding this out and integrating by parts, we have

$$-\int_W \mathrm{Tr}\ (F \wedge \star F + D\phi_0 \wedge \star D\phi_0) = -2\int_{\partial W} \mathrm{Tr}\,\phi_0 F. \tag{10.2}$$

The boundary term vanishes on a component of ∂W on which either ϕ_0 or F vanishes. In Figure 15(a), we have $F = 0$ at one end and $\phi_0 = 0$ at the other, so the boundary term vanishes at each end. But the left hand side of (10.2) is positive semi-definite and can only vanish if $F = D\phi_0 = 0$. So, given that E_- is flat, any solution of the Bogomolny equations with these boundary conditions and without 't Hooft operators is given by a flat connection that is pulled back from C, with $\phi_0 = 0$.

In particular, E_+ is isomorphic to E_-, and $\mathcal{Z}$ is a point, of dimension zero. One can further verify that, in expanding around such a trivial solution with the boundary condition of Figure 15(a), H^2 vanishes. Thus, in this case the virtual dimension of $\mathcal{Z}$ is the same as the actual dimension, namely zero. Now, what happens if 't Hooft operators are included? Each singularity associated with an 't Hooft operator $T({}^L R)$ shifts the virtual dimension of the moduli space by an amount that only depends on the representation ${}^L R$ of the dual group (and not the details of the three-manifold in which the 't Hooft operator is inserted, or the topology of the bundle E, or the possible presence of other 't Hooft operators). This follows from general excision properties of index theory [123], as shown by Pauly [124]. Moreover, Pauly computed, in our language, the contribution of an 't Hooft operator to the dimension of moduli space for $G = PSU(2) = SO(3)$, ${}^L G = SU(2)$. The answer[41] is that an 't Hooft operator associated with a representation of highest weight ${}^L w = (a/2, -a/2)$, with positive integer a, shifts the virtual (complex) dimension of the moduli space $\mathcal{Z}$ of solutions of the Bogomolny equations by a.

The fact that the singularity increases the complex dimension by a agrees with the result we found in section 9.1 by considering the space of Hecke modifications. There we showed that for $U(2)$ weight $(a, 0)$, or equivalently $SU(2)$ weight $(a/2, -a/2)$, the space of Hecke modifications has complex dimension a.

[41] This is Theorem I of [124], with k corresponding to what we call $a/2$. The theorem is stated for even a, as the gauge group is taken to be $G = SU(2)$ rather than $PSU(2)$, but the proof also works for odd a and in fact for all G, as we will discuss.

Pauly's proof actually generalizes immediately to give the result for any compact Lie group G. To explain this, we must review the technique behind Pauly's proof, which was introduced by Kronheimer [110]. The basic idea is to consider instantons on the four-manifold[42] $\mathbb{C}^2 \cong \mathbb{R}^4$ that are invariant under the action of $\mathcal{F} = U(1)$ on $\mathbb{C}^2$ by $(z_1, z_2) \to (\exp(i\theta)z_1, \exp(i\theta)z_2)$. The quotient $\mathbb{C}^2/\mathcal{F}$ is isomorphic to $\mathbb{R}^3$. $\mathcal{F}$ does not act freely – it has a fixed point at the origin, $z_1 = z_2 = 0$ – but nonetheless the quotient $\mathbb{C}^2/\mathcal{F}$ is a manifold $\mathbb{R}^3$. The map from $(z_1, z_2) \in \mathbb{C}^2$ to $\vec{x} \in \mathbb{R}^3$ is

$$\vec{x} = \overline{z}\vec{\sigma}z, \tag{10.3}$$

where $\vec{\sigma}$ are the 2×2 traceless hermitian matrices (normalized to $\mathrm{Tr}\,\sigma_i\sigma_j = 2\delta_{ij}$ and known as the Pauli spin matrices). This suggests that $\mathcal{F}$-invariant instantons on $\mathbb{C}^2$ might be related to interesting objects on $\mathbb{R}^3$, but we should expect something special to happen at the origin of $\mathbb{R}^3$, which corresponds to the fixed point at the origin of $\mathbb{C}^2$.

The description of $\mathcal{F}$-invariant instantons on $\mathbb{C}^2$ is somewhat subtle. If the action of $\mathcal{F}$ on $\mathbb{C}^2$ is lifted to an action on a G-bundle $\widehat{E} \to \mathbb{C}^2$, then in particular $\mathcal{F}$ acts on the fiber $\widehat{E}_0$ of $\widehat{E}$ at the fixed point. Such an action is characterized by a homomorphism $\rho : \mathcal{F} \cong U(1) \to G$. We recall that such a homomorphism can be interpreted as a weight of ${}^LG$ and determines an 't Hooft operator $T(\rho)$. Kronheimer considers $\mathcal{F}$-invariant instantons on $\mathbb{C}^2$ with a given choice of ρ, and shows that they are equivalent to solutions of the Bogomolny equations on $\mathbb{R}^3$ with a singularity at the origin which in our language represents the insertion of the operator $T(\rho)$.

Pauly then shows that the contribution Δ_ρ of a singularity of type ρ to the virtual dimension of $\mathcal{Z}$ can be computed from the contribution of the fixed point at the origin in $\mathbb{C}^2$ to the $\mathcal{F}$-equivariant index of the linear operator that computes the deformations of instantons on $\mathbb{C}^2$. Let $\mathrm{ad}(\widehat{E})$ be the adjoint bundle derived from $\widehat{E}$, and let $\mathrm{ad}(\widehat{E})_0$ denote its fiber at the origin. The fixed point contribution to the index involves a trace in $\mathrm{ad}(\widehat{E})_0$. (The adjoint representation comes in because deformation theory of instantons involves an elliptic operator acting on the adjoint bundle.) In view of Pauly's computation in section 4.2, the result can be described as follows. The action of $\mathcal{F} \cong U(1)$ on $\mathrm{ad}(\widehat{E})_0$, which is obtained by composing

[42] Kronheimer actually considers a more general situation with $\mathbb{C}^2$ replaced by a more general four-dimensional hyper-Kahler manifold with $U(1)$ action, and gets a description of solutions of the Bogomolny solutions on $\mathbb{R}^3$ with a more general set of singularities.

$\rho : \mathcal{F} \to G$ with the adjoint representation of G, decomposes as a sum of characters. Any character of $\mathcal{F}$ takes the form $\exp(i\theta) \to \exp(im\theta)$ for some integer m. As the adjoint representation of G is real, the nonzero integers appearing in the decomposition of ρ into characters come in pairs $c_k, -c_k$, where we can take c_k to be positive. Then the contribution of the 't Hooft operator $T(\rho)$ to the virtual dimension of $\mathcal{Z}$ is

$$\Delta_\rho = \sum_k c_k. \tag{10.4}$$

This generalizes the case of $G = PSU(2)$, in which only a single integer c appears. Pauly's analysis immediately extends from $SU(2)$ to any G because the equivariant index theorem works for any G. The computation involves a trace in the Lie algebra, which leads to (10.4).

With no 't Hooft operators at all, the virtual dimension with our boundary conditions is zero. With a single 't Hooft operator $T(\rho)$, the virtual dimension is therefore Δ_ρ. But with a single 't Hooft operator, $\mathcal{Z}$ is just the space $\mathcal{Y}(\rho)$ of Hecke modifications of type ρ. The formula (10.4) for the dimensions of spaces of Hecke modifications agrees with the result (9.25) for $U(N)$ (or $SU(N)$ or $PSU(N)$) and generalizes it to arbitrary compact G.

Since the contribution of an 't Hooft operator to the index theory is local, we can immediately write the general result for the dimension of the moduli space $\mathcal{Z}(\rho_1, \dots, \rho_n)$ of solutions of the Bogomolny equations with an arbitrarily prescribed set of 't Hooft operator insertions. If there are no 't Hooft operators, the dimension of $\mathcal{Z}$ is zero. Each 't Hooft operator $T(\rho_i)$ associated with a homomorphism $\rho_i : U(1) \to G$ contributes Δ_{ρ_i} to the dimension, so in general the dimension of $\mathcal{Z}(\rho_1, \dots, \rho_n)$ is

$$\Delta = \sum_i \Delta_{\rho_i}. \tag{10.5}$$

10.2. Monopole Bubbling

In section 9.2, we found that the natural compactification $\overline{\mathcal{Y}}({}^L w)$ of the space $\mathcal{Y}({}^L w)$ of Hecke modifications of some given type ${}^L w$ is achieved by including Hecke modifications of various associated types with smaller weights.

We can now see why this is true from the point of view of the Bogomolny equations. Consider a solution of the Bogomolny equations on $\mathbb{R}^3$ with a specified singularity at the origin. It corresponds to an equivariant instanton

on $\mathbb{C}^2 \cong \mathbb{R}^4$. The instanton equations on $\mathbb{R}^4$ are scale-invariant, so a phenomenon can occur that is called "bubbling" in the mathematical literature: an instanton can shrink to a point. In the case of an $\mathcal{F}$-equivariant instanton, this point will automatically be the origin in $\mathbb{C}^2$.

When interpreted in three dimensions using the relation [110] between equivariant instantons and solutions of the Bogomolny equations in three dimensions with point singularities (which we interpret in terms of 't Hooft operators) this gives a phenomenon that we might call monopole bubbling. Monopole bubbling (which was also discovered in unpublished work by Kronheimer related to [110]) is relatively unfamiliar because it does not occur for smooth solutions of the Bogomolny equations on $\mathbb{R}^3$ (or other complete Riemannian manifolds) with the usual boundary conditions. On the contrary, smooth monopoles on $\mathbb{R}^3$ have a discrete topological classification and a characteristic size, which involves the asymptotic behavior at infinity on $\mathbb{R}^3$. Monopole bubbling can occur only in the presence of an 't Hooft operator; it is a process in which a quantized unit of charge becomes concentrated at a point (the position of the 't Hooft operator) and disappears.

We will now very briefly use the ADHM construction [125] of instantons on $\mathbb{R}^4 \cong \mathbb{C}^2$ to determine how bubbling of equivariant instantons – or in other words monopole bubbling – affects the weight ${}^L w$ of an 't Hooft operator for the case of gauge group $G = SU(N)$. (A similar analysis could be made for other groups for which there is an ADHM construction, namely the classical groups $SO(N)$ and $Sp(N)$. But for exceptional groups there is no ADHM construction.) We will see that bubbling changes the weight in precisely the expected fashion.

We will follow [126], section 3.3. To describe $SU(N)$ instantons on $\mathbb{R}^4$, let U be a two-dimensional vector space (which arises as one of the spin bundles of $\mathbb{R}^4$) on which $\mathcal{F}$ acts by the matrix

$$\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}. \tag{10.6}$$

For any vector space V with $\mathcal{F}$ action, we write $\chi(V)$ for the character of $\mathcal{F}$; for instance, $\chi(U) = e^{i\theta} + e^{-i\theta}$. To construct instantons of instanton number k, let $\mathcal{H}$ be a k-dimensional vector space. And let E_∞ be a vector space of rank N (which turns out to be the fiber at infinity of the instanton bundle). The instanton is built from a special kind of complex

$$0 \longrightarrow \mathcal{H} \xrightarrow{\alpha(z)} \mathcal{H} \otimes U \oplus E_\infty \xrightarrow{\beta(z)} \mathcal{H} \longrightarrow 0, \tag{10.7}$$

where $\beta(z)\alpha(z) = 0$. For our limited purposes here, we do not need the details. $\alpha(z)$ and $\beta(z)$ depend linearly on the complex coordinates z_1, z_2 of $\mathbb{C}^2$. For a smooth instanton, the only nonzero cohomology group of this complex, for any $z = (z_1, z_2) \in \mathbb{C}^2$, is $H^1(z) = \ker \beta(z)/\mathrm{im}\, \alpha(z)$. The fiber at z of the instanton bundle E derived from the ADHM data is $E_z = H^1(z)$.

To get an $\mathcal{F}$-equivariant instanton, we pick an action of $\mathcal{F}$ on E_∞ and $\mathcal{H}$, and choose the complex (10.7) to be $\mathcal{F}$-equivariant. The fiber of the instanton bundle at the origin, which we denote E_0, is in particular the first cohomology group of the corresponding complex:

$$0 \longrightarrow \mathcal{H} \xrightarrow{\alpha(0)} \mathcal{H} \otimes U \oplus E_\infty \xrightarrow{\beta(0)} \mathcal{H} \longrightarrow 0, \tag{10.8}$$

For an equivariant complex $0 \to V_0 \xrightarrow{\alpha} V_1 \xrightarrow{\beta} V_2 \to 0$ whose only cohomology is H^1, the character of H^1 can be expressed as $\chi(H^1) = -\chi(V_0) + \chi(V_1) - \chi(V_2)$. So in particular

$$\chi(E_0) = \chi(E_\infty) + \chi(\mathcal{H})(e^{i\theta} + e^{-i\theta} - 2). \tag{10.9}$$

$\chi(E_0)$ is important because, according to Kronheimer's construction, it determines the weights of the 't Hooft operator that appears at the origin of $\mathbb{R}^3 = \mathbb{C}^2/\mathcal{F}$ when the equivariant instanton is interpreted in three dimensions. If $\chi(E_0) = \sum_{i=1}^N \exp(im_i\theta)$, then the weights are ${}^L w = (m_1, m_2, \dots, m_N)$.

Instanton bubbling occurs when α and β are varied in such a way that, for some decomposition $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$, the complex (10.7) decomposes as the direct sum of a zero complex

$$0 \longrightarrow \mathcal{H}_1 \xrightarrow{0} \mathcal{H}_1 \otimes U \xrightarrow{0} \mathcal{H}_1 \longrightarrow 0 \tag{10.10}$$

and a complex

$$0 \longrightarrow \mathcal{H}_2 \xrightarrow{\alpha(z)} \mathcal{H}_2 \otimes U \oplus E_\infty \xrightarrow{\beta(z)} \mathcal{H}_2 \longrightarrow 0 \tag{10.11}$$

with cohomology only in dimension 1. The "bubbled" instanton bundle E', whose instanton number is reduced by the rank of $\mathcal{H}_1$, has for its fiber the first cohomology group of this second complex, and by analogy with (10.9), the character of its fiber at the origin is $\chi(E_0') = \chi(E_\infty) + \chi(\mathcal{H}_2)(e^{i\theta} + e^{-i\theta} - 2)$. Hence

$$\chi(E_0') = \chi(E_0) + \chi(\mathcal{H}_1)(2 - e^{i\theta} - e^{-i\theta}). \tag{10.12}$$

This formula determines the change in weights under instanton bubbling. The weights m_i of E_0 and the weights m_i' of E_0' are related by

$$\sum_{i=1}^{N} \exp(im_i'\theta) = \sum_{i=1}^{N} \exp(im_i\theta) + \chi(\mathcal{H}_1)(2 - e^{i\theta} - e^{-i\theta}). \tag{10.13}$$

For example, if $\mathcal{H}_1$ is one-dimensional with character $\exp(im\theta)$ for some integer m, then the list of weights of E_0 must include $m+1$ and $m-1$ (since $\chi(E_0')$ is a Fourier series with nonnegative coefficients). (10.13) implies that precisely two weights change in this bubbling process, by $(m+1, m-1) \to (m, m)$. This corresponds to the simplest degeneration of 't Hooft/Hecke operators that we discussed in section 9.2. For another example, suppose $\mathcal{H}_1$ has rank two, with character $\exp(im\theta) + \exp(im'\theta)$. If $|m - m'| > 1$, we just get a pair of moves of the kind just described, with weights $(m+1, m-1, m'+1, m'-1)$ replaced by (m, m, m', m'). But if $m' = m+1$, then (10.13) allows another move with weights $(m+2, m-1)$ replaced by $(m+1, m)$. This corresponds to another expected degeneration of 't Hooft/Hecke operators. More generally, if $\mathcal{H}_1$ has rank s and weights $m, m+1, \dots, m+s-1$, then (10.13) leads to a move[43] in which weights $(m+s, m-1)$ of E_0 are replaced by weights $(m+s-1, m)$ of E_0'. The expected degenerations of 't Hooft/Hecke operators for $G = SU(N)$ are all compositions of moves of this type. For a general G, the expected degenerations can be generated by bubbling of $SU(2)$ instantons embedded in various subgroups of G.

Note in particular that while in the absence of $\mathcal{F}$-invariance, instantons can be assumed to bubble one at a time, this is not so once one imposes $\mathcal{F}$-invariance. There are irreducible processes that involve the simultaneous bubbling of an arbitrarily large number of $\mathcal{F}$-invariant instantons. See [112,113] for detailed description of some Bogomolny moduli spaces on $\mathbb{R}^3$ in which monopole bubbling phenomena can be seen, in a suitable limit.

Hyper-Kahler Structure At Singularities In section 9.2, we noted in an example that although the compactified space $\overline{\mathcal{Y}}({}^L w)$ of Hecke modifications of a given type ${}^L w$ is not a hyper-Kahler manifold, its singularities do have a hyper-Kahler structure. Now we can explain why. The singularities come from monopole bubbling. Monopole bubbling is governed by the

[43] A complex of the appropriate ADHM form to realize this move can be achieved with $\mathcal{H} = \mathcal{H}_1$, $\mathcal{H}_2 = 0$ and E' being the trivial bundle with fiber E_∞. With the characters of E_0' and $\mathcal{H}_1$ given in the text, there is an essentially unique $\mathcal{F}$-invariant choice of τ_1, τ_2, σ, and π that obeys eqn. (3.3.11) of [126] and gives a smooth instanton bundle E, with instanton number s, that can degenerate to E' by bubbling.

moduli space of $\mathcal{F}$-equivariant instantons on $\mathbb{C}^2 = \mathbb{R}^4$. But instanton moduli spaces on $\mathbb{R}^4$ have a hyper-Kahler structure inherited from the hyper-Kahler structure of $\mathbb{R}^4$ (and manifest in the ADHM construction). Since the group $\mathcal{F}$ preserves the hyper-Kahler structure of $\mathbb{R}^4$, the space of $\mathcal{F}$-equivariant instantons is hyper-Kahler (and again, the ADHM construction makes this manifest).

We can also argue in three-dimensional terms, without using the relation to equivariant instantons in four dimensions, that monopole bubbling, and hence the singularities of $\overline{\mathcal{Z}}({}^L w_1, \dots, {}^L w_n)$, is governed by a hyper-Kahler moduli space. Indeed, the moduli space of solutions of the Bogomolny equations on $\mathbb{R}^3$ is hyper-Kahler [109].[44] The spaces of $\overline{\mathcal{Z}}({}^L w_1, \dots, {}^L w_n)$ do not come from solutions of the Bogomolny equations on $\mathbb{R}^3$, but on the more complicated three-manifold $W = I \times C$ with rather particular boundary conditions. So $\overline{\mathcal{Z}}({}^L w_1, \dots, {}^L w_n)$ is not hyper-Kahler. However, monopole bubbling is a local phenomenon. When a unit of monopole charge shrinks to a point, the local structure does not depend on the details of the three-manifold in which this is occurring. So the structure near the singularity is hyper-Kahler, just as if one is on $\mathbb{R}^3$.

We give further details presently on the deformation and resolution of these hyper-Kahler singularities.

10.3. Kahler Structure And The Moment Map

Our next goal is to exhibit $\mathcal{Z}({}^L w_1, \dots, {}^L w_n)$ as a Kahler manifold, with a complex structure that agrees with the appropriate moduli space of Hecke modifications, as described in section 8. This analysis will make many things clearer. It will give a new motivation for the choice of boundary conditions that we used in defining $\mathcal{Z}$; it will give a more precise framework for understanding the relation between 't Hooft and Hecke operators; and it will make more transparent the properties of the singularities of the space of Hecke modifications.

As in the introduction to section 8, we take the metric of the three-manifold $W = I \times C$ to be of the form $ds^2 = h|dz|^2 + dy^2$, where $z = x^2 + ix^3$

[44] This can be shown using the same sort of arguments we use momentarily to exhibit a Kahler structure for the moduli space of solutions of the Bogomolny equations on $W = I \times C$. Those arguments also work if I is replaced by $\mathbb{R}$. If in addition $C = \mathbb{R}^2$, then $W = \mathbb{R}^3$ can be written as $\mathbb{R}^2 \times \mathbb{R}$ in many different ways, leading to a hyper-Kahler structure, not just the Kahler structure that we obtain in section 10.3.

is a local complex coordinate on C and y parametrizes I. We consider the Bogomolny equations $F = \star D\phi_0$, where F is the curvature of a connection A on a G-bundle E. At the left end of W, we specify a flat G-bundle E_-, and divide only by gauge transformations that equal 1 on the boundary; at the right end, we require $\phi_0 = 0$ with and divide by all gauge transformations.

We expand the connection as $A = A_z dz + A_{\overline{z}} d\overline{z} + A_y dy = A_2 dx^2 + A_3 dx^3 + A_y dy$. By rewriting (9.7) and (9.8), the Bogomolny equations in these coordinates can be written as a complex equation

$$[D_y - i\phi_0, D_{\overline{z}}] = 0 \tag{10.14}$$

and a real equation

$$F_{23} = hD_y\phi_0. \tag{10.15}$$

On the space $\mathcal{C}$ of fields (A, ϕ_0), we introduce a complex structure $\mathcal{I}$ in which the complex coordinates are $A_{\overline{z}}$ and $B = A_y - i\phi_0$. The rationale for this choice is that the complex equation in (10.14) is holomorphic in this complex structure. Furthermore, we endow $\mathcal{C}$ with the symplectic structure

$$\omega = \int_W \mathrm{Tr}\Big(\delta A_2 \wedge \delta A_3 - h\,\delta A_1 \wedge \delta\phi_0\Big) dx^2\, dx^3\, dy. \tag{10.16}$$

The complex structure $\mathcal{I}$ and symplectic structure ω combine to a Kahler structure, the Kahler metric being

$$-\int_W \mathrm{Tr}\Big(\delta A_2 \otimes \delta A_2 + \delta A_3 \otimes \delta A_3 + h\,\delta A_y \otimes \delta A_y + h\,\delta\phi_0 \otimes \delta\phi_0\Big) dx^2\, dx^3\, dy. \tag{10.17}$$

Now we want to find the moment map μ for the action of the gauge group. As usual, the moment map $\mu(v)$ for a symplectic vector field v is characterized by

$$d\mu(v) = \iota_v \omega, \tag{10.18}$$

where ι_v is the operation of "contracting with v." For the vector field characterized by the infinitesimal gauge transformation that acts by $\delta A = -D\epsilon$ and $\delta\phi_0 = [\epsilon, \phi_0]$, with $\epsilon \in \Omega^0(C, \mathrm{ad}(E))$, we find the moment map to be

$$\mu(\epsilon) = \int_W \mathrm{Tr}\,(\epsilon(\,F_{23} - hD_y\phi_0))\, dx^2\, dx^3\, dy + \int_{\partial W} \mathrm{Tr}\,(\epsilon\phi_0)\, h\, dx^2\, dx^3. \tag{10.19}$$

The calculation is similar to that in deriving eqn. (4.11), except that a boundary term appears in integration by parts.

In the presence of 't Hooft operators at points $p_i \in W$, we modify the definitions slightly. A and ϕ_0 are required to have prescribed singularities near p_i. And ϵ is constrained near p_i to generate a gauge transformation that leaves fixed the chosen singularity. Our formulas for the moment map remain valid in this situation.

We want to interpret the real Bogomolny equation (10.15) in terms of the vanishing of the moment map. In fact, for $\mu(\epsilon)$ to vanish does give us (10.15) together with another condition: $\mathrm{Tr}\,\epsilon\phi_0$ must vanish on the boundary. On any given boundary component $C_\pm$ of W, there are two simple ways to ensure that $\mathrm{Tr}\,\epsilon\phi_0 = 0$. We can set $\epsilon = 0$ on the boundary, in other words we can allow only gauge transformations that are trivial on this boundary component. Or we can set $\phi_0 = 0$. To define the moduli space $\mathcal{Z}$, we have used (Figure 16) one of these boundary conditions on C_-, and the other on C_+. Thus we have a new rationale for our boundary conditions: they are the simplest ones that allow us to interpret the Bogomolny equations in terms of a holomorphic equation together with the vanishing of the moment map.

Now we can invoke the usual correspondence between symplectic quotients and complex quotients. Let $\mathcal{C}_0$ be the subspace of $\mathcal{C}$ characterized by vanishing of the holomorphic Bogomolny equation (10.14). The moduli space $\mathcal{Z}$ is defined by setting μ to zero and dividing by the group $\mathcal{G}$ of G-valued gauge transformations. We can compare it to $\mathcal{Z}' = \mathcal{C}_0/\mathcal{G}_{\mathbb{C}}$, the quotient of $\mathcal{C}_0$ by the complexified group $\mathcal{G}_{\mathbb{C}}$ of $G_{\mathbb{C}}$-valued gauge transformations.

Of course, because of the boundary conditions, $\mathcal{G}$ is the group of gauge transformations that are trivial at the left end of W, and likewise $\mathcal{G}_{\mathbb{C}}$ is the group of $G_{\mathbb{C}}$-valued gauge transformations that are trivial at that end.

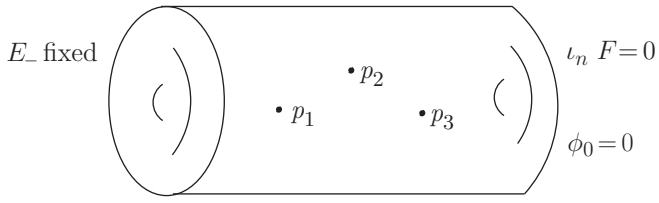

Figure 16: Comparing 't Hooft and Hecke operators on a three-manifold $W = I \times C$. On the left, a connection A and bundle E_- are specified, and the gauge transformations are required to be trivial. On the right, one sets $\phi_0 = 0$ and solves for E_+.

Usually, in comparing a symplectic quotient to the corresponding complex quotient, one has to worry about possible unstable and semistable orbits. In the present case, however, because the gauge transformations are restricted to be trivial at one end of W, such issues are avoided. $\mathcal{G}_{\mathbb{C}}$ acts freely and with closed orbits, so all orbits are stable and $\mathcal{Z} = \mathcal{Z}'$.

It is straightforward to analyze $\mathcal{Z}'$. In fact, we have essentially done so in section 9.1. For any $y \in I$ away from the location of an 't Hooft operator, let $E_y = E|_{\{y\}\times C}$. The complex Bogomolny equation says that the holomorphic type of E_y is independent of y, as long as one does not cross the position of an 't Hooft operator. Moreover, this holomorphic type is the only $\mathcal{G}_{\mathbb{C}}$-invariant in an interval without 't Hooft operators. In jumping across 't Hooft operators, E_y undergoes Hecke modifications. So $\mathcal{Z}'$ is the space of Hecke modifications of E_- due to the prescribed singularities.

This comparison of complex and symplectic quotients thus gives a precise framework for understanding the relationship between 't Hooft and Hecke operators, which we compared more informally in section 9.0. In addition, it will enable us to understand better some properties of the spaces of Hecke modifications that were described in section 9.1.

Changing the Complex And Symplectic Structures Let us denote the points p_i in $W = I \times C$ at which 't Hooft operators are inserted as $y_i \times z_i$, with $y_i \in I$ and $z_i \in C$. We want to understand the dependence of the moduli space $\mathcal{Z}$ on the points p_i.

It is convenient to introduce the modified connection $A' = A - i\phi_0 dy$, the idea being that if we expand $A' = dz\, A'_z + d\overline{z}\, A'_{\overline{z}} + dy\, A'_y$, then A'_y and $A'_{\overline{z}}$ are holomorphic in complex structure $\mathcal{I}$, although A'_z is not. Since A is real, that is, it takes values in the real Lie algebra of G, A'_z (which equals A_z) is the complex conjugate of $A'_{\overline{z}}$; of course, A'_y obeys no reality condition. To describe a holomorphic group action on the space $\mathcal{C}$ of pairs (A, ϕ_0), which is the same as the space of modified connections A', it suffices to describe how $A'_{\overline{z}}$ and A'_y transform. The transformation of A'_z is then determined by the fact that it is the complex conjugate of $A'_{\overline{z}}$.

Consider $W = I \times C$ as a fiber bundle over C, and let D_I be the group of diffeomorphisms of W that preserve this fibration. Such a diffeomorphism, while keeping z fixed, maps y to $\widetilde{y}(y; z, \overline{z})$, where $\widetilde{y}$ agrees with y on the boundaries of W and $\partial\widetilde{y}/\partial y > 0$. We define an action of D_I on $\mathcal{C}$ by simply saying that $A'_{\overline{z}}$ and A'_y transform in the natural way, by the "pullback."

Note that we cannot transform A' by pullback; for $\beta \in D_I$, the operation $A' \to \beta^*(A')$ would not preserve the fact that A'_z and $A'_{\overline{z}}$ are complex

conjugates. Instead we define $\beta(A') = \beta^*(A') + dz\,\varepsilon$, where ε is determined to retain the condition that A'_z and $A'_{\overline{z}}$ are complex conjugates.

This action of D_I on the space $\mathcal{C}$ of pairs (A, ϕ_0) preserves the complex structure $\mathcal{I}$. It also preserves the complex Bogomolny equation, so it preserves the space $\mathcal{C}_0$ of solutions of that equation. These statements depend on the fact that diffeomorphisms in D_I keep z fixed. They therefore preserve the foliation generated by $\partial/\partial y$ and $\partial/\partial\overline{z}$, and the assertion of flatness along leaves of this foliation – which is the content of the complex Bogomolny equation – is invariant. In addition, D_I maps the group $\mathcal{G}_{\mathbb{C}}$ of $G_{\mathbb{C}}$-valued gauge transformations of $\mathcal{C}$ to itself. So $\mathcal{Z}$, defined as a complex manifold by $\mathcal{Z} = \mathcal{C}_0/\mathcal{G}_{\mathbb{C}}$, is invariant under D_I.

However, the symplectic form ω on $\mathcal{C}$ is definitely not D_I-invariant, and accordingly neither is the moment map for the $\mathcal{G}$ action, or the real Bogomolny equation. So $\mathcal{Z}$, understood as the symplectic quotient of $\mathcal{C}_0$ by the action of $\mathcal{G}$, using the symplectic structure ω, is not invariant under D_I.

The conclusion is that the complex structure of $\mathcal{Z}$ is D_I-invariant, but not its symplectic structure.

Now let us consider 't Hooft operators at points $p_i = y_i \times z_i$. The group D_I can slide the p_i up and down in the y direction, in a fashion that is arbitrary except for one restriction: if two p_i are at the same value of z, then their ordering in the y direction is D_I-invariant. The conclusion is that sliding the points p_i in the y direction, without letting two points at the same z cross each other, changes the symplectic structure of $\mathcal{Z}$ but not its complex structure.

There is also a reverse version of this. The definition of the symplectic structure of $\mathcal{C}$ used the area form of C, but not its complex structure. The same is, therefore, true for the moment map μ. So just given the area form of C, we can define the symplectic quotient $\mathcal{C}//\mathcal{G} = \mu^{-1}(0)/\mathcal{G}$. This is an infinite-dimensional symplectic manifold whose symplectic structure depends only on the area form of C. Now in general, in a symplectic manifold, a family of symplectic submanifolds defined by a varying family of equations has a fixed symplectic structure. In our application, the varying family of equations are the complex Bogomolny equations, which depend on the complex structure of C. So as the complex structure and the positions of the points vary, the complex structure of $\mathcal{Z}$ may change, but the symplectic structure does not.

If 't Hooft operators are inserted at point $p_i = y_i \times z_i \in I \times C$, then changing the z_i with the y_i fixed, and without changing whether or not $z_i = z_j$ for a given pair i, j, is equivalent to a special case of changing the

complex structure of C. So changing the z_i in this way changes the complex structure of $\mathcal{Z}$ but not its symplectic structure.

Comparison To Holomorphic Results We can use these observations to get a new perspective on some facts that we described in section 9.2.

Consider $W = I \times C$ with insertion of two 't Hooft operators $T({}^L w, p)$ and $T({}^L w', p')$ of the indicated weights, inserted at $p = y \times z$ and $p' = y' \times z'$ (Figure 17). We write $\overline{\mathcal{Z}}({}^L w, p; {}^L w', p')$, or more briefly $\overline{\mathcal{Z}}(p; p')$, for the corresponding compactified moduli space with some specified bundle E_- at the left end of W, and likewise we write $\overline{\mathcal{Z}}({}^L w, p)$ for a similar Bogomolny moduli space with only one singularity.

We can vary the symplectic structure of $\overline{\mathcal{Z}}(p; p')$ by moving the points in the I direction keeping the projection to C fixed. Or we can vary the complex structure by moving the points in the C direction keeping the projection to I fixed.

Generically, $\overline{\mathcal{Z}}(p; p')$ is singular. Even the moduli space of Hecke modifications associated with a single 't Hooft operator, such as $T({}^L w_1)$ or $T({}^L w_2)$, is generically singular, as we discussed in section 9.2. Exceptions correspond to minuscule representations of ${}^L G$.

So $\overline{\mathcal{Z}}(p; p')$ is typically singular even if $p \neq p'$. But it becomes more singular (or develops a singularity if it was previously smooth) when $p' \to p$. Let us denote as $\overline{\mathcal{Z}}^*$ the singular limit at $p' \to p$. We can partially resolve the singularities of $\overline{\mathcal{Z}}^*$ by displacing the points in the I direction, and we can partially deform the singularities by displacing the points in the C direction (Figure 18). This partial deformation of singularities and partial resolution of singularities are topologically equivalent, since we can simultaneously move p and p' in both the I and C directions.

If $z \neq z'$, then $\overline{\mathcal{Z}}(p; p')$ is as a complex manifold a simple product $\overline{\mathcal{Z}}({}^L w, p) \times \overline{\mathcal{Z}}({}^L w', p')$, parameterizing independent Hecke transformations at the two points z and z' in C. If $z = z'$ and $y < y'$, it is a fiber bundle over

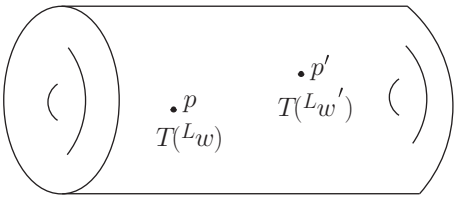

Figure 17: Insertion of two 't Hooft operators of weights ${}^L w$ and ${}^L w'$ at points $p, p' \in W$.

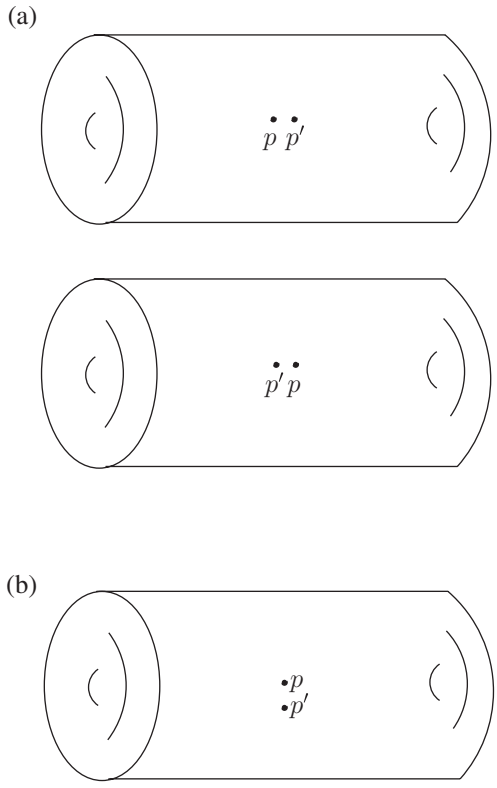

Figure 18: (a) Two partial resolutions of singularity. The moduli space is a fiber bundle parameterizing successive Hecke modifications at the same point $z \in C$. (b) A partial deformation of singularity. The moduli space is a product, parameterizing independent Hecke modifications at distinct points $z, z' \in C$.

$\overline{\mathcal{Z}}({}^L w, p)$ with fiber $\overline{\mathcal{Z}}({}^L w', p')$, describing successive Hecke modifications of E_- at the point $z \in C$ with weights ${}^L w$ and ${}^L w'$. If $z = z'$ and $y' < y$, then $\overline{\mathcal{Z}}(p; p')$ is again a fiber bundle, but the roles of ${}^L w$ and ${}^L w'$ are reversed.

What is the space $\overline{\mathcal{Z}}^*$ that has the simultaneous deformations and resolutions just described? It is in fact $\overline{\mathcal{Z}}({}^L \overline{w}, p')$, the compactified moduli space of solutions of the Bogomolny equations in the presence of a single 't Hooft operator of weight ${}^L \overline{w} = {}^L w + {}^L w'$.

From a physical point of view, this statement is part of the operator product expansion of the operators $T({}^L w, p)$ and $T({}^L w', p')$. We discussed some general properties of this expansion in section 6.3. Simple arguments that we explained there suffice to show that the product $T({}^L w, p) \cdot T({}^L w', p')$

has an expansion as $\sum_{i\in\mathcal{R}} c_i T({}^L w_i, p')$, for some set $\mathcal{R}$ of weights ${}^L w_i$ and coefficients c_i. S-duality predicts the detailed form of this expansion (which should match the representation ring of ${}^L G$), but it is difficult to verify this prediction directly. However, the simple arguments of section 6.3 do suffice to show that the form of the expansion as $p \to p'$ is

$$T({}^L w, p)\cdot T({}^L w, p') \longrightarrow T({}^L \overline{w}, p') + \sum_{i\in\mathcal{R}'} c_i T({}^L w_i, p'), \tag{10.20}$$

where ${}^L\overline{w} = {}^L w + {}^L w'$, and the weights ${}^L w_i$, $i \in \mathcal{R}'$ are the dominant weights such that ${}^L\overline{w} - {}^L w_i$ is also dominant. This means that ${}^L w_i$ is what we have called an associated weight of ${}^L\overline{w}$.

The formula (10.4) shows that the dimension of $\overline{\mathcal{Z}}({}^L\overline{w}, p')$ precisely equals the dimension of $\overline{\mathcal{Z}}({}^L w, p; {}^L w', p')$. This is consistent with the statement that the former is the limit of the latter for $p \to p'$. Other spaces $\overline{\mathcal{Z}}({}^L w_i, p')$, $i \in \mathcal{R}'$ have strictly smaller dimension. In fact, in a certain sense, the spaces $\overline{\mathcal{Z}}({}^L w_i, p')$ all appear in the limit of $\overline{\mathcal{Z}}({}^L w, p; {}^L w', p')$ for $p \to p'$. Since the ${}^L w_i$ are associated weights of ${}^L\overline{w}_i$, the natural compactification $\overline{\mathcal{Y}}({}^L\overline{w})$ of the space of Hecke modifications of type ${}^L\overline{w}$ includes the spaces $\overline{\mathcal{Y}}({}^L w_i)$ of Hecke modifications of types ${}^L w_i$, for all $i \in \mathcal{R}'$. In terms of Bogomolny moduli spaces, this means that the spaces $\overline{\mathcal{Z}}({}^L w_i, p')$ are all contained in the singular "monopole bubbling" locus of $\overline{\mathcal{Z}}({}^L\overline{w}, p')$.

This is why it is difficult to determine the operator product expansion for 't Hooft operators; the subleading terms in (10.20) all come entirely from singularities of $\overline{\mathcal{Z}}({}^L\overline{w}, p')$. In section 10.4, we will explain certain mathematical results [96-98] which from a physical point of view are equivalent to determining the full operator product expansion.

First, however, we want to recall an example analyzed in section 9.2 that provides a good illustration of these arguments. Take $G = PSU(2)$, ${}^L G = SU(2)$, ${}^L w = {}^L w' = (1/2, -1/2)$, and ${}^L\overline{w} = {}^L w + {}^L w' = (1, -1)$. The weights ${}^L w$ and ${}^L\overline{w}$ correspond to the two-dimensional and the adjoint representations of ${}^L G = SU(2)$, respectively. As we learned in section 9.2, the spaces $\overline{\mathcal{Z}}({}^L w, z)$ and $\overline{\mathcal{Z}}({}^L w', z)$ are copies of $\mathbb{CP}^1$. The space $\overline{\mathcal{Z}}^* = \overline{\mathcal{Z}}({}^L\overline{w}, z)$ is the weighted projective plane $\mathbb{WCP}^2(1,1,2)$ and has an orbifold singularity. We have found that this singular space can be deformed into a product $\mathbb{CP}^1 \times \mathbb{CP}^1$ and also resolved into a Hirzebruch surface which is a $\mathbb{CP}^1$ bundle over $\mathbb{CP}^1$. We also saw that the resolution and deformation are compatible, in the sense that the Hirzebruch surface can be deformed into $\mathbb{CP}^1 \times \mathbb{CP}^1$. All this is in accord with the above general arguments. (However, in this particular example, ${}^L w = {}^L w'$, so the two resolutions are equivalent.)

In this example, there is a unique way to write ${}^L\overline{w}$ as a sum of nonzero positive weights. In general, there will be several ways to do this and thus several choices of partial resolutions and desingularizations of $\overline{\mathcal{Z}}({}^L\overline{w}, z)$ along these lines. For a simple Lie group ${}^LG$ of rank r, there are r fundamental weights ${}^Lw_\gamma$, $\gamma = 1, \dots, r$, such that any dominant weight ${}^L\overline{w}$ can be uniquely expanded ${}^L\overline{w} = \sum_{\gamma=1}^r t_\gamma {}^Lw_\gamma$ with some non-negative integers t_γ. (For example, if ${}^LG = SU(r+1)$, the ${}^Lw_\gamma$ are the highest weights of the minuscule representations $\wedge^\gamma \mathsf{V}$, $\gamma = 1, \dots r$. These are called the fundamental representations.) The maximum partial resolution and desingularization of $\overline{\mathcal{Y}}({}^L\overline{w}, z)$ that we can make using the Bogomolny equations is to relate it to a moduli space with t_γ insertions of 't Hooft operators of type ${}^Lw_\gamma$, for $\gamma = 1, \dots, r$. For $SU(r+1)$, since the fundamental representations are minuscule, this gives a complete resolution or deformation of singularities of the space of Hecke modifications of any given weight.

For other groups, the best we can do is to reduce to the singularities of a product of 't Hooft operators $T({}^Lw_\gamma)$ with fundamental weights. The space of Hecke modifications for such an operator has singularities, associated with monopole bubbling, that can be locally modeled by a hyper-Kahler manifold, but there is no obvious hyper-Kahler resolution of these singularities.

10.4. Operator Product Expansion Of 't Hooft Operators

Let us write the tensor product of representations of ${}^LG$ as

$$ {}^LR \otimes {}^LR' = \bigoplus_{\alpha \in \mathcal{R}} \left(N_\alpha \otimes {}^LR_\alpha \right). \tag{10.21} $$

Here N_α is the vector space $N_\alpha = \mathrm{Hom}_G({}^LR \otimes {}^LR', {}^LR_\alpha)$, where Hom_G is the space of G-invariant linear transformations. The set $\mathcal{R}$ consists of the representations for which N_α is nonzero. It is convenient to also write n_α for the integer which is the dimension of the vector space N_α.

As we discussed in section 6.4, the tensor product of representations is precisely mirrored in the operator product expansion of Wilson operators. In eqn. (6.27), we wrote the operator product expansion for Wilson *loop* operators in terms of c-number coefficients n_α. However, for a static Wilson *line* operator, which we regard as part of the problem of quantization, rather than an operator acting on quantum states, we can write a closely related and more precise statement:

$$ W({}^LR, p) \cdot W({}^L\widetilde{R}, p') \longrightarrow \bigoplus_{\alpha \in \mathcal{R}} \left(N_\alpha \otimes W({}^LR_\alpha, p') \right). \tag{10.22} $$

We are here taking the gauge group to be ${}^LG$, and p and p' are the positions of the static Wilson operators. (10.22) is a recipe for constructing the quantum Hilbert space in the presence of two Wilson operators labeled by ${}^LR$ and ${}^LR'$, as a direct sum of contributions; each contribution is the tensor product with a vector space N_α of the quantum Hilbert space obtained by quantization in the presence of a single Wilson operator labeled by ${}^LR_\alpha$. The basis for the statement (10.22) is that the effect of a static Wilson operator in the quantum theory is to introduce an external charge in the appropriate representation. If there are two Wilson static Wilson operators, there are two external charges; after decomposing the tensor product of the appropriate representations in a direct sum of irreducibles, one arrives at (10.22).

From S-duality, we expect the corresponding operator product expansion of 't Hooft line operators to take the same form

$$T({}^LR,p)\cdot T({}^LR',p')\longrightarrow\bigoplus_{\alpha\in\mathcal{R}}\left(N_\alpha\otimes T({}^LR_\alpha,p')\right). \tag{10.23}$$

As in section 6.4, even without S-duality, since the 't Hooft operators are the only line operators at $\Psi=0$, a relation such as (10.23) must hold with some set $\mathcal{R}$ of representations (which depends on ${}^LR$ and ${}^LR'$) and some vector spaces N_α. Moreover, simple arguments suffice to determine the set $\mathcal{R}$, which depends on ${}^LR$ and ${}^LR'$. So the problem is to describe the N_α.

If we specialize to the case of ordinary 't Hooft loop operators which may act on quantum states (as opposed to static line operators that modify the definition of the quantum Hilbert space), the coefficients in the operator product expansion are ordinary c-numbers n_α rather than vector spaces N_α. One way to make this specialization is to take the "time" direction to be a circle. This leads to a trace in all of the quantum Hilbert spaces, and the vector space N_α is replaced by its integer dimension n_α. The static 't Hooft line operators become circular 't Hooft loop operators (which can act on quantum states if we interpret a different direction as the "time" direction for quantization), and we learn that the operator product expansion coefficients of 't Hooft loop operators, or of any loop operators in a topological field theory, are integers.

To make the point in a slightly different way, let us go back to the case of Wilson operators. A closed Wilson loop operator is defined via the trace of the holonomy. Taking the trace replaces a vector space by its dimension. But an open Wilson operator (the important example for us is a static Wilson line operator) is defined in terms of the holonomy itself, regarded as an

operator on a vector space. In decomposing the tensor products of such operators, the "coefficients" are vector spaces N_α, not integers n_α. For some purposes, the distinction is not important and one can replace N_α by its dimension.

Some Useful Branes Once the scene is properly set, the determination of the N_α is a consequence of a mathematical theory that originated with Lusztig's work [96] on the relation of multiplicities in tensor products of irreducible representations of G to the convolution of intersection cohomology complexes of affine Schubert varieties on ${}^LG$, and was further refined by Ginzburg [97] and Mirkovic and Vilonen [98] to an equivalence of tensor categories. In physical terms, this refinement is related to a description of the operator product expansion of 't Hooft operators in which the "coefficients" are vector spaces N_α (as in our equation (10.23)), as opposed to a more conventional description in which only closed loop operators are considered and the N_α are replaced by their integer dimensions n_α (as in our equation (6.27)). We will set the context for the analysis, get as far as we can with purely physical reasoning, and then attempt to elucidate the mathematical theory.

To determine the N_α, it suffices to consider any convenient and suitably rich class of computations in which static 't Hooft operators play a role. For this, we consider two or more 't Hooft operators in our familiar three-manifold $W = I \times C$ and with suitable branes $\mathcal{B}$ and $\mathcal{B}'$ chosen at the two ends. We want to pick branes that will simplify the computation (fig. 19).

't Hooft operators preserve supersymmetry of type (B, A, A), so it is most natural to choose $\mathcal{B}$ and $\mathcal{B}'$ to be branes of that type. For this, we can take $\mathcal{B}$ and $\mathcal{B}'$ to be supported on subvarieties U and U' of Hitchin's moduli space $\mathcal{M}_H$ that are holomorphic in complex structure I and Lagrangian with respect to the symplectic form Ω_I that is of type $(2, 0)$ in this complex

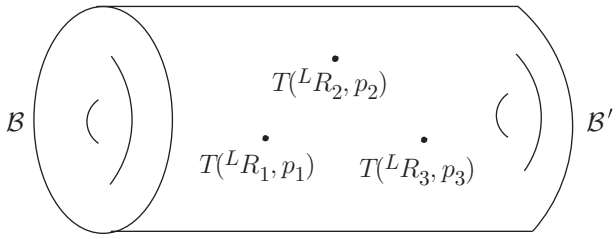

Figure 19: Insertion of several 't Hooft operators $T({}^LR_i, p_i)$, with a brane $\mathcal{B}$ on the left that specifies an "initial" bundle E_-, and a brane $\mathcal{B}'$ on the right that sets $\varphi = 0$.

structure. (More briefly, U and U' are complex Lagrangian submanifolds in complex structure I.) We endow U and U' with trivial Chan-Paton line bundles.

We take U to parametrize Hitchin pairs (E, φ), where we require E to have a specified holomorphic type E_- and arbitrary φ. We require E_- to be stable, but otherwise the choice of E_- does not matter. And we take U' to parametrize pairs $(E, 0)$ with any E but with $\varphi = 0$.

The first point is that U and U' intersect transversely at a single point, corresponding to the Higgs bundle $(E_-, 0)$. Moreover, U and U' are complex symplectic manifolds in complex structure I, so $\mathcal{B}$ and $\mathcal{B}'$ are branes of type (B, A, A).

Finally, we have chosen $\mathcal{B}$ and $\mathcal{B}'$ so that the space of supersymmetric configurations with these boundary conditions and with some 't Hooft operators included is precisely the familiar moduli space $\mathcal{Z}({}^L R_1, p_1; \dots; {}^L R_k, p_k)$ of solutions of the Bogomolny equations.

A few subtleties go into this statement. First of all, supersymmetric configurations in general have $\varphi \neq 0$ and must be described using the extended Bogomolny equation that we described at the outset of section 9, not the ordinary Bogomolny equations. We have not yet even analyzed the extended Bogomolny equations; this will be the subject of section 10.5. When we do so, we will see that if φ vanishes at one end of W – and it does here, because of the choice of U' – then φ vanishes everywhere, and the extended Bogomolny equations reduce to ordinary ones.

Second, in describing the branes $\mathcal{B}$ and $\mathcal{B}'$, we only described the boundary conditions on the fields (A, φ) of the sigma-model with target $\mathcal{M}_H$. The field ϕ_0, which is an important part of the Bogomolny equations, is present in the gauge theory, but not in the sigma-model (except at singularities) because in reduction to two dimensions it is "massive." By extending the analysis in section 12, the branes $\mathcal{B}$ and $\mathcal{B}'$ can both be described in gauge theory [26], and in that context it can be shown that the appropriate boundary conditions for the field ϕ_0 are the ones we use.

The Space Of Physical States In the A-model, in the absence of instanton corrections (they are absent here for reasons we discussed at the beginning of section 9), the space of physical states is the cohomology of the space of time-independent supersymmetric configurations. In fact, the A-model as understood physically leads most naturally to de Rham cohomology.

In other words, with branes $\mathcal{B}$ and $\mathcal{B}'$, and 't Hooft operators $T({}^L R_i, p_i)$, the space of physical states is the cohomology of our friend, the moduli space $\overline{\mathcal{Z}}({}^L R_1, p_1; \dots; {}^L R_k, p_k)$ of solutions of the Bogomolny equations with

the appropriate singularities. As $\overline{\mathcal{Z}}$ is generically singular, we need to explain here what kind of refinement of de Rham cohomology is meant. But first let us just formally state what sort of result we expect.

For every finite set of representations ${}^LR_i$, $i = 1, \dots, k$, the cohomology of the moduli space $\overline{\mathcal{Z}} = \overline{\mathcal{Z}}({}^LR_1, p_1; \dots; {}^LR_k, p_k)$ gives us a space of physical states $\mathcal{H}({}^LR_1, {}^LR_2, \dots, {}^LR_k) = H^*(\overline{\mathcal{Z}})$. Notice that because of the underlying four-dimensional topological quantum field theory (TQFT), the choices of the points p_i do not matter. To be more exact, the TQFT says that locally in p_i the space of physical states is independent of the p_i (there is a flat connection[45] on the space of distinct points p_i). Since the space of n-tuples in the three-manifold $W = I \times C$ is simply-connected, there is also no room for monodromies, and thus the p_i are irrelevant. So we have labeled the space of physical states $\mathcal{H}({}^LR_1, \dots, {}^LR_k)$ just by the representations, not by the points.

Modulo questions about singularities, we do not really need quantum field theory for this argument. In the case that the representations ${}^LR_i$ are minuscule (and the p_i are distinct), $\overline{\mathcal{Z}}({}^LR_1, p_1; \dots; {}^LR_k, p_k)$ is a family of smooth manifolds. The cohomology of such a family is locally constant, just as expected from the TQFT. In the general case with singularities, we need to make sure to use a refinement of de Rham cohomology in which this sort of argument is valid. The property of $\overline{\mathcal{Z}}({}^LR_1, p_1; \dots; {}^LR_k, p_k)$ that makes this possible is that, although it is singular, its singular structure is completely independent of the p_i. The singularities come from degenerations of individual Hecke transformations (or monopole bubbling at positions of individual 't Hooft operators), independent of the positions of the others.

Now the operator product expansion says that we can replace a product of 't Hooft operators $T({}^LR_1) \cdot T({}^LR_2)$ by a sum $\oplus_{\alpha \in \mathcal{R}} C_\alpha \otimes T({}^LR_\alpha)$, with some set $\mathcal{R}$ and some vector spaces C_α. In the present context, for spaces of physical states in the presence of the branes we have considered, this means we must have a family of isomorphisms

$$\mathcal{H}({}^LR_1, {}^LR_2, \dots, {}^LR_k) \cong \oplus_{\alpha \in \mathcal{R}} \left(C_\alpha \otimes \mathcal{H}({}^LR_\alpha, {}^LR_3, \dots, {}^LR_k) \right). \tag{10.24}$$

Since the left hand side only depends on the unordered set of representations ${}^LR_i$, the isomorphisms in (10.24) are automatically compatible with associativity, that is with taking repeated operator products to further reduce the number of representations.

[45] Moving the points is a special case of changing the metric on W. The change in metric is trivial in the BRST cohomology, a fact which gives the flat connection.

S-duality says that the set $\mathcal{R}$ and vector spaces C_α should be the ones that arise in the tensor product of representations of ${}^LR$:

$$ {}^LR_1 \otimes {}^LR_2 = \oplus_{\alpha\in\mathcal{R}} \left(N_\alpha \otimes {}^LR_\alpha\right). \tag{10.25}$$

We actually can be much more specific if we recall that[46] $\mathcal{H}({}^LR_1, {}^LR_2, \dots, {}^LR_k)$ is the cohomology of $\overline{\mathcal{Z}}({}^LR_1, {}^LR_2, \dots, {}^LR_k)$ and that topologically $\overline{\mathcal{Z}}({}^LR_1, {}^LR_2, \dots, {}^LR_k)$ is a simple product $\prod_{i=1}^k \overline{\mathcal{Z}}({}^LR_i)$. (And $\overline{\mathcal{Z}}({}^LR_i)$ is the same as the compactified space $\overline{\mathcal{Y}}({}^LR_i)$ of Hecke modifications of type ${}^LR_i$.) The de Rham cohomology of a product is the product of the cohomologies of the factors, so we get

$$ \mathcal{H}({}^LR_1, {}^LR_2, \dots, {}^LR_k) = \otimes_{i=1}^k \mathcal{H}({}^LR_i). \tag{10.26}$$

This implies that (10.24) must really come from a family of isomorphisms

$$ \mathcal{H}({}^LR_1) \otimes \mathcal{H}({}^LR_2) = \oplus_{\alpha\in\mathcal{R}} N_\alpha \otimes \mathcal{H}({}^LR_\alpha). \tag{10.27}$$

Comparing (10.25) to (10.27), we see that the $\mathcal{H}({}^LR_i)$ obey the same algebra of tensor products as the ${}^LR_i$. This suggests that there is an isomorphism between $\mathcal{H}({}^LR_i)$ and ${}^LR_i$ that produces the automorphism between the two algebras. Indeed, it can be shown [127] that this is the only way for the families of finite-dimensional vector spaces ${}^LR_i$ and $\mathcal{H}({}^LR_i)$ to have the same algebra of tensor products. The isomorphism between $\mathcal{H}({}^LR_i)$ and ${}^LR_i$ is unique up to conjugation by an element of ${}^LG$.

Grading By Ghost Number Since ${}^LG$ certainly acts on the ${}^LR_i$, in a way compatible with the tensor products, this means that there is a corresponding action of ${}^LG$ on $\mathcal{H}({}^LR_i)$ (defined up to conjugation). What this means is mysterious from a physical point of view, since ${}^LG$ is not a symmetry of the full A-model of gauge group G at $\Psi = 0$, but only of the particular piece of it that we have looked at to analyze the operator product expansion. (Things might become clearer if one could understand the S-duals of the branes $\mathcal{B}$ and $\mathcal{B}'$ that were used in the construction.) However, the full A-model does have one symmetry that is relevant.

This is the "ghost number" symmetry $\mathcal{K}$, whose origin in topological twisting of $\mathcal{N} = 4$ super Yang-Mills theory was explained in section 3.1. In the A-model, the ghost number is, roughly speaking, the degree or dimension

[46] These assertions do not depend on the points $p_i \in W$, so we omit them.

of a cohomology class.[47] There is, however, one key subtlety. For a complex manifold X of complex dimension n, the degree of a differential form is usually understood to run from 0 to $2n$. But the space of physical states of the A-model has a fermion conjugation symmetry[48] that exchanges a class of dimension d with one of dimension $2n-d$ and maps $\mathcal{K} \to -\mathcal{K}$. The ghost number of a state that is related to a cohomology class of degree d is actually[49]

$$\mathcal{K} = d - \dim_{\mathbb{C}}(X). \tag{10.28}$$

In the map from ${}^LR_i$ to $\mathcal{H}({}^LR_i)$, the symmetry $\mathcal{K}$ of the A-model must correspond to a derivation (that is, the generator of an automorphism) of the tensor algebra of representations of ${}^LG$. Any such derivation is actually an element of the complexified Lie algebra ${}^L\mathfrak{g}$. The element of ${}^L\mathfrak{g}$ corresponding to $\mathcal{K}$ is well-defined up to conjugation.

For given ${}^LG$, one can find the right conjugacy class by considering a specific example. For example, for ${}^LG = SU(N)$, consider the N-dimensional representation V. The corresponding space of Hecke modifications is $\overline{\mathcal{Y}}(\mathsf{V}) \cong \mathbb{CP}^{N-1}$, as we learned in section 9.1. The de Rham cohomology of $\mathbb{CP}^{N-1}$ is N-dimensional. The fact that this coincides with the dimension of V is an example of the correspondence ${}^LR \leftrightarrow \mathcal{H}({}^LR)$. $\mathbb{CP}^{N-1}$ has complex dimension $N-1$ and has cohomology in dimensions $d = 0, 2, 4, \dots, 2N-2$. So the element of

[47] For a Kahler target, the Hodge decomposition $H^d(X) = \oplus_{p+q=d} H^{p,q}(X)$ gives a further refinement of the A-model, in general. However, this is not relevant for our present discussion, as the cohomology of the spaces $\overline{\mathcal{Z}}({}^LR_1, \dots, {}^LR_k)$ is all of type (p,p).

[48] The symmetry in question is two-dimensional CPT symmetry. It is an exact symmetry of the physical supersymmetric sigma-model with $(2,2)$ supersymmetry and target X. On a non-flat Riemann surface Σ, this symmetry is spoiled by the topological twisting that defines the A-model. But the space of physical states is obtained by quantizing the sigma-model on the flat manifold $\mathbb{R} \times S^1$, which is unaffected by the twisting.

[49] For completeness, let us note that if X is a Calabi-Yau manifold, then the A-model mapping from states to operators involves a spectral flow that adds $\dim_{\mathbb{C}}(X)$ to $\mathcal{K}$. So if ψ is a class in the d-dimensional cohomology of X, then the A-model state corresponding to ψ has $\mathcal{K} = d - \dim_{\mathbb{C}}(X)$, but the corresponding operator $\mathcal{O}_\psi$ has $\mathcal{K} = d$. However, the spaces $\overline{\mathcal{Z}}({}^LR_1, \dots, {}^LR_k)$ are not Calabi-Yau, and more importantly our discussion of branes and 't Hooft operators refers to the physical states, not to corresponding operators.

${}^L\mathfrak{g}$ corresponding to $\mathcal{K}$ has eigenvalues $N-1, N-3, N-5, \dots, -(N-1)$ in the representation V:

$$\mathcal{K} = \begin{pmatrix} N-1 & & & & & \\ & N-3 & & & & \\ & & N-5 & & & \\ & & & \ddots & & \\ & & & & -(N-1) \end{pmatrix}. \tag{10.29}$$

For ${}^LG = SU(2)$, the conjugacy class of $\mathcal{K}$ is simply

$$\mathcal{K} \cong \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{10.30}$$

Intersection Cohomology A mathematical theory that was initiated by Lusztig [96], and later refined by Ginzburg [97] and Mirkovic and Vilonen [98], in effect determines the "coefficients" in the operator product expansion of 't Hooft operators. Describing these results in detail is a task better left to others. Instead, here we will try to express a few of the ideas in a language that physicists might find illuminating.

First of all, the cohomology theory used in this theorem is intersection cohomology [128,129]. Roughly speaking, and modulo standard conjectures, the intersection cohomology of an algebraic variety X is the $\mathbb{L}^2$ cohomology of the smooth part of X. One considers square-integrable differential forms on the smooth part of X. The intersection cohomology $IH^*(X)$ is the cohomology of the exterior derivative operator d in the space of such square-integrable forms. (It cannot necessarily be expressed in terms of harmonic forms.) To match intersection cohomology with quantum field theory, it is useful to make the same shift in dimensions that arises naturally in the A-model: we define a grading of $IH^*(X)$ such that an $\mathbb{L}^2$ cohomology class that is a d-form is considered to have degree $\mathcal{K} = d - \dim_{\mathbb{C}}(X)$.

To be more precise, the $\mathbb{L}^2$ cohomology of the smooth part of X is the intersection cohomology of X relative to its standard intersection cohomology complex. We might denote this $IH^*(X, IC_X)$. One also defines other intersection cohomology complexes on X. We will really only need the simplest construction. If $Y \subset X$ is a subvariety, one defines an intersection cohomology complex IC_Y which is supported on Y and such that $H^*(X, IC_Y)$ is the $\mathbb{L}^2$ cohomology of the smooth part of Y, shifted in dimension by

$-\dim_{\mathbb{C}}(Y)$. Thus, IC_Y is the standard intersection cohomology complex of Y, but embedded in X (and "extended by zero" away from $Y \subset X$).

In the case of the spaces $\overline{\mathcal{Y}}({}^L\overline{w})$ of Hecke modifications, the subvarieties of interest are the spaces $\overline{\mathcal{Y}}({}^Lw_\alpha)$, where ${}^Lw_\alpha$ is a weight associated to ${}^L\overline{w}$. So for each associated weight, one defines an intersection cohomology complex $IC({}^Lw_\alpha)$, supported on the subspace of $\overline{\mathcal{Y}}({}^L\overline{w})$ on which monopole bubbling has reduced the weight from ${}^L\overline{w}$ down to ${}^Lw_\alpha$. We also denote the standard intersection cohomology complex of $\overline{\mathcal{Y}}({}^L\overline{w})$ as $IC({}^L\overline{w})$.

It will help at this point to recall our knowledge of the spaces of Hecke modifications and consider a simple example (whose relevant properties were described in detail in section 9.2). Let ${}^LG = SU(2)$ and let ${}^L\overline{w} = (1,-1)$ be the weight associated to the adjoint representation. The compactified space of Hecke modifications is then a weighted projective space $\overline{\mathcal{Y}}({}^L\overline{w}) = \mathbb{WCP}^2(1,1,2)$. This space is an orbifold, so its de Rham cohomology is well-defined (with no need to resort to $\mathbb{L}^2$ cohomology). The cohomology is three-dimensional, with generators in degrees $d = 0, 2, 4$. Hence the values of $\mathcal{K} = d - \dim_{\mathbb{C}}(\mathbb{WCP}^2(1,1,2)) = d-2$ are $\mathcal{K} = 2, 0, -2$. This is in accord with expectations; the adjoint representation of $SU(2)$ is three-dimensional, and the element $\mathcal{K}$ of its Lie algebra that is defined in the two-dimensional representation in (10.30) indeed has eigenvalues $2, 0, -2$ in the adjoint representation.

But so far we have not said anything that depends on the singularity of the weighted projective space. To see its role, let ${}^Lw = (1/2, -1/2)$ be the weight corresponding to the two-dimensional representation. The operator product expansion of the corresponding Wilson operator with itself should read

$$T({}^Lw)\cdot T({}^Lw) = T({}^L\overline{w})\oplus T(0), \tag{10.31}$$

where 0 is the zero weight, and $T(0) = 1$ is the corresponding trivial 't Hooft operator. (The coefficients are really one-dimensional vector spaces, but we have omitted them.) Let us see how this comes about. The space of Hecke modifications of type ${}^Lw$ is isomorphic to $\mathbb{CP}^1$. If (Figure 20) we make repeated Hecke modifications of type ${}^Lw$ at the same point $z \in C$ but different points in I, we get a moduli space $\overline{\mathcal{Z}}({}^Lw, z, y; {}^Lw, z, y')$ that is a $\mathbb{CP}^1$ bundle over $\mathbb{CP}^1$. Its cohomology is four-dimensional, with classes in dimensions $d = 0, 2, 2, 4$, so the eigenvalues of $\mathcal{K}$ are $-2, 0, 0, 2$.

Now let us try to take the limit as $y \to y'$. For any $y < y'$, we can use Hodge theory; the space of harmonic forms on $\overline{\mathcal{Z}}$ is four-dimensional. At $y = y'$, an A_1 orbifold singularity develops; the local structure looks

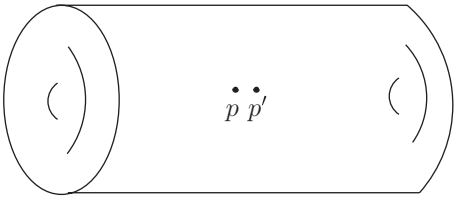

Figure 20: A pair of Hecke modifications made at the same value of z but at distinct points in I.

like $\mathbb{R}^4/\mathbb{Z}_2$. For y slightly less than y' (compared to the length of the interval I), this singularity is resolved. The resolution can be locally modeled by a hyper-Kahler manifold W (discovered in [130]) that in one of its complex structures can be identified with $T^*\mathbb{CP}^1$. The metric on W is complete, and moreover, is asymptotic at infinity to the flat metric on $\mathbb{R}^4/\mathbb{Z}_2$.

Crucially for us, the space of $\mathbb{L}^2$ harmonic forms on W is one-dimensional, generated [130] by a two-form ω. (Note that here the $\mathbb{L}^2$ condition has to do with the behavior at infinity, while in the definition of intersection cohomology it has to do with the behavior near singularities.) Being in $\mathbb{L}^2$, this form has support near the exceptional divisor, and as $y \to y'$ (and the exceptional divisor shrinks), it converges to a form with delta function support at the singularity. Having $d = 2$, it corresponds to a cohomology class with $\mathcal{K} = 0$, the value corresponding to the trivial representation of ${}^LG = SU(2)$. The $\mathbb{L}^2$ harmonic two-form on W exists for topological reasons that can be understood from arguments in [131].

Now we can understand the meaning of the operator product relation (10.31). The left hand side is associated with the space $\overline{\mathcal{Z}}$ of repeated Hecke modifications. For $y < y'$, it is a $\mathbb{CP}^1$ bundle over $\mathbb{CP}^1$, and has a four-dimensional space of harmonic forms, which is the physical Hilbert space $\mathcal{H}$. A three-dimensional subspace of this space consists of harmonic forms that, as $y \to y'$, converge to $\mathbb{L}^2$ harmonic forms on the smooth part of $\overline{\mathcal{Z}}^* = \mathbb{WCP}^2(1,1,2)$ (which is the limit of $\overline{\mathcal{Z}}$ for $y \to y'$). They reflect the contribution of $T({}^L\overline{w})$ to $\mathcal{H}$. The fourth harmonic form can be approximated for y near y' by the $\mathbb{L}^2$ harmonic form on W and is supported near the exceptional divisor. The support of this fourth harmonic form converges as $y \to y'$ to the singular point of $\overline{\mathcal{Z}}^*$, which is the space of Hecke modifications of weight ${}^Lw = 0$. It is the contribution of $T(0)$ to $\mathcal{H}$. The fact that the coefficient of $T(0)$ is 1 expresses the fact that the space of harmonic two-forms that converge to the singularity as $y \to y'$ is precisely one-dimensional.

To be more exact, the coefficient of $T(0)$ is this one-dimensional space of harmonic forms.

More generally, suppose we are given any pair of weights ${}^Lw$, ${}^Lw'$. We want to describe the operator product $T({}^Lw)\cdot T({}^Lw')$, which is expected to take the form

$$T({}^Lw)\cdot T({}^Lw') = T({}^L\overline{w})\oplus\big(\oplus_{\alpha\in\mathcal{R}'}N_\alpha\otimes T({}^Lw_\alpha)\big), \tag{10.32}$$

where ${}^L\overline{w} = {}^Lw + {}^Lw'$, and the sum runs over weights associated to ${}^L\overline{w}$. As $y\to y'$, $\overline{\mathcal{Z}}({}^Lw,y,z;{}^Lw,y',z)$ converges as a space to $\overline{\mathcal{Z}}({}^L\overline{w},y,z)$ (whose dependence on y and z is inessential). Some of the $\mathbb{L}^2$ harmonic forms on the moduli space converge to such forms on the smooth part of $\overline{\mathcal{Z}}({}^L\overline{w})$. These are the contribution of $T({}^L\overline{w})$ in the operator product expansion. Others converge to have delta function support on $\overline{\mathcal{Z}}({}^Lw_\alpha)$ for some associated weight ${}^Lw_\alpha$ of ${}^L\overline{w}$. The forms that for $y\to y'$ have support on $\overline{\mathcal{Z}}({}^Lw_\alpha)$ and not on $\overline{\mathcal{Z}}({}^Lw_\beta)$ for any weight ${}^Lw_\beta$ associated to ${}^Lw_\alpha$ are the contribution of $T({}^Lw_\alpha)$ in the operator product.

Let us simply conclude by describing how this idea has been expressed mathematically [96-98]. The moduli space $\overline{\mathcal{Z}}({}^Lw,z,y;{}^Lw',z,y')$ is a fibration with fiber $\overline{\mathcal{Z}}({}^Lw',z,y')$ and base $\overline{\mathcal{Z}}({}^Lw,z,y)$, as we explained in section 10.2. As $y\to y'$, this moduli space is "blown down" to make the more singular space $\overline{\mathcal{Z}}({}^L\overline{w},z,y')$, whose dependence on z and y' is inessential. The blowdown is a rational map

$$\pi:\overline{\mathcal{Z}}({}^Lw,z,y;{}^Lw',z,y')\longrightarrow\overline{\mathcal{Z}}({}^L\overline{w}). \tag{10.33}$$

We write $IC({}^Lw)\star IC({}^Lw')$ for the standard intersection cohomology complex on $\overline{\mathcal{Z}}({}^Lw,z,y;{}^Lw',z,y')$. (The notation is motivated by the fiber bundle structure of this space.) There is a pushdown map π_* in intersection cohomology. As the sheaf $IC({}^Lw)\star IC({}^Lw')$ is invariant under the group called $GL(N,\mathcal{O})$ in section 9.2 or its analog $G(\mathcal{O})$ for other G, and π_* commutes with this action, the pushdown $\pi_*(IC({}^Lw)\star IC({}^Lw'))$ is likewise invariant under $G(\mathcal{O})$. It therefore is possible to expand it as a sum of the standard intersection cohomology complex $IC({}^L\overline{w})$ of $\overline{\mathcal{Z}}({}^L\overline{w})$, and the complexes $IC({}^Lw_\alpha)$ that are supported on the singular orbits with weights ${}^Lw_\alpha$ that are associated to ${}^L\overline{w}$. (These are the only $G(\mathcal{O})$-invariant intersection cohomology complexes on $\overline{\mathcal{Z}}({}^L\overline{w})$.)

So we must have[50]

$$\pi_*\left(IC({}^L w)\star IC({}^L w')\right) = IC({}^L\overline{w})\oplus\left(\bigoplus_\alpha \widetilde{N}_\alpha\otimes IC({}^L w_\alpha)\right). \tag{10.34}$$

The "coefficient" of the leading term $IC({}^L\overline{w})$ is a trivial one-dimensional vector space (which we can omit), since the map π is an isomorphism over the smooth part of the moduli space. The nature of the leading term has an elementary explanation explained in the discussion of eqn. (6.29). The other "coefficients" are vector spaces $\widetilde{N}_\alpha$. This formulation is analogous to the general structure (6.29) or (10.23) of the product of 't Hooft operators, with the intersection complex $IC({}^L w_\alpha)$ representing the contribution of the 't Hooft operator $T({}^L w_\alpha)$. Of course, the hard part of the proofs [96-98] is to describe the spaces $\widetilde{N}_\alpha$ and establish their isomorphism with the analogous spaces N_α that appear in (10.25).

10.5. The Extended Bogomolny Equations

So far, in our study of the action of 't Hooft operators on branes, we have considered only supersymmetric solutions that are time-independent and time-reversal invariant. With this restriction, the BPS equations (9.1) reduce to the ordinary Bogomolny equations, and the 't Hooft operators reduce to the usual geometric Hecke operators.

The assumption of time-independence can actually be justified in general, but requiring time-reversal invariance puts a severe restriction on the allowed boundary conditions. It is only adequate if branes are chosen (as we did in section 10.3 in discussing the product of 't Hooft operators) to ensure that the relevant solutions have $\varphi = 0$. The most important A-branes for the geometric Langlands duality (the generic fibers of the Hitchin fibrations and the c.c. brane described in section 11) do not have their support limited to $\varphi = 0$, so in order to study the action of 't Hooft operators on them, we need to drop the condition of time-reversal invariance. This is the purpose of the following discussion. Considerably more detail will appear elsewhere [26].

We consider again the BPS equations on a four-manifold $M = \mathbb{R}\times W$ and write the Higgs field ϕ as $\phi_0\, dx^0 + \pi^*\widetilde{\phi}$, where π is the projection to W, and $\widetilde{\phi}$ is a one-form Higgs field on W. Likewise, we write the gauge field as

[50] We have shifted all complexes in dimension to be symmetric between positive and negative dimensions. The map π_* preserves this symmetry. So we do not need to be concerned about further shifts.

$A = A_0\,dx^0 + \widetilde{A}$, where $\widetilde{A}$ is a three-dimensional connection with curvature $\widetilde{F}$. The time-independent BPS equations for $t = 1$ read

$$\begin{aligned} \widetilde{F} - \widetilde{\phi}\wedge\widetilde{\phi} &= \star\left(D\phi_0 - [A_0,\widetilde{\phi}]\right), \\ \star D\widetilde{\phi} &= [\phi_0,\widetilde{\phi}] + DA_0, \\ D^*\widetilde{\phi} + [A_0,\phi_0] &= 0. \end{aligned} \tag{10.35}$$

The exterior derivative D, the Hodge star operator $\star$, and the operator $D^* = \star D\star$ are all understood in the three-dimensional sense.

These equations can be greatly simplified using suitable vanishing theorems. In fact, one can show that $\phi_1 = A_0 = 0$ using the same arguments, and the same type of assumptions, as in the time-reversal invariant case (recall eqn. (9.4)).

So we set $\phi_1 = A_0 = 0$, choose a local complex coordinate $z = x^2 + ix^3$ on C, denote $y = x^1$, and work in the gauge $\widetilde{A}_y = 0$. We also now drop the tildes from our notation. And, as in section 9.0, we take the metric on W to be $ds^2 = h|dz|^2 + dy^2$, for some function h. Then the extended Bogomolny equations (10.35) reduce to

$$\begin{aligned} D_{\overline{z}}\phi_z &= 0, \\ \partial_y A_{\overline{z}} &= -iD_{\overline{z}}\phi_0, \\ \partial_y\phi_z &= -i[\phi_z,\phi_0], \\ F_{z\overline{z}} - [\phi_z,\phi_{\overline{z}}] &= \frac{ih}{2}\partial_y\phi_0. \end{aligned} \tag{10.36}$$

The first of the equations (10.36) says that $\varphi = \phi_z\,dz$, restricted to $C_y = \{y\}\times C$, is a holomorphic section of $\mathrm{End}(E)\otimes K_C$ for any $y\in I$. So denoting as E the holomorphic bundle over C_y defined by the $\overline{\partial}$ operator $\overline{D} = d\overline{z}(\partial_{\overline{z}} + A_{\overline{z}})$, the pair (E,φ) is a Higgs bundle for any y. The next two equations have an interpretation that should be easy to anticipate based on our experience with the ordinary Bogomolny equations in section 9.1: they say that evolution in the y direction changes the Higgs pair (E,φ) only by a gauge transformation. Thus until we encounter the singularity associated with an 't Hooft operator, the point in $\mathcal{M}_H$ determined by a solution of the extended Bogomolny equations is independent of y.

Now suppose that there is an 't Hooft operator of weight ${}^L w$ at, say, $y = y_0$, $z = z_0$. When y is increased past $y = y_0$, the bundle E undergoes an ordinary Hecke modification, as described in section 9.1. In fact, the evolution

equation for $A_{\overline{z}}$ is the same as it was in section 9.1, and so is the singularity due to the 't Hooft operator. The pair (E, φ) likewise undergoes what we will call a Hecke modification. The role of φ is, however, quite different from the role of E.

A holomorphic bundle can be modified at a single point, while preserving holomorphy, and this is what the Hecke or 't Hooft operators do. There is a moduli space of possible such modifications, the familiar space $\overline{\mathcal{Y}}({}^L w; z_0)$ parameterizing singular solutions of the Bogomolny equations (and isomorphic to a Schubert variety). But a holomorphic field, in this case a holomorphic section φ of $\mathrm{ad}(E)\otimes K_C$, cannot be modified at a point while preserving holomorphy. Hence, if φ is given away from $z = 0$, it has at most a unique holomorphic extension over $z = 0$.

So including φ does not cause the Hecke modification due to the 't Hooft operator to depend on additional parameters. On the contrary, including φ generically eliminates some parameters. The reason is that even if φ is holomorphic at $z = z_0$ for $y < y_0$, it will generically have a pole at $z = z_0$ for $y > y_0$. To see how this happens, let us take $G = U(N)$ and consider an 't Hooft operator with weight ${}^L w = (m_1, \dots, m_N)$. The holomorphic type of the bundle E jumps in crossing $y = y_0$; we write E_- and E_+ for the bundles "before" and "after." Relative to some local decomposition $E_- \cong \oplus_{i=1}^N \mathcal{L}_i$, the bundle E_+ has a local decomposition as $\oplus_{i=1}^N \mathcal{L}_i \otimes \mathcal{O}(z_0)^{m_i}$. That is, if $\mathcal{L}_i$ is locally generated by a holomorphic section s_i, $i = 1, \dots, N$, then E_+ is generated by holomorphic sections

$$t_i = (z - z_0)^{-m_i} s_i. \tag{10.37}$$

Here we used a natural isomorphism between E_+ and E_- for $z \neq 0$.

Suppose that φ is holomorphic at z_0 for $y < y_0$. The products $\varphi\, s_i$ are holomorphic sections of $E_- \otimes K_C$ (where K_C will play no essential role as we can trivialize it near $z = z_0$), so they can be expanded in terms of the s_i:

$$\varphi \cdot s_i = \sum_j f_i{}^j\, s_j. \tag{10.38}$$

Holomorphy of φ at $z = z_0$ for $y < y_0$ says that the $f_i{}^j$ (which are sections of K_C) are holomorphic at $z = z_0$. They are subject to no other constraint.

Similarly, for $y > y_0$, we can expand $\varphi\, t_i$ as a linear combination of the t_j. Given the relation (10.37) between the s_i and the t_i, we can immediately

write the form of the expansion:

$$\varphi \cdot t_i = \sum_j (z - z_0)^{m_j - m_i} f_i{}^j \, t_j. \tag{10.39}$$

For φ to be holomorphic at $y > y_0$, the functions $(z-z_0)^{m_j-m_i} f_i{}^j$ must all be regular at $z = z_0$. If we order the m_i in the usual way $m_1 \geq m_2 \geq \ldots \geq m_N$, then the number of conditions on the Taylor series coefficients of the $f_i{}^j$ near $z = z_0$ is $\sum_{i<j}(m_i - m_j)$.

This is the same as the dimension of the space of Hecke modifications of bundles as determined in eqn. (9.25). This strongly suggests that, in sharp contrast to the case of Hecke modifications of bundles, the Hecke modifications of a Higgs bundle (E, φ) are generically a finite set. In fact, this is the case if $\varphi(z_0)$ is regular and semisimple (which means for $G = U(N)$ that it is diagonalizable with distinct eigenvalues). The most systematic description of the result is to say that φ determines a vector field on the variety $\overline{\mathcal{Y}}({}^L w, z_0)$ that parametrizes Hecke modifications of E of type ${}^L w$ at the point z_0; the zeroes of this vector field are the possible Hecke modifications of the pair (E, φ). (The vector field is not quite canonically determined, but its zeroes are.)

Deferring a detailed explanation of this to [26], we will briefly describe the simple example with $G = U(2)$, ${}^L w = (1, 0)$. In this example, as we learned in section 9.1, if we take the bundle E_- to be trivial, then E_+ can be characterized by saying that a section of E_+ holomorphic at $z = 0$ has the form

$$s_+ = (z - z_0)^{-1} u + (h_1(z), h_2(z)), \tag{10.40}$$

where u is some nonzero element of $\mathbb{C}^2$ and h_1 and h_2 are holomorphic near $z = z_0$. Up to scaling by $\mathbb{C}^*$, u parametrizes the space of possible Hecke modifications of E, which is $\mathbb{CP}^1$ in this example. φ is holomorphic at $z = z_0$ for $y > y_0$ if and only if the product $\varphi \cdot s_+$ is of the same form, or in other words if and only if φu is a multiple of u. In the regular semisimple case, this means that u must be one of the two eigenvectors of φ.

Going back to the extended Bogomolny equations (10.36), an immediate consequence of the relation $\partial_y \phi_z = -i[\phi_z, \phi_0]$ is that ϕ_z is everywhere zero if it zero for some y. We assumed this in our analysis of the product of 't Hooft operators in section 10.3. A further consequence of the same equation is that the characteristic polynomial of φ is independent of y, even in crossing an 't Hooft operator. As a result, Hecke transformations of Higgs bundles map each fiber of the Hitchin fibration to itself.

Now consider a brane wrapped on a fiber of the Hitchin fibration, with a flat Chan-Paton bundle. These are the branes that according to S-duality must be magnetic eigenbranes. From what we have said in the last paragraph, the support of such a brane is invariant under Hecke transformations. This observation is a step toward proving directly, without relying on S-duality, that such branes are magnetic eigenbranes, as will be shown in detail elsewhere [26].

Hecke Correspondence For Higgs Bundles Just as in our discussion of the ordinary Bogomolny equations, we can consider the possible Hecke modifications of a given Higgs bundle (E, φ) – which has been our point of view so far – or we can describe the initial and final states more symmetrically and describe a Hecke correspondence $\mathcal{Q}$ on the space of Higgs bundles. We define $\mathcal{Q}$ by solving the extended Bogomolny equations on $W = I \times C$ with suitable boundary conditions at the two ends and in the presence of specified 't Hooft operators. We specify both A and φ (but not ϕ_0) at both ends of W, and divide only by gauge transformations that are trivial at each end.

Let $\mathcal{W}_0$ be the space of all pairs (A, φ) consisting of a gauge field and Higgs field on C satisfying $\overline{D}\varphi = 0$. Then $\mathcal{W}_0$ is a gauge theory version of the "stack" Higgs_G of all Higgs bundles, just as the space of all connections is a gauge theory version of Bun_G, the stack of all G-bundles. By restricting a solution of the extended Bogomolny equations to its boundary values, $\mathcal{Q}$ has a pair of maps to $\mathcal{W}_0$, or in other words, a map to $\mathcal{W}_0 \times \mathcal{W}_0$. This is the Hecke correspondence for Higgs bundles.

$\mathcal{Q}$ is very complicated, in general, with a variety of components of different dimension. However, if we limit ourselves to a good region of $\mathcal{W}_0$ (where each Higgs bundle has only finitely many possible Hecke modifications of the chosen type up to a gauge transformation), then $\mathcal{Q}$ can be regarded as a submanifold of $\mathcal{W}_0 \times \mathcal{W}_0$. In this good region, moreover, we can divide by gauge transformations and $\mathcal{Q}$ becomes a middle-dimensional submanifold of $\mathcal{M}_H \times \mathcal{M}_H$.

't Hooft operators preserve supersymmetry of type (B, A, A), as we explained at the end of section 6.3. This means that it must be possible to interpret $\mathcal{Q}$ as a brane of type (B, A, A) in $\mathcal{M}_H \times \mathcal{M}_H$. In particular, it must be holomorphic in complex structure I and Lagrangian[51] with respect to complex structures J and K or equivalently with respect to the holomorphic symplectic form $\Omega_I = \omega_J + i\omega_K$ defined in eqn. (4.9). More

[51] Above the middle dimension, there are A-branes that are not Lagrangian [80]. But in the good region, $\mathcal{Q}$ is middle-dimensional and must be Lagrangian.

exactly, if $\pi_i\colon \mathcal{Q} \to \mathcal{M}_H \times \mathcal{M}_H$, $i = 1, 2$, are the two projections, then $\mathcal{Q}$ must be Lagrangian with respect to the holomorphic symplectic form $\Omega = \pi_1^*(\Omega_I) - \pi_2^*(\Omega_I)$ on $\mathcal{M}_H \times \mathcal{M}_H$. The minus sign here ensures that the diagonal in $\mathcal{M}_H \times \mathcal{M}_H$, which is the trivial Hecke correspondence with no 't Hooft operators, is Lagrangian.

Letting C_+ and C_- denote the two ends of $W = I \times C$, we can write the restriction of Ω to $\mathcal{Q}$ as

$$\frac{1}{\pi}\left(\int_{C_+} - \int_{C_-}\right) |d^2z|\,\mathrm{Tr}\,(\delta\phi_z \wedge \delta A_{\overline{z}}) = \frac{1}{\pi}\int_I dy \int_C |d^2z|\,\partial_y \mathrm{Tr}\,(\delta\phi_z \wedge \delta A_{\overline{z}})\,. \tag{10.41}$$

To compute the restriction of this form to $\mathcal{Q}$, we can assume that the variations $\delta\phi_z$ and $\delta A_{\overline{z}}$ satisfy the linearization of (10.36). The linearizations of the first three equations in (10.36) read

$$\begin{aligned} D_{\overline{z}}\delta\phi_z &= -[\delta A_{\overline{z}}, \phi_z], \\ \partial_y \delta A_{\overline{z}} &= -iD_{\overline{z}}\delta\phi_0 - i[\delta A_{\overline{z}}, \phi_0] \\ \partial_y \delta\phi_z &= -i[\delta\phi_z, \phi_0] - i[\phi_z, \delta\phi_0]. \end{aligned} \tag{10.42}$$

Using these equations, the Jacobi identity, and integration by parts, one can verify that (10.41) vanishes, so that $\mathcal{Q}$ is complex-Lagrangian, as expected.

11. *A*-Branes and $\mathcal{D}$-modules

A crucial ingredient is missing from what we have presented so far. According to the geometric Langlands program, there is a natural correspondence between homomorphisms of $\pi_1(C)$ into ${}^L G$ and Hecke eigensheaves on $\mathcal{M}(G, C)$ which are also $\mathcal{D}$-modules. A $\mathcal{D}$-module is a module for the sheaf of differential operators; the concept is further elucidated below.

Concerning the left hand side of the correspondence, we learned in section 7 that homomorphisms of $\pi_1(C)$ to ${}^L G$ correspond to electric eigenbranes. Their duals, which are branes of type $\boldsymbol{F}$ – that is, branes wrapped on a fiber of the Hitchin fibration, with a flat Chan-Paton bundle – must therefore be magnetic eigenbranes. This statement follows from S-duality and can also be verified directly [26], using the abelianization that is provided by the spectral cover construction.

However, the magnetic eigenbranes that we get this way are branes of type (A, B, A) on $\mathcal{M}_H(G, C)$, while in the conventional statement of the geometric Langlands correspondence, the right hand side is supposed to involve instead a $\mathcal{D}$-module on, roughly speaking, $\mathcal{M}(G, C)$, the moduli space of G-bundles.

This section is devoted to supplying the missing link. We will show that every A-brane in complex structure K on $\mathcal{M}_H(G, C)$ automatically gives rise to, roughly speaking, a $\mathcal{D}$-module on $\mathcal{M}(G, C)$.

One can regard $\mathcal{D}$-modules on $\mathcal{M}(G, C)$ as B-branes on a noncommutative deformation of the cotangent bundle of $\mathcal{M}(G, C)$. It has been argued previously [41] that for a certain class of holomorphic symplectic manifolds, the category of A-branes is equivalent to the category of B-branes on the noncommutative deformation of the same manifold. The connection between A-branes on $\mathcal{M}_H(G, C)$ and $\mathcal{D}$-modules on $\mathcal{M}(G, C)$ can be viewed as a special case of this.

Actually, in the geometric Langlands program, it is not sufficient (nor is it possible, because of the way the Hecke operators act) to work only on $\mathcal{M}(G, C)$, the moduli space of stable bundles. One instead constructs a $\mathcal{D}$-module on $\mathrm{Bun}_G(C)$, the "stack" of all G-bundles on C. The gauge theory analog of the stack of all G-bundles is simply the space of all connections. The analog in our construction of the usual statement that the $\mathcal{D}$-modules are defined on $\mathrm{Bun}_G(C)$ is that the branes of interest can be defined in the underlying gauge theory, and not only in the sigma-model with target space $\mathcal{M}_H(G, C)$. Showing this for the important branes is in fact the goal of section 12.

11.1. The Canonical Coisotropic *A*-Brane

Our starting point is the observation made in [80] that for a general symplectic manifold X there may exist A-branes which are supported not on a Lagrangian submanifold but on a more general coisotropic submanifold. A coisotropic submanifold Y of a symplectic manifold X is a submanifold defined locally by first-class constraints, in other words it is defined locally by the vanishing of functions f_i, $i = 1, \dots, d$ that are Poisson-commuting. The maximal number of independent Poisson-commuting functions on X is one half the dimension of X, so the dimension of Y is at least half the dimension of X. If the dimension of Y is precisely half the dimension of X, then Y is Lagrangian, and the A-branes supported on Y are the usual Lagrangian A-branes. Coisotropic A-branes with support above the middle dimension also exist, in general, though the conditions for their existence are rather special.

We will need only the case that $Y = X$ and the A-brane has rank one. Thus its Chan-Paton bundle is a $U(1)$ bundle $\mathcal{L} \to X$. The Chan-Paton gauge field takes values in the Lie algebra of $U(1)$, which is a one-dimensional real vector space. So this gauge field is locally an ordinary real one-form A with covariant derivative $D = d + A$; its curvature is defined as $F = dA$. In a unitary representation of $U(1)$, the element 1 of the Lie algebra $\mathfrak{u}(1)$ acts by the 1×1 anti-hermitian matrix $-in$ for some integer n, and parallel transport is $\exp(in\int A)$. We interpret the Chan-Paton bundle over X as a line bundle associated to the representation $n = 1$. For a sigma-model map $\Phi : \Sigma \to X$, the factor in the worldsheet path integral associated with the Chan-Paton bundle is the holonomy of $\Phi^*(A)$ around $\partial\Sigma$ in the representation with $n = 1$, or $\exp(i \oint_{\partial\Sigma} \Phi^*(A))$.

In order for a brane with $Y = X$ and Chan-Paton line bundle $\mathcal{L}$ to be an A-brane, its curvature F must satisfy

$$(\omega^{-1}F)^2 = -1. \tag{11.1}$$

That is, the $(1,1)$ tensor $\omega^{-1}F$ is an almost complex structure. The forms F and ω are closed, and together with (11.1) this implies that the almost complex structure $\widehat{N} = \omega^{-1}F$ is integrable [80]. It is easy to see that both ω and F are of type $(2,0)\oplus(0,2)$ with respect to $\widehat{N}$. Furthermore, the complex form $\omega - iF$ is of type $(2,0)$ and so is a holomorphic symplectic form on X. Thus a coisotropic A-brane of this kind may exist only if ω is the real part of a holomorphic symplectic form. See also [67] for a construction of branes of this type in generalized complex geometry.

We are interested in the A-model on $\mathcal{M}_H(G,C)$ with the symplectic structure $\omega = (\mathrm{Im}\ \tau)\,\omega_K$. Then there is an obvious solution to eqn. (11.1):

$$F = (\mathrm{Im}\ \tau)\,\omega_J, \quad \widehat{N} = \omega_K^{-1}\omega_J = I. \tag{11.2}$$

(There is also a second solution with the sign of F and $\widehat{N}$ reversed.) The two-form ω_J is exact, as we explained in section 4.0, so F is the curvature two-form of a connection on a trivial line bundle $\mathcal{N} \to \mathcal{M}_H(G,C)$. The connection form is

$$\alpha = \frac{\mathrm{Im}\ \tau}{2\pi}\int_C |d^2z|\ \mathrm{Tr}\,(\phi_z\delta A_{\overline{z}} + \phi_{\overline{z}}\delta A_z) \tag{11.3}$$

and $F = d\alpha$. Although a more general two-form

$$(\mathrm{Im}\ \tau)\,(a\omega_J + b\omega_I)\,, \quad a^2 + b^2 = 1, \tag{11.4}$$

solves eqn. (11.1), it is not exact for $b \neq 0$, and for general values of τ does not have periods which are integral multiples of 2π. So in general, this form is not the curvature two-form of any connection, and cannot be used to construct an A-brane.

We will call the A-brane on $\mathcal{M}_H$ with $F = (\mathrm{Im}\,\tau)\,\omega_J$ the canonical coisotropic brane, or c.c. brane, for short. We also sometimes denote it as $\mathcal{B}_{c.c.}$. If G is simply-connected, then so is $\mathcal{M}_H(G, C)$, and the c.c. brane is unique. Otherwise, it is unique up to twisting its Chan-Paton bundle by a flat line bundle.

Our first observation about the c.c. brane is that it is an (A, B, A)-brane, i.e. it is an A-brane in complex structures I and K and a B-brane in complex structure J. To check that it is an A-brane in complex structure I, we only need the identity

$$(\omega_I^{-1}\omega_J)^2 = -1. \tag{11.5}$$

This means that the c.c. brane obeys the conditions of a coisotropic brane in complex structure I, just as it does in complex structure K. To check that it is a B-brane in complex structure J, it is sufficient to note that the curvature of the gauge field on the brane is a multiple of ω_J, which has type $(1, 1)$ in complex structure J.

These formulas really only define the c.c. brane on the smooth part of $\mathcal{M}_H$. However, the c.c. brane can be defined by a boundary condition in the underlying gauge theory; see section 12.4 for details. This statement is the quantum field theory equivalent of saying that the construction works on the whole "stack" of G-bundles over C.

Next, we would like to understand topological open strings with both endpoints on the c.c. brane. They are represented by BRST-invariant boundary operator insertions. In general, as we sketched in Figure 6 of section 6.2, for any brane $\mathcal{B}$, the $(\mathcal{B}, \mathcal{B})$ strings form an associative, but not necessarily commutative, algebra. In mathematical terms, this is the endomorphism algebra of the brane $\mathcal{B}$ regarded as an object of the category of branes on $\mathcal{M}_H(G, C)$. We would like to compute this endomorphism algebra in the case of the c.c. brane.

As a first step, we will argue that worldsheet instantons do not contribute to the computation. The boundary condition defining coisotropic branes in the sigma-model [80] says that strings ending on a space-filling coisotropic brane such as we consider here must obey

$$G_{IJ}\partial_1\Phi^J = -iF_{IJ}\partial_0\Phi^J, \tag{11.6}$$

where ∂_1 and ∂_0 are normal and tangential derivatives at the worldsheet boundary. Imposing also the instanton equation (that is, the map Φ should be holomorphic in complex structure K), to relate $\partial_1\Phi$ to $\partial_0\Phi$, we get

$$(\omega - iF)\partial_0\Phi = 0. \tag{11.7}$$

Since both ω and F are nondegenerate, the instanton must be constant on the boundary. But then, by analyticity, it must be constant everywhere, hence trivial.

Since there are no instanton corrections, anything interesting will have to come from worldsheet perturbation theory. But worldsheet perturbation theory, to any order in $\hbar$, generates corrections to sigma-model operators and couplings that are local in the target space. Hence, at least to all orders in world-sheet perturbation theory, it is very natural to consider boundary observables that may be defined only locally in target space. Such observables form a sheaf over the target space, so instead of an algebra of observables, we can get, to all orders in perturbation theory, something more powerful: a sheaf of algebras. (In the closed-string case, such localization in target space has been discussed in [133–135], interpreting earlier mathematical work [136].) Along with analyzing this sheaf in perturbation theory, we will discuss what happens in the exact theory.

The space of BRST-invariant boundary operators for an arbitrary coisotropic A-brane Y has been determined in the classical limit in [80,132]. For our purposes, it will be sufficient to consider boundary operators of ghost number zero. Then, in the case $Y = X$, such observables are simply holomorphic functions on X, where X is regarded as a complex manifold with complex structure $\widehat{N}$. Specializing to this case and localizing in target space, we conclude that additively (ignoring the ring structure) the sheaf of topological boundary operators for the c.c. brane is the sheaf of functions on $\mathcal{M}_H(G,C)$ which are holomorphic in complex structure I. Note that while we are working in the A-model in complex structure K, the complex structure which is relevant here is $\widehat{N} = I$. Locally the sheaf of boundary operators has no higher cohomology and cannot be deformed, so it cannot have quantum corrections. But globally it has quantum corrections, which we will analyze.

The sheaf of holomorphic functions on $\mathcal{M}_H$ has an obvious commutative algebra structure, which is the correct one in the classical limit. In the context of "physical" string theory, in the presence of the gauge field strength F, the algebra of open strings becomes noncommutative [137].

Let us first discuss informally how this would work in the present context. The usual noncommutative parameter, called θ in [137], is the antisymmetric

part of $(g+F)^{-1}$. In the present context, this becomes a multiple of ω_J^{-1}. ω_J is the real part of Ω_I, the holomorphic $(2,0)$-form in complex structure I. When we multiply observables derived from functions that are holomorphic in complex structure I (which are the topological observables), ω_J^{-1} can be replaced by Ω_I^{-1}. So the multiplication of such observables is deformed via the Poisson brackets derived from Ω_I, as we will find shortly.

In physical string theory, for any $F \neq 0$, the multiplication of open strings contains noncommutative phases coming from $(g+F)^{-1}$, but generally there are many other stringy effects. A precise description in terms of a noncommutative deformation of the algebra of functions on the target space usually arises [137] only when $|F| \gg |g|$. We are not in that limit in the present problem; rather, we have $|F| = |g|$. However, once we restrict to topological observables, we will get a simplification similar to what usually occurs in physical string theory for $|F| \gg |g|$. In the context of the B-type topological strings, this was first discussed in [71]. In essence, the elimination of stringy excitations that in physical string theory occurs only if $|F| \gg |g|$ arises here because of restricting to topological observables.

Now we turn to a more precise analysis for the topological theory. The action for the A-model on a disk can be written in the form

$$S = \int_\Sigma \Phi^*(\omega - iF) + \int_\Sigma d^2\sigma\{Q, V\}, \tag{11.8}$$

where Q is the BRST operator or topological supercharge. In our case, the form $\omega - iF = -i\Omega_I \mathrm{Im}\ \tau$ is exact (and of course this is always true locally). If we introduce the potential ϖ for Ω_I, so that $\Omega_I = d\varpi$, then the action takes the form

$$S = -i\,\mathrm{Im}\ \tau \int_{\partial\Sigma} \Phi^*(\varpi) + \int_\Sigma d^2\sigma\{Q, V\}. \tag{11.9}$$

Further, ϖ is proportional to the canonical holomorphic one-form on $\mathcal{M}_H(G, C)$:

$$\varpi = \frac{1}{\pi} \int |d^2 z|\, \mathrm{Tr}\,(\phi_z \delta A_{\overline{z}})\,. \tag{11.10}$$

If we restrict to $T^*\mathcal{M}(G, C)$, we can be more explicit. If q^α are local holomorphic coordinates on $\mathcal{M}(G, C)$ and p_α are linear functions on the fibers of $T^*\mathcal{M}(G, C)$ that are canonically conjugate to the ϕ_α, then $\varpi = \sum_\alpha p_\alpha\, dq^\alpha$ and $S = -i\,\mathrm{Im}\,\tau \int p_\alpha\, dq^\alpha + \{Q, \ldots\}$. Thus, up to Q-exact terms, the A-model

action is the holomorphic version of the quantum mechanical action for a particle on $\mathcal{M}(G,C)$ (with zero Hamiltonian).

We undoubtedly lose some important structure by specializing to $T^*\mathcal{M} \subset \mathcal{M}_H$, since the c.c. brane is defined on all of $\mathcal{M}_H$, and even in the four-dimensional gauge theory, not just on $T^*\mathcal{M}$. But let us see what structure remains after we make this specialization.

The form of the action in (11.9) is standard, and the result is familiar: to first order in $1/\mathrm{Im}\,\tau$, the effect of the perturbation is to deform the commutative algebra of locally-defined holomorphic functions by the Poisson brackets derived from this action. So in perturbation theory in $1/\mathrm{Im}\,\tau$, the c.c. strings (or more fastidiously, the $(\mathcal{B}_{c.c.},\mathcal{B}_{c.c.})$ strings) form a sheaf of noncommutative algebras. But precisely what sheaf do we get?

A powerful tool is the $\mathbb{C}^*$ group $\mathcal{U}$ that was introduced at the end of section 4.1 and used at various points in this paper. It acts by

$$\phi_z \longrightarrow \lambda\phi_z \tag{11.11}$$

and the associated grading on the space of observables was used in section 5.3 to analyze the action of S-duality on zerobranes. ϖ is of degree 1 under this scaling; therefore, $\mathcal{U}$ can be promoted to a symmetry of the quantum theory if we assign degree 1 to $1/\mathrm{Im}\,\tau$, that is to Planck's constant $\hbar$ in the analog quantum-mechanical problem.

At the classical level, the multiplication of observables preserves the grading by $\mathcal{U}$; the product of observables of degrees m and n is an observable of degree $m+n$. At the quantum level, there can also be terms of lower degree in the product, if they are accompanied by positive powers of Planck's constant. Thus the algebra of quantum observables is only filtered, not graded.

In order to effectively use the $\mathbb{C}^*$ symmetry to analyze the algebra, we need to consider observables that are defined in a $\mathbb{C}^*$-invariant open set. For this, we pick an open set $\mathcal{V} \subset \mathcal{M}$ and consider observables that are defined on $T^*\mathcal{V} \subset \mathcal{M}_H$. Thus, we are only going to partly sheafify the space of strings ending on the c.c. brane, allowing only open sets in $\mathcal{M}_H$ of this form. The advantage of this partial sheafification is that it works nonperturbatively, as we will see shortly. A complete sheafification in which one defines an associative algebra of observables for an arbitrary open set $\mathcal{W} \subset \mathcal{M}_H$ works fine to all orders in $1/\mathrm{Im}\,\tau$, but it is not realistic to expect it to work in the exact theory. For a general discussion of such matters, see [138].

To understand why in general there is an obstruction to sheafification of deformation quantization beyond perturbation theory, consider the simple example of deformation quantization of $\mathbb{C}^2$ with coordinates x, y and

symplectic form $dx \wedge dy$. Deformation quantization in this case can be carried out exactly using the Moyal-Wigner product[52] if one restricts oneself to polynomial functions on $\mathbb{C}^2$. To sheafify this situation, one picks an open set $\mathcal{W} \subset \mathbb{C}^2$, and tries to define a Moyal-Wigner product of functions holomorphic in $\mathcal{W}$. There is no problem to any finite order in $\hbar$, but the fact that the Moyal-Wigner product involves an infinite sum causes a problem in getting an exact result. To see the problem, try to evaluate at the origin the Moyal-Wigner product of the functions $1/(x-a)$ and $1/(y-b)$, each of which are holomorphic in suitable open sets in $\mathbb{C}^2$.

So we limit ourselves to partial sheafification in which we consider an open set in $\mathcal{M}_H$ of the form $T^*\mathcal{V}$, $\mathcal{V} \subset \mathcal{M}$. Moreover, we restrict ourselves to functions on $T^*\mathcal{V}$ that have polynomial growth along the fibers of the projection $T^*\mathcal{V} \to \mathcal{V}$. The algebra of such functions is generated by elements of degree less than or equal to 1, that is, by functions that are at most linear along the fibers.

Elements of degree 0 are simply holomorphic functions on $\mathcal{V}$. There is no possibility to deform the algebra of multiplication of such functions, because there are no observables in $T^*\mathcal{V}$ of negative degree. As $\hbar$ has degree 1, a correction to the multiplication of functions of degree 0 would have to involve an observable of negative degree.

Elements of degree 1 are linear functions on the fibers of the cotangent bundle. So if q^α are local coordinates on $\mathcal{V}$ and p_α are the conjugate linear functions on the fibers, then an element of degree 1 takes the form $\kappa = \sum_\alpha p_\alpha V^\alpha(q)$ and is in natural correspondence with a holomorphic vector field $\sum_\alpha V^\alpha \frac{\partial}{\partial q^\alpha}$ on $\mathcal{M}$. In multiplying an object f of degree 0 and an object κ of degree 1, there is only "room" for a first order deformation – since terms of order $\hbar^2$ in the product would again multiply operators of negative degree. The first order deformation is characterized by the Poisson bracket

$$[\kappa, f] = -i(\mathrm{Im}\ \tau)^{-1}\kappa(f). \tag{11.12}$$

To proceed, we write $\widehat{w}$ for the open string state corresponding to a holomorphic function w on $T^*\mathcal{V}$. The formula (11.12) determines the $\star$ product (that is, the algebra of open string multiplication) of the ring elements $\widehat{f}$, $\widehat{\kappa}$ corresponding to f and κ, but not quite uniquely, since (while preserving the way the algebra is filtered) we could add to $\widehat{\kappa}$ an observable of degree

[52] For reprints of their papers and many other original papers, along with a survey, see [139].

zero. If for $\kappa = V^\alpha(q)p_\alpha$, we define $\widehat{\kappa} = \widehat{V}^\alpha \star \widehat{p}_\alpha$, then the algebra is uniquely determined and is

$$\begin{aligned} \widehat{f} \star \widehat{\kappa} &= \widehat{f\kappa} \\ \widehat{\kappa} \star \widehat{f} &= \widehat{f\kappa} - i(\mathrm{Im}\ \tau)^{-1}\widehat{\kappa(f)}. \end{aligned} \tag{11.13}$$

Of course, $\widehat{f\kappa}$ and $\widehat{\kappa(f)}$ are the ring elements that correspond to the functions $f\kappa$ and $\kappa(f) = V^\alpha\, \partial f/\partial q^\alpha$. We still have not completely used the freedom to redefine $\widehat{\kappa}$; the remaining freedom is that for the linear functions p_α on the cotangent bundle, we could replace

$$\widehat{p}_\alpha \longrightarrow \widehat{p}_\alpha + \widehat{A}_\alpha \tag{11.14}$$

for some functions $A_\alpha(q)$.

When we multiply two functions κ and κ' that are each of degree 1, something new happens. There is a first order deformation, which is given again by the Poisson bracket, but there now is also room for a second order deformation. The Poisson bracket of p_α and p_β vanishes, but the commutator $[\widehat{p}_\alpha, \widehat{p}_\beta]$ in the $\star$ algebra might be $\hbar^2 F_{\alpha\beta}(q)$ for some functions $F_{\alpha\beta}$. One can show using associativity of the algebra that the two-form $F_{\alpha\beta}dq^\alpha\, dq^\beta$ is closed, and hence can be set to zero (for a small open set $\mathcal{V}$) by the transformation (11.14).

The algebra associated with $T^*\mathcal{V}$ for a small open set $\mathcal{V} \subset \mathcal{M}_H$ is therefore isomorphic simply to the algebra of holomorphic differential operators on $\mathcal{V}$, with functions on $\mathcal{M}$ acting by multiplication and

$$\widehat{p}_\alpha = -\frac{i}{\mathrm{Im}\,\tau}\frac{\partial}{\partial q^\alpha}. \tag{11.15}$$

Now suppose that we cover $\mathcal{M}$ with small open sets $\mathcal{V}^{(i)}$. In overlap regions $\mathcal{V}^{(i)} \cap \mathcal{V}^{(j)}$, the degree zero ring elements $\widehat{f}$ must agree, but the degree 1 elements $\widehat{p}_\alpha^{(i)}$ and $\widehat{p}_\alpha^{(j)}$ may differ by a transformation (11.14). Since we require $[\widehat{p}_\alpha, \widehat{p}_\beta] = 0$ in each open set, they must be related by

$$\widehat{p}_\alpha^{(i)} = \widehat{p}_\alpha^{(j)} + \frac{1}{\mathrm{Im}\,\tau}\widehat{\frac{\partial s^{(ij)}}{\partial q^\alpha}} \tag{11.16}$$

for some local holomorphic functions $s^{(ij)}$ defined on $\mathcal{V}^{(i)} \cap \mathcal{V}^{(j)}$. Consistency of this relation implies that in triple overlaps the quantities $c^{(ijk)} = s^{(ij)} + s^{(jk)} + s^{(ki)}$ are complex constants. Moreover, they automatically obey a cocycle condition and define an element of $H^2(\mathcal{M}, \mathbb{C})$.

If those constants take values in $2\pi\mathbb{Z}$, then the quantities $\exp(-is^{(ij)})$ are the transition functions of a complex line bundle $\mathcal{L} \to \mathcal{M}_H$. In this case, the sheaf of algebras of open strings is the sheaf of differential operators acting on sections of the line bundle $\mathcal{L}$. We call this sheaf of algebras $\mathcal{D}_{\mathcal{L}}$.

$\mathcal{L}$ is not quite uniquely determined. In general, the sheaf of algebras $\mathcal{D}_{\mathcal{L}}$ is invariant under twisting $\mathcal{L}$ by a flat line bundle, since a flat line bundle has constant transition functions that commute with the momenta $\widehat{p}_\alpha$.

There is no reason that the constants $c^{(ijk)}$ must be properly quantized, that is, must take values in $2\pi\mathbb{Z}$. For example, beginning with a properly quantized case leading to the sheaf $\mathcal{D}_{\mathcal{L}}$ of differential operators on a line bundle $\mathcal{L}$, we can define a new sheaf $\mathcal{D}_{\mathcal{L}^\gamma}$ by multiplying the $s^{(ij)}$ by a complex number γ. If $\mathcal{L}$ is topologically nontrivial, a line bundle $\mathcal{L}^\gamma$ does not exist (except for integer or perhaps rational values of γ) but the corresponding sheaf of algebras $\mathcal{D}_{\mathcal{L}^\gamma}$ does exist.

Since $H^2(\mathcal{M}(G,C),\mathbb{C})$ is one-dimensional, in this specific example the most general sheaf of algebras that could arise from this sort of construction is of the form $\mathcal{D}_{\mathfrak{L}^\gamma}$ where γ is a complex number and $\mathfrak{L}$ is the fundamental line bundle over $\mathcal{M}$, which we loosely call the determinant line bundle. If $\mathcal{M}$ is replaced by a more general space X, a slight generalization is possible involving a product of complex powers of line bundles.

At any rate, as we will see, the basic case of the geometric Langlands program involves the differential operators on a certain ordinary line bundle. But a generalization that we consider in section 11.3 does involve the complex powers of a line bundle.

Notice that in the above analysis, there is no question of whether perturbation theory converges, since everything is determined by what happens in second order. Had we tried to completely sheafify, associating an algebra with every open set in $\mathcal{M}_H$, we would have had trouble with convergence of perturbation theory.

Time Reversal Symmetry These somewhat abstract considerations do not quite determine the sheaf of algebras of c.c. strings. It is the sheaf of differential operators on some line bundle (or complex power of a line bundle), but which one? To determine it, we will use time-reversal symmetry. It is natural to get an extra constraint from time-reversal symmetry, because as we see in eqn. (3.56), this is a symmetry of the twisted topological field theory precisely at $t = 1$, the value at which we are working in the present discussion.

In gauge theory, time-reversal, which we call $\mathcal{T}$, reverses the time coordinate of the four-manifold M while acting trivially on the space coordinates. If $M = \mathbb{R} \times I \times C$, then $\mathcal{T}$ just acts as -1 on $\mathbb{R}$. In the effective

two-dimensional theory, $\mathcal{T}$ is an orientation-reversing symmetry of the string worldsheet Σ, which maps the boundary of Σ to itself while reversing the orientation of the boundary. Since our considerations are local, we can assume that Σ or M has such a symmetry.

Under $\mathcal{T}$, the bosons (A, ϕ) transform to $(\mathcal{T}^* A, -\mathcal{T}^*\phi)$, as we saw in eqn. (3.57). The terms in (11.8) and (11.9) that involve ω, F, and ϖ would be odd under reversal of worldsheet orientation if $\mathcal{T}$ were to map ϕ to $\mathcal{T}^*\phi$. But with the extra minus sign, these terms are invariant.

Time-reversal reverses the order of operator insertions on the worldsheet boundary, so it maps the algebra of c.c. strings to its "opposite" algebra, in which the elements are the same but multiplication goes in the opposite order. Thus, to every element x of an algebra C, the opposite algebra C^{op} has an element x^{op}, with $(x \star y)^{op} = y^{op} \star^{op} x^{op}$. Here $\star$ is the multiplication in C and $\star^{op}$ is the multiplication in C^{op}.

The opposite of the algebra of differential operators on a line bundle $\mathcal{L}$ over a complex manifold X is [140] isomorphic to the algebra of differential operators on $\mathcal{L}^{-1} \otimes K_X$, where K_X is the canonical line bundle of X (that is, the bundle of holomorphic forms of top degree). To see this, let s and u be compactly supported (not holomorphic) sections, respectively, of $\mathcal{L}$ and $\mathcal{L}^{-1} \otimes K_X$, and let F be a holomorphic differential operator acting on $\mathcal{L}$. Then $s\,\mathsf{F}u$ is a section of K_X. A section of K_X is a differential form of middle dimension, and can be integrated over any real slice Z of X. Moreover, by integrating by parts, we can form a "transpose" operator F^t, acting on sections of $\mathcal{L}^{-1} \otimes K_X$, and obeying $\int_Z u\,\mathsf{F}s = \int_Z (\mathsf{F}^t u)s$ for all u and s. F^t is a differential operator with holomorphic coefficients, just like F, and does not depend on the choice of Z. The map $\mathsf{F} \to \mathsf{F}^t$ reverses the order of multiplication, showing that the opposite sheaf of algebras to $\mathcal{D}_{\mathcal{L}}$ is $\mathcal{D}_{\mathcal{L}^{-1}\otimes K_X}$.

Time-reversal symmetry of the c.c. brane means that the sheaf of algebras obtained from open strings ending on the c.c. brane must be isomorphic to its opposite algebra. This, plus the fact that it is the sheaf of differential operators acting on some line bundle, implies that this algebra must be $\mathcal{D}_{K_{\mathcal{M}}^{1/2}}$, the sheaf of differential operators acting on $K_{\mathcal{M}}^{1/2}$, the square root of the canonical bundle of $\mathcal{M}$. $\mathcal{M}$ actually is a spin manifold, so $K_{\mathcal{M}}^{1/2}$ is an ordinary line bundle. The spin structure of $\mathcal{M}$ is not necessarily unique, but since the sheaf of differential operators acting on a line bundle is invariant under twisting by a flat line bundle, the sheaf of algebras $\mathcal{D}_{K_{\mathcal{M}}^{1/2}}$ does not depend on the choice of $K_{\mathcal{M}}^{1/2}$.

Global Algebra of Topological Strings Now we want to discuss the *global* c.c. strings.

How do we get the algebra of global observables from the sheaf of observables? We simply take the global sections of the sheaf of observables, or more generally, the cohomology of $\mathcal{M}$ with values in this sheaf. To keep things simple, we will just consider global sections, that is observables of ghost number or cohomological degree zero (but any scaling degree under the $\mathbb{C}^*$ symmetry $\mathcal{U}$).

Classically, the sheaf of observables is just the sheaf of holomorphic functions on $\mathcal{M}_H$, and its global sections are simply the global holomorphic functions on $\mathcal{M}_H$. They are the commuting Hamiltonians of the integrable system, which we described in section 4.2.

Likewise quantum mechanically, to get the global observables, we must take the global sections of the sheaf of local observables. But there is a quantum correction to the gluing law that defines this sheaf, relative to what we have classically. The quantum correction entered our analysis explicitly in eqn. (11.16). So there is potentially a quantum correction to the description of the global observables.

Having understood that the sheaf of c.c. strings is the sheaf of differential operators on $\mathcal{M}_H$, we can understand concretely what this correction means. Suppose that we are given a holomorphic function on $\mathcal{M}_H$, for example a function of degree three that once we pick local complex coordinates on $\mathcal{M}$ looks like $w = \sum_{\alpha,\beta,\gamma} f^{\alpha\beta\gamma}(q) p_\alpha p_\beta p_\gamma$. To promote it to a global $(\mathcal{B}_{c.c.}, \mathcal{B}_{c.c.})$ string, we need to find a third order differential operator $\widehat{w}$ on $\mathcal{M}$ whose "symbol" is w. In other words, we want $\widehat{w} = (-i/\mathrm{Im}\,\tau)^3 f^{\alpha\beta\gamma} \partial^3/\partial q^\alpha \partial q^\beta \partial q^\gamma + \ldots$, where the ellipses refer to a differential operator of order two or less. Because of the absence of a canonical system of local complex coordinates q^α on $\mathcal{M}$, such an operator is not unique locally, and may not exist globally.

If a holomorphic differential operator $\widehat{w}$ with symbol w does exist, we say that the function w can be "quantized." The commuting Hamiltonians on $\mathcal{M}_H$ can all be quantized. This was first proved for functions of degree two by Hitchin [141] and for functions of any degree by Beilinson and Drinfeld [5]; this result is part of the rationale for the title of their paper.[53] We will give an argument that involves further use of the time-reversal symmetry $\mathcal{T}$. This argument also shows that the resulting differential operators commute, as first proved in the above-cited references.

[53] Their argument is based on two-dimensional conformal field theory, more precisely on current algebra of the group G at level $k = -h$, with h the dual Coxeter number of G.

We use the fact that $(\mathcal{B}_{c.c}, \mathcal{B}_{c.c.})$ strings correspond to boundary operators that can be inserted at a boundary labeled by the c.c. brane. Moreover, these boundary observables can be derived from operators of the four-dimensional gauge theory, which is a powerful tool because it has properties of locality that are lost upon reduction to two dimensions. This will enable us to give a simple argument. In fact, as we recall from section 4.2, the commuting Hamiltonians take the form $\int_C \alpha\, \mathcal{P}(\varphi)$, with $\mathcal{P}$ a homogeneous gauge-invariant polynomial of some degree k, and $\alpha \in H^1(C, K_C^{1-k})$. To show that such an operator is Q-invariant, it suffices to show that the local operator $\mathcal{P}(\varphi)$ is Q-invariant. We do this in section 12.4, simply by classifying the possible operators and using time-reversal symmetry. From this point of view, it is clear that the differential operators obtained by quantizing two classical expressions $\int_C \alpha\, \mathcal{P}(\varphi)$ and $\int_C \widetilde{\alpha}\, \widetilde{\mathcal{P}}(\varphi)$ commute, since one can take α and $\widetilde{\alpha}$ to have disjoint support.

To illustrate the nontrivial nature of the fact that the classical holomorphic functions on $\mathcal{M}$ can be quantized to get differential operators on $K_{\mathcal{M}}^{1/2}$, let us note [141,5] that they cannot be quantized to get differential operators on any line bundle other than $K_{\mathcal{M}}^{1/2}$. The fact that they can be quantized to get *commuting* differential operators means that Hitchin's classical integrable system on $\mathcal{M}_H$ can be "quantized" to make a quantum integrable system on $\mathcal{M}$.

11.2. $\mathcal{D}$-modules Corresponding to A-Branes

The existence of the c.c. brane means that every A-brane in complex structure K is a module for the sheaf of differential operators on $K_{\mathcal{M}}^{1/2}$. In fact, in general, if $\mathcal{B}$ and $\mathcal{B}'$ are any two branes, then the $(\mathcal{B}, \mathcal{B}')$-strings form a module for the algebra of $(\mathcal{B}, \mathcal{B})$ strings. This idea is illustrated in Figure 6 of section 6.3, and is part of the usual axioms of open-closed topological field theory [142,143]. To the extent that sheafification is possible, the $(\mathcal{B}, \mathcal{B})$ strings form a sheaf of algebras, not just an algebra, and the $(\mathcal{B}, \mathcal{B}')$ strings form a sheaf of modules for this sheaf of algebras. This statement just means that the ring and module structures can be defined for open strings that are regular only in a suitable open set in the target space.

We want to apply this construction for the case that $\mathcal{B} = \mathcal{B}_{c.c.}$ is the canonical coisotropic brane. In this case, we claim that for any brane $\mathcal{B}$ that is of A-type in complex structure K, there are no instanton corrections to $(\mathcal{B}_{c.c}, \mathcal{B}')$ strings. The argument also applies to higher order topological couplings (such as cubic Yukawa couplings) involving the c.c. brane. The absence of instantons can be argued in much the same way as for strings

ending entirely on the c.c. brane. The relevant disk instantons have a part of their boundary on the c.c. brane, and on this part of the boundary the instanton must be constant. But then, by analyticity, the instanton must be constant everywhere.

We also know, from our previous investigation, what sort of sheafification is possible. We can associate an algebra and a module to an open set in $\mathcal{M}_H$ of the form $T^*\mathcal{V}$, $\mathcal{V} \subset \mathcal{M}$, but not necessarily to more general open sets. So an A-brane on $\mathcal{M}_H(G,C)$ in complex structure K gives a sheaf of modules for the sheaf of algebras $\mathcal{D}_{K_{\mathcal{M}}^{1/2}}$ over $\mathcal{M}$.

Now we will give a few examples of this. (For some examples worked out in detail of open string quantization involving coisotropic branes, see [144].)

Our first example is an A-brane $\mathcal{B}'$ defined by the condition $\varphi = 0$, with trivial Chan-Paton bundle. This submanifold is a copy of $\mathcal{M}(G,C)$ and is Lagrangian in complex structures K and J and complex in complex structure I. Thus $\mathcal{B}'$ is an example of a (B,A,A)-brane. (We used this brane in section 10.3 in analyzing the operator product expansion of 't Hooft operators.) We take the flat Chan-Paton connection on $\mathcal{B}'$ to be trivial, for simplicity. As described in (11.3), the c.c. brane has a Chan-Paton line bundle $\mathcal{N}$ that is topologically trivial, but endowed with a non-trivial connection.

To compute the spectrum of open strings on the classical level, one can reduce the supersymmetric sigma-model to supersymmetric quantum mechanics, i.e. retain only the zero modes. In this approximation, open-string states with $(\mathcal{B}_{c.c.},\mathcal{B}')$ boundary conditions are sections of the tensor product with $\mathcal{N}^{-1}$ of a vector bundle obtained by quantizing the space of fermion zero modes. The connection on $\mathcal{N}$ is zero when restricted to $\varphi = 0$, so we can omit the factor of $\mathcal{N}^{-1}$. The space of fermion zero modes is the tangent space to $\mathcal{M} \subset \mathcal{M}_H$ (fermions associated with normal directions to $\mathcal{M}$ obey opposite boundary conditions at the two ends of a string and have no zero modes). When we quantize the space of fermion zero modes, we get the spinors on $\mathcal{M}$. Viewing $\mathcal{M}$ as a complex manifold, its spin bundle is the same as $K_{\mathcal{M}}^{1/2} \otimes (\oplus_{i=0}^{\dim \mathcal{M}} \Omega^{0,j}(\mathcal{M}))$. Here $\Omega^{0,j}$ is the sheaf of $(0,j)$-forms on $\mathcal{M}$.

The BRST operator or topological supersymmetry Q is in this situation simply the $\overline{\partial}$ operator acting on the $(0,j)$-forms with values in $K_{\mathcal{M}}^{1/2}$. The sheaf of physical $(\mathcal{B}_{c.c.},\mathcal{B}')$ strings is therefore the sheaf of holomorphic sections of[54] $K_{\mathcal{M}}^{1/2}$. This is of course a sheaf of modules for the sheaf of rings

[54] The Ω^j-forms with $j > 0$ do not contribute to the cohomology over a small open set $\mathcal{V}$, so they can be omitted here.

$\mathcal{D}_{K_{\mathcal{M}}^{1/2}}$; indeed, it is the sheaf of modules by which this sheaf of rings was defined. This construction gives, possibly, a more direct explanation of why the sheaf of rings derived from the c.c. brane is precisely $\mathcal{D}_{K_{\mathcal{M}}^{1/2}}$, rather than $\mathcal{D}_{\mathcal{L}}$ for some other $\mathcal{L}$.

It may appear that we have implicitly used again the time-reversal symmetry $\mathcal{T}$ in claiming that quantization of the space of fermion zero modes gives precisely the spinors on $\mathcal{M}$, rather than spinors with values in some line bundle $\mathcal{L}$. Actually, we can give an alternative argument for this point, though it is an argument that uses another discrete symmetry of the theory. The branes $\mathcal{B}_{c.c.}$ and $\mathcal{B}'$ are physically sensible, unitary branes in the four-dimensional supersymmetric gauge theory defined on $M = \mathbb{R} \times I \times C$; here $\mathbb{R}$ is the time direction, and I is an interval with boundary conditions at the two ends defined respectively by $\mathcal{B}_{c.c.}$ and $\mathcal{B}'$. To define the theory on M, there is some twisting to preserve supersymmetry in the compactification on C, but no twisting that involves the time direction. So along with the topological supersymmetry Q, the theory on M is also invariant under the adjoint supersymmetry $Q^\dagger$ (they obey $\{Q, Q^\dagger\} = H$, where H is the Hamiltonian). The physical theory has a "charge conjugation" symmetry that exchanges Q and $Q^\dagger$. Invariance under this symmetry implies that the Chan-Paton bundle obtained by quantizing the fermion zero modes is precisely the spin bundle of $\mathcal{M}$, not the tensor product of the spin bundle with an additional line bundle.[55] This type of argument was first made in [145] in quantizing solitons.

Note that in this computation it was important that the support of $\mathcal{B}'$ is not only a Lagrangian submanifold with respect to ω_K, but is also a complex submanifold with respect to the complex structure $\widehat{N} = I$ determined by the c.c. brane. Otherwise, the topological supercharge or BRST operator would not reduce to the $\overline{\partial}$ operator in complex structure I.

Our second example is the case that is important for the geometric Langlands program: a brane of type $\boldsymbol{F}$, that is a brane $\mathcal{B}_{\boldsymbol{F}}$ supported on a fiber $\boldsymbol{F}$ of the Hitchin fibration with a flat Chan-Paton line bundle. This is a brane of type (A, B, A), so in particular it is an A-brane in complex structure K.

[55] If $\pi_1(G)$ is nontrivial, we can modify the construction by taking the Chan-Paton bundle on the brane $\mathcal{B}'$ to be a flat line bundle. Then we get spinors on $\mathcal{M}$ with values in a flat line bundle; as a special case of this, we get all the possible spin structures on $\mathcal{M}$. Because of the relation of D-branes to K-theory [102], the choice of spin structure of $\mathcal{M}$ is really part of a careful description of the Chan-Paton bundle of the brane $\mathcal{B}'$.

Therefore, it gives rise to a sheaf of modules for $\mathcal{D}_{K_{\mathcal{M}}^{1/2}}$. We will explain at the end of section 11.3, by a further elementary argument, how to convert this to a sheaf of modules for $\mathcal{D}$, the sheaf of ordinary differential operators on $\mathcal{M}$. So the brane of type $\boldsymbol{F}$ has the two key properties: it is a magnetic eigenbrane because it is the S-dual of a zerobrane; and it gives rise to a sheaf of $\mathcal{D}$-modules over $\mathcal{M}$. These are the basic claims of the geometric Langlands program.

It is difficult to explicitly describe the sheaf of $\mathcal{D}$ or $\mathcal{D}_{K_{\mathcal{M}}^{1/2}}$-modules that comes from a brane of type $\boldsymbol{F}$. But we can do this in the the abelian case $G = U(1)$. In this special case, the sigma-model is a free field theory, and the boundary conditions are linear, so the computation of the spectrum and module structure of the open strings is straightforward.

For $G = U(1)$, we can think of A and ϕ as real one-forms (A is only defined up to gauge transformations, of course, while ϕ is gauge-invariant). The Hitchin equations for A and ϕ decouple, and in complex structure I, the Hitchin moduli space is a product of the Jacobian $\mathrm{Jac}(C)$ of C (which is the moduli space of topologically trivial holomorphic line bundles on C) and the vector space $\boldsymbol{B} = H^0(C, \Omega^1)$. $\mathrm{Jac}(C)$ is a complex torus of dimension g_C, the genus of C, and the Hitchin moduli space $\mathrm{Jac}(C) \times \boldsymbol{B}$ can be identified with its cotangent bundle. The Hitchin fibration is the projection to $\boldsymbol{B}$.

Given $p \in \boldsymbol{B}$ and the corresponding fiber $\boldsymbol{F}_p \simeq \mathrm{Jac}(C)$ of the Hitchin fibration, we can compute the space of $(\mathcal{B}_{c.c.}, \mathcal{B}_{\boldsymbol{F}_p})$ strings much as in the previous example. One difference is that although the Chan-Paton line bundle $\mathcal{L} \to \boldsymbol{F}_p$ of the brane $\mathcal{B}_{\boldsymbol{F}_p}$ is still topologically trivial, we now allow an arbitrary unitary flat connection on it. Another difference is that the restriction of the Chan-Paton line bundle $\mathcal{N}$ of the c.c. brane to $\boldsymbol{F}_p$ is holomorphically nontrivial, in general. On the other hand, the canonical bundle $K_{\boldsymbol{F}_p}$ is trivial as $\boldsymbol{F}_p$ is a torus. Consequently, BRST-invariant open string states with ghost number zero are holomorphic sections of $T_p = \mathcal{L} \otimes \mathcal{N}^{-1}|_{\boldsymbol{F}_p}$.

But how does the sheaf of $(\mathcal{B}_{c.c.}, \mathcal{B}_{c.c.})$ strings acts on sections of T_p? We will answer this question next. We work on $\mathbb{R} \times I$ with the left boundary on the brane $\mathcal{B}_{c.c.}$ and the right boundary on $\mathcal{B}_{\boldsymbol{F}_p}$. The fiber p of the Hitchin fibration is defined by $\phi = \upsilon$, where $\upsilon = \upsilon_z\, dz + \upsilon_{\overline{z}}\, d\overline{z}$ is a real harmonic one-form. The Chan-Paton line bundle $\mathcal{L} \to \boldsymbol{F}_p$ of the brane $\mathcal{B}_{\boldsymbol{F}_p}$ has a flat unitary connection that can be conveniently represented as

$$\beta = \frac{1}{2\pi} \int_C |d^2 z|\, (a_z \delta A_{\overline{z}} + a_{\overline{z}} \delta A_z) \tag{11.17}$$

where $a = a_z\,dz + a_{\overline{z}}\,d\overline{z}$ is a real harmonic one-form on C. Similarly, the connection on the line bundle $\mathcal{N}$ on the c.c. brane is[56]

$$\alpha = -\frac{\operatorname{Im}\tau}{2\pi}\int_C |d^2z|\,(\phi_z\delta A_{\overline{z}} + \phi_{\overline{z}}\delta A_z)\,. \tag{11.18}$$

Finally, the symplectic form $\omega = (\operatorname{Im}\tau)\omega_K$ is the exterior derivative of a one-form

$$\zeta = -\frac{i\operatorname{Im}\tau}{2\pi}\int_C |d^2z|\,(\phi_{\overline{z}}\delta A_z - \phi_z\delta A_{\overline{z}})\,. \tag{11.19}$$

The action of the A-model, up to Q-exact terms, is an integral over the boundary of Σ. The contribution to the action from the right boundary, which we call $\partial_R\Sigma$, is

$$\begin{aligned}\int_{\partial_R\Sigma}\Phi^*(\zeta - i\beta) = \frac{i\operatorname{Im}\tau}{2\pi}\int_{\partial_R\Sigma}\Big(&\left(\upsilon_z - (\operatorname{Im}\tau)^{-1}a_z\right)\dot{A}_{\overline{z}}\\ &-\left(\upsilon_{\overline{z}} + (\operatorname{Im}\tau)^{-1}a_{\overline{z}}\right)\dot{A}_z\Big)\,ds.\end{aligned} \tag{11.20}$$

Here we used that on the right boundary, $\phi = \upsilon$. Also, s is a "time" coordinate on $\Sigma = \mathbb{R}\times I$, and $\dot{A} = \partial A/\partial s$. The left boundary $\partial_L\Sigma$, having the opposite orientation, contributes

$$-\int_{\partial_L\Sigma}\Phi^*(\zeta - i\alpha) = -\frac{i\operatorname{Im}\tau}{\pi}\int_{\partial_L\Sigma}\phi_z\dot{A}_{\overline{z}}\,ds. \tag{11.21}$$

The supersymmetric string states come from zero modes along the string, so to determine them, the distinction between $\partial_L\Sigma$ and $\partial_R\Sigma$ is unimportant and we can just add the two contributions to the action. The total action is accordingly

$$\begin{aligned}&\frac{i\operatorname{Im}\tau}{\pi}\int\left(-\phi_z + \frac{1}{2}(\upsilon_z - (\operatorname{Im}\tau)^{-1}a_z)\right)\dot{A}_{\overline{z}}\,ds - \frac{i\operatorname{Im}\tau}{2\pi}\int(\upsilon_{\overline{z}} + (\operatorname{Im}\tau)^{-1}a_{\overline{z}})\dot{A}_z\,ds.\end{aligned} \tag{11.22}$$

[56] To get this formula and the next one, we use eqn. (4.8) for ω_J and ω_K. In eqn. (4.8), Tr represents a trace in the N-dimensional representation of $U(N)$, and A and ϕ are understood to be skew-hermitian. Since here we consider A and ϕ as real one-forms, we must include a minus sign.

We see that the action depends on the parameters $v_z, v_{\overline{z}}, a_z, a_{\overline{z}}$ in the combinations

$$v_z - (\operatorname{Im}\tau)^{-1} a_z, \;\; v_{\overline{z}} + (\operatorname{Im}\tau)^{-1} a_{\overline{z}}. \tag{11.23}$$

The first term in (11.22), upon quantization, tells us how ϕ_z acts:

$$\phi_z \to -i\pi(\operatorname{Im}\tau)^{-1}\left(\frac{\delta}{\delta A_{\overline{z}}} + \frac{i\operatorname{Im}\tau}{2\pi} v_z - \frac{i}{2\pi} a_z\right). \tag{11.24}$$

This is a covariant ∂ operator on a topologically trivial line bundle over $\mathrm{Jac}(C)$. The second term indicates that the wavefunctions are sections of a topologically trivial but holomorphically nontrivial line bundle $T_p \to \mathrm{Jac}(C)$, which is, in fact, the same line bundle $T_p = \mathcal{L} \otimes \mathcal{N}^{-1}|_{\boldsymbol{F}_p}$ that we identified before. Indeed, this term is a total derivative, so if we absorb it into the initial and final state wavefunctions, we find that the wavefunctions are no longer independent of A_z but are annihilated by the operator

$$\frac{\delta}{\delta A_z} - \frac{i\operatorname{Im}\tau}{2\pi} v_{\overline{z}} - \frac{i}{2\pi} a_{\overline{z}}. \tag{11.25}$$

This is a covariant $\overline{\partial}$ operator on a trivial line bundle over $\mathrm{Jac}(C)$. Obviously, these two operators define a flat connection on $\mathrm{Jac}(C)$. It is unitary if and only if $v = 0$.

So in short, quantization of the $(\mathcal{B}_{c.c.}, \mathcal{B}_{\boldsymbol{F}_p})$ strings has given us a $\mathcal{D}$-module associated to a choice of complex flat connection on a trivial line bundle over $\mathrm{Jac}(C)$. On the other hand, S-duality identifies this family of A-branes with the set of all zerobranes on $\mathcal{M}_H(U(1), C)$, which in complex structure J is the moduli space of flat $\mathbb{C}^*$-bundles over C. In other words, in the abelian case S-duality establishes a natural correspondence between gauge-equivalence classes of flat $\mathbb{C}^*$ connections on C and those on $\mathrm{Jac}(C)$.

This correspondence can be seen directly. A flat $\mathbb{C}^*$ connection on a manifold M is the same as a homomorphism from $H_1(M, \mathbb{Z})$ to $\mathbb{C}^*$. Thus it is sufficient to show that $H_1(C, \mathbb{Z})$ is isomorphic to $H_1(\mathrm{Jac}(C), \mathbb{Z})$. But since $\mathrm{Jac}(C) \simeq H^1(C, \mathbb{R})/H^1(C, \mathbb{Z})$, this is the same as proving that $H_1(C, \mathbb{Z}) \simeq H^1(C, \mathbb{Z})$, which is Poincaré duality.

11.3. Generalizations of the c.c. Brane and Twisted Differential Operators

We have seen that geometric Langlands duality can be understood as a result of the S-transformation exchanging $\Psi = 0$ and $\Psi = \infty$. More generally, the

transformation S maps a gauge theory with gauge group G and canonical parameter Ψ to the gauge theory with gauge group ${}^LG$ and the canonical parameter $-1/n_{\mathfrak{g}}\Psi$ (where, as we explained in section 2.1, $n_{\mathfrak{g}} = 1$ if G is simply-laced, and otherwise $n_{\mathfrak{g}} = 2$ or 3). Our goal here is to describe what sort of generalization of the geometric Langlands program arises when we depart from $\Psi = 0, \infty$. (See [9], sections 6.3 and 8.6, for a discussion of this generalization via conformal field theory, and [146] for the abelian case.)

The generalization of the geometric Langlands program that we will explore is perhaps most interesting for the case of rational Ψ, since, as we saw in section 6.2, this is the case in which there are interesting line operators. Moreover, the full S-duality group, which as we showed in section 3.5, is generated by $T : \Psi \to \Psi + 1$ along with $S : \Psi \to -1/n_{\mathfrak{g}}\Psi$, maps rational values of Ψ to rational values. In fact, all rational values of Ψ are related by this group, a fact which also ensures that at a rational value of Ψ there are plenty of branes, just as there are at $\Psi = 0$ and $\Psi = \infty$.

The duality nonetheless works for all complex-valued Ψ. For simplicity, however, here we will take Ψ to be real. One can obtain an arbitrary real value of Ψ by letting $t = 1$ and keeping the θ-angle arbitrary:

$$\Psi = \frac{\theta}{2\pi} = \mathrm{Re}\ \tau. \tag{11.26}$$

S-duality in general transforms $\mathrm{Re}\,\tau = \theta$, $\mathrm{Im}\,\tau = 4\pi/e^2$, and t. But Ψ is the only important parameter, so as long as Ψ is real, we can think of S-duality as acting on θ only. In this set-up, the "electric" and "magnetic" theories look much more symmetric: they both have $t = 1$ but their θ-angles are inversely related. Upon compactification on the Riemann surface C, one obtains an A-model in complex structure K with target $\mathcal{M}_H(G, C)$ and B-field

$$B = -\omega_I \mathrm{Re}\ \tau, \tag{11.27}$$

as explained in eqn. (4.19). The S-transformation is implemented by replacing Re τ by $-1/n_{\mathfrak{g}}$Re τ and G by ${}^LG$.

The B-field given by (11.27) is not flat along the fibers of the Hitchin fibration; indeed, it is a multiple of the Kahler form in complex structure I and hence is nondegenerate when restricted to a fiber. An important consequence is that the fibers of the Hitchin fibration are no longer valid A-branes for general Ψ. To explain this, recall that the sigma-model with target X in

the presence of a D-brane wrapped on a submanifold Y is invariant under a gauge transformation

$$B \longrightarrow B - d\lambda, \quad A \longrightarrow A + \lambda|_Y, \quad \lambda \in \Omega^1(X). \tag{11.28}$$

The gauge-invariant combination is $F + B|_Y$. Now suppose X is symplectic and Y is Lagrangian. If $B = 0$, then a single D-brane wrapped on Y can be an A-brane only if $F = 0$. The gauge-invariant generalization of this condition, which is also valid for $B \neq 0$, is

$$F + B|_Y = 0. \tag{11.29}$$

If the cohomology class of $B|_Y/2\pi$ is not integral, then this equation cannot be satisfied, for any $U(1)$ gauge field, and therefore a single A-brane cannot be wrapped on Y. In our case, this means that for $\theta \neq 2\pi n$, the fibers of the Hitchin fibration are not valid A-branes, regardless of the choice of Chan-Paton bundle.

For non-integer Ψ, the objects of study will therefore not be branes of rank one supported on a fiber of the Hitchin fibration. As a consistency check, note that their naive T-duals, i.e. points on $\mathcal{M}_H({}^L G, C)$, are also not valid topological branes at $t = 1$ and $\theta \neq 0$. In fact, a point is never an A-brane, since the support of an A-brane is always at least middle-dimensional. For rational Ψ, there are A-branes of finite rank supported on a fiber of the Hitchin fibration; we can find what they are by following the duality starting at $\Psi = 0$. In any case, S-duality still maps A-branes for $\mathcal{M}_H(G, C)$ to A-branes for $\mathcal{M}_H({}^L G, C)$, even though examples of A-branes may be scarce.

But what can we say about A-branes when $\Psi \neq 0$ that generalizes what we have found at $\Psi = 0$? In the remainder of this subsection we will answer this question. We will perform all computations near the classical limit $e^2 \to 0$, $\operatorname{Im} \tau \to \infty$, but since the theory depends only on $\Psi = \operatorname{Re} \tau$, the results will be valid for all values of e^2.

We will see that an analogue of the c.c. brane exists for all Ψ. We will call this brane $\mathcal{B}^{\Psi}_{c.c.}$. Its algebra of open strings will turn out to be $\mathcal{D}_{K_{\mathcal{M}}^{1/2} \otimes \mathfrak{L}^{\Psi}}$, that is, the sheaf of holomorphic differential operators on $K_{\mathcal{M}}^{1/2} \otimes \mathfrak{L}^{\Psi}$, where $\mathfrak{L}$ is the determinant line bundle on $\mathcal{M}(G, C)$ (see the discussion of eqn. (4.10)). Given any other A-brane $\mathcal{B}'$ on $\mathcal{M}_H(G, C)$, the sheaf of $(\mathcal{B}^{\Psi}_{c.c.}, \mathcal{B}')$ strings is a module for $\mathcal{D}_{K_{\mathcal{M}}^{1/2} \otimes \mathfrak{L}^{\Psi}}$. So an A-brane at general Ψ gives a sheaf of modules for this algebra. Just as at $\Psi = 0$, the construction really comes from four-dimensional gauge theory, which in algebraic geometry means that

these twisted $\mathcal{D}$-modules can be extended to the stack of all G-bundles. S-duality then maps twisted $\mathcal{D}$-modules for G to those of ${}^LG$, with the exchange $\Psi \to -1/n_{\mathfrak{g}}\Psi$.

The gauge field on a space-filling A-brane must solve the equation

$$(\omega^{-1}(F+B))^2 = -1, \tag{11.30}$$

where $\omega = (\mathrm{Im}\tau)\,\omega_K$. An obvious solution is

$$F = \mathrm{Im}\,\tau\,(\cos\mathfrak{q})\,\omega_J, \quad \sin\mathfrak{q} = -\frac{\mathrm{Re}\ \tau}{\mathrm{Im}\,\tau}. \tag{11.31}$$

We take this solution to define the brane $\mathcal{B}^{\Psi}_{c.c.}$. Note that, as $\mathrm{Im}\,\tau$ can be arbitrarily large, $\mathfrak{q}$ is extremely small, and ultimately the interesting effects will be linear in $\Psi = \mathrm{Re}\ \tau$, which is of order $\mathfrak{q}$.

If we ignore the algebra structure, then the sheaf of open strings is locally isomorphic to the sheaf of holomorphic functions on $\mathcal{M}_H(G,C)$ in complex structure[57]

$$I(\Psi) = \omega^{-1}(F+B) = I\cos\mathfrak{q} - J\sin\mathfrak{q}. \tag{11.32}$$

The corresponding holomorphic coordinates on $\mathcal{M}_H(G,C)$ are

$$A'_{\overline{z}} = A_{\overline{z}} - i\tan\frac{\mathfrak{q}}{2}\,\phi_{\overline{z}}, \quad \phi'_z = \phi_z + i\tan\frac{\mathfrak{q}}{2}\,A_z. \tag{11.33}$$

As usual, A is a connection on a G-bundle $E \to C$, and ϕ is a one-form valued in $\mathrm{ad}(E)$. What we have in section 4.2 called Hitchin's second fibration is the forgetful map

$$\pi_0 : (A_{\overline{z}}, \phi_z) \longrightarrow A_{\overline{z}} \tag{11.34}$$

from $\mathcal{M}_H$ to $\mathcal{M}$. It is plain that π_0 is not holomorphic with respect to $I(\Psi)$ except for $\Psi = 0$. To rectify the situation, let us define a new $\overline{\partial}$-connection on E by

$$D'_{\overline{z}} = D_{\overline{z}} - i\tan\frac{\mathfrak{q}}{2}\,\phi_{\overline{z}} = \partial_{\overline{z}} + A_{\overline{z}} - i\tan\frac{\mathfrak{q}}{2}\,\phi_{\overline{z}}, \tag{11.35}$$

[57] Comparing to eqn. (4.6), we see that $I(\Psi)$ is the same as I_w with $w = i\tan(\mathfrak{q}/2)$.

and map the pair (A, ϕ) into the holomorphic bundle on C defined by $D'_{\overline{z}}$. Let us denote this map π_Ψ; it is obviously holomorphic with respect to $I(\Psi)$ and for $\Psi = 0$ reduces to π_0. As usual, we will restrict this fibration to the subset of Higgs bundles for which E, endowed with the modified $\overline{\partial}$ operator, is stable.

What does the fiber of the map π_Ψ look like? To understand this, we note that the obvious candidate for the fiber coordinate, namely $\phi_z + i \tan \frac{\mathfrak{q}}{2} A_z$, transforms inhomogeneously under gauge transformations. The object which transforms homogeneously is

$$D'_z = \phi_z + i \tan \frac{\mathfrak{q}}{2} D_z, \tag{11.36}$$

where D_z is the holomorphic covariant derivative on E. Up to a factor $i \tan \frac{\mathfrak{q}}{2}$, D'_z is a ∂-connection over the bundle E. We choose not to divide by $i \tan \frac{\mathfrak{q}}{2}$, because we would like the formulas to make sense even for $\Psi = 0$. Then D'_z is what is called a holomorphic λ-connection on E, where $\lambda = i \tan \frac{\mathfrak{q}}{2}$. For any complex manifold X, a holomorphic λ-connection on a holomorphic bundle $E \to X$ is a linear map $\nabla : \Gamma(E) \to \Gamma(E \otimes T^*X)$ that commutes with the $\overline{\partial}$ operator of E and such that

$$\nabla(f \cdot s) = f \nabla s + \lambda \partial f \wedge s, \tag{11.37}$$

for any function f and any section s of E. For $\lambda = 1$, a λ-connection is an ordinary ∂ operator on E, while for $\lambda = 0$ it is a section of $\mathrm{End}(E) \otimes T^*X$.

By virtue of Hitchin's equations, D'_z commutes with $D'_{\overline{z}}$ (this is closely related to the discussion of complex structure I_w at the end of section 4.1) and therefore defines a holomorphic λ-connection on the holomorphic bundle $E' = \pi_\Psi(E, \phi)$. By the usual logic, this implies that the fiber over E' is the moduli space of holomorphic λ-connections on E'.

The space of holomorphic λ-connections on a fixed vector bundle $E' \to C$ is naturally an affine space modeled on the space of holomorphic sections of $\mathrm{End}(E') \otimes K_C$, where K_C is the canonical line bundle of C. (This statement means that to a holomorphic λ-connection, we can add a holomorphic section of $\mathrm{End}(E') \otimes K_C$, and any two holomorphic λ-connections are related in this way.) The space $H^0(C, \mathrm{End}(E') \otimes K_C)$ can be identified with the cotangent space to $\mathcal{M}(G, C)$ at the point E'. Thus in complex structure $I(\Psi)$, the moduli space $\mathcal{M}_H(G, C)$ looks like an affine deformation of $T^*\mathcal{M}(G, C)$; in other words, it is a bundle of affine spaces, with the underlying bundle of vector spaces being $T^*\mathcal{M}(G, C)$. To summarize, if we ignore the algebra structure,

the sheaf of $(\mathcal{B}^{\Psi}_{c.c.}, \mathcal{B}^{\Psi}_{c.c.})$ strings is the sheaf of holomorphic functions on a "twisted" cotangent bundle of $\mathcal{M}(G, C)$.

It remains to determine the algebra structure of the sheaf of $(\mathcal{B}^{\Psi}_{c.c}, \mathcal{B}^{\Psi}_{c.c.})$ strings. We restrict ourselves to functions of polynomial growth along the fibers of the affine bundle.[58] We can follow the same logic as at $\Psi = 0$. Locally, the fact that we are quantizing a twisted cotangent bundle rather than an ordinary one is irrelevant. Globally, it might (and as we will see, it does) modify the transition functions in (11.16). So the sheaf of c.c. strings at any Ψ is the sheaf of holomorphic differential operators acting on a power of some holomorphic line bundle over $\mathcal{M}(G, C)$.

We can be more precise if Ψ is an integer. Shifting Ψ to $\Psi + n$ shifts the B-field, or more precisely $B/2\pi$, by the first Chern class of the line bundle $\mathfrak{L}^{-n}$. (We can see this from (4.19); the shift $\Psi \to \Psi + n$, which is $\theta \to \theta + 2\pi n$, shifts B by $-n\omega_I$, where $\omega_I/2\pi$ is the first Chern class of $\mathfrak{L}$.) This observation enables us to define a convenient A-brane in complex structure K. At $\Psi = 0$, one of the important branes was the brane $\mathcal{B}_0$ supported at $\phi = 0$ with trivial Chan-Paton bundle. At $\Psi = n$, we likewise can consider a brane $\mathcal{B}_n$ supported at $\phi = 0$, but now with Chan-Paton bundle $\mathfrak{L}^n$, endowed with its natural connection. This obeys (11.29), and so is an A-brane in complex structure K.

Since we have set $\Psi = n$, we write $\mathcal{B}^n_{c.c.}$ instead of $\mathcal{B}^{\Psi}_{c.c.}$. Now we quantize the $(\mathcal{B}^n_{c.c.}, \mathcal{B}_n)$ strings. Everything is as before, except that we must include the Chan-Paton bundle of the brane $\mathcal{B}_n$. So the physical states are now sections of $K_{\mathcal{M}}^{1/2} \otimes \mathfrak{L}^n \otimes \left(\oplus_{j=0}^n \Omega^{0,j}(\mathcal{M})\right)$. This is a module for $\mathcal{D}_{K_{\mathcal{M}}^{1/2} \otimes \mathfrak{L}^n}$, and not for $\mathcal{D}_{K_{\mathcal{M}}^{1/2} \otimes \mathfrak{L}^m}$ for any other m. So the sheaf of algebras of $(\mathcal{B}^n_{c.c.}, \mathcal{B}^n_{c.c.})$ strings is precisely $\mathcal{D}_{K_{\mathcal{M}}^{1/2} \otimes \mathfrak{L}^n}$.

What happens if Ψ is not an integer? We cannot hope to find for general Ψ a brane that leads to a module consisting of sections of $K_{\mathcal{M}}^{1/2} \otimes \mathfrak{L}^{\Psi}$, since there is no such module unless Ψ is an integer. However, we know that the

[58] This restriction amounts to picking a particular algebraic structure on $\mathcal{M}_H$ endowed with the complex structure $I(\Psi)$. In this algebraic structure, it is called the moduli space of G-bundles with λ-connection. There is a second algebraic structure that is compatible with the same complex structure, namely one in which $\mathcal{M}_H$ is equivalent to the moduli space of stable homomorphisms $\pi_1(C) \to G_{\mathbb{C}}$. In the second algebraic structure, the traces of holonomies are regarded as the algebraic functions. What we have just encountered is the one point in the present paper in which we have to choose between the different algebraic structures compatible with a fixed complex structure on $\mathcal{M}_H$. This choice plays a more prominent role in other approaches to the geometric Langlands program.

algebra of c.c. strings at any Ψ is the sheaf of differential operators acting on $K_{\mathcal{M}}^{1/2} \otimes \mathfrak{L}^{f(\Psi)}$ for some function $f(\Psi)$. So if we can prove that $f(\Psi)$ is a linear function, this, together with the fact that $f(\Psi) = \Psi$ for integer Ψ, implies that $f(\Psi) = \Psi$ for all Ψ.

To show that $f(\Psi)$ is linear, consider the Euclidean action of the sigma-model:

$$S = \int_\Sigma \Phi^*(\omega - iF - iB) = -i\,\mathrm{Im}\tau \int_\Sigma \Phi^* \left(\cos \mathsf{q}\,\omega_J + \sin \mathsf{q}\,\omega_I + i\omega_K\right). \tag{11.38}$$

(We have omitted terms of the form $\{Q, V\}$.) For $\Psi = 0$, we have $\mathsf{q} = 0$, and the two-form in the integrand becomes an exact two-form on $T^*\mathcal{M}(G, C)$. The corresponding one-form potential ϖ is, up to a numerical factor, the canonical one-form $p_\alpha\, dq^\alpha$. For $\Psi \neq 0$, the best we can do is to write this two-form as an exact form on $T^*\mathcal{M}(G, C)$ plus the pull-back of a closed two-form from $\mathcal{M}(G, C)$:

$$\Omega' = \frac{\mathrm{Im}\tau}{\pi} \delta \int_C \mathrm{Tr}\,(\phi_z \delta A'_{\bar z}) - i\,\mathrm{Re}\,\tau \int_C \mathrm{Tr}\,(\delta A'_z \delta A'_{\bar z}), \tag{11.39}$$

where we defined A'_z as minus the Hermitian conjugate of $A'_{\bar z}$. The second term in this formula is $\mathrm{Re}\,\tau$ times a $(1,1)$-form on $\mathcal{M}(G, C)$ which is closed but not exact. It is a multiple of the curvature of the natural line bundle $\mathfrak{L} \to \mathcal{M}$. To compute the transition functions (11.16), we do perturbation theory using a propagator that is the inverse of the form Ω'. (11.16) was deduced from the commutator $[\widehat{p}_\alpha, \widehat{p}_\beta]$, which is proportional to $\hbar^2$ or $1/(\mathrm{Im}\,\tau)^2$. So we need to do perturbation theory up to order $1/(\mathrm{Im}\,\tau)^2$. In that order, perturbation theory is linear in $\Psi = \mathrm{Re}\,\tau$, since the inverse of Ω' can be expanded as a series in $\mathrm{Re}\,\tau$ and $\mathrm{Im}\,\tau$ that schematically is of the form $\sum_{k\geq 0} a_k (\mathrm{Re}\,\tau)^k/(\mathrm{Im}\,\tau)^{k+1}$. This completes the argument.

The canonical line bundle over $\mathcal{M}(G, C)$ is isomorphic to $\mathfrak{L}^{-2h}$, where h is the dual Coxeter number of G [141]. Thus we can also say that the algebra of c.c. strings at any given value of Ψ is the sheaf of holomorphic differential operators on the line bundle $\mathfrak{L}^{\Psi - h}$. In particular, if we let $\Psi = h$, the algebra of c.c. strings is the sheaf of ordinary differential operators, that is, differential operators acting on functions. It follows that by combining S-duality with a transformation by T^h, which shifts Ψ by h, we can map a zerobrane on $\mathcal{M}_H({}^L G, C)$ to an ordinary (untwisted) $\mathcal{D}$-module on $\mathcal{M}(G, C)$. This is a more standard formulation of the geometric Langlands duality. For certain G, it is also a more precise statement, as one sees if one considers carefully [5,26] the dependence on the spin structure of C.

12. Branes From Gauge Theory

So far we have always considered branes in the effective two-dimensional sigma-model with target $\mathcal{M}_H$. Our goal in this concluding section is to show that the important branes actually can be defined in four-dimensional gauge theory.

Branes of any interest always preserve at least one supersymmetry. For example, they are B-branes in complex structure J or A-branes in complex structure K. But the branes important for geometric Langlands duality have special properties with respect to all three complex structures. For example, the zerobrane on $\mathcal{M}_H({}^LG, C)$ is a (B, B, B)-brane: it is a B-brane with respect to complex structures I, J, K. Its mirror is a fiber of the Hitchin fibration, which is complex with respect to I and Lagrangian with respect to J and K. Thus it is a (B, A, A)-brane. The canonical coisotropic brane is an (A, B, A)-brane, and so is its mirror. Branes of any of these types preserve two topological supercharges (linear combinations of which give, for example, for a brane of type (B, B, B), the requisite B-type supersymmetry in complex structures I, J, K).

We will focus on these four kinds of brane and describe the corresponding boundary conditions in the gauge theory. The lift to gauge theory is not necessarily unique; in fact, in some cases we describe several distinct boundary conditions in the gauge theory which upon reduction to two dimensions apparently become equivalent, at least away from singularities.

To discuss the twisted theory as a topological field theory, one replaces four-dimensional Minkowski space by a general four-manifold M. In topological field theory, it is most convenient to take M to have Euclidean signature, for the following reason. In going to a general four-manifold, in order to preserve some supersymmetry, one "twists" by interpreting the first four scalar fields $\phi_0, \dots, \phi_3$ as a section of the tangent bundle of M. Since the scalar fields naturally have a positive signature, this twisting is much more natural with Euclidean signature on M. (Alternatively, one could possibly make a Wick rotation in ϕ space, but this seems unfelicitous.) Additionally, any compact four-manifold admits a positive signature metric, while admitting a metric of Lorentz signature is a severe topological restriction.

For the present discussion, we will consider branes in a more restricted situation. As usual, we compactify to two dimensions on a Riemann surface C. This gives an effective two-dimensional theory, and we want to understand how branes in this theory can be interpreted in the underlying four-dimensional gauge theory. Near the boundary, there is a natural "time"

direction (normal to C and tangent to the boundary) and no major advantage in a Wick rotation. On the contrary, we wish to avoid the Wick rotation in order to make clear that all branes we consider are physically sensible and unitary.

This being so, whenever a topological supercharge Q is preserved by our boundary conditions, its adjoint $Q^\dagger$ is also preserved. They obey a physical supersymmetry algebra $\{Q, Q^\dagger\} = H$, with H the Hamiltonian. Since our boundary conditions preserve two topological supercharges, they preserve a total of four supersymmetries. This is one-fourth of the 16 global supersymmetries of $\mathcal{N} = 4$ super Yang-Mills theory, so such branes are called 1/4 BPS branes. From the point of view of the sigma model, which has only one-half as much supersymmetry, they are called 1/2 BPS branes. Going back to gauge theory, however, in some instances our boundary conditions in the gauge theory preserve eight supersymmetries locally along C, and the reduction to four supersymmetries comes from the twisting and curvature of C.

To describe a brane in the two-dimensional effective theory, we formulate this theory on (say) $\mathbb{R}^2_+$, with coordinates x^0, x^1, where $-\infty < x^0 < \infty$, $0 \leq x^1 < \infty$. A supersymmetric brane is obtained by specifying a supersymmetric boundary condition on $x^1 = 0$. To obtain such a construction from the underlying four-dimensional gauge theory, we simply consider this theory on $M = \mathbb{R}^2_+ \times C$, and again describe supersymmetric boundary conditions at $x^1 = 0$.

We write x^2, x^3 for local coordinates on C. As we take $\mathbb{R}^2_+$ to be flat, the twisting only affects the two scalar fields ϕ_2, ϕ_3 (or A_6, A_7; the notation was explained in section 2.1). The supersymmetry left unbroken by the compactification and twisting is generated by those spinors ϵ that are covariantly constant on C in the appropriate sense. The condition, in the twisted theory, is simply that ϵ should obey

$$\Gamma_{2367}\epsilon = \epsilon, \tag{12.1}$$

as in eqn. (5.3).

12.1. General Properties Of Boundary Conditions

First let us recall a few generalities about what a boundary condition is supposed to be.

A boundary condition constrains the values of the fields (and their normal derivatives, in the case of bosons that obey second order equations of

motion) in such a way that the boundary terms in the equations of motion vanish. For example, for gauge fields on an n-manifold M with metric g of signature $-++\ldots+$, the minimal action is

$$\boldsymbol{I}_B = \frac{1}{2e^2}\int_M d^n x\sqrt{g}\,\mathrm{Tr}\,F_{IJ}F^{IJ}. \tag{12.2}$$

M is an n-manifold with x^0 as the "time" direction and with a boundary defined by $x^1 = 0$. The boundary terms in the variation $\delta\boldsymbol{I}_B$ of the action are

$$(\delta\boldsymbol{I}_B)_{bdry} = \frac{2}{e^2}\int_{\partial M} d^{n-1}x\sqrt{g}\,\mathrm{Tr}\left(g^{IJ}\sum_{I,J\neq 1}\delta A_I F_{1J}\right). \tag{12.3}$$

A boundary condition must set to zero a linear combination of the boundary values of F_{I1} and δA_I, for $I \neq 1$, such that $(\delta\boldsymbol{I}_B)_{bdry} = 0$. What it means to set $\delta A_I = 0$ for some I is to specify the boundary values of A_I or in other words to impose Dirichlet boundary conditions on A_I, with some prescribed boundary values. Conversely, in a gauge with $A_1 = 0$ near the boundary of M, the condition that $F_{I1} = 0$ means that the normal derivative of A_I vanishes. So $F_{I1} = 0$ amounts to a gauge-invariant version of Neumann boundary conditions on A_I. For quantization, we do not want to overconstrain the boundary values, so we set to zero a minimal linear combination of δA_I and F_{I1} such that $(\delta\boldsymbol{I}_B)_{bdry} = 0$. (The boundary conditions therefore define, in a suitable sense, a Lagrangian submanifold of the boundary data.) It is also possible to add boundary terms to $\boldsymbol{I}_B$, thus modifying $(\delta\boldsymbol{I}_B)_{bdry}$; the boundary conditions always define a maximal subspace of allowed boundary values (of A_I and F_{I1}, $I \neq 1$), such that $(\delta\boldsymbol{I}_B)_{bdry} = 0$. We call such a boundary condition a hyperbolic boundary condition.

There is, of course, a similar story for fermions. The minimal action is

$$\boldsymbol{I}_F = -\frac{i}{e^2}\int_M d^n x\sqrt{g}\,\mathrm{Tr}\,\overline{\lambda}\Gamma^I D_I\lambda, \tag{12.4}$$

to which one may add additional boundary terms. Using the minimal action, the boundary term in the variation of $\boldsymbol{I}_F$ is

$$\delta\boldsymbol{I}_F = -\frac{i}{e^2}\int_{\partial M} d^{n-1}x\sqrt{g}\,\mathrm{Tr}\,\overline{\lambda}\Gamma_1\delta\lambda. \tag{12.5}$$

For fermions, a suitable boundary condition sets to zero the boundary values of one-half the components of λ (and therefore of $\delta\lambda$) such that $\overline{\lambda}\Gamma_1\delta\lambda = 0$.

In the bulk theory, supersymmetry means (for M flat or after suitable twisting for M curved) that there is a conserved supercurrent J_I. For supersymmetric Yang-Mills theory,

$$J_I = \frac{1}{2}\Gamma^{KL}F_{KL}\Gamma_I\lambda. \tag{12.6}$$

In the presence of a boundary, the condition that the boundary condition is compatible with supersymmetry is that the normal component of J vanishes at the boundary, so that the flux of the supercurrent does not disappear at the boundary. More precisely, the condition is

$$\overline{\epsilon}J_1 = 0 \tag{12.7}$$

for suitable supersymmetry generators ϵ.

This automatically ensures that the boundary conditions are consistent with the supersymmetric variations of both bosons and fermions. If, therefore, one knows what boundary conditions one wants on bosons and fermions, the condition $J_1 = 0$ is the basic one. In practice, since we do not know to begin with what boundary conditions we want, we will start by postulating a simple boundary condition on the fermions (chosen to ensure that $\overline{\lambda}\Gamma_1\delta\lambda = 0$), deduce the corresponding boundary conditions on bosons to ensure that the boundary conditions on the fermions are preserved by supersymmetry, and then verify that $J_1 = 0$.

If $J_1 = 0$, we have found boundary conditions that will enable us to define conserved supercharges (if M is flat or in a suitably twisted theory if M is curved). However, we want more. A boundary condition that forces supersymmetry to be spontaneously broken will not be of much interest if our goal is to study supersymmetric states or topological field theory (in which only the supersymmetric states are of interest). Dirichlet boundary conditions for bosons are generically incompatible with unbroken supersymmetry. Dirichlet boundary conditions mean, as we have discussed, that the boundary values of some components A_I of the gauge fields, and hence (possibly) of some components F_{IJ} of the field strength, are specified. We are mainly interested in boundary conditions in which the boundary values of F_{IJ} are fixed in a way that is compatible with unbroken supersymmetry, that is with the vanishing of the supersymmetry variation

$$\delta\lambda = \frac{1}{2}\Gamma^{IJ}F_{IJ}\epsilon \tag{12.8}$$

for suitable supersymmetry generators ϵ.

12.2. Branes of Type (B, B, B)

Our branes will always be constructed to preserve those supersymmetries whose generators obey a suitable condition $T\epsilon = \epsilon$ as well as[59] $\Gamma_{2367}\epsilon = \epsilon$. These combined conditions will allow four supersymmetries. T will commute with Γ_{2367}. In some examples, T will be chosen so that $T^2 = 1$. Then the condition $T\epsilon = \epsilon$ would, by itself, allow eight supersymmetries, and the boundary condition will preserve those eight supersymmetries. Even when that is so, the compactification to two dimensions and the "twisting" will impose the second condition $\Gamma_{2367}\epsilon = \epsilon$ on unbroken supersymmetries, leaving only four supersymmetries in the low energy theory.

We will explain illustrative choices of T that lead to branes of the various types. We certainly do not claim to describe all interesting constructions of branes in the four-dimensional gauge theory, only some simple ones that are sufficient to exhibit in the gauge theory the branes that will be most important in the present paper.

To obtain branes of type (B, B, B), we consider first $T = \Gamma_{04}$ with the boundary condition

$$T\lambda| = \lambda|. \tag{12.9}$$

For any field Φ, we write $\Phi|$ for the restriction of Φ to $x^1 = 0$. (However, we omit this symbol when confusion seems unlikely.) This boundary condition on λ is only compatible with supersymmetries such that the supersymmetric transformation $\lambda \to \lambda + \delta\lambda$ preserves the condition $T\lambda| = \lambda|$. The requirement is that $(T-1)\delta\lambda = 0$. Since $\delta\lambda = \frac{1}{2}\Gamma^{IJ}F_{IJ}\epsilon$ and $T\epsilon = \epsilon$, we have

$$(T-1)\delta\lambda = \frac{1}{2}[T, \Gamma^{IJ}F_{IJ}]\epsilon. \tag{12.10}$$

Upon evaluating the commutator, we find that we require

$$\sum_{I\neq 0,4} (\Gamma_{0I}F_{4I} - \Gamma_{4I}F_{0I})\epsilon = 0. \tag{12.11}$$

[59] Replacing the condition $T\epsilon = \epsilon$ with $T\epsilon = -\epsilon$ would lead to nothing essentially new, as there always will be symmetries of the theory that reverse the sign of T. For example, in our first example, $T = \Gamma_{04}$, such a symmetry would be a reflection in the 49 plane.

Since ϵ is constrained precisely by $\Gamma_0\Gamma_4\epsilon = \epsilon$, this condition is equivalent to

$$F_{0I} + F_{4I} = 0 \quad \text{for } I \neq 0, 4. \tag{12.12}$$

To pick a hyperbolic boundary condition compatible with what has just been described, we set (in a suitable gauge) the boundary values of $A_0 + A_4$ to zero, and for A_I, $I \neq 0, 1, 4$, we impose Dirichlet boundary conditions with some specified, time-independent, boundary values. Finally, (12.12) tells us that $F_{01} + F_{41} = 0$, corresponding to gauge-invariant Neumann boundary conditions on $A_0 + A_4$. (Since $A_0 + A_4$ corresponds to a "null" direction, it is possible in a hyperbolic boundary condition to set to zero both $A_0 + A_4$ and its normal derivative. This depends on the minus sign in the signature, the fact that $g^{IJ} = \mathrm{diag}(-1, 1, 1, \ldots, 1)$ in (12.3).)

So far, starting with the boundary conditions on the fermions, we have guessed what the boundary conditions on the bosons must be. At this stage, we can verify the basic condition that the normal component of the supercurrent vanishes on the boundary:

$$\overline{\epsilon}\Gamma^{MN}F_{MN}\Gamma_1\lambda = 0. \tag{12.13}$$

Because $\Gamma_{04}\lambda = \lambda$ and $\Gamma_{04}\epsilon = \epsilon$, the only components of F_{MN} that contribute are those with precisely one of M and N equal to 0 or 4; using this, one can verify that (12.13) is obeyed if $F_{0I} + F_{4I} = 0$, $I \neq 0, 4$.

This implies that the boundary conditions on both fermions and bosons are compatible with supersymmetry. We have already seen this for the fermions, and we can readily verify it for the bosons. The supersymmetry variation of the bosons

$$\delta A_I = i\overline{\epsilon}\Gamma_I\lambda \tag{12.14}$$

implies, with $(\Gamma_{04} - 1)\lambda = (\Gamma_{04} - 1)\epsilon = 0$, that $\delta A_I = 0$ for $I \neq 0, 4$, and $\delta A_0 + \delta A_4 = 0$, whence $\delta(F_{0I} + F_{4I}) = 0$ for $I \neq 0, 4$. Hence our condition $F_{0I} + F_{4I} = 0$, $I \neq 0, 4$, is consistent with supersymmetry.

So far, we have obtained a boundary condition that ensures the existence of conserved supercharges. We also would like the boundary condition to allow for existence of a supersymmetric state. For this, the Dirichlet boundary conditions on A_I, $I \neq 0, 1, 4$, are highly constrained. Given the supersymmetric variation $\delta\lambda = \frac{1}{2}\sum_{IJ}\Gamma^{IJ}F_{IJ}\epsilon$, we want to pick the boundary

values such that

$$\sum_{I,J\neq 0,1,4} \Gamma^{IJ} F_{IJ}\epsilon = 0. \tag{12.15}$$

If this were supposed to be so for all ϵ, we would have to set $F_{IJ} = 0$, $I, J \neq 0, 1, 4$. Since ϵ is constrained to $\Gamma_{2367}\epsilon = T\epsilon = \epsilon$, the constraint on F is less severe. It suffices to take F to be selfdual in the subspace generated by x^2, x^3, x^6, and x^7. Thus, we must impose the Hitchin equations, which in this notation read

$$\begin{aligned} F_{23} - F_{67} &= 0 \\ F_{27} - F_{36} &= 0 \\ F_{26} - F_{73} &= 0. \end{aligned} \tag{12.16}$$

Thus, the boundary values of the gauge fields define a point p in $\mathcal{M}_H$, the moduli space of Higgs bundles. Once we impose (12.16), (12.15) is obeyed if in addition $F_{IJ}| = 0$ unless both I and J are in the set $\mathcal{R} = \{2, 3, 6, 7\}$. In particular, vanishing of $F_{IJ}|$ for $I \in \mathcal{R}$, $J \notin \mathcal{R}$ implies that $A_J|$, for $J = 5, 8, 9$ is covariantly constant (on the Riemann surface C with local coordinates x^2, x^3) and commutes with the Higgs field, whose components are $\phi_2 = A_6$ and $\phi_3 = A_7$. If p is a smooth point in $\mathcal{M}_H$ corresponding to an irreducible Higgs bundle, these conditions imply that $A_J| = 0$ for $J = 5, 8, 9$. This furthermore implies the vanishing of $F_{IJ}|$ for $I, J \notin \mathcal{R}$. So for an irreducible Higgs bundle, the boundary condition (of this type) is uniquely determined. If p corresponds to a reducible Higgs bundle, we only learn in general that A_5, A_8, and A_9 generate symmetries of the Higgs bundle and commute with each other.

The boundary conditions we have obtained so far correspond to a zerobrane on $\mathcal{M}_H$ supported at the point p. This is a basic type of (B, B, B)-brane. As expected, such a brane is unique if p corresponds to a smooth point in $\mathcal{M}_H$.

Another Construction Of The Zerobrane Somewhat surprisingly, there is another gauge theory construction of zerobranes on $\mathcal{M}_H$. It seems to lead to the same brane in the low energy theory as long as one is at a smooth point in $\mathcal{M}_H$, but may lead to something new in the case of a zerobrane supported at a singular point in $\mathcal{M}_H$.

Let us replace T by $\widetilde{T} = T\Gamma_{2367} = \Gamma_{023467}$. The supersymmetries preserved by the boundary condition will now be those whose generators obey $\widetilde{T}\epsilon = \epsilon$. So the boundary condition in the four-dimensional gauge theory

will preserve a different set of eight supercharges than what we had before. However, the unbroken supersymmetries in the low energy theory are those that obey the combined conditions $T\epsilon = \Gamma_{2367}\epsilon = \epsilon$. These conditions are invariant under replacing T by $\widetilde{T}$, so the change in boundary conditions will not change the unbroken supersymmetry in the low energy theory and the construction will give a new (B, B, B)-brane.

We will choose the boundary condition on the fermions to be $\widetilde{T}\lambda| = -\lambda|$. Asking for this condition to be invariant under supersymmetry, we require that $\{\Gamma^{IJ}F_{IJ}, \widetilde{T}\}\epsilon = 0$. Let us assume in addition that $\Gamma_{2367}\epsilon = \epsilon$. Then the supersymmetry condition is satisfied if

$$\sum_{I,J\in\{0,2,3,4,6,7\}} \Gamma^{IJ}F_{IJ}|\epsilon = 0 \tag{12.17}$$

and

$$\sum_{I,J\in\{1,5,8,9\}} \Gamma^{IJ}F_{IJ}|\epsilon = 0. \tag{12.18}$$

(12.17) is obeyed if the A_I obey Dirichlet boundary conditions with $A_4| = 0$, and with the boundary values of A_I, $I \in \{2, 3, 6, 7\}$ constrained to obey the Hitchin equations (12.16).

Note that if ϵ does not obey $\Gamma_{2367}\epsilon = \epsilon$, the supersymmetry constraint is not satisfied, in general. This means that even in flat spacetime, this (B, B, B)-brane preserves only four supercharges.

To obey (12.18), we *cannot* ask for $F_{IJ}| = 0$, $I, J \in \{1, 5, 8, 9\}$ as these conditions are overdetermined. Indeed, $F_{15}| = F_{18}| = 0$ means that A_5 and A_8 obey Neumann boundary conditions and are free to fluctuate on the boundary; this being so, we cannot also ask for $F_{58}| = 0$. However, the condition $\widetilde{T}\epsilon = \epsilon$, together with the chirality condition $\widehat{\Gamma}\epsilon = \epsilon$ of ten-dimensional super Yang-Mills theory (here $\widehat{\Gamma} = \Gamma_0\Gamma_1\cdots\Gamma_9$), implies $\Gamma_{1589}\epsilon = -\epsilon$. With this constraint on ϵ, (12.18) is equivalent to an anti-selfduality condition on the boundary values in the 1589 plane:

$$\begin{aligned} F_{15}| + F_{89}| &= 0 \\ F_{18}| + F_{95}| &= 0 \\ F_{19}| + F_{58}| &= 0. \end{aligned} \tag{12.19}$$

This is a modified version of Neumann boundary conditions on A_5, A_8, and A_9, and is *not* overdetermined.

Since A_2, A_3, A_6, and A_7 obey Dirichlet boundary conditions determined by a specific solution of the Hitchin equations, what we obtain in this way is a zerobrane supported at a point p in $\mathcal{M}_H$. What distinguishes it from the previous construction of the zerobrane is the boundary condition on A_4, A_5, A_8, and A_9, as well as the fact that in flat spacetime the former construction preserves half as many supersymmetries as the latter. If p is a smooth point in $\mathcal{M}_H$ corresponding to an irreducible Higgs bundle, the fields A_4, A_5, A_8, and A_9 have no zero modes and vanish in the low energy theory, regardless of the microscopic boundary conditions. Hence the new construction of the zerobrane should be equivalent to the old one as long as p is a smooth point in $\mathcal{M}_H$. If p corresponds to a reducible Higgs bundle whose stabilizer is a subgroup of G of positive rank, then A_4, A_5, A_8, and A_9 do have zero modes – which generate symmetries of the Higgs bundle – and there is no reason to expect the two zerobranes to be equivalent.

One can get yet another (B,B,B)-brane by taking the boundary condition for the fermions to be $\widetilde{T}\lambda| = \lambda|$. It turns out that it corresponds to a space-filling brane with a flat gauge field. Since this brane does not play a major role in the present paper, we omit the details.

12.3. Branes Of Type (B, A, A)

We can get branes of type (B, A, A) by applying S-duality to branes of type (B, B, B). Though it may be in general unclear how to transform a particular brane under S-duality, there is no problem in transforming the unbroken supersymmetries under S-duality. In view of eqn. (2.25), under the transformation S, for $\mathrm{Re}\,\tau = 0$, the generators $\epsilon_{\ell,r}$ of supersymmetry transform as

$$\epsilon_\ell \longrightarrow \epsilon_\ell \cdot \exp\left(-\frac{i\pi}{4}\right), \quad \epsilon_r \longrightarrow \epsilon_r \cdot \exp\left(\frac{i\pi}{4}\right), \tag{12.20}$$

or equivalently

$$\epsilon \longrightarrow \epsilon' = \frac{1+\widehat{\Gamma}}{\sqrt{2}}\epsilon \tag{12.21}$$

where $\widehat{\Gamma} = \Gamma_0\Gamma_1\Gamma_2\Gamma_3$ measures the four-dimensional chirality. If $T\epsilon = \epsilon$, then $\epsilon' = \frac{1+\widehat{\Gamma}}{\sqrt{2}}\epsilon = \frac{1+\widehat{\Gamma}}{\sqrt{2}}T\epsilon = \frac{1+\widehat{\Gamma}}{\sqrt{2}}T\frac{1-\widehat{\Gamma}}{\sqrt{2}}\epsilon'$, so $T'\epsilon' = \epsilon'$ where

$$T' = \frac{1+\widehat{\Gamma}}{\sqrt{2}}T\frac{1-\widehat{\Gamma}}{\sqrt{2}}. \tag{12.22}$$

If $T = \Gamma_0\Gamma_4$, then

$$T' = \Gamma_1\Gamma_2\Gamma_3\Gamma_4. \tag{12.23}$$

We impose the boundary condition $T'\lambda| = -\lambda|$ on the fermions. Then supersymmetry requires that $\{T', \Gamma^{IJ}F_{IJ}\}\epsilon = 0$, or

$$\left(\sum_{I,J\in\{1,2,3,4\}} \Gamma^{IJ}F_{IJ} + \sum_{I,J\notin\{1,2,3,4\}} \Gamma^{IJ}F_{IJ} \right) \epsilon = 0. \tag{12.24}$$

This condition implies that $F_{0I}| = 0$ for $I > 4$, so we impose Dirichlet boundary conditions on the A_I of $I > 4$. We simply set $A_I| = 0$ for $I = 5, 8, 9$, and we look for a boundary condition that gives a nonzero value for the "Higgs fields" $\phi_2 = A_6$, $\phi_3 = A_7$. Assuming that (12.24) is supposed to be true for all ϵ obeying $\Gamma_{1234}\epsilon = \epsilon$ (without the further condition $\Gamma_{2367}\epsilon = \epsilon$), it implies that $[\phi_2, \phi_3] = F_{67} = 0$. Given this, the gauge-invariant content of ϕ_2 and ϕ_3 is in the characteristic polynomial of $\phi_z = \frac{1}{2}(\phi_2 - i\phi_3)$. We pick Dirichlet boundary conditions in which this characteristic polynomial is specified.

On the other hand, A_2, A_3, and A_4 obey modified Neumann boundary conditions

$$\begin{aligned} F_{12}| - F_{34}| &= 0 \\ F_{13}| - F_{42}| &= 0 \\ F_{14}| - F_{23}| &= 0, \end{aligned} \tag{12.25}$$

which follow from imposing (12.24) with $\Gamma_{1234}\epsilon = \epsilon$.

We also want to require the supersymmetries which satisfy $\Gamma_{2367}\epsilon = \epsilon$ to be unbroken. This puts constraints on the boundary values of the fields which satisfy Dirichlet conditions. Requiring the vanishing of the supersymmetry variations of the fermions on the boundary, we find

$$\begin{aligned} F_{26}| - F_{73}| &= 0, \\ F_{27}| - F_{36}| &= 0. \end{aligned} \tag{12.26}$$

These equations are equivalent to $D_{\overline{z}}\phi_z = 0$, which is the "complex" Hitchin equation.

These boundary conditions mean that in the Higgs bundle E determined by a pair $(A_{\overline{z}}, \phi_z)$, the field $A_{\overline{z}}$ which determines the holomorphic structure of E is allowed to fluctuate, while keeping fixed the characteristic polynomial

of ϕ_z. Presumably, the corresponding brane is supported on a fiber of the Hitchin map $\pi : \mathcal{M}_H \to \boldsymbol{B}$. This gives us an archetypical (B, A, A)-brane.

The condition $[\phi_2, \phi_3] = 0$ that arose in this classical analysis is not one of the Hitchin equations defining $\mathcal{M}_H$. Instead, we should have $[\phi_2, \phi_3] - F_{23} = 0$. Presumably, the extra term arises as a sort of quantum correction in the renormalization group flow to the infrared.

In the limit that the gauge theory reduces to the two-dimensional sigma-model, S-duality maps a (B, A, A)-brane wrapping the fiber of the Hitchin fibration to a (B, B, B) zerobrane. In the abelian case, one can verify directly that the boundary conditions that we have described for these branes are S-dual.

We can construct another (B, A, A)-brane that seemingly is likewise supported on a fiber of the Hitchin fibration by replacing T' with $T'' = T'\Gamma_{2367} = -\Gamma_{1467}$. Imposing $T''\epsilon = \epsilon$ and $T''\lambda| = \lambda|$, we get

$$[\Gamma^{IJ} F_{IJ}, T'']\epsilon = 0, \tag{12.27}$$

which implies again Dirichlet boundary conditions for $A_6 = \phi_2$ and $A_7 = \phi_3$ and Neumann boundary conditions for A_2 and A_3.

Of course, for a given fiber $\boldsymbol{F}$ of the Hitchin fibration $\pi : \mathcal{M}_H \to \boldsymbol{B}$, there should be many branes, associated with the possible flat line bundles on $\boldsymbol{F}$. We have given two constructions, but because of time-reversal symmetry, each leads to a trivial Chan-Paton line bundle (or at most one that is of order two). It hopefully is possible to modify the gauge theory construction to make this bundle vary.

12.4. Branes Of Type (A, B, A)

Our next goal is to construct from the gauge theory some branes of type (A, B, A). The most interesting of these is the canonical coisotropic brane. As a prelude, we introduce the abbreviation $\Gamma_{\pm I \pm J} = (\pm\Gamma_I \pm \Gamma_J)/\sqrt{2}$. Similarly, we take $F_{\pm I \pm J, K} = (\pm F_{IK} \pm F_{JK})/\sqrt{2}$, etc.

Let us define

$$T' = -\Gamma_{15}\Gamma_{2-7}\Gamma_{3+6}. \tag{12.28}$$

As we will see shortly, this choice leads to branes of type (A, B, A). One point of this definition is that T' is invariant under simultaneous rotation of the $x^2 - x^3$ and $x^6 - x^7$ planes. Thus in the twisted theory, T' does not depend on the choice of the coordinates x^2, x^3, provided they are orthogonal.

For the same reason, if we consider the twisted theory on $M = \mathbb{R}^2_+ \times C$, T' depends only on the metric on C, not on the specific coordinate choice. The branes we will construct using T' depend only on the conformal class of the metric, i.e. the complex structure on C.

We can again construct two kinds of branes, with boundary conditions $T'\lambda| = \lambda|$ or $T'\lambda| = -\lambda|$. In the former case, supersymmetry gives in the usual way $[\Gamma^{IJ}F_{IJ}, T']\epsilon = 0$. This tells us that $F_{IJ} = 0$ if one of I, J is in the space generated by the $1, 5, 2-7$, and $3+6$ directions and the other is orthogonal to it. In particular, $F_{0I}| = 0$ for $I = 5, 2-7, 3+6$, and we impose Dirichlet boundary conditions for such values of I. But $F_{1I} = 0$ for $I = 2+7, 3-6, 4, 8, 9$, and we impose Neumann boundary conditions on A_I for such I.

In such a brane, $A_2 - \phi_3$ and $A_3 + \phi_2$ have specified boundary values, while $A_2 + \phi_3$ and $A_3 - \phi_2$ are free to fluctuate. To understand this result, it helps to recall that in the complex structure J, the holomorphic coordinates are $2(A_z + i\phi_z) = (A_2 - iA_3) + i(\phi_2 - i\phi_3) = (A_2 + \phi_3) - i(A_3 - \phi_2)$ and $2(A_{\overline{z}} + i\phi_{\overline{z}}) = (A_2 + iA_3) + i(\phi_2 + i\phi_3) = (A_2 - \phi_3) + i(A_3 + \phi_2)$. So our result means that for this type of brane, $\mathcal{A}_{\overline{z}} = A_{\overline{z}} + i\phi_{\overline{z}}$ is fixed and $\mathcal{A}_z = A_z + i\phi_z$ is unconstrained. This defines a submanifold W' of $\mathcal{M}_H$ which is clearly a complex submanifold in the complex structure J. In addition, W' is Lagrangian from the point of view of the holomorphic symplectic form $\Omega_J = (-i/4\pi)\int_C \mathrm{Tr}\,\delta\mathcal{A}\wedge\delta\mathcal{A}$. Indeed, W' is middle-dimensional, and in addition Ω_J vanishes on W', since W' is characterized by $\delta\mathcal{A}_{\overline{z}} = 0$. So the brane supported on W', with a trivial Chan-Paton bundle, is an (A, B, A)-brane, as expected. W' can be characterized as the submanifold of $\mathcal{M}_H$ consisting of flat $G_{\mathbb{C}}$ bundles with a specified holomorphic structure.

For the other type of brane, with $T'\lambda| = -\lambda|$, supersymmetry requires $\{\Gamma^{IJ}F_{IJ}, T'\}\epsilon = 0$. This leads to two conditions. The first is that $F_{IJ} = 0$ if I, J are in the subspace perpendicular to directions $1, 5, 2-7$, and $3+6$. In particular, $F_{0I} = 0$ if direction I is in this subspace, and we impose Dirichlet boundary conditions for such I. Thus, in particular, $A_2+\phi_3$ and $A_3-\phi_2$ obey Dirichlet boundary conditions. The second consequence of $\{\Gamma^{IJ}F_{IJ}, T'\}\epsilon = 0$ is that F must be anti-selfdual when restricted to the subspace generated by directions $1, 5, 2-7, 3+6$. This leads to modified Neumann boundary conditions on A_5, A_{2-7}, and A_{3+6}:

$$\begin{aligned} F_{15} + F_{2-7,3+6} &= 0 \\ F_{1,2-7} + F_{3+6,5} &= 0 \\ F_{1,3+6} + F_{5,2-7} &= 0. \end{aligned} \tag{12.29}$$

For this type of brane, A_{2+7} and A_{3-6} are fixed on the boundary while A_{2-7} and A_{3+6} are free to fluctuate. So, from a low energy point of view, the roles of $\mathcal{A}_z$ and $\mathcal{A}_{\overline{z}}$ are exchanged, relative to the branes with $T'\lambda| = \lambda|$. Thus we get a brane supported on a submanifold W'' of $\mathcal{M}_H$ that parametrizes flat $G_{\mathbb{C}}$ bundles with a fixed antiholomorphic structure. This again is a brane of type (A, B, A).

There is another and more trivial way to exchange the roles of $\mathcal{A}_z$ and $\mathcal{A}_{\overline{z}}$: we could reverse the sign of ϕ_2 and ϕ_3 or in other words replace T' by $T'' = -\Gamma_{15}\Gamma_{2+7}\Gamma_{3-6}$. So we really have two families of branes supported on manifolds parameterizing bundles with fixed holomorphic or antiholomorphic structure. The two families differ by the conditions placed on modes that are massive away from singularities of $\mathcal{M}_H$; they may differ at singularities of $\mathcal{M}_H$.

***S*-Duality And The Canonical Coisotropic Brane** By applying S-duality to the examples that we just constructed, we can obtain new branes, which will also be branes of type (A, B, A). We know how to apply S-duality to the condition for unbroken supersymmetry: $T'\epsilon = \epsilon$ is replaced by $\widehat{T}\epsilon = \epsilon$ where

$$\widehat{T} = \frac{1+\widehat{\Gamma}}{\sqrt{2}} T' \frac{1-\widehat{\Gamma}}{\sqrt{2}} = -\frac{\Gamma_{15}}{2}\Big((\Gamma_{26}+\Gamma_{37}) - \widehat{\Gamma}(\Gamma_{23}+\Gamma_{67})\Big). \tag{12.30}$$

To get new (A, B, A)-branes in this way, we must construct boundary conditions that preserve the supersymmetries with $\widehat{T}\epsilon = \Gamma_{2367}\epsilon = \epsilon$.

We look for a supersymmetric brane with $\widehat{T}\lambda| = -\lambda|$. After a somewhat lengthy calculation, we deduce from the condition $\{\Gamma^{IJ}F_{IJ}, \widehat{T}\}\epsilon = 0$ for supersymmetry that the boundary conditions on the bosons must be

$$\begin{aligned} F_{15} + F_{26} + F_{37} &= 0 \\ F_{12} + F_{06} - F_{56} &= 0 \\ F_{13} + F_{07} - F_{57} &= 0 \\ F_{16} + F_{52} - F_{02} &= 0 \\ F_{17} + F_{53} - F_{03} &= 0, \end{aligned} \tag{12.31}$$

along with $F_{0J} = 0$ for $J = 4, 8, 9$.

We set the massive fields A_J, $J = 4, 8, 9$ to zero on the boundary. Otherwise, all fields obey modified Neumann boundary conditions and are free to fluctuate on the boundary, so the (A, B, A)-brane obtained this way is a space-filling brane whose target space is all of $\mathcal{M}_H$.

We have seen mixed Neumann-Dirichlet boundary conditions before, but there is a key difference here. In previous examples, the boundary conditions become purely Dirichlet or purely Neumann if one sets to zero the fields A_4, A_5, A_8, and A_9 that play no role in the low energy theory away from singularities, so they are irrelevant in the effective two-dimensional field theory of the brane. In the present example, this is not the case. After discarding the "massive" fields and choosing the gauge $A_1 = 0$, the boundary conditions become

$$\begin{aligned} D_0\phi_2 + \partial_1 A_2 &= 0, \\ D_0\phi_3 + \partial_1 A_3 &= 0, \\ F_{02} - \partial_1\phi_2 &= 0, \\ F_{03} - \partial_1\phi_3 &= 0. \end{aligned} \tag{12.32}$$

Mixed Dirichlet-Neumann boundary conditions on the sigma-model fields arise when the gauge field on the brane is not flat; specifically, the boundary condition is

$$f_{MN}\partial_0 X^N + g_{MN}\partial_1 X^N = 0, \tag{12.33}$$

where f is the curvature of the target-space gauge field, and g is the target-space metric. Comparing with (12.32) we find

$$f = \frac{1}{2\pi}\int_C |d^2z| \mathrm{Tr}\ (\delta\phi_z \wedge \delta A_{\overline{z}} + \delta\phi_{\overline{z}} \wedge \delta A_z) = \omega_J. \tag{12.34}$$

This is the target-space gauge field for the canonical coisotropic (A, B, A)-brane in the sigma model with target $\mathcal{M}_H(G, C)$. So that is the brane that we have found.

Boundary Observables for the Canonical Coisotropic Brane Now we want to make a simple observation about boundary observables for the canonical coisotropic brane. We consider local operators $\mathcal{O}$ inserted at a point P of the boundary of M. The boundary, of course, looks like $\mathbb{R} \times C$. We will consider the cohomology of the topological supercharge Q of the A-model in complex structure K acting on local operators inserted at a point $s \times p \in \mathbb{R} \times C$. The metric of C is irrelevant mod $\{Q, \ldots\}$, so we can assume C to be flat near p.

We can classify local operators by their dimension and also by the "spin" with which they transform under rotations of C around the point p. (The considerations will be local, so it will not matter that the global structure of

C typically breaks the rotation symmetry.) The natural dimensions in the twisted theory are 1 for $A, \phi, \psi, \widetilde{\psi}$, and 2 for $\chi, \eta, \widetilde{\eta}$.

We consider Q-invariant operators of dimension n and spin n. To have these quantum numbers, a gauge-invariant operator must be constructed only from ϕ_z, ψ_z, $\widetilde{\psi}_z$, and the covariant derivative D_z. But for an operator to be Q-invariant, D_z cannot appear, since A_z is not Q-invariant even on the boundary. The boundary condition $\widehat{T}\psi| = -\psi|$ implies that

$$\begin{aligned} \psi_2| &= -i\widetilde{\psi}_3| \\ \psi_3| &= i\widetilde{\psi}_2|. \end{aligned} \tag{12.35}$$

With $\psi_z = (\psi_2 - i\psi_3)/2$, this leads to $\psi_z| = \widetilde{\psi}_z|$. So Q-invariant boundary observables of dimension n and spin n are functions just of ϕ_z and ψ_z, without any derivatives.

The fields ϕ_z and ψ_z have ghost number $\mathcal{K} = 0$ and $\mathcal{K} = 1$, respectively. So an operator of this type that also has $\mathcal{K} = 0$ is a gauge-invariant function of ϕ_z only, such as $\mathcal{O} = \mathrm{Tr}\,\phi_z^n$ for some n. A typical operator of $\mathcal{K} = 1$ is $\mathcal{O}' = \mathrm{Tr}\,\phi_z^{n-1}\psi_z$.

Classically, the boundary conditions ensure that on the boundary $[Q, \phi_z\,] = \{Q, \psi_z\,\} = 0$, so all operators $\mathcal{O}$ and $\mathcal{O}'$ of the sort just described are nonzero elements of the cohomology of Q. What happens quantum mechanically? Could there be a quantum correction to the action of Q, such that the exact quantum formula would be $[Q, \mathcal{O}] = \epsilon\mathcal{O}'$ for some ϵ? If so, at the quantum level, $\mathcal{O}$ and $\mathcal{O}'$ would pair up and disappear from the cohomology.

That this does not occur can be argued using time-reversal symmetry. We define an orientation-reversing symmetry $\mathcal{T}$ that reverses the sign of the time coordinate x^0 and acts trivially on other coordinates of M. Allowing for the minus sign with which $\mathcal{T}$ acts on ϕ (see eqn. (3.57)), $\mathcal{T}$ is a symmetry of the boundary conditions that define the canonical coisotropic brane. $\mathcal{T}$ commutes with the topological supercharge of the A-model in complex structure K, since $t = 1$ is a fixed point of $\mathcal{T}$ (see eqn. (3.56)). Also, ϕ_z is odd under $\mathcal{T}$, in view of (3.57), but (3.58) plus the boundary condition $\psi_z| = \widetilde{\psi}_z|$ implies that $\psi_z|$ is even under $\mathcal{T}$. So the operators $\mathcal{O}$ and $\mathcal{O}'$ cannot pair up under the action of Q, and must survive in the cohomology. More generally, consider a local boundary operator of dimension and spin n and ghost number k that is Q-invariant at the classical level. Such an operator is determined by a gauge-invariant function $F(\phi_z, \psi_z)$ that is of degree $n-k$ in ϕ_z and degree k in ψ_z. It transforms as $(-1)^{n-k}$ under $\mathcal{T}$. For Q to act non-trivially on such an operator at the quantum level, it would have to leave n unchanged and

increase k by 1. This would reverse the eigenvalue of $\mathcal{T}$, and so is impossible. Thus, the cohomology of Q in the space of local operators of dimension and spin n coincides with the classical result, and is given simply by the space of gauge-invariant functions $F(\phi_z, \psi_z)$.

References

[1] R. Langlands, "Problems In The Theory Of Automorphic Forms," in Lect. Notes in Math. **170** (Springer-Verlag, 1970), pp. 18-61; "Where Stands Functoriality Today?" in *Representation Theory And Automorphic Forms*, Proc. Symp. Pure Math. **61** (American Mathematical Society, 1997), pp. 457–471.

[2] V. Drinfeld, "Langlands Conjecture For $GL(2)$ Over Function Field," Proc. Int. Congress. Math. (Helsinki, 1978), pp. 565-574; "Two-Dimensional l-Adic Representations Of The Fundamental Group Of A Curve Over A Finite Field And Automorphic Forms On $GL(2)$," Amer. J. Math. **105** (1983), 85–114.

[3] L. Lafforgue, "Chtoucas de Drinfeld et Correspondance de Langlands," Invent. Math. **147** (2002), 1–241.

[4] G. Laumon, "Correspondance de Langlands Geometrique Pour Les Corps de Fonctions," Duke. Math. J. **54** (1987), 309–359.

[5] A. Beilinson and V. Drinfeld, "Quantization Of Hitchin's Integrable System And Hecke Eigensheaves," preprint (ca. 1995), http://www.math.uchicago.edu/arinkin/langlands/.

[6] E. Frenkel, D. Gaitsgory and K. Vilonen, "On The Geometric Langlands Conjecture," Journal of the AMS **15** (2001), 367–417.

[7] D. Gaitsgory, "On A Vanishing Conjecture Appearing In The Geometric Langlands Correspondence," Ann. of Math. (2) **160** (2004), 617–182, math.AG/0204081.

[8] A. Braverman and R. Bezrukavnikov, "Geometric Langlands Correspondence For $\mathcal{D}$-Modules in Prime Characteristic: The $GL(n)$ Case," math.AG/0602255.

[9] E. Frenkel, "Lectures On The Langlands Program And Conformal Field Theory," arXiv:hep-th/0512172.

[10] P. Goddard, J. Nuyts, and D. I. Olive, "Gauge Theories And Magnetic Charge," Nucl. Phys. **B125** (1977), 1–28.

[11] F. Englert and P. Windey, "Quantization Condition For 't Hooft Monopoles In Compact Simple Lie Groups," Phys. Rev. **D14** (1976), 2728–2731.

[12] C. Montonen and D. I. Olive, "Magnetic Monopoles As Gauge Particles?" Phys. Lett. **B72** (1977), 117–120.

[13] D. I. Olive and E. Witten, "Supersymmetry Algebras That Include Topological Charges," Phys. Lett. **B78** (1978), 97–101.

[14] H. Osborn, "Topological Charges For $\mathcal{N} = 4$ Supersymmetric Gauge Theories And Monopoles Of Spin 1," Phys. Lett. **B83** (1979), 321–326.

[15] J. Cardy and E. Rabinovici, "Phase Structure Of $\mathbb{Z}_p$ Models In The Presence Of A Theta Parameter," Nucl. Phys. **B205** (1982), 1–16; J. Cardy, "Duality And The Theta Parameter In Abelian Lattice Models," Nucl. Phys. **B205** (1982), 17–26.

[16] A. Shapere and F. Wilczek, "Selfdual Models With Theta Terms," Nucl. Phys. **B320** (1989), 669–695.

[17] J. H. Schwarz and A. Sen, "Duality Symmetries Of $4 - D$ Heterotic Strings," Phys. Lett. **312B** (1993), 105–114, "Duality Symmetric Actions," Nucl. Phys. **B411** (1994), 35–63, arXiv:hep-th/9304154.

[18] M. Bershadsky, A. Johansen, V. Sadov, and C. Vafa, "Topological Reduction Of $4-D$ SYM To $2-D$ Sigma Models," Nucl. Phys. **B448** (1995), 166–186, arXiv:hep-th/9501096.

[19] J. A. Harvey, G. W. Moore, and A. Strominger, "Reducing S Duality To T Duality," Phys. Rev. **D52** (1995), 7161–7167.

[20] N. Hitchin, "The Self-Duality Equations On A Riemann Surface," Proc. London Math. Soc. (3) **55** (1987), 59–126

[21] T. Hausel and M. Thaddeus, "Mirror Symmetry, Langlands Duality, And The Hitchin System," Invent. Math. **153** (2003), 197–229, math.AG/0205236.

[22] R. Donagi and D. Gaitsgory, "The Gerbe Of Higgs Bundles," Transform. Groups **7** (2002), 109–153, math.AG/0005132.

[23] R. Donagi and T. Pantev, "Torus Fibrations, Gerbes, and Duality," math.AG/0306213.

[24] D. Arinkin, "Moduli Of Connections With A Small Parameter On A Curve," math.AG/0409373.

[25] E. Witten, "Topological Quantum Field Theory," Commun. Math. Phys. **117** (1988), 353–386.

[26] E. Witten, *Gauge Theory And The Geometric Langlands Program*, to appear.

[27] E. Frenkel, "Affine Algebras, Langlands Duality, And Bethe Ansatz," in the Proceedings of the XIth International Congress of Mathematical Physics, Paris, 1994, ed. D. Iagolnitzer (International Press, 1995) 606–642, q-alg/9506003.

[28] A. Beilinson and V. Drinfeld, *Chiral Algebras*, American Mathematical Society Colloquium Publications **51** (American Mathematical Society, 2004).

[29] E. Frenkel and D. Ben-Zvi, *Vertex Algebras And Algebraic Curves*, Mathematical Surveys And Monographs **88**, second edition (American Mathematical Society, 2004).

[30] E. Frenkel and D. Gaitsgory, "Local Geometric Langlands Correspondence And Affine Kac-Moody Algebras," math.RT/0508382.

[31] E. Witten, "Free Fermions On An Algebraic Curve," in *The Mathematical Heritage Of Hermann Weyl*, ed. R. Wells (American Mathematical Society, Durham, 1988), pp. 329–344, "Quantum Field Theory, Grassmannians, And Algebraic Curves," Commun. Math. Phys. **113** (1988), 529–600.

[32] A. Neveu and J. Scherk, "Parameter-Free Regularization Of One-Loop Unitary Dual Diagram," Phys. Rev. **D1** (1970), 2355–2359.

[33] D. J. Gross, A. Neveu, J. Scherk, and J. H. Schwarz, "Renormalization And Unitarity In The Dual-Resonance Model," Phys. Rev. **D2** (1970), 697–710.

[34] P. Goddard and D. Olive, "Algebras, Lattices, And Strings," in *Vertex Operators In Mathematics And Physics* (1985), ed. J. Lepowsky et al.

[35] J. Harvey and G. Moore, "Algebras, BPS States, And Strings," Nucl. Phys. **B463** (1996), 315–368, arXiv:hep-th/9510182, "Exact Gravitational Threshold Correction in the FHSV Model," Phys. Rev. **D57** (1998), 2329–2336, arXiv:hep-th/9611176.

[36] R. Dijkgraaf, E. Verlinde, and H. Verlinde, "Counting Dyons In $\mathcal{N} = 4$ String Theory," Nucl. Phys. **B484** (1997), 543–561, arXiv:hep-th/9607026.

[37] M. B. Green and S. Sethi, "Supersymmetry Constraints On Type IIB Supergravity," Phys. Rev. **D59** (1999) 046006, hep-th/9808061.

[38] H. Nicolai, B. Pioline, J. Plefka and A. Waldron, "R^4 Couplings, the Fundamental Membrane and Exceptional Theta Correspondences," JHEP **03** (2001), 036, hep-th/0102123.

[39] M. Gunaydin, A. Neitzke, B. Pioline, and A. Waldron, "BPS Black Holes, Quantum Attractor Flows, And Automorphic Forms," arXiv: hep-th/0512296.

[40] R. Schimmrigk, "The Langlands Program and String Modular K3 Surfaces," arXiv:hep-th/0603234.

[41] A. Kapustin, "A-Branes And Noncommutative Geometry," arXiv:hep-th/0502212.

[42] L. Brink, J. H. Schwarz, and J. Scherk, "Supersymmetric Yang-Mills Theories," Nucl. Phys. **B121** (1977), 77–92.

[43] W. Nahm, "Supersymmetries And Their Representations," Nucl. Phys. **B135** (1978), 149–166.

[44] S. Weinberg, *Quantum Theory of Fields, II* (Cambridge University Press, 1996).

[45] C. Vafa and E. Witten, "A Strong Coupling Test Of S Duality," Nucl. Phys. **B431** (1994), 3–77, arXiv:hep-th/9408074.

[46] N. Dorey, C. Fraser, T. J. Hollowood and M. A. C. Kneipp, "S-Duality in $\mathcal{N} = 4$ Supersymmetric Gauge Theories With Arbitrary Gauge Group," Phys. Lett. B **383** (1996), 422–428, arXiv:hep-th/9605069.

[47] P. Argyres, A. Kapustin, N. Seiberg, "On S-Duality For Non-Simply-Laced Gauge Groups," arXiv:hep-th/0603048.

[48] C. Vafa, "Geometric Origin of Montonen-Olive Duality," Adv. Theor. Math. Phys. **1**, 158–166 (1997), arXiv:hep-th/9707131.

[49] E. Witten, "Baryons and Branes in Anti de Sitter Space," JHEP **9807**, 006 (1998), arXiv:hep-th/9805112.

[50] E. Witten, "Dyons Of Charge $e\theta/2\pi$," Phys. Lett. **B86** (1979), 283–287.

[51] S. Weinberg, *Gravitation And Cosmology* (WIley, 1972).

[52] J. P. Yamron, "Topological Actions From Twisted Supersymmetric Theories," Phys. Lett. **B213** (1988), 325–330.

[53] N. Marcus, "The Other Topological Twisting Of $\mathcal{N} = 4$ Yang-Mills," Nucl. Phys. **B452** (1995), 331–345, arXiv:hep-th/9506002.

[54] M. Blau and G. Thompson, "Aspects Of $N_T \geq 2$ Topological Gauge Theories And D-Branes," Nucl. Phys. **B492** (1997), 545–590, arXiv:hep-th/9612143.

[55] J. M. F. Labastida and C. Lozano, "Mathai-Quillen Formulation Of Twisted $\mathcal{N} = 4$ Supersymmetric Gauge Theories In Four Dimensions," Nucl. Phys. **B502** (1997), 741–790.

[56] C. Lozano, "Duality In Topological Quantum Field Theories," arXiv: hep-th/9907123.

[57] J. M. F. Labastida and C. Lozano, "Duality In Twisted $\mathcal{N} = 4$ Supersymmetric Gauge Theories In Four Dimensions," Nucl. Phys. **B537** (1999), 203–242.

[58] K. Corlette, "Flat G-Bundles With Canonical Metrics," J. Diff. Geom. **28** (1988), 361–382.

[59] E. Witten, "Introduction To Cohomological Field Theories," Int. J. Mod. Phys. **A6** (1991), 2775–2792.

[60] D. Quillen, "Determinants Of Cauchy-Riemann Operators Over A Riemann Surface," Funct. Anal. Appl. **19** (1985), 31–34.

[61] N. J. Hitchin, A. Karlhede, U. Lindstrom, and M. Rocek, "Hyperkahler Metrics and Supersymmetry," Commun. Math. Phys. **108** (1987) 535–589.

[62] N. Hitchin, "Stable Bundles And Integrable Systems," Duke Math. J. **54** (1987), 91–114.

[63] A. Gorsky and N. Nekrasov, "Elliptic Calogero-Moser System From Two-Dimensional Current Algebra," arXiv:hep-th/9401021; "Relativistic Calogero-Moser Model As Gauged WZW Theory," Nucl. Phys. **B436** (1995), 582–608.

[64] R. Donagi and E. Markman, "Spectral Covers, Algebraically Completely Integrable Hamiltonian Systems, And Moduli Of Bundles," in *Integrable Systems And Quantum Groups*, Lecture Notes in Math. **1620** (Springer-Verlag, 1996) pp. 1–119, alg-geom/9507017.

[65] L. Alvarez-Gaumé and D. Z. Freedman, "Geometrical Structure And Ultraviolet Finiteness In The Supersymmetric Sigma Model," Commun. Math. Phys. **80** (181), 443.

[66] S. J. Gates, C. M. Hull, and M. Rocek, "Twisted Multiplets And New Supersymmetric Nonlinear Sigma Models," Nucl. Phys. **B248** (1984), 157–186.

[67] M. Gualtieri, "Generalized Complex Geometry," D. Phil. thesis, Oxford University, math.DG/0401221.

[68] V. Pestun, "Topological Strings In Generalized Complex Space," arXiv:hep-th/0603145.

[69] N. Hitchin, "Generalized Calabi-Yau Manifolds," Q. J. Math. **54** (2003), 281–308, math.DG/0209099.

[70] E. Witten, "Mirror Manifolds And Topological Field Theory," in *Mirror Symmetry I*, S. T. Yau, ed. (American Mathematical Society, 1998), pp. 121–160, arXiv:hep-th/9112056.

[71] A. Kapustin, "Topological Strings On Noncommutative Manifolds," Int. J. Geom. Meth. Mod. Phys. **1** (2004), 49–81, arXiv:hep-th/0310057.

[72] A. Kapustin and Y. Li, " Topological Sigma Models With H Flux And Twisted Generalized Complex Manifolds," arXiv:hep-th/0407249.

[73] U. Lindstrom, R. Minasian, A. Tomasiello and M. Zabzine, "Generalized Complex Manifolds and Supersymmetry," Commun. Math. Phys. **257** (2005), 235, hep-th/0405085.

[74] A. Bredthauer, U. Lindstrom, J. Persson and M. Zabzine, "Generalized Kahler Geometry from Supersymmetric Sigma Models," hep-th/0603130.

[75] M. Grana, R. Minasian, M. Petrini, and A. Tomasiello, "Supersymmetric Backgrounds From Generalized Calabi-Yau Manifolds," JHEP 0408:046 (2004), arXiv:hep-th/0406137; "Type II Strings And Generalized Calabi-Yau Manifolds," Comptes Rendus Physique **5** (2004), 979–986, arXiv:hep-th/0409176.

[76] M. Grana, R. Minasian, M. Petrini, and A. Tomasiello, "Generalized Structures Of $\mathcal{N} = 1$ Vacua," JHEP 0511:020 (2005), arXiv:hep-th/0505212.

[77] M. Grana, J. Louis, and D. Waldram, JHEP 0601:008 (2006), arXiv:hep-th/0505264.

[78] W. Lerche, C. Vafa, and N. Warner, "Chiral Rings In $\mathcal{N} = 2$ Superconformal Theories," Nucl. Phys. **B324** (1989), 427–474

[79] A. Strominger, S.-T. Yau, and E. Zaslow, "Mirror Symmetry Is T Duality," Mucl. Phys. **B479** (1996), 243–259.

[80] A. Kapustin and D. Orlov, "Remarks On A-Branes, Mirror Symmetry, and the Fukaya Category," J. Geom. Phys. **48** (2003), 84–99, arXiv:hep-th/019098

[81] N. Hitchin, "Langlands Duality And G_2 Spectral Curves," math.AG/0611524.

[82] R. Donagi and T. Pantev, "Langlands Duality For Hitchin Systems," math.AG/0604617.

[83] G. Faltings, "Stable G-Bundles And Projective Connections," J. Alg. Geom. **2** (1993), 507.

[84] M. Blau and G. Thompson, "Aspects Of $N_T \geq 2$ Topological Gauge Theories And D-Branes," Nucl. Phys. **B492** (1997), 545–590, arXiv:hep-th/9612143.

[85] L. Baulieu, A. Losev, and N. Nekrasov, "Chern-Simons And Twisted Supersymmetry In Various Dimensions," Nucl. Phys. **B522** (1998), 82–104, arXiv:hep-th/9707174.

[86] J. Maldacena, "Wilson Loops In Large N Field Theories," Phys. Rev. Lett. **80** (1998), 4859–4862, arXiv:hep-th/9803002.

[87] S.-J. Rey and J.-T. Yee, "Macroscopic Strings As Heavy Quarks In Large N Gauge Theory And Anti-de Sitter Supergravity," Eur. Phys. J. **C22** (2001), 379–394, hep-th/9803001.

[88] K. Zarembo, "Supersymmetric Wilson Loops," Nucl. Phys. **B643** (2002), 157–171, arXiv:hep-th/0205160.

[89] E. Witten, "Lectures On Quantum Field Theory," Lecture 10, in P. Deligne et. al, eds., *Quantum Fields And Strings: A Course For Mathematicians*, vol. 2 (American Mathematical Society, 1999).

[90] A. Kapustin, "Wilson-'t Hooft Operators In Four-Dimensional Gauge Theories And S-Duality," arXiv:hep-th/0501015.

[91] M. F. Atiyah and R. Bott, "Yang-Mills Equations Over Riemann Surfaces," Phil. Trans. R. Soc. Lond. **A308** (1983), 523–615.

[92] M. Henningson, "Wilson-'t Hooft Operators And The Theta Angle," arXiv:hep-th/0603188.

[93] E. Wong and I. Affleck, "Tunneling In Quantum Wires: A Boundary Conformal Field Theory Approach," Nucl. Phys. **B417** (1994) 403–438.

[94] M. Oshikawa and I. Affleck, "Boundary Conformal Field Theory Approach To The Critical Two-Dimensional Ising Model With A Defect Line," Nucl. Phys. **B495** (1997), 533–582, cond-mat/9612187.

[95] C. Bachas, J. de Boer, R. Dijkgraaf, and H. Ooguri, "Permeable Conformal Walls And Holography," JHEP **6** (2002), 1–32, arXiv:hep-th/0111210.

[96] G. Lusztig, "Singularities, Character Formula, And A q-Analog Of Weight Multiplicities," in *Analyse et Topologie Sur Les Espace Singuliers II-III*, Asterisque vol. 101–2 (1981), 208–229.

[97] V. Ginzburg, "Perverse Sheaves On A Loop Group And Langlands Duality," alg-geom/9511007.

[98] I. Mirkovic and K. Vilonen, "Perverse Sheaves On Affine Grassmannians And Langlands Duality," Math. Res. Lett. **7** (2000), 13–24, math.AG/9911050.

[99] I. Runkel, J. Fuchs, and C. Schweigert, "Categorification And Correlation Functions In Conformal Field Theory," math.CT/0602079.

[100] G. Moore, "K-Theory From A Physical Perspective," hep-th/0304018.

[101] G. 't Hooft, "On The Phase Transition Towards Permanent Quark Confinement," Nucl. Phys. **B138** (1978), 1–25, "A Property Of Electric And Magnetic Flux In Nonabelian Gauge Theories," Nucl. Phys. **B153** (1979), 141–160.

[102] E. Witten, "D-Branes And K-Theory," JHEP 9812:019,1998, arXiv: hep-th/9810188.

[103] M. R. Douglas, B. Fiol, and C. Romelsberger, "Stability And BPS Branes," JHEP **006** (2005), 1–14, arXiv:hep-th/0002037.

[104] M. R. Douglas, "D-Branes, Categories, and $\mathcal{N} = 1$ Supersymmetry," J. Math. Phys. **42** (2001), 2818–2843, arXiv:hep-th/0011017.

[105] E. B. Bogomolny, "Stability Of Classical Solutions," Sov. J. Nucl. Phys. **24** (1976), 449–454.

[106] M. K. Prasad and C. Sommerfield, "An Exact Classical Solution For The 't Hooft Monopole And The Julia-Zee Dyon," Phys. Rev. Lett. **35** (1975), 760–762.

[107] E. J. Weinberg, "Parameter Counting For Multi-Monopole Solutions," Phys. Rev. **D20** (1979), 936–944, "Fundamental Monopoles And Multi-Monopole Solutions For Arbitrary Simple Gauge Groups," Nucl. Phys. **B167** (1980), 500–524.

[108] C. H. Taubes, "Stability In Yang-Mills Theories," Commun. Math. Phys. **91** (1983), 235–263.

[109] M. F. Atiyah and N. Hitchin, *The Geometry And Dynamics Of Magnetic Monopoles* (Princeton University Press, 1988).

[110] P. Kronheimer, MSc. thesis (Oxford University, 1986), unpublished.

[111] E. J. Weinberg, "A Continuous Family Of Magnetic Monopole Solutions," Phys. Lett. **B119** (1982), 151–154.

[112] K.-M. Lee, E. J. Weinberg, and P. Yi, "Massive And Massless Monopoles With Nonabelian Magnetic Charges, Phys. Rev. **D54** (1996), 6351–6371; E. J. Weinberg, and P. Yi, "Explicit Multimonopole Solutions In $SU(N)$ Gauge Theory," Phys. Rev. **D58** (1998), 046001, arXiv:hep-th/9803164.

[113] C. J. Houghton and E. J. Weinberg, "Multicloud Solutions With Massless And Massive Monopoles," Phys. Rev. **D66** (2002), 125002, arXiv:hep-th/0207141.

[114] F. A. Bais and B. J. Schroers, "Quantization Of Monopoles With Nonabelian Magnetic Charge," Nucl. Phys. **B512** (1998), 250–294, arXiv:hep-th/9708004.

[115] A. Hanany and E. Witten, "Type IIB Superstrings, BPS Monopoles, And Three-Dimensional Gauge Dynamics," Nucl. Phys. **B492** (1997), 152–190, arXiv:hep-th/9611230.

[116] S. A. Cherkis and A. Kapustin, "Singular Monopoles And Supersymmetric Gauge Theories In Three-Dimensions," Nucl. Phys. **B525** (1998), 215–234, hep-th/9711145.

[117] S. A. Cherkis and A. Kapustin, "$D(k)$ Gravitational Instantons And Nahm Equations," Adv. Theor. Math. Phys. **2** (1999), 1287–1306, hep-th/9803112.

[118] S. A. Cherkis and A. Kapustin, "Periodic Monopoles With Singularities And $N = 2$ Super QCD," Commun. Math. Phys. **234** (2003), 1–35, arXiv:hep-th/0011081.

[119] R. Donagi, unpublished.

[120] A.Levin, M.Olshanetsky, and A.Zotov, "Hitchin Systems – Symplectic Hecke Correspondence and Two-dimensional Version," Commun. Math. Phys. **236** (2003), 93–133, nlin.SI/0110045.

[121] P. Kronheimer, "The Construction Of ALE Spaces As Hyper-Kahler Quotients," J. Diff. Geom. **29** (1989), 665–683.

[122] A. Pressley and G. Segal, *Loop Groups,* (Oxford University Press, 1988).

[123] M. F. Atiyah and I. M. Singer, "The Index Of Elliptic Operators (I)," Annals of Mathematics **87** (1968), 484–530.

[124] M. Pauly, "Monopole Moduli Spaces For Compact 3-Manifolds," Math. Ann. **311** (1998), 125–146.

[125] M. F. Atiyah, V. Drinfeld, N. J. Hitchin, and Yu. I. Manin, "Construction Of Instantons," Phys. Lett. **65A** (1978), 185–187.

[126] S. Donaldson and P. B. Kronheimer, *The Geometry Of Four-Manifolds* (Oxford University Press, 1990).

[127] P. Deligne and J. Milne, "Tannakian Categories," Lect. Notes in Math **900** (1982), 101–228.

[128] M. Goresky and R. MacPherson, "Intersection Homology Theory," Topology **19** (1980), 135–162.

[129] F. Kirwan, *An Introduction To Intersection Homology Theory* (Longman, Harlow, 1988).

[130] T. Eguchi and A. J. Hansen, "Selfdual Solutions To Euclidean Gravity," Ann. Phys. **120** (1979), 82–106.

[131] G. Segal and A. Selby, "The Cohomology Of The Space Of Magnetic Monopoles," Commun. Math. Phys. **177** (1996), 775–787; G. Segal, "Topology Of The Space Of $SU(2)$-Monopoles In $\mathbb{R}^3$," in *Geometry And Physics*, ed. J. E. Anderson et al. (Marcel Dekker, New York, 1997), 141–147.

[132] A. Kapustin and Y. Li, "Open String BRST Cohomology for Generalized Complex Branes," arXiv:hep-th/0501071.

[133] E. Witten, "Two-Dimensional Models With (0,2) Supersymmetry: Perturbative Aspects," arXiv:hep-th/0504078.

[134] A. Kapustin, "Chiral de Rham Complex and the Half-Twisted Sigma-Model," arXiv:hep-th/0504074.

[135] N. Nekrasov, "Lectures On Curved Beta-Gamma Systems, Pure Spinors, And Anomalies," arXiv:hep-th/0511008.

[136] F. Malikov, V. Schechtman, and A. Vaintrob, "Chiral de Rham Complex," Commun. Math. Phys. **204** (1999), 439–473, math.AG/9803041.

[137] N. Seiberg and E. Witten, "String Theory And Noncommutative Geometry," JHEP 9909:032, 1999, arXiv:hep-th/9908142.

[138] M. Kontsevich, "Deformation Quantization Of Algebraic Varieties," Lett. Math. Phys. **56** (2001), 271–294, math.AG/0106006.

[139] C. K. Zachos, D. B. Fairlie, and T. L. Curtwright, *Quantum Mechanics In Phase Space: An Overview With Selected Papers* (World-Scientific, 2005).

[140] A. Beilinson and J. Bernstein, "A Proof of Jantzen's Conjectures," I. M. Gelfand Seminar, 1–50, Adv. Sov. Math. 16, Part 1, AMS, 1993.

[141] N. J. Hitchin, "Flat Connections And Geometric Quantization," Commun. Math. Phys. **131** (1990), 347–380.

[142] C. I. Lazaroiu, "On the Structure of Open-Closed Topological Field Theory in Two Dimensions," Nucl. Phys. B **603**, 497–530 (2001), arXiv:hep-th/0010269.

[143] G. Moore and G. Segal, "D-Branes And K-Theory In 2D Topological Field Theory," hep-th/0609042; see also lectures by G. Moore, at http://online.itp.ucsb.edu/online/mp01.

[144] M. Aldi and E. Zaslow, "Coisotropic Branes, Noncommutativity, and the Mirror Correspondence," JHEP **0506** (2005), 019, arXiv:hep-th/0501247.

[145] R. Jackiw and C. Rebbi, "Solitons With Fermion Number 1/2," Phys. Rev. D **13** (1976), 3398–3409.

[146] A. Polishchuk and M. Rothstein, "Fourier Transform For $\mathcal{D}$-Algebras," Duke Math. J. **109** (2001), 123–146.

E-mail addresses

Anton Kapustin: kapustin@theory.caltech.edu
Edward Witten: witten@ias.edu