

Gauge Theory, Ramification, And The Geometric Langlands Program

Sergei Gukov and Edward Witten

In the gauge theory approach to the geometric Langlands program, ramification can be described in terms of “surface operators,” which are supported on two-dimensional surfaces somewhat as Wilson or ’t Hooft operators are supported on curves. We describe the relevant surface operators in $\mathcal{N} = 4$ super Yang-Mills theory, and the parameters they depend on, and analyze how S -duality acts on these parameters. Then, after compactifying on a Riemann surface, we show that the hypothesis of S -duality for surface operators leads to a natural extension of the geometric Langlands program for the case of tame ramification. The construction involves an action of the affine Weyl group on the cohomology of the moduli space of Higgs bundles with ramification, and an action of the affine braid group on A -branes or B -branes on this space.

CONTENTS

1. Introduction	36
2. Monodromy And Surface Operators	39
2.1. Definition Of Surface Operators	39
2.2. Geometric Interpretation Of Parameters	42
2.3. Theta Angles	45
2.4. Electric-Magnetic Duality	46
2.5. Shifting The Theta Angle	52
2.6. Levi Subgroups And More General Surface Operators	59
3. More On The Classical Geometry	63
3.1. Hyper-Kahler Quotient	63
3.2. Complex Viewpoint	67
3.3. The Non-Regular Case	71
3.4. Parabolic Bundles	78
3.5. Topology Of $\mathcal{M}_{\mathbb{L}}(\alpha; p)$	81
3.6. Topology Of \mathcal{M}_H	91

3.7.	Action Of The Affine Weyl Group	97
3.8.	Nahm's Equations And Local Singularity Of \mathcal{M}_H	102
3.9.	The Hitchin Fibration	108
4.	Geometric Langlands With Tame Ramification	110
4.1.	Review Of Unramified Case	110
4.2.	Sigma Model With Ramification	111
4.3.	Branes	114
4.4.	Twisted \mathcal{D} -Modules	116
4.5.	Line Operators And Monodromies	124
4.6.	Representations And Branes	131
5.	Line Operators And Ramification	135
5.1.	General Framework	136
5.2.	The B -Model	139
5.3.	The A -Model	141
6.	Local Models And Realizations By String Theory	153
6.1.	Overview	153
6.2.	Linear Sigma Model For $G_{\mathbb{C}} = SL(2, \mathbb{C})$	155
6.3.	Instantons And The Local Singularity	161
6.4.	Some String Theory Constructions	164
	Appendix A. Review Of Duality	168
	Appendix B. Index Of Notation	175
	References	177

1. Introduction

The Langlands program of number theory [1] relates representations of the Galois group of a number field to automorphic forms (such as ordinary modular forms of $SL(2, \mathbb{Z})$). It has also a geometric analog, involving ordinary Riemann surfaces instead of number fields. This geometric analog has turned out to be intimately related to quantum field theory. It has been extensively studied using two-dimensional conformal field theory [2]–[6] and more recently via four-dimensional gauge theory and electric-magnetic duality [8]. Additional explanation and references can be found in the introduction to [8] and in a recent review article [9].

The simplest version of the geometric Langlands correspondence involves, on one side, a flat connection on a Riemann surface C , and, on the other side, a more sophisticated structure known as a \mathcal{D} -module. The problem has a generalization in which one omits finitely many points p_1, p_2, \dots, p_n from C , and begins with a flat connection on $C' = C \setminus \{p_1, \dots, p_n\}$ that has a prescribed singularity near the given points. This situation gives a geometric analog of what in number theory is called ramification. Since ramification is almost inescapable in number

theory, the extension of the geometric Langlands program to the ramified case is an important part of the analogy between number theory and geometry. There is also an important local version of the problem, in which one focuses on the behavior near a ramification point.

The goal of the present paper is to extend the gauge theory approach to the geometric Langlands program to cover the ramified case. The basic idea is that allowing ramification in the Langlands program corresponds, in gauge theory, to introducing surface operators, somewhat analogous to Wilson or 't Hooft operators, but supported on a two-manifold rather than a one-manifold. The relevant surface operators are defined by specifying a certain type of singularity on a codimension two submanifold. Codimension two singularities in gauge theory of roughly the relevant type have been considered in various contexts, including the theory of cosmic strings [10], Donaldson theory [11], topological field theory (see section 5 of [12]), the dynamics of gauge theory and black holes [13]–[15], the relation of instantons to Seiberg-Witten theory and integrable systems [16, 17], and string theory, where special cases of the operators we consider can be constructed via intersecting branes [18], as we will describe in section 6.4.

Now we will briefly indicate how these gauge theory singularities are related to the theory of Higgs bundles. The gauge theory approach to geometric Langlands is based on $\mathcal{N} = 4$ super Yang-Mills theory twisted and compactified on a Riemann surface. The theory so compactified reduces at low energies [19, 20] to a sigma model in which the target space is a hyper-Kähler manifold that is Hitchin's moduli space of Higgs bundles [21]. What codimension two singularities can be incorporated in this picture? The appropriate singularities, in the basic case that the flat connection on C has only a simple pole, have been described and analyzed by Simpson [22]. Higgs bundles with a singularity of this type are what we will call ramified Higgs bundles. (They are also called Higgs bundles with parabolic structure.) The associated hyper-Kähler moduli spaces have been constructed by Konno [23] and their topology clarified by Nakajima [24]. A corresponding theory for Higgs bundles with poles of higher order has been developed by Biquard and Boalch [25]. The cases of a simple pole or a pole of higher order correspond respectively to what is called tame ramification¹ and wild ramification in the context of the Langlands program. In this paper, we concentrate on tame ramification.

In sections 2, we consider in the context of four-dimensional gauge theory the singularity corresponding to a simple pole. We make a natural proposal for how S -duality acts on the parameters. We further

¹Sometimes, the term “tame ramification” is used more narrowly to refer to the case of flat bundles with unipotent monodromy. We will not make this restriction and consider connections and Higgs bundles with arbitrary simple poles, as described in eqn. (2.2) and section 3.3.

explore the classical geometry in section 3. The highlight of this section is the action of the affine braid group on the cohomology of the moduli space of ramified Higgs bundles; this action is extended in section 4 to an action of the affine braid group on A -branes and B -branes. These phenomena are close cousins of a number of structures found in representation theory, including the Springer representations of the Weyl group [26], the Kazhdan-Lusztig theory of representations of the affine Hecke algebra [27, 28], and the recent extension of this by Bezrukavnikov to an action of the affine braid group on the derived category of the Springer resolution [29]. For an exposition of some of this material, see the book by Chriss and Ginzburg [30]. Interpreting such results in terms of the parameters of a hyper-Kähler resolution (as we will do in the case of Higgs bundle moduli space) was first suggested by Atiyah and Bielawski [31] in the context of coadjoint orbits and Slodowy slices.

Based on our duality proposal, we make in section 4 a proposal for what the geometric Langlands program should say in the case of tame ramification. (A similar proposal has been made mathematically, based on [29] and other results cited in the last paragraph. For an exposition, see section 9.4 of the survey [7]. Some particular cases have been studied in detail in forthcoming work [32].) The extension to wild ramification will be considered elsewhere [33].

In section 5, we use gauge theory to describe the operators (generalizing the 't Hooft/Hecke operators studied in [8]) that can act on branes at a ramification point. This gives a more down-to-earth approach to some topics treated in section 4. In section 6, we give a more local description of some aspects of the behavior at a ramification point in terms of a sigma model whose target is a complex coadjoint orbit endowed with a hyper-Kähler metric. Such metrics were first constructed for semi-simple or nilpotent orbits in [34, 35] and generalized to arbitrary orbits in [36, 37]. This and the related analysis in section 3.8 are the closest we come to analyzing the local case of geometric Langlands. Finally, in section 6, we also briefly describe some string theory constructions of surface operators of the type considered in this paper.

Using conformal field theory, a proposal has been made [6] for a unified approach to the geometric Langlands program allowing poles of any order. This work is surveyed in [7]. Unfortunately, we make contact here neither with the use of conformal field theory nor with this unified statement. We hope, of course, to eventually understand more.

Some background in group theory is reviewed in Appendix A. An index of notation appears in Appendix B. Many facts about Montonen-Olive duality, Hitchin's moduli space, etc., that are used in the present paper are explained more fully in [8].

We thank A. Braverman, D. Gaiatsgory, E. Frenkel, and D. Kazhdan for patient and extremely helpful explanations. We also thank J. Andersen, P. Aspinwall, M.F. Atiyah, D. Ben-Zvi, R. Bezrukavnikov,

R. Bielawski, R. Dijkgraaf, R. Donagi, N. Hitchin, L. Jeffrey, A. Kapustin, P. Kronheimer, Y. Laszlo, H. Nakajima, C. Sorger, and M. Thaddeus, among others, for a wide variety of helpful comments and advice. Research of SG was partly supported by DOE grant DE-FG03-92-ER40701. Research of EW was partly supported by NSF Grant PHY-0503584.

2. Monodromy And Surface Operators

2.1. Definition Of Surface Operators. Our basic technique in this paper will be to study how $\mathcal{N} = 4$ super Yang-Mills theory can be modified along a codimension two submanifold in spacetime. Thus, we consider $\mathcal{N} = 4$ super Yang-Mills theory on a four-manifold M , but modified along a two-dimensional submanifold D in such a way that the four-dimensional fields will have singularities along D . The construction is thus an analog for surface operators of the usual construction of 't Hooft operators via codimension three singularities. For general background see [38] or section 6.2 of [8].

Our focus will be on the GL-twisted version of $\mathcal{N} = 4$ super Yang-Mills theory, which is the basis of the gauge theory approach to the geometric Langlands program. The gauge group is a compact Lie group G , which we will generally assume to be simple. The most important bosonic fields are the gauge field A , which is a connection on a G -bundle $E \rightarrow M$, and an $\text{ad}(E)$ -valued one-form field ϕ . Our gauge theory conventions are those of [8]. In particular, A and ϕ take values in the real Lie algebra of G (and so in a unitary representation of G are represented by anti-hermitian matrices), the covariant derivative is $D = d_A = d + A$, and the holonomy is $P \exp(-\int A)$.

The fields that will be singular along D are simply the restrictions of A and ϕ to the normal bundle to D . Locally, we can model our four-manifold as $M = D \times D'$, where D' is the fiber to the normal bundle, and the singularity will be at a point in D' , say the origin.

The supersymmetric equations of GL-twisted $\mathcal{N} = 4$ super Yang-Mills theory depend on a parameter t , as explained in [8]. But upon reduction to two dimensions this parameter disappears, and we are left with Hitchin's equations:

$$(2.1) \quad \begin{aligned} F - \phi \wedge \phi &= 0 \\ d_A \phi &= 0 \\ d_A \star \phi &= 0. \end{aligned}$$

(Here \star is the Hodge star operator.) Therefore, we will define a surface operator by describing an isolated singularity that can arise in a solution of Hitchin's equations.

In fact, for this paper, we will only require the simplest type of singularity. To begin with, let us take $D' = \mathbb{R}^2$, with Euclidean coordinates

x^1, x^2 , such that $\star(dx^1) = dx^2$, $\star(dx^2) = -dx^1$. We also introduce polar coordinates with $x^1 + ix^2 = re^{i\theta}$. We write \mathbb{T} for a maximal torus of G , and we write $\mathfrak{g}, \mathfrak{t}$ for the Lie algebras of G and \mathbb{T} , respectively. To describe a solution of Hitchin's equations with an isolated singularity at the origin, we pick elements $\alpha, \beta, \gamma \in \mathfrak{t}$, and take

$$(2.2) \quad \begin{aligned} A &= \alpha d\theta \\ \phi &= \beta \frac{dr}{r} - \gamma d\theta. \end{aligned}$$

Since α, β , and γ commute (as \mathfrak{t} is abelian), and the one-forms $d\theta$ and dr/r are closed and co-closed, Hitchin's equations are obeyed. We define our surface operator exactly as one usually defines 't Hooft operators (see for example [38] or section 6.2 of [8]): we require that near $r = 0$, A and ϕ behave as in (2.2). In a global situation, for a two-dimensional submanifold $D \subset M$, we define such a surface operator by requiring that at each point in D , the fields in the normal plane look (in some gauge) like (2.2), with the specified values of α, β , and γ .

Clearly, we can act with the Weyl group \mathcal{W} of G on the trio (α, β, γ) without changing the theory in an essential way. So the surface operator that we have defined depends on (α, β, γ) only modulo a Weyl transformation. But there is an additional freedom. If $u \in \mathfrak{t}$ is such that $\exp(2\pi u) = 1$, then a gauge transformation by the \mathbb{T} -valued function

$$(2.3) \quad (r, \theta) \rightarrow \exp(\theta u)$$

shifts α by u . Modulo this transformation, the only invariant of α is the \mathbb{T} -valued monodromy of the connection A around a circle of constant r ; this monodromy is $\exp(-2\pi\alpha)$. Thus, the trio (α, β, γ) takes values in $\mathbb{T} \times \mathfrak{t} \times \mathfrak{t}$, modulo the action of \mathcal{W} .

We will often but not always use an additive notation for α . This corresponds to thinking of \mathbb{T} as the quotient \mathfrak{t}/Λ for some lattice Λ . To identify this lattice, note that if $\mathbb{T} = \mathfrak{t}/\Lambda$, then $\Lambda = \pi_1(\mathbb{T}) = \text{Hom}(U(1), \mathbb{T})$. We call this the cocharacter lattice of G , denoted Λ_{cochar} . (See Appendix A for more information.) In fact, $\text{Hom}(U(1), \mathbb{T})$ parametrizes \mathbb{T} -valued magnetic charges, or equivalently, by the basic GNO duality [39], electric charge of the dual group ${}^L G$. Λ_{cochar} is a sublattice of the coweight lattice Λ_{cowt} . Their quotient is the center $\mathcal{Z}(G)$ of G :

$$(2.4) \quad \mathcal{Z}(G) = \Lambda_{\text{cowt}}/\Lambda_{\text{cochar}}.$$

Extensions Of Bundles

Let us try to compute the curvature at the origin of the singular connection $A = \alpha d\theta$. Since $d(d\theta) = 2\pi\delta_D$ (δ_D is a two-form delta function supported at the origin in D' and Poincaré dual to D), we seem to get

$$(2.5) \quad F = 2\pi\alpha\delta_D.$$

This, however, cannot be a natural formula, since we are free to shift α by a lattice vector. What has gone wrong with the definition of curvature? We have introduced A as a connection on a G -bundle E , but because of the singularity along D , this bundle is only naturally defined on the complement of D in M . It is possible to pick an extension of E over D , but there is no natural extension. The different possible extensions of E over D correspond to different ways to lift α from $\mathbb{T} = \mathfrak{t}/\Lambda_{\text{cochar}}$, where it naturally takes values, to \mathfrak{t} . Once we pick an extension, it makes sense to compute the curvature at the origin, and the result is (2.5).

The gauge transformation $(r, \theta) \rightarrow \exp(\theta u)$ that shifts α by a lattice vector, because of its singularity at the origin, maps one extension of E over D to another. Though there is no natural way to extend E over D as a G -bundle, we can do the following. Near D , the structure group of E naturally reduces to the subgroup \mathbb{T} that commutes with the singular part of A and ϕ . The singular gauge transformation $\theta \rightarrow \exp(\theta u)$ acts trivially on \mathbb{T} , and hence, though there is no natural G -bundle over D , there is a natural \mathbb{T} -bundle over D . We will assume that the restriction of A to D is a connection on this \mathbb{T} -bundle, and that the curvature F , when restricted to D , is likewise \mathfrak{t} -valued.

Suppose that the gauge group G is not simply-connected; for example, it may be a group of adjoint type. Then the gauge transformation $\exp(\theta u)$ may not lift to a single-valued gauge transformation in the simply-connected cover \overline{G} of G ; rather, under $\theta \rightarrow \theta + 2\pi$, it is multiplied by an element $y \in \mathcal{Z}(\overline{G})$, the center of \overline{G} . When this is the case, this gauge transformation changes the topology of the bundle E , by shifting the characteristic class $\xi \in H^2(M, \pi_1(G))$ that measures the obstruction to lifting E to a \overline{G} -bundle. If we use an additive notation for $\mathcal{Z}(\overline{G})$, the shift is

$$(2.6) \quad \xi \rightarrow \xi + y[D],$$

where $[D]$ is the class Poincaré dual to D . Thus, gauge theories with different values of ξ and suitably related values of α are equivalent.

A variant of this is as follows. Suppose that the gauge group is in fact \overline{G} or some other form in which the gauge transformation by $\exp(\theta u)$ is not single-valued. This gauge transformation nevertheless makes sense locally as a symmetry of $\mathcal{N} = 4$ super Yang-Mills theory (in which all fields are in the adjoint representation). If the global topology is such that a gauge transformation that looks locally like $\exp(\theta u)$ near D can be extended globally over M , then $\mathcal{N} = 4$ super Yang-Mills is invariant to $\alpha \rightarrow \alpha + u$. Here u may be any element of Λ_{cowt} for which $y[D] = 0$. More generally, if D is a union of disjoint components D_i near which we make gauge transformations by $\exp(\theta_i u_i)$ with $y_i = \exp(2\pi u_i)$, then

the condition is

$$(2.7) \quad \sum_i y_i [D_i] = 0.$$

Non-Trivial Normal Bundle

In motivating this construction, we began with the special case of a product $M = D \times D'$. However, more generally, for an arbitrary embedded two-manifold $D \subset M$, we consider gauge fields with a singularity like (2.2) in each normal plane. For simplicity, in this paper, we only consider the case that M and D are oriented. D may have a non-trivial normal bundle, and hence a nonzero self-intersection number $D \cap D$, which can be characterized as²

$$(2.8) \quad D \cap D = \int_M \delta_D \wedge \delta_D.$$

When the self-intersection number is nonzero, it is not possible globally for α to have arbitrary values. We explain this first for $G = U(1)$. A connection A which has a singularity $A = \alpha d\theta$ near D will have the property $\int_D F/2\pi = \alpha D \cap D \bmod \mathbb{Z}$. Since the integrated first Chern class $\int_D F/2\pi$ is always an integer, it will always be that

$$(2.9) \quad \alpha D \cap D \in \mathbb{Z}.$$

For any G , the generalization of this is simply that the same statement holds in any $U(1)$ subgroup of \mathbb{T} . So if $\alpha \rightarrow f(\alpha)$ is any real-valued linear function on \mathfrak{t} that takes integer values on the lattice Λ_{cochar} , then

$$(2.10) \quad f(\alpha) D \cap D \in \mathbb{Z}.$$

It is also true that if $D \cap D \neq 0$, the twisted gauge transformation that is defined in the normal plane in (2.3) cannot always be defined globally along D . Only those gauge transformations that shift α in a way compatible with (2.10) can actually be defined globally.

Singularities along surfaces with $D \cap D \neq 0$ are important in four-manifold theory [11], but will be less important in our applications here, since the geometric Langlands program, in its usual form, deals with a situation in which $D \cap D = 0$.

2.2. Geometric Interpretation Of Parameters. We have defined a surface operator supported on a two-manifold $D \subset M$ by requiring that the fields behave near D as in (2.2). In general, quantum mechanically, Hitchin's equations or even the second order classical field equations of the theory will not necessarily be obeyed; the definition of the surface operator only requires that they are obeyed near D . However, to understand better the meaning of the parameters α, β, γ in classical geometry, let us consider a situation in which we *do* want to

²For this and some other assertions made momentarily, see the description of the Thom class in [40].

solve Hitchin's equations on a Riemann surface C (which corresponds to D' in the above analysis), with an isolated singularity of the above-described type near some point $p \in C$. (We will here explain only the facts about the classical geometry that are needed to motivate the duality conjecture of section 2.4. We will reconsider the classical geometry in much more detail in section 3.)

Solutions of Hitchin's equations with the type of singularity described in (2.2) have been analyzed in [22]–[24]. Just like smooth solutions of Hitchin's equations, the moduli space of such solutions is a hyper-Kähler manifold \mathcal{M}_H , which we will call the moduli space of ramified Higgs bundles, also known as the moduli space of Higgs bundles with parabolic structure. (We refer to it as $\mathcal{M}_H(G)$, $\mathcal{M}_H(C)$, $\mathcal{M}_H(\alpha, \beta, \gamma; p)$, etc., if we want to make explicit the gauge group, the Riemann surface, or the nature and location of a singularity.)

Because of the hyper-Kähler structure, solutions of Hitchin's equations can be viewed in different ways. From the standpoint of one complex structure, usually called I , a solution of Hitchin's equations on a Riemann surface C describes a Higgs bundle, that is a pair (E, φ) , where E is a holomorphic G -bundle and φ is a holomorphic section of $K_C \otimes \text{ad}(E)$. (K_C denotes the canonical bundle of C .) The Higgs bundle is constructed as follows starting with a solution (A, ϕ) of Hitchin's equations. One interprets the $(0, 1)$ part of the gauge-covariant exterior derivative $d_A = d + A$ as a $\bar{\partial}_A$ operator that gives the bundle E a holomorphic structure. And one defines φ as the $(1, 0)$ part of ϕ ; that is, one decomposes ϕ as $\varphi + \bar{\varphi}$, where φ is of type $(1, 0)$ and $\bar{\varphi}$ is of type $(0, 1)$. Then Hitchin's equations imply that φ is holomorphic, and the pair (E, φ) is a Higgs bundle.

In the present context, setting $z = x^1 + ix^2$, we find from (2.2) that

$$(2.11) \quad \varphi = \frac{1}{2}(\beta + i\gamma) \frac{dz}{z}.$$

Thus, from the point of view of complex structure I , the surface operator introduces in the Higgs field a pole with residue³ $(1/2)(\beta + i\gamma)$. The characteristic polynomial of φ varies holomorphically in complex structure I , and in particular this is so for the conjugacy class of the pole in φ . (2.11) shows that the conjugacy class of this pole is holomorphic in $\beta + i\gamma$, and independent of α . This is part of a more general statement; the complex structure I varies holomorphically with $\beta + i\gamma$, and is independent of α . On the other hand, the corresponding Kähler form ω_I that is of type $(1, 1)$ in complex structure I has a cohomology class that is independent of β and γ and is a linear function of α . These

³This statement holds if $\beta + i\gamma$ is a “regular” element of the Lie algebra $\mathfrak{g}_{\mathbb{C}}$ (the subalgebra of $\mathfrak{g}_{\mathbb{C}}$ that commutes with it is precisely $\mathfrak{t}_{\mathbb{C}}$). What happens otherwise is described in section 3.3.

statements, along with some similar ones below, follow from the construction of the moduli space as a hyper-Kähler quotient [23, 24], as we will explain in section 3.2. So if we look at \mathcal{M}_H from the vantage point of complex structure I , then $\beta + i\gamma$ is a complex parameter (on which I depends holomorphically) and α is a Kähler parameter.

In fact, though we will not require the details in this paper, α has a natural meaning in pure algebraic geometry (without mentioning Kähler metrics), but this meaning is a little subtle. \mathcal{M}_H in complex structure I parametrizes pairs (E, φ) , where φ has a simple pole whose conjugacy class is determined by β and γ (as in (2.11)) and moreover the pair (E, φ) is “stable.” The appropriate notion of stability [22] depends on α .

In complex structure J , the natural complex variable is the $G_{\mathbb{C}}$ -valued connection $\mathcal{A} = A + i\phi$, which is flat according to Hitchin’s equations. The monodromy of $\mathcal{A} = (\alpha - i\gamma)d\theta$ around the singularity at p is

$$(2.12) \quad U = \exp(-2\pi(\alpha - i\gamma)).$$

It depends holomorphically on $\gamma + i\alpha$, and is independent of β . Indeed, in complex structure J , $\gamma + i\alpha$ is a complex parameter and β is a Kähler parameter.

In complex structure J , \mathcal{M}_H parametrizes flat $G_{\mathbb{C}}$ -bundles over C with a monodromy around the point p that is in the conjugacy class⁴ containing U . Like α in complex structure I , β can be interpreted in complex structure J in purely holomorphic terms (without mentioning Kähler metrics), but this interpretation is a little elusive (and will play only a slight role in the present paper). According to [22], β determines in complex structure J the weights of a monodromy-invariant weighted filtration of the flat bundle over $C \setminus p$ (that is, C with the point p omitted) whose connection is \mathcal{A} .

Finally, in complex structure $K = IJ$, the natural complex variable is the $G_{\mathbb{C}}$ -valued connection $\tilde{\mathcal{A}} = A + i \star \phi$. It is again flat by virtue of Hitchin’s equations. Its monodromy around the singularity at p is $\exp(-2\pi(\alpha + i\beta))$. In complex structure K , $\alpha + i\beta$ is a complex structure parameter and γ is a Kähler parameter. The interpretation of γ in complex structure K is just like the interpretation of β in complex structure J .

These statements are summarized in Table 1. In the table, one sees a cyclic symmetry under permutations of I, J, K together with α, β, γ . (This cyclic symmetry is part of an $SO(3)$ symmetry that appears in a closely related context; see section 3.8.) If $G = U(N)$, then (α, β, γ) are called $(\alpha, 2b, 2c)$ in the table on p. 720 of [22]. We have adjusted a factor of 2 to get the cyclic symmetry.

⁴As in the previous footnote, this statement holds if U is regular; we postpone a discussion of the more general case to section 3.3.

Model	Complex Modulus	Kahler Modulus
I	$\beta + i\gamma$	α
J	$\gamma + i\alpha$	β
K	$\alpha + i\beta$	γ

TABLE 1. Complex and Kahler moduli for \mathcal{M}_H in complex structures I , J , and K . Complex structure I , for example, depends holomorphically on $\beta + i\gamma$, while the corresponding Kahler structure depends on α .

2.3. Theta Angles. Quantum mechanically, in addition to α , β , and γ , an additional parameter is present. One may guess that this will occur, because so far we have described in each complex structure only a real Kahler parameter (listed in the last column of Table 1), but in supersymmetric theories, the Kahler parameters are usually complexified.

In explaining this, let us assume for the moment that the trio (α, β, γ) is regular, meaning that the subgroup of G that leaves this trio invariant is precisely \mathbb{T} . Requiring the behavior (2.2) in each normal plane to a two-manifold $D \subset M$ means, in particular, that along D we are given a reduction of the structure group of E from G to \mathbb{T} . Therefore, along D we are doing abelian gauge theory, with gauge group \mathbb{T} . (We explained at the end of section 2.1 that along D , there is a natural \mathbb{T} -bundle, though there is no natural G -bundle.)

In abelian gauge theory in two dimensions, an important role is played by the “theta-angle.” For example, let the gauge group be $U(1)$. A $U(1)$ -bundle $\mathcal{L} \rightarrow D$ is classified topologically by its degree $d = \int_D c_1(\mathcal{L})$, where $c_1(\mathcal{L})$ is the first Chern class.⁵ The theta angle enters the theory by a phase $\exp(i\theta d)$ that is included as an extra factor in the path integral. Here θ takes values in $\mathbb{R}/2\pi\mathbb{Z} \cong U(1)$.

Now let us return to our problem, involving surface operators in nonabelian gauge theory. Suppose that G has rank r . Then its maximal torus \mathbb{T} is isomorphic to $U(1)^r$, and a two-dimensional gauge theory with gauge group G will have r theta-angles, taking values in an r -dimensional torus. Let us see exactly which torus this is. A \mathbb{T} -bundle over a two-manifold D can be constructed uniquely (from a topological point of view) by starting with a $U(1)$ bundle of degree 1 and then mapping this to \mathbb{T} via some homomorphism $\rho : U(1) \rightarrow \mathbb{T}$. So \mathbb{T} -bundles over D are classified topologically by a characteristic class \mathfrak{m}

⁵For simplicity, we assume D to be closed. Otherwise, in defining the quantum field theory, one needs a suitable boundary condition on the boundary of D (or at infinity). With some care, one can then give a suitable definition of $c_1(\mathcal{L})$ and of θ .

that takes values in the lattice $\Lambda_{\text{cochar}} = \text{Hom}(U(1), \mathbb{T})$. Therefore, the theta-angle of \mathbb{T} gauge theory is a homomorphism $\eta : \Lambda_{\text{cochar}} \rightarrow U(1)$.

In other words, η takes values in $\text{Hom}(\Lambda_{\text{cochar}}, U(1))$. We claim that

$$(2.13) \quad \text{Hom}(\Lambda_{\text{cochar}}, U(1)) = {}^L\mathbb{T},$$

where ${}^L\mathbb{T}$ is the maximal torus of the dual group ${}^L G$. In fact, by Pontryagin duality,⁶ since $\Lambda_{\text{cochar}} = \text{Hom}(U(1), \mathbb{T})$, (2.13) is equivalent to

$$(2.14) \quad \text{Hom}(U(1), \mathbb{T}) = \text{Hom}({}^L\mathbb{T}, U(1)).$$

But this is a standard characterization of the relation between the group and the dual group. Indeed, the left hand side classifies magnetic charges of G , and the right hand side classifies electric charges of ${}^L G$. The equality of the two is the basic GNO duality [39].

Just as $\mathbb{T} = \mathfrak{t}/\Lambda_{\text{cochar}}$, we have ${}^L\mathbb{T} = {}^L\mathfrak{t}/\Lambda_{\text{char}}$, with ${}^L\mathfrak{t}$ the Lie algebra of ${}^L\mathbb{T}$, and $\Lambda_{\text{char}} = \text{Hom}(T, U(1))$ the lattice of electric charges of G . ${}^L\mathfrak{t}$ coincides with \mathfrak{t}^\vee , the dual of \mathfrak{t} . We frequently use an additive notation for η , thinking of it as an element of $\mathfrak{t}^\vee/\Lambda_{\text{char}}$.

2.4. Electric-Magnetic Duality. $\mathcal{N} = 4$ super Yang-Mills theory has a large discrete group Γ of duality symmetries. Optimistically assuming that the class of surface operators that we have described is mapped to itself by Γ , let us determine how Γ acts on the parameters $(\alpha, \beta, \gamma, \eta)$.

First we consider the fundamental electric-magnetic duality S . It acts on the coupling parameter $\tau = \theta/2\pi + 4\pi i/g^2$ of the gauge theory by $S : \tau \rightarrow -1/n_{\mathfrak{g}}\tau$, where $n_{\mathfrak{g}}$ is 1 for simply-laced G and otherwise is 2 or 3 (the ratio of the length squared of the long and short roots of G). It also maps G to ${}^L G$. How does it transform the parameters of a surface operator?

Transformation Of β and γ

S acts on β and γ in a very simple way, because β and γ determine a pole in the characteristic polynomial of the Higgs field, which transforms very simply under duality (see [8], section 5.4; however, we will adopt a different normalization from the one used there).

Since β and γ take values in \mathfrak{t} , while ${}^L\beta$ and ${}^L\gamma$ take values in ${}^L\mathfrak{t}$, the comparison between them depends on a choice of map from \mathfrak{t} to ${}^L\mathfrak{t}$. The vector spaces \mathfrak{t} and ${}^L\mathfrak{t}$ are dual (so that we also denote ${}^L\mathfrak{t}$ as \mathfrak{t}^\vee), and acted on by the same Weyl group. Any choice of a Weyl-invariant metric on \mathfrak{t} gives a Weyl-invariant identification between them.

⁶Pontryagin duality says that if W is a locally compact abelian group and $V = \text{Hom}(W, U(1))$, then $W = \text{Hom}(V, U(1))$. In our application, $W = \Lambda_{\text{cochar}} = \text{Hom}(U(1), \mathbb{T})$, and $V = {}^L\mathbb{T}$.

To prepare the ground for the application to geometric Langlands in section 4, we will describe in some detail the identification we will use.⁷ First of all, it is convenient to introduce an invariant quadratic form on \mathfrak{g} , normalized so that a short coroot is of length squared 2. We write this quadratic form as $(x, y) = -\text{Tr } xy$, for $x, y \in \mathfrak{g}$. (The notation is motivated by the fact that for $G = SU(N)$, Tr is the trace in the N -dimensional representation.) Similarly, we introduce a quadratic form on the Lie algebra ${}^L\mathfrak{g}$ of LG , normalized so that a short root of G has length squared 2. We write this as $({}^Lx, {}^Ly) = -{}^L\text{Tr } {}^Lx {}^Ly$. The quadratic form on \mathfrak{g} , when restricted to \mathfrak{t} , gives a Weyl-invariant map from $x \in \mathfrak{t}$ to $x^* \in {}^L\mathfrak{t}$ (such that $x^*(y) = -\text{Tr } xy$), and likewise, the quadratic form on ${}^L\mathfrak{g}$, restricted to ${}^L\mathfrak{t}$, gives a Weyl-invariant map from ${}^Lx \in {}^L\mathfrak{t}$ to ${}^Lx^* \in \mathfrak{t}$. As explained in Appendix A, the composition of the two maps is multiplication by $n_{\mathfrak{g}}$, that is,

$$(2.15) \quad (x^*)^* = n_{\mathfrak{g}}x,$$

or equivalently

$$(2.16) \quad {}^L\text{Tr } x^* y^* = n_{\mathfrak{g}} \text{Tr } xy$$

for any $x, y \in \mathfrak{g}$. This relation is symmetric between G and LG .

Now as in eqn. (2.8) of [8], we normalize the scalar fields ϕ of $\mathcal{N} = 4$ super Yang-Mills theory with gauge group G so that their kinetic energy in Lorentz signature is

$$(2.17) \quad \frac{\text{Im } \tau}{4\pi} \int d^4x \text{Tr } D_\mu \phi D^\mu \phi,$$

where $\text{Im } \tau = 4\pi/e^2$. Likewise, for the S -dual theory with gauge group LG , we normalize the scalars ${}^L\phi$ so that their kinetic energy is

$$(2.18) \quad \frac{\text{Im } {}^L\tau}{4\pi} \int d^4x \text{Tr } D_\mu {}^L\phi D^\mu {}^L\phi,$$

with

$$(2.19) \quad {}^L\tau = -\frac{1}{n_{\mathfrak{g}}\tau}.$$

In general, there is no local identification between ϕ and ${}^L\phi$ (but only between gauge-invariant local operators constructed from these fields). However, for the sake of finding how β and γ transform under duality, we can abelianize the problem, as in section 5.1 of [8], and go to a vacuum in which G or LG is spontaneously broken to its maximal torus by expectation values of scalar fields. In such an abelian vacuum, the light scalar fields takes values in \mathfrak{t} or ${}^L\mathfrak{t}$, and duality acts on them simply by a linear transformation that we can choose to be Weyl-invariant (and

⁷For gauge groups G_2 or F_4 , the convention we are about to describe differs from the most common one in the physics literature [41, 42]. The relation between the two approaches is explained at the end of Appendix A. See also, for example, [71].

hence a multiple of the operation $\phi \rightarrow \phi^*$). This transformation must preserve the kinetic energy, so

$$(2.20) \quad \frac{\text{Im } \tau}{4\pi} \int d^4x \text{Tr } D_\mu \phi D^\mu \phi = \frac{\text{Im } L\tau}{4\pi} \int d^4x \text{Tr } D_\mu L\phi D^\mu L\phi.$$

Together with (2.19) and (2.16), this implies that in an abelian vacuum the relation between ϕ and $L\phi$ is

$$(2.21) \quad L\phi = |\tau| \phi^*,$$

a relation that is completely symmetric between G and LG , as one can verify using (2.15) and (2.19). Since β and γ parametrize singularities of ϕ , and likewise $L\beta$ and $L\gamma$ parametrize singularities of $L\phi$, these parameters transform in the same way, that is $L\beta = |\tau| \beta^*$, $L\gamma = |\tau| \gamma^*$, or more briefly

$$(2.22) \quad (L\beta, L\gamma) = |\tau| (\beta^*, \gamma^*).$$

Since the parameters characterize the operator, not the vacuum, this formula must hold in general, not just in an abelian vacuum.

The basic case of the geometric Langlands program is conveniently studied by setting $\text{Re } \tau = 0$, in which case (2.22) can be more conveniently written

$$(2.23) \quad (L\beta, L\gamma) = (\text{Im } \tau) (\beta^*, \gamma^*).$$

We can also invert this relation:

$$(2.24) \quad (\beta, \gamma) = (\text{Im } L\tau) (L\beta^*, L\gamma^*).$$

The two formulas are compatible, since $(x^*)^* = n_{\mathfrak{g}} x$ and (for imaginary τ) $\text{Im } L\tau = 1/n_{\mathfrak{g}} \text{Im } \tau$.

β and γ are manifestly unaffected by a shift in the theta-angle, which as we discuss later gives the second generator $T : \tau \rightarrow \tau + 1$ of the duality group Γ . So their full transformation under Γ is determined by (2.22). The result can be described particularly simply if G is simply-laced, in which case, as explained in Appendix A, the difference between (β, γ) and (β^*, γ^*) is inessential. In that case, (2.22) together with invariance of (β, γ) under $\tau \rightarrow \tau + 1$ implies that for a general element

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \cong SL(2, \mathbb{Z}), \text{ the transformation is } (\beta, \gamma) \rightarrow |c\tau + d| (\beta, \gamma).$$

A similar result can be written for any G . If one restricts to an index 2 subgroup of Γ that maps G to itself (rather than LG), (2.22) implies that the pair (β, γ) transforms by rescaling by a positive factor. Other elements of Γ map (β, γ) to a positive multiple of (β^*, γ^*) .

Transformation Of α and η

The other parameters of a surface operator are $\alpha \in \mathbb{T}$ and $\eta \in L\mathbb{T}$. Since S exchanges G and LG , it exchanges \mathbb{T} and $L\mathbb{T}$, strongly suggesting that it exchanges α and η . This is much more interesting

than the relatively trivial transformation of β and γ . It will be our basic hypothesis in the present paper.

In fact, S^2 is a central element of the duality group Γ . It acts trivially on τ , and acts on other fields and parameters by charge conjugation. So S^2 maps (α, η) to $(-\alpha, -\eta)$. Together with the fact that S exchanges \mathbb{T} and ${}^L\mathbb{T}$, it follows that up to sign, S must act by

$$(2.25) \quad S : (\alpha, \eta) \rightarrow (\eta, -\alpha).$$

Since the duality group Γ contains the central element S^2 that reverses the sign of the pair (α, η) , the overall sign in (2.25) depends on precisely how we lift S from a symmetry of the upper half τ -plane to an element of Γ . We specify our lifting in eqn. (2.52) below.

The Abelian Case

For nonabelian G , we cannot prove (2.25), but regard it as the natural extension to surface operators of the Montonen-Olive duality conjecture.⁸ In the case $G = U(1)$, however, we can directly demonstrate that S does act in this fashion, as we will now explain. We will follow the approach to abelian S -duality in [43] (which in turn was modeled on a similar approach to T -duality in two dimensions [44, 45]).

In abelian gauge theory, the gauge field is locally a real one-form A (which we can think of as a connection on a principal $U(1)$ -bundle⁹ $\mathcal{R} \rightarrow M$) with curvature $F = dA$. We take the action to be¹⁰

$$(2.26) \quad \begin{aligned} \mathbf{I} &= \frac{1}{8\pi} \int_M d^4x \sqrt{h} \left(\frac{4\pi}{g^2} F_{mn} F^{mn} - \frac{i\theta}{2\pi} \frac{1}{2} \epsilon_{mnpq} F^{mn} F^{pq} \right) \\ &= -\frac{i}{8\pi} \int_M d^4x \sqrt{h} \left(\tau F_{mn}^+ F^{+mn} - \bar{\tau} F_{mn}^- F^{-mn} \right), \end{aligned}$$

where h is a metric on M , $\tau = \theta/2\pi + 4\pi i/g^2$, ϵ_{mnpq} is the Levi-Civita antisymmetric tensor, and $F^\pm = \frac{1}{2}(F \pm \star F)$ are the selfdual and anti-selfdual projections of F . As explained in [43], \mathbf{I} is invariant mod $2\pi i\mathbb{Z}$ under $\tau \rightarrow \tau + 2$ for any closed four-manifold M , and under $\tau \rightarrow \tau + 1$ if M is a spin manifold. Quantum theory depends on the action only mod $2\pi i\mathbb{Z}$ (since the action enters the path integral via a factor $\exp(-\mathbf{I})$), so the quantum theory possesses the symmetry $\tau \rightarrow \tau + 1$ or $\tau \rightarrow \tau + 2$, depending on M .

⁸In some special cases, (2.25) follows from broader string theory duality conjectures, in view of the constructions explained in section 6.4.

⁹The Lie algebra of $U(1)$ is real, so a connection on a principal $U(1)$ -bundle is naturally represented locally by a real one-form. (By contrast, if we view A as a connection on a unitary complex line bundle \mathcal{L} , coming from a unitary representation of $U(1)$, we would represent it locally by an imaginary one-form.)

¹⁰To agree better with conventions used in much of the physics literature as well as [8] and the present paper (but at the cost of some tension with conventions usually used in Donaldson theory), we have reversed the sign of θ relative to [43]. This has the effect of transforming $\tau \rightarrow -\bar{\tau}$.

Our interest here is in the more subtle symmetry $\tau \rightarrow -1/\tau$. First we will review how to see this symmetry in the absence of surface operators. We add a two-form field \mathbf{k} (called G in [43]) which we assume to be invariant under the usual abelian gauge symmetry $A \rightarrow A - d\epsilon$ (for a zero-form ϵ). But we ask for the extended gauge symmetry

$$(2.27) \quad \begin{aligned} A &\rightarrow A + b \\ \mathbf{k} &\rightarrow \mathbf{k} + db, \end{aligned}$$

where b is any connection on a principal $U(1)$ -bundle \mathcal{T} , and db is its curvature. (The ordinary Maxwell gauge symmetry is a special case of this with $b = -d\epsilon$.) If A is a connection on a principal $U(1)$ -bundle \mathcal{R} , then $A+b$ is a connection on $\mathcal{R} \otimes \mathcal{T}$, so to get invariance under (2.27), we will need to sum over all possible choices of \mathcal{R} . A transformation (2.27) can shift the periods of \mathbf{k} by integer multiples of 2π ; thus, if $D \subset M$ is a two-cycle, we can have

$$(2.28) \quad \int_D \mathbf{k} \rightarrow \int_D \mathbf{k} + 2\pi m, \quad m \in \mathbb{Z}.$$

An obvious way to find a Lagrangian with the invariance (2.27) is to set $\mathcal{F} = F - \mathbf{k}$ and replace F everywhere in the Maxwell Lagrangian by \mathcal{F} . But the resulting theory is trivial, and certainly not equivalent to Maxwell theory. To get something interesting, we introduce another field \mathbf{v} which is a connection on a principal $U(1)$ -bundle $\tilde{\mathcal{R}}$, with curvature $\mathbf{V} = d\mathbf{v}$. We add to the action a term

$$(2.29) \quad \tilde{\mathbf{I}} = \frac{i}{8\pi} \int_M d^4x \sqrt{h} \epsilon^{mnpq} \mathbf{V}_{mn} \mathbf{k}_{pq} = \frac{i}{2\pi} \int_M \mathbf{V} \wedge \mathbf{k}.$$

To check the symmetry (2.27), note that if \mathcal{T} is topologically trivial, then the connection b is globally-defined as a one-form, and an integration by parts shows that $\tilde{\mathbf{I}}$ is invariant under $\mathbf{k} \rightarrow \mathbf{k} + db$. We have chosen the coefficient in (2.29) so that $\tilde{\mathbf{I}}$ is invariant mod $2\pi i\mathbb{Z}$ even if \mathcal{T} is topologically non-trivial.

We now define an extended theory with field variables $\mathbf{v}, \mathbf{k}, A$ and action

$$(2.30) \quad \hat{\mathbf{I}}(\mathbf{v}, \mathbf{k}, A) = \frac{i}{8\pi} \int d^4x \sqrt{h} (\epsilon^{mnpq} \mathbf{V}_{mn} \mathbf{k}_{pq} - \tau \mathcal{F}_{mn}^+ \mathcal{F}^{+mn} + \bar{\tau} \mathcal{F}_{mn}^- \mathcal{F}^{-mn}).$$

From what has been said, the invariance of $\hat{\mathbf{I}}$ under (2.27) mod $2\pi i\mathbb{Z}$ should be clear. The proof of S -duality of abelian gauge theory comes by comparing two ways of studying this extended theory.

One approach is to perform first the path integral over \mathbf{v} . The part of the integral that depends on \mathbf{v} is

$$(2.31) \quad \sum_{\tilde{\mathcal{R}}} \int D\mathbf{v} \exp\left(-\frac{i}{2\pi} \int_M \mathbf{V} \wedge \mathbf{k}\right)$$

where \mathbf{V} is the curvature of \mathbf{v} . We must sum over $U(1)$ -bundles $\tilde{\mathcal{R}}$, and for each such bundle we must integrate over the space of all connections on it. The integral gives a delta function setting $d\mathbf{k} = 0$. The sum over bundles gives a delta function stating that the periods of \mathbf{k} take values in $2\pi\mathbb{Z}$. The combined conditions precisely say that \mathbf{k} can be set to zero by a transformation (2.27). After setting \mathbf{k} to zero, we are left with the original abelian gauge theory (2.26) with A as the only field variable.

So the extended theory with action (2.30) is equivalent to the original theory with coupling parameter τ . Next, let us consider another way to study the same theory. We use the extended gauge invariance (2.27) to set $A = 0$. After doing this, as the action depends quadratically on \mathbf{k} without any derivatives of \mathbf{k} (and the term quadratic in \mathbf{k} is nondegenerate), we can “integrate out” \mathbf{k} by simply solving the Euler-Lagrange equations to determine \mathbf{k} in terms of \mathbf{V} . (In [43], this process is described somewhat more precisely at the quantum level.) In this way, we get an action for \mathbf{V} , which is

$$(2.32) \quad -\frac{i}{8\pi} \int_M d^4x \sqrt{h} \left(\left(-\frac{1}{\tau} \right) (\mathbf{V}^+)^2 - \left(-\frac{1}{\bar{\tau}} \right) (\mathbf{V}^-)^2 \right).$$

This is the original abelian gauge theory, but with the connection \mathbf{v} instead of A , and the coupling parameter τ replaced by $-1/\tau$. So comparing the two ways to analyze the extended theory (2.30) gives us the $\tau \rightarrow -1/\tau$ symmetry of abelian gauge theory.

Now let us introduce a surface operator, supported on a two-manifold $D \subset M$. To keep things simple, we will consider the special case that this surface operator has $\alpha = 0$, $\eta \neq 0$. This means that in the path integral of the underlying abelian gauge theory, we want to include a factor¹¹

$$(2.33) \quad \exp \left(i\eta \int_D F \right).$$

This is equivalent to adding to the action a term $-i\eta \int_D F$. To incorporate the surface operator in the extended theory, we replace F by \mathcal{F} and add

$$(2.34) \quad -i\eta \int_D \mathcal{F} = -i\eta \int_M \delta_D \wedge \mathcal{F}$$

to the action (2.30). Here δ_D is a two-form delta function that is Poincaré dual to D .

The extra term does not depend on \mathbf{v} , so if we first integrate over \mathbf{v} , we get back to the abelian gauge theory with the surface operator. But what happens if we instead use the extended gauge symmetry to

¹¹Though η is an angular variable, we have normalized it to take values in \mathbb{R}/\mathbb{Z} , rather than $\mathbb{R}/2\pi\mathbb{Z}$, to avoid unnatural-looking factors of 2π in the transformations under S -duality. The alternative is to modify the definition of α , β , and γ by a factor of 2π .

set $A = 0$, and then solve for \mathbf{k} ? The action has two terms linear in \mathbf{k} , which combine to

$$(2.35) \quad \frac{i}{2\pi} \int_M (\mathbf{V} - 2\pi\eta\delta_D) \wedge \mathbf{k}.$$

To get this formula, we used the last expressions given in (2.29) and (2.34). Moreover, this part of the action is the only part that contains \mathbf{V} . (The rest of the action is quadratic in \mathbf{k} and independent of \mathbf{V} .) So the effect of having $\eta \neq 0$ is precisely to replace \mathbf{V} by $\widehat{\mathbf{V}} = \mathbf{V} - 2\pi\eta\delta_D$. Hence when we integrate out \mathbf{k} again, we will get the same action as in (2.32), but with \mathbf{V} replaced by $\widehat{\mathbf{V}}$:

$$(2.36) \quad -\frac{i}{8\pi} \int_M d^4x \sqrt{h} \left(\left(-\frac{1}{\tau} \right) (\widehat{\mathbf{V}}^+)^2 - \left(-\frac{1}{\bar{\tau}} \right) (\widehat{\mathbf{V}}^-)^2 \right).$$

This action is potentially divergent because of the delta function term in $\widehat{\mathbf{V}}$. Since the action is quadratic in $\widehat{\mathbf{V}}$, with positive definite real part, the only way to make the action finite is for \mathbf{V} to be such as to cancel the delta function contribution in $\widehat{\mathbf{V}}$. Thus, the connections that contribute to the path integral must obey

$$(2.37) \quad \mathbf{V} = 2\pi\eta\delta_D + \dots$$

where the ellipses refer to terms that are regular near D . But this is precisely the characterization of a surface operator with parameter $\alpha = \eta$, as should become clear upon comparing (2.37) to (2.5).

So we have shown that the transformation $S : \tau \rightarrow -1/\tau$ maps a surface operator with parameters $(\alpha, \eta) = (0, \eta)$ to one with parameters $(\eta, 0)$. This is a special case of (2.25). To get the general case, one replaces \mathcal{F} in (2.30) by $\mathcal{F} - 2\pi\alpha\delta_D$, and then repeats the calculation. After gauging A to zero and integrating out \mathbf{k} , one gets back to an action of the same kind, with τ replaced by $-1/\tau$ and α and η exchanged as in (2.25).

2.5. Shifting The Theta Angle. We now return to the case that G is a simple non-abelian gauge group. Apart from $S : \tau \rightarrow -1/n_{\mathfrak{g}}\tau$, which we have considered in section 2.4, the other generator of the duality group is a shift in the four-dimensional theta-angle, which enters the four-dimensional action via a term

$$(2.38) \quad \mathbf{I}_\theta = -i\theta\mathbf{N}$$

where

$$(2.39) \quad \mathbf{N} = -\frac{1}{8\pi^2} \int_M \text{Tr} F \wedge F$$

is the instanton number. Tr is a negative-definite quadratic form on \mathfrak{g} such that $\text{Tr} x^2 = -2$ for x a short coroot. (The notation is motivated by the fact that for $G = SU(N)$, Tr is the trace in the N -dimensional

representation.) The normalization ensures that if M is a closed four-manifold without surface operators, and G is simply-connected, then N is integer-valued. For example, if $G = SU(N)$, then $N = -\int_M c_2(E)$, where c_2 is the second Chern class. When N is integer-valued, there is a symmetry $T : \theta \rightarrow \theta + 2\pi$, or $T : \tau \rightarrow \tau + 1$. This expresses the familiar fact that quantum field theory depends on the action \mathbf{I} only modulo $2\pi i\mathbb{Z}$.

If G is not simply-connected, then N takes values in $\frac{1}{k}\mathbb{Z}$ for some integer k , and instead of T we consider the symmetry $T^k : \tau \rightarrow \tau + k$.

Assuming for notational simplicity that G is simply-connected, let us study the symmetry T in the presence of a surface operator supported on a two-manifold D . First, in the presence of the singularity associated with the surface operator, we have to define precisely what we mean by the integral defining N . The integral has a bulk contribution, coming from the integration over the complement of D :

$$(2.40) \quad N_0 = -\frac{1}{8\pi^2} \int'_M \text{Tr } F \wedge F.$$

The symbol \int'_M means that we integrate over the complement of D , ignoring possible delta function contributions at D . According to (2.5), once we pick an extension of the bundle E over D , there is also a delta function contribution to the integral. This contribution is

$$(2.41) \quad N' = -\frac{1}{2\pi} \int_D \text{Tr } \alpha F - \frac{1}{2} D \cap D \text{Tr } \alpha^2.$$

To obtain (2.41) (which corresponds to Proposition 5.7 in [11]), we use the fact that the delta function contribution to F near D is $2\pi\alpha\delta_D$. We have also used (2.8).

The sum

$$(2.42) \quad N = N_0 + N' = -\frac{1}{8\pi^2} \int'_M \text{Tr } F \wedge F - \frac{1}{2\pi} \int_D \text{Tr } \alpha F - \frac{1}{2} D \cap D \text{Tr } \alpha^2$$

is integer-valued. However, N is not natural in the sense that it depends on a choice of lifting of α from $\mathfrak{t}/\Lambda_{\text{cochar}}$, where it naturally takes values, to \mathfrak{t} . There is a simple reason for this; the integer-valued invariant $\int_M c_2(E)$ (or its analog for groups other than $SU(N)$) is not determined by the restriction of E to $M \setminus D$, but depends on the choice of an extension of E over D . Integrality of N , however, implies that upon reduction mod \mathbb{Z} , we get

$$(2.43) \quad N_0 = -N' = \frac{1}{2\pi} \int_D \text{Tr } \alpha F + \frac{1}{2} D \cap D \text{Tr } \alpha^2 \pmod{\mathbb{Z}}.$$

Since N_0 does not depend on a lifting of α , the same must be true mod \mathbb{Z} of the right hand side of (2.43). We can verify this as follows. The restriction of F to D is \mathfrak{t} -valued, and its integral $\mathfrak{m} = \int_D F/2\pi$

is a “magnetic charge,” an element of the lattice Λ_{cochar} . This lattice actually coincides with the root lattice of ${}^L G$ (since we are assuming G to be simply-connected) or in other words the coroot lattice of G . Because of the way the trace was normalized, the bilinear function $\mathfrak{m}, \mathfrak{m}' \rightarrow \text{Tr } \mathfrak{m} \mathfrak{m}'$ takes integer values for $\mathfrak{m}, \mathfrak{m}'$ in this lattice, and takes even integer values if $\mathfrak{m} = \mathfrak{m}'$. Once an extension of E is picked, α takes values in \mathfrak{t} , which is the same as $\Lambda_{\text{cochar}} \otimes_{\mathbb{Z}} \mathbb{R}$. So we can write

$$(2.44) \quad \mathbf{N}' = -\text{Tr } \alpha \mathfrak{m} - \frac{1}{2} D \cap D \text{Tr } \alpha^2$$

and hence

$$(2.45) \quad \mathbf{N}_0 = \text{Tr } \alpha \mathfrak{m} + \frac{1}{2} D \cap D \text{Tr } \alpha^2 \pmod{\mathbb{Z}}.$$

This statement is invariant under shifts of α by a lattice vector, given the integrality properties of the trace, together with (2.10).

We want to define \mathbf{I}_θ to be independent of any choice of extension of the bundle. The only obvious way to do this is to omit the delta function contribution from D , and set

$$(2.46) \quad \mathbf{I}_\theta = -i\theta \mathbf{N}_0.$$

This does not mean that we will ignore the delta function contribution. Rather, we will have such a contribution from the two-dimensional theta-like parameter η that was introduced in section 2.3. In the same notation, we take the contribution of η to the action to be

$$(2.47) \quad \mathbf{I}_\eta = -2\pi i \text{Tr } \eta \mathfrak{m} - \pi i D \cap D \text{Tr } \alpha \eta.$$

The term $\text{Tr } \eta \mathfrak{m}$ is the expected term for the theta-like angles η . To this, we have added a c -number term that depends only on α and η and not on any of the field variables of the theory.

The sum of the two contributions to the action is therefore

$$(2.48) \quad \widehat{\mathbf{I}} = \mathbf{I}_\theta + \mathbf{I}_\eta = -i\theta \mathbf{N}_0 - 2\pi i \text{Tr } \eta \mathfrak{m} - \pi i D \cap D \text{Tr } \alpha \eta.$$

At this stage, we can more fully justify our definition of \mathbf{I}_θ . Adding to \mathbf{I}_θ a multiple of $\text{Tr } \alpha \mathfrak{m}$ would have no essential effect, since one can compensate for this by shifting η . So we may as well define \mathbf{I}_θ as we have.

We will now see that the choice we have made leads to a simple result for how α and η must transform under $\theta \rightarrow \theta + 2\pi$. The change in \mathbf{I}_θ is

$$(2.49) \quad \Delta \mathbf{I}_\theta = -2\pi i \mathbf{N}_0 = -2\pi i \text{Tr } \alpha \mathfrak{m} - \pi i D \cap D \text{Tr } \alpha^2 \pmod{2\pi i \mathbb{Z}},$$

where (2.45) has been used.

To get a symmetry of the theory, $\widehat{\mathbf{I}}$ must be invariant mod $2\pi i\mathbb{Z}$. For this, we let $T : \theta \rightarrow \theta + 2\pi$ act on (α, η) by

$$(2.50) \quad \begin{aligned} \eta &\rightarrow \eta - \alpha \\ \alpha &\rightarrow \alpha. \end{aligned}$$

The variation of \mathbf{I}_η then precisely cancels the variation of \mathbf{I}_θ mod $2\pi i\mathbb{Z}$, and $\widehat{\mathbf{I}}$ is invariant.

The constant term $-\pi i D \cap D \operatorname{Tr} \alpha \eta$ that we included in the action is not invariant mod $2\pi i$ under lattice shifts of α or η . However, the non-invariance is, like this term itself, independent of the quantum fields. Geometrically, this means that when $D \cap D$ is non-zero, the partition function is not a complex-valued function of α and η but a section of a complex line bundle over $\mathbb{T} \times {}^L\mathbb{T}$. Of course, it would be possible to omit the c -number term from the action and instead claim that the symmetry $T : \tau \rightarrow \tau + 1$ holds up to a c -number.

This discussion generalizes straightforwardly to the case that G is not simply-connected and the instanton number takes values in $\frac{1}{k}\mathbb{Z}$. One considers the symmetry $T^k : \tau \rightarrow \tau + k$, and the same derivation shows that it acts as

$$(2.51) \quad \begin{aligned} \eta &\rightarrow \eta - k\alpha \\ \alpha &\rightarrow \alpha. \end{aligned}$$

Let us combine the result (2.50) with our previous result (2.25) for the action of electric-magnetic duality. First we consider the case that G is selfdual and simply-laced, so that in particular the duality group is simply $\Gamma = SL(2, \mathbb{Z})$. The only simple Lie group that actually has these properties is E_8 . In this situation, Γ is generated by the elements

$$(2.52) \quad S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

The formulas (2.25) and (2.50) tell us that for \mathbf{M} equal to S or T , α and η transform under \mathbf{M} by

$$(2.53) \quad (\alpha, \eta) \rightarrow (\alpha, \eta)\mathbf{M}^{-1}.$$

This is therefore true for any $\mathbf{M} \in SL(2, \mathbb{Z})$. In particular, the pair (α, η) transform naturally under $SL(2, \mathbb{Z})$, and our results for the action of S and T , though motivated independently, are compatible with each other.

The generalization to any simple G is as follows. First recall that electric charge takes values in the character lattice¹² Λ_{char} of G , and

¹²In much of the physics literature, root and coroot lattices are taken here, because the theory is considered only on $\mathbb{R}^{3,1}$ or \mathbb{R}^4 , where the physical electric and magnetic charges take values in the root and coroot lattices. However, the fact that the theory could be probed with Wilson and 't Hooft operators as external charges shows that it must be possible to refine the usual discussion of duality to the case that

magnetic charge takes values in the cocharacter lattice Λ_{cochar} . So the full set of charges takes values in the lattice $\widehat{\Lambda} = \Lambda_{\text{cochar}} \oplus \Lambda_{\text{char}}$, which has a natural non-degenerate skew pairing since Λ_{char} and Λ_{cochar} are dual lattices. The duality group Γ acts linearly on this lattice, preserving the skew pairing, as well as acting naturally on $\tau = \theta/2\pi + 4\pi i/g^2$. For example, for E_8 we have $\widehat{\Lambda} = \Lambda_{\text{char}} \otimes \mathbb{Z}^2$, where \mathbb{Z}^2 is a rank two lattice, and $\Gamma = SL(2, \mathbb{Z})$ acts on $\widehat{\Lambda}$ via its natural action on \mathbb{Z}^2 . At any rate, Γ always acts on the full set of electric and magnetic charges, and thus on $\widehat{\Lambda}$. The details are a little complex, however, especially [41, 42] if G is not simply-laced.

At any rate, the linear action of Γ on $\widehat{\Lambda}$ induces an action on $\widehat{\Lambda} \otimes_{\mathbb{Z}} \mathbb{R}$, and hence on $(\widehat{\Lambda} \otimes_{\mathbb{Z}} \mathbb{R})/\widehat{\Lambda}$. But $(\widehat{\Lambda} \otimes_{\mathbb{Z}} \mathbb{R})/\widehat{\Lambda}$ is the same as $\mathbb{T} \times {}^L\mathbb{T}$. So the action of Γ on the charges determines an action on $\mathbb{T} \times {}^L\mathbb{T}$, where the pair (α, η) take values. Thus it determines a natural action on (α, η) . For any G , the meaning of (2.25) and (2.50) is that the action of Γ on (α, η) is precisely the natural action determined by its action on the electric and magnetic charges. The hypothesis (2.25) asserts that this is true for S , and it is true for $T : \tau \rightarrow \tau + 1$ since the derivation of (2.50) was actually a close cousin of the computation [46] of the action of $T : \theta \rightarrow \theta + 2\pi$ on the charges.

The Dirac String

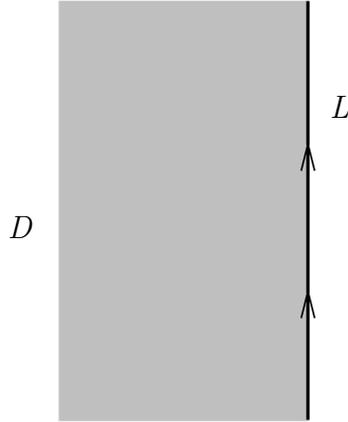


FIGURE 1. A surface operator whose support D has a boundary L , which turns out to be the world-line of a magnetic monopole or dyon.

the charges take values in the character and cocharacter lattices. This refinement is the one relevant here, roughly because the same topological issues arise either by allowing Wilson and 't Hooft operators, working on a general four-manifold, or admitting surface operators such as those considered here.

Now, leaving some details to the reader, and making use of some ideas in [14], we are going to explain an informal interpretation of this result.

Let us ask whether the support D of a surface operator can have a boundary L as in figure 1. A little thought will show that for this to occur, L must be the world-line of a magnetic monopole with magnetic charge α . The gauge field holonomy around D must unwind at the boundary L , and this unwinding of the holonomy characterizes magnetic charge.

However, we have not imposed Dirac quantization on the magnetic charge α . As we consider it to be defined modulo a lattice vector, we are really interested in the case that Dirac quantization is *not* obeyed. According to Dirac, in this case, for gauge-invariance, L must be the boundary of the world-volume of what is commonly called a “Dirac string.” In our context, the surface D can be regarded as the world-volume of the string. Thus, our surface operator can be regarded as representing the Dirac string associated with improperly quantized magnetic charge. Dirac defined the Dirac string in terms of the monodromy of the gauge field around it, so our surface operator indeed has the right property to be the world-volume of a Dirac string.

More generally, if $\eta \neq 0$, the monopole at the end of the string also carries electric charge. It is thus a dyon.

Our claim that (α, η) transform under S -duality just like magnetic and electric charge can be interpreted as a statement that Dirac strings associated with improperly quantized charges transform under duality just like properly quantized charges.

We conclude with a related comment, also anticipated in [14]. Physically, a surface operator might arise if $\mathcal{N} = 4$ super Yang-Mills theory (or any gauge theory of interest, such as the Standard Model) is embedded in some more complete theory that reduces to it at low energies. The embedding in a more complete theory might give rise to what in other contexts are called cosmic strings. Suppose that such a string is heavy enough that we can consider it to be frozen in position, with a known orbit in spacetime. It is then appropriate to consider the “low energy” $\mathcal{N} = 4$ dynamics in the presence of the string. This will involve studying the $\mathcal{N} = 4$ theory coupled to a surface operator of some kind. The details depend on the particular type of cosmic string considered. Strings that produce an Aharonov-Bohm effect, as first explored in [10] will lead to surface operators of the sort considered in the present paper. In section 6.4, we will consider some surface operators defined by embedding $\mathcal{N} = 4$ super Yang-Mills theory in a more complete theory (with the more complete theory being Type IIB superstring theory). In the theory of cosmic strings, it is familiar that some kinds of string can break, terminating on magnetic monopoles [47].

An Illustration

We will now explain in more detail how S -duality acts on the charges in the case of $G = SU(2)$, ${}^L G = SO(3)$. This will enable us to spell out a few points that have been hidden in our analysis above. Following [48] we explain two different ways to view the problem. This discussion is not used in the rest of the paper, and we will be rather brief on some points.

$SU(2)$ is simply-connected, so for $SU(2)$ gauge theory, the instanton number is integer-valued and there is a symmetry $T : \tau \rightarrow \tau + 1$, acting as in (2.50). On the other hand, for $SO(3)$, the instanton number takes values in $\frac{1}{4}\mathbb{Z}$, or in $\frac{1}{2}\mathbb{Z}$ if M is a spin manifold. So the basic shift symmetry of the theta-angle is $T^4 : \tau \rightarrow \tau + 4$, or $T^2 : \tau \rightarrow \tau + 2$ if M is spin. These symmetries act as in eqrefzimax. In addition, there is the symmetry $S : \tau \rightarrow -1/\tau$, which exchanges $SU(2)$ and $SO(3)$. The group of duality symmetries of $SU(2)$ gauge theory is generated if M is not spin by

$$(2.54) \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad ST^4S^{-1} = \begin{pmatrix} 1 & 0 \\ -4 & 1 \end{pmatrix}.$$

(ST^4S^{-1} is a duality symmetry for $SU(2)$, since S^{-1} maps $SU(2)$ to $SO(3)$, T^4 is a duality transformation of $SO(3)$, and S maps back to $SU(2)$.) This duality group, which is known as $\Gamma_0(4)$, is a congruence subgroup of $SL(2, \mathbb{Z})$. It is a group of duality symmetries of $SU(2)$ gauge theory on a general M , and acts on α and η according to (2.53). The group of duality symmetries of $SO(3)$ gauge theory is likewise a copy of $\Gamma_0(4)$, generated by STS^{-1} and T^4 . And finally, of course, we have the symmetry S that exchanges the two groups. If M is spin, one can replace 4 by 2 everywhere, and the duality groups are isomorphic to $\Gamma_0(2)$.

$G = SU(2)$ and ${}^L G = SO(3)$ have the same Lie algebra, so we can think of $\mathbb{T} = \mathfrak{t}/\Lambda_{\text{rt}}$ and ${}^L\mathbb{T} = \mathfrak{t}/\Lambda_{\text{wt}}$ as quotients of the same space by different lattices, which are respectively the root and weight lattices Λ_{rt} and Λ_{wt} . Λ_{rt} is of index two in Λ_{wt} . The identity map on \mathfrak{t} therefore projects to a natural, two-to-one map from \mathbb{T} , where α takes values, to ${}^L\mathbb{T}$, where η takes values. This map is implicit in the transformation $T : (\alpha, \eta) \rightarrow (\alpha, \eta + \alpha)$. The identity map on \mathfrak{t} does not project to a natural map of ${}^L\mathbb{T}$ to \mathbb{T} , but the map of multiplication by 2 (or any even integer) does so project. So there is a natural map $ST^2S^{-1} : (\alpha, \eta) \rightarrow (\alpha - 2\eta, \eta)$, which is realized as a duality transformation if M is spin. Similarly the operation $ST^4S^{-1} : (\alpha, \eta) \rightarrow (\alpha - 4\eta, \eta)$, which appears as a duality transformation for any M , is naturally defined.

This way of describing things is in some tension with the physics literature, where the duality group is generally considered to be $SL(2, \mathbb{Z})$ for a simply-laced group such as $SU(2)$. However, the usual analysis

is made only for $M = \mathbb{R}^4$, and without detailed consideration of Wilson and 't Hooft operators. Under these circumstances, the $SO(3)$ and $SU(2)$ gauge theories coincide. One can distinguish the two theories on \mathbb{R}^4 by adding a Wilson operator in the two-dimensional representation to specialize to $SU(2)$, or an 't Hooft operator with minimal charge to specialize to $SO(3)$. (One cannot add operators of both types simultaneously, as they are not mutually local.) If the $SU(2)$ and $SO(3)$ theories are elaborated in this way, the appropriate duality groups on \mathbb{R}^4 are precisely those that we have just described on a general spin manifold.

Actually, there is another formulation of the problem, of which we will only give an outline, that really gives more information, and in which the full group $SL(2, \mathbb{Z})$ plays a role. In this approach, we always take the gauge group to be the adjoint group $SO(3)$ (not its cover $SU(2)$), but we specify the second Stiefel-Whitney class $\xi = w_2(E)$ of the bundle E . Summing over all $SO(3)$ -bundles with fixed ξ , we define a partition function¹³ $Z_\xi(\tau)$ for each ξ . One then shows [48] that the $Z_\xi(\tau)$ transform in a unitary representation of $SL(2, \mathbb{Z})$.

To incorporate in this approach a surface operator with parameters α, η , supported on a surface D , one defines α and η to take values in \mathbb{T} , the maximal torus of the simply-connected group $SU(2)$. Then we have for each ξ a partition function $Z_\xi(\tau; \alpha, \eta)$, and the claim is that this family of partition functions transforms as a representation of $SL(2, \mathbb{Z})$ (which acts on τ in the usual fashion, on ξ as described in [48], and on α and η by the natural action described in (2.53)).

The description by parameters ξ, α, η is slightly redundant. To describe the redundancy, we use a multiplicative notation for the maximal torus \mathbb{T} (where α and η take values) and recall that this torus contains the element -1 of the center of $SU(2)$. Then shifting α to $-\alpha$ is equivalent to replacing ξ with $\xi + [D]$ according to (2.6), and shifting η to $-\eta$ multiplies the partition function $Z_\xi(\tau; \alpha, \eta)$ by $(-1)^{(\xi, D)}$ (this is essentially explained at the end of section 4.2). This redundancy is compatible with the action of $SL(2, \mathbb{Z})$, but it seems hard to eliminate it without making the action of $SL(2, \mathbb{Z})$ look less natural.

2.6. Levi Subgroups And More General Surface Operators.

Now we will describe a simple but important generalization of our definition of surface operators in which the maximal torus \mathbb{T} is replaced by a more general subgroup \mathbb{L} of G that contains \mathbb{T} . We assume also that \mathbb{L} can be characterized as the subgroup of G that commutes with some $\alpha \in \mathfrak{t}$. Such an \mathbb{L} is called a subgroup of G of Levi type. We consider

¹³Along with a full set of operators, correlation functions, quantum states, branes, etc., making up a quantum field theory. In a fuller description, the whole quantum field theory, not just the set of partition functions, is transformed by $SL(2, \mathbb{Z})$.

two such groups to be equivalent if they are conjugate in G . The usual case is that $\alpha \neq 0$ and \mathbb{L} is a proper subgroup of G , but we also will consider the case $\alpha = 0$ and $\mathbb{L} = G$. Any Levi subgroup contains \mathbb{T} , so \mathbb{T} is a minimal Levi subgroup. A typical example of a non-minimal Levi subgroup is the subgroup of $SU(3)$ of the form

$$(2.55) \quad \begin{pmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & * \end{pmatrix}.$$

This is isomorphic to $U(2)$, and is actually a next-to-minimal Levi subgroup, since a smaller Levi subgroup would have to be \mathbb{T} itself. In general, for any G of rank r , a next-to-minimal Levi subgroup is isomorphic to $SU(2) \times U(1)^{r-1}$ or a quotient of this by \mathbb{Z}_2 .

We now will define what we will call a surface operator of type \mathbb{L} . In this language, the surface operator that we defined originally is an operator of type \mathbb{T} ; we will also call it the generic surface operator. To define a surface operator of type \mathbb{L} , we consider a two-manifold D with a gauge theory singularity labeled classically by the usual parameters (α, β, γ) . But now we require the parameters to be \mathbb{L} -invariant. Moreover, when we perform the quantum path integral, we divide by gauge transformations that are \mathbb{L} -valued when restricted to D , not just those that are \mathbb{T} -valued. It is because of this last step that the surface operator of type \mathbb{L} is not just a special case – for \mathbb{L} -invariant parameters – of the surface operator that we defined originally, the operator of type \mathbb{T} in which (α, β, γ) are simply \mathfrak{t} -valued. The surface operator of type \mathbb{L} is something new, associated with a different path integral. Actually, we will learn in section 3.6 that if \mathbb{L} contains \mathbb{T} as a proper subgroup, then the surface operator of type \mathbb{T} becomes singular when the parameters become \mathbb{L} -invariant. So for such special parameters, the operator that makes sense is the new surface operator of type \mathbb{L} .

At the quantum level, there is an additional parameter η , the two-dimensional theta-angle η of section 2.3. In the present case, with the group \mathbb{T} replaced by \mathbb{L} , η takes values not in ${}^L\mathbb{T}$ but in a subgroup. In fact, we want to define η as a theta-angle of the abelian part of \mathbb{L} . To give a convenient description of where η takes values, it is useful to observe that Levi subgroups $\mathbb{L} \subset G$ are in natural correspondence with Levi subgroups ${}^L\mathbb{L} \subset {}^L G$. One way to describe the correspondence is to make use of the fact that the Weyl group \mathcal{W} of G naturally coincides with the Weyl group ${}^L\mathcal{W}$ of ${}^L G$. (See Appendix A.) Moreover, the Weyl groups of \mathbb{L} and ${}^L\mathbb{L}$ are subgroups of \mathcal{W} and ${}^L\mathcal{W}$. The correspondence between \mathbb{L} and ${}^L\mathbb{L}$ is simply that they have the same Weyl group, which we will denote as $\mathcal{W}_{\mathbb{L}}$. Another way to state the correspondence between \mathbb{L} and ${}^L\mathbb{L}$ is to say that the coroots of ${}^L\mathbb{L}$ (which are a subset of the coroots of ${}^L G$) are multiples of the coroots of \mathbb{L} – with different multiples for long

or short coroots. With all this understood, a surface operator of type \mathbb{L} depends on parameters $(\alpha, \beta, \gamma, \eta)$ which take values in the $\mathcal{W}_{\mathbb{L}}$ -invariant part of $\mathbb{T} \times \mathfrak{t} \times \mathfrak{t} \times {}^L\mathbb{T}$. We are using here the fact that \mathbb{L} -invariance of (α, β, γ) is equivalent to $\mathcal{W}_{\mathbb{L}}$ -invariance. The formulation in terms of $\mathcal{W}_{\mathbb{L}}$ -invariance has an important advantage: it enables us to treat η on the same footing as the other variables. This is more satisfactory than saying that (α, β, γ) are \mathbb{L} -invariant and η is ${}^L\mathbb{L}$ -invariant. The condition that η must be $\mathcal{W}_{\mathbb{L}}$ -invariant is the right one, because it means that η is a character of the abelian magnetic fluxes of the gauge bundle E restricted to D ; the structure group of this bundle is \mathbb{L} , and its characteristic classes are $\mathcal{W}_{\mathbb{L}}$ -invariant.

We extend the duality conjecture of section 2.4 to say that a surface operator of type \mathbb{L} maps under duality to a surface operator of type ${}^L\mathbb{L}$, with the parameters transforming in the familiar fashion $(\alpha, \beta, \gamma, \eta) \rightarrow (\eta, \beta, \gamma, -\alpha)$.

Some More Group Theory

Now we describe some more group theory that will be useful in the rest of the paper.

Levi subgroups are closely related to what are called parabolic subgroups of $G_{\mathbb{C}}$. Let us pick a particular $\alpha \in \mathfrak{t}$ which commutes precisely with \mathbb{L} . We say that such an α is \mathbb{L} -regular (and if $\mathbb{L} = \mathbb{T}$, we simply say that α is regular). We let \mathcal{P} be the subgroup of $G_{\mathbb{C}}$ whose Lie algebra \mathfrak{p} is spanned by elements $\psi \in \mathfrak{g}$ that obey

$$(2.56) \quad [\alpha, \psi] = i\lambda\psi, \quad \lambda \geq 0.$$

A group of this form is called a parabolic subgroup of $G_{\mathbb{C}}$. If in (2.56) we replace the condition $\lambda \geq 0$ by $\lambda > 0$, we get the Lie algebra \mathfrak{n} of a subgroup $\mathcal{N} \subset \mathcal{P}$ that is known as the unipotent radical of \mathcal{P} .

For example, for $G_{\mathbb{C}} = SL(3, \mathbb{C})$ and $\mathbb{L} = \mathbb{T}$, we can take $\alpha = i \operatorname{diag}(y_1, y_2, y_3)$ with the y_i real and $y_1 > y_2 > y_3$. In this case, the parabolic group we get is the group of upper triangular matrices

$$(2.57) \quad \begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \end{pmatrix},$$

and is called a Borel subgroup \mathcal{B} . In this example, the unipotent radical \mathcal{N} consists of matrices of this form:

$$(2.58) \quad \begin{pmatrix} 1 & * & * \\ 0 & 1 & * \\ 0 & 0 & 1 \end{pmatrix},$$

A different choice of \mathbb{T} -regular (or simply regular) α would be obtained from the one we used by permuting the eigenvalues by a Weyl

transformation. This leads to a Weyl-conjugate Borel subgroup \mathcal{B}' . More generally, for any G , the parabolic subgroup associated with a pair (\mathbb{T}, α) is called a Borel subgroup \mathcal{B} , and is unique up to a Weyl transformation. \mathcal{B} is a minimal parabolic subgroup, just as \mathbb{T} is a minimal Levi subgroup.

A non-minimal Levi subgroup \mathbb{L} may be associated with several inequivalent parabolic subgroups. For instance, in the example of (2.55), we can take

$$(2.59) \quad \alpha = iy \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix},$$

with real nonzero y . If y is positive, we get the parabolic subgroup \mathcal{P} of matrices of the form

$$(2.60) \quad \begin{pmatrix} * & * & * \\ * & * & * \\ 0 & 0 & * \end{pmatrix},$$

but if y is negative, we get an inequivalent parabolic subgroup \mathcal{P}' of matrices of the form

$$(2.61) \quad \begin{pmatrix} * & * & 0 \\ * & * & 0 \\ * & * & * \end{pmatrix}.$$

The unipotent radicals \mathcal{N} and \mathcal{N}' consist of matrices of the form

$$(2.62) \quad \begin{pmatrix} 1 & 0 & * \\ 0 & 1 & * \\ 0 & 0 & 1 \end{pmatrix},$$

or

$$(2.63) \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ * & * & 1 \end{pmatrix}.$$

Maximal Levi Subgroup

If $\alpha = 0$, we get a special case of the above construction in which $\mathbb{L} = G$, $\mathcal{P} = G_{\mathbb{C}}$, and $\mathcal{N} = 1$. It is most convenient to allow this special case and to regard G itself as a maximal Levi subgroup.

In the above, we have been a little imprecise about whether α and η are Lie algebra-valued or torus-valued. The torus-valued case leads to a more general construction, since the space of $\mathcal{W}_{\mathbb{L}}$ -invariant α or η may not be connected. We will spell this out for the case $\mathbb{L} = G$. The condition for α to be G -invariant says that it takes values in the

center of G , which we denote $\mathcal{Z}(G)$. Thus, what it means to have a surface operator of type G supported on a surface D is that the gauge field is defined on $M \setminus D$ and has monodromy around D labeled by a prescribed element $\alpha \in \mathcal{Z}(G)$. The dual of the choice of α is that a G -bundle on D is classified topologically by a characteristic class $\xi \in H^2(D, \pi_1(G)) \cong \pi_1(G)$, and the path integral can be weighted by the choice of a discrete η -angle valued in $\text{Hom}(\pi_1(G), U(1))$. The finite groups $\mathcal{Z}(G)$ and $\text{Hom}(\pi_1(G), U(1))$ are exchanged by duality, and the natural duality conjecture states that the discrete versions of α and η are exchanged in the usual way.

Though our duality conjecture for surface operators of type \mathbb{L} is meant to include the disconnected components, we will not consider them in any detail in this paper.

3. More On The Classical Geometry

In this section, we will reconsider the construction of section 2.2 and describe aspects that require understanding the classical geometry of \mathcal{M}_H in more depth. We aim for a relatively simple – but certainly not mathematically complete – introduction to aspects that are or may become useful for the geometric Langlands program. Most of the topics we consider have been treated much more fully in the mathematical literature, but a few points may be new, notably the action of the affine Weyl group on the cohomology of \mathcal{M}_H (section 3.7), the linearity of the cohomology classes of the symplectic forms (section 3.2), and the nature of the local singularities of \mathcal{M}_H at non-regular points (section 3.6).

Many of the most significant phenomena that we will describe have local models involving hyper-Kähler metrics on complex coadjoint orbits, constructed in [34]–[37]. We defer this to section 3.8.

3.1. Hyper-Kähler Quotient. One of the most basic properties of the moduli space $\mathcal{M}_H(G, C)$ of Higgs bundles on a Riemann surface C is that it can be constructed as a hyper-Kähler quotient [21]. We recall first the construction in the absence of singularities. Picking a smooth G -bundle $E \rightarrow C$, we denote as \mathcal{W} the space of pairs (A, ϕ) , with A a connection on E and $\phi \in \Omega^1(C, \text{ad}(E))$. \mathcal{W} is a flat hyper-Kähler manifold; for a detailed account see section 4.1 of [8]. Let \mathcal{G} be the group of gauge transformations of E . Then \mathcal{G} acts on \mathcal{W} , preserving the hyper-Kähler structure, with hyper-Kähler moment map $\vec{\mu} = (\mu_I, \mu_J, \mu_K)$:

$$(3.1) \quad \begin{aligned} \mu_I &= -\frac{1}{2\pi} \int_C \text{Tr } \epsilon(F - \phi \wedge \phi), \\ \mu_J &= -\frac{1}{2\pi} \int_C |d^2 z| \text{Tr } \epsilon(D_z \phi_{\bar{z}} + D_{\bar{z}} \phi_z), \\ \mu_K &= -\frac{i}{2\pi} \int_C |d^2 z| \text{Tr } \epsilon(D_z \phi_{\bar{z}} - D_{\bar{z}} \phi_z). \end{aligned}$$

Here $\epsilon \in \Omega^0(C, \text{ad}(E))$ is an element of the Lie algebra of \mathcal{G} ; $\vec{\mu}$ is linear in ϵ . Given a hyper-Kähler manifold \mathcal{W} and a group \mathcal{G} that acts on \mathcal{W} , preserving the hyper-Kähler structure, with moment map $\vec{\mu}$, the hyper-Kähler quotient $\mathcal{W}///\mathcal{G}$ is defined in general [49] as $\vec{\mu}^{-1}(0)/\mathcal{G}$. This hyper-Kähler quotient is a hyper-Kähler manifold. In the present case, the equations $\vec{\mu} = 0$ (for all ϵ) are precisely Hitchin's equations (2.1). The hyper-Kähler quotient $\mathcal{W}///\mathcal{G}$ is hence precisely the moduli space \mathcal{M}_H of solutions of Hitchin's equations.

This story can be repeated for Higgs bundles with an isolated singularity of the familiar sort [23], [24]. We pick a point $p \in C$, and a reduction of the structure group of the bundle E at p to a torus \mathbb{T} , and we denote as \mathcal{G}_p the subgroup of \mathcal{G} consisting of gauge transformations whose restriction to p lies in \mathbb{T} . We also pick a trio¹⁴ $(\alpha, \beta, \gamma) \in \mathfrak{t}$, the Lie algebra of \mathbb{T} . Picking coordinates r, θ near p , we denote as $\mathcal{W}(\alpha, \beta, \gamma; p)$ the space of pairs (A, ϕ) with the familiar sort of singularity near p :

$$(3.2) \quad \begin{aligned} A &= \alpha d\theta + \dots \\ \phi &= \beta \frac{dr}{r} - \gamma d\theta + \dots \end{aligned}$$

The ellipses refer to terms less singular than $1/r$. And we denote as \mathcal{G}_p the group of gauge transformations of the bundle E which at p take values in \mathbb{T} . These are the gauge transformations that preserve the condition (3.2). Then the moduli space $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ is the hyper-Kähler quotient $\mathcal{W}(\alpha, \beta, \gamma; p)///\mathcal{G}_p$. We call this the moduli space of Higgs bundles with ramification at the point p . One can similarly construct as a hyper-Kähler quotient the analogous moduli space of Higgs bundles with ramification at several points, that is, the moduli space of solutions of Hitchin's equations with singularities at several points $p_1, \dots, p_s \in C$, labeled by parameters $(\alpha_i, \beta_i, \gamma_i) \in \mathfrak{t}$, $i = 1, \dots, s$. To keep the notation simple, we formulate the present section mainly for the case of ramification at one point, but all statements have direct analogs for the more general case.

A hyper-Kähler manifold, with complex structures I, J , and K , has corresponding Kähler forms ω_I, ω_J , and ω_K . In the present case, these

¹⁴As explained in section 2.1, \mathcal{M}_H really depends on α only via its image in $\mathfrak{t}/\Lambda_{\text{cochar}} = \mathbb{T}$.

can be written, just as in the absence of singularities,¹⁵

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

Just as in the absence of singularities (see section 4.1 of [8]), ω_I is cohomologous to

$$(3.4) \quad \omega'_I = -\frac{i}{2\pi} \int_C |d^2z| \operatorname{Tr} (\delta A_{\bar{z}} \wedge \delta A_z) = -\frac{1}{4\pi} \int_C \operatorname{Tr} (\delta A \wedge \delta A),$$

which depends only on A and so is a pullback from the moduli space \mathcal{M} of G -bundles. Indeed, let $\lambda_I = \frac{1}{4\pi} \int_C \operatorname{Tr} \phi \wedge \delta \phi$. λ_I is gauge-invariant, and vanishes (after integration by parts and use of Hitchin's equations) if contracted with a generator of gauge-transformations, $\delta A = -d_A \epsilon$, $\delta \phi = [\epsilon, \phi]$. So it is the pullback of a one-form on \mathcal{M}_H , and the formula $\omega_I - \omega'_I = \delta \lambda_I$ shows that ω_I and ω'_I are cohomologous.

In the absence of singularities, the cohomology classes of ω_J and ω_K vanish. For example, one proves this for ω_K by writing $\omega_K = \delta \lambda_K$, where $\lambda_K = \frac{1}{2\pi} \int_C \operatorname{Tr} \phi \wedge \delta A$. This still works in the presence of singularities as long as $\gamma = 0$, but for $\gamma \neq 0$, λ_K does not vanish if contracted with the generator of a gauge transformation so is not the pullback of a one-form on \mathcal{M}_H . In fact, contraction with the generator of a gauge transformation maps ω_K to $\frac{1}{2\pi} \int_C \operatorname{Tr} \phi \wedge (-d_A \epsilon) = \operatorname{Tr} \gamma \epsilon(p)$, where we have integrated by parts and used the fact that $d_A \phi = -2\pi \gamma \delta_p$. So the cohomology class of ω_K is nonzero for $\gamma \neq 0$. Similarly, the cohomology class of ω_J is nonzero for $\beta \neq 0$. We will describe all of these cohomology classes in section 3.6.

Generalization Of Type \mathbb{L}

We can readily extend this to incorporate a general Levi subgroup \mathbb{L} .

We restrict the parameters (α, β, γ) to be \mathbb{L} -invariant. We define \mathcal{G}_p to be the group of gauge transformations whose restriction to p lies in \mathbb{L} (rather than in \mathbb{T} , as above). The hyper-Kähler quotient $\mathcal{W}(\alpha, \beta, \gamma; p) // \mathcal{G}_p$, carried out exactly as above, now gives us what we will call $\mathcal{M}_{H, \mathbb{L}}$, the moduli space of Higgs bundles with ramified structure

¹⁵As in [8], we write δ for the exterior derivative on \mathcal{W} , and d for the exterior derivative on C .

of type \mathbb{L} . See [23] for a rigorous construction. When we do not indicate the ramified type explicitly, this means that we are taking $\mathbb{L} = \mathbb{T}$.

One has to be careful in defining $\mathcal{M}_{H,\mathbb{L}}$ and require that the solution deviate from the asymptotic form determined by (α, β, γ) by terms that are small compared to $1/r \ln r$, not just small compared to $1/r$. The purpose of this is to avoid the type of asymptotic behavior shown in eqn. (3.101) below.

The reason that we have introduced a special notation for $\mathcal{M}_{H,\mathbb{L}}(\alpha, \beta, \gamma; p)$ is that, as will hopefully be clear in section 3.6, it is not the limit of $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ (defined for a regular triple (α, β, γ)) as (α, β, γ) approach \mathbb{L} -invariant values. The two spaces do not even have the same dimension. Rather, their relation turns out to be that, when the trio (α, β, γ) becomes \mathbb{L} -invariant, $\mathcal{M}_{H,\mathbb{L}}$ is a locus of singularities of \mathcal{M}_H .

A special case of this definition is the case $\mathbb{L} = G$. Ramified structure of type G can be no ramification at all, since the group of gauge transformations that at p take values in G is simply the group of all gauge transformations. However, as discussed at the end of section (2.6), if G has a non-trivial center \mathcal{Z} , our definitions lead to a slight generalization. Indeed, the triple $(\alpha, \beta, \gamma) \in \mathbb{T} \times \mathfrak{t} \times \mathfrak{t}$ is G -invariant precisely if $\beta = \gamma = 0$ and α is an element of \mathcal{Z} . Thus $\mathcal{M}_{H,G}$ is a union of components labeled by \mathcal{Z} . One of these components (corresponding to the identity element of \mathcal{Z}) is the moduli space \mathcal{M}_H of ordinary Higgs bundles with no singularity at p .

The opposite extreme is $\mathbb{L} = \mathbb{T}$. In this case, $\mathcal{M}_{H,\mathbb{L}}(\alpha, \beta, \gamma; p)$ is what we usually denote simply as $\mathcal{M}_H(\alpha, \beta, \gamma; p)$. This is the generic case that we have in mind when we speak of ramified Higgs bundles without specifying \mathbb{L} .

Our point of view is that a surface operator supported on a surface D is defined by the choice of the group \mathbb{L} of Levi type. A particular surface operator leads to a particular quantum field theory problem and a particular moduli space $\mathcal{M}_{H,\mathbb{L}}$ of ramified Higgs bundles. In the theory of $\mathcal{M}_{H,\mathbb{L}}$, an emphasis is sometimes placed on the parabolic structure, a notion that we will explain in section 3.4. As we will see (and as one may anticipate from the definition of parabolic subgroups in section (2.6)), the interpretation in terms of parabolic structure depends on α having an \mathbb{L} -regular value. One can [24] interpolate from one parabolic type to another, keeping \mathbb{L} fixed, without meeting a singularity, by varying β and γ as well as α . So the same surface operator can lead to different kinds of parabolic structure, and it is better to label the surface operators by the Levi type rather than the parabolic type. That is why we prefer to speak of the moduli space of ramified Higgs bundles, rather than the moduli space of parabolic Higgs bundles. We will, however, speak of parabolic Higgs bundles when emphasizing the role of complex structure I , where this terminology is standard.

3.2. Complex Viewpoint. One can often [49] learn more about a hyper-Kähler quotient by focusing on one of the complex structures, for example I . The function $\nu_I = \mu_J + i\mu_K$ is holomorphic in complex structure I . The hyper-Kähler quotient $\mathcal{W}///\mathcal{G}_p$ is the same as the symplectic quotient of $\nu_I^{-1}(0)$ by \mathcal{G}_p . (This symplectic quotient is defined by setting to zero the ordinary moment map μ_I , restricted to $\nu_I^{-1}(0)$, and then dividing by \mathcal{G}_p . At this point, one has set to zero both ν_I and μ_I , giving the hyper-Kähler quotient.) Via the usual relation of symplectic and complex analytic quotients, the symplectic quotient of $\nu_I^{-1}(0)$ by \mathcal{G}_p can also be obtained as a geometric invariant theory quotient $\nu_I^{-1}(0)/\mathcal{G}_{p,\mathbb{C}}$, where $\mathcal{G}_{p,\mathbb{C}}$ is the complexification of \mathcal{G}_p . (Because \mathcal{W} is an affine space with linear action of \mathcal{G}_p , the complexification $\mathcal{G}_{p,\mathbb{C}}$ of \mathcal{G}_p has a natural action on \mathcal{W} once one selects one of the complex structures of \mathcal{W} .) All of these statements have analogs if complex structure I is replaced by any one of the complex structures that form part of the hyper-Kähler structure of \mathcal{W} .

In the present case, $\nu_I^{-1}(0)/\mathcal{G}_{p,\mathbb{C}}$ can be given a holomorphic description, as a moduli space of stable parabolic Higgs bundles (E, φ) . This is explained in [22], [23]. For our purposes, we do not really need the details, but we do need the fact that such a description exists. Roughly, a parabolic Higgs bundle, with parabolic structure at a specified point $p \in C$, is a Higgs bundle (E, φ) , where E is a holomorphic $G_{\mathbb{C}}$ -bundle over C , and φ is a section of $K_C \otimes \text{ad}(E)$ that is holomorphic away from p , and has a simple pole at p that obeys a certain condition. If $\sigma = \frac{1}{2}(\beta + i\gamma)$ is regular, meaning that it commutes precisely with \mathbb{T} (or more generally with \mathbb{L}), the condition can be anticipated from eqn. (2.11) and is that the polar part of φ is conjugate to $\sigma dz/z$. (Here z is a local holomorphic parameter near p .) We postpone to section 3.3 an explanation of what happens if σ is not regular.

The condition for a parabolic Higgs bundle to be stable depends on α in general, but is independent of α if σ is generic [24]. In Kähler geometry, one would expect wall-crossing phenomena at special values of α , but in hyper-Kähler geometry this is avoided for generic σ , as we explain in section 3.6.

The moduli space of stable parabolic Higgs bundles, which we temporarily denote $\tilde{\mathcal{M}}_H$, is a complex symplectic manifold, with holomorphic symplectic form $\Omega_I = \omega_J + i\omega_K$. Moreover, the complex structure of $\tilde{\mathcal{M}}_H$ and the cohomology class of Ω_I are manifestly holomorphic in $\beta + i\gamma$. They are also manifestly independent of α (because α only affects the stability condition) if σ is generic.

Since $\mathcal{M}_H(\alpha, \beta, \gamma; p)$, viewed as a complex manifold in complex structure I , is the same as $\tilde{\mathcal{M}}_H$, all of these statements have immediate implications for $\mathcal{M}_H(\alpha, \beta, \gamma; p)$. They explain certain claims made in the table in section 2.2. In particular, complex structure I on \mathcal{M}_H varies

holomorphically with $\beta + i\gamma$, and is independent of α . Moreover, if we write $[\omega]$ for the cohomology class of a closed differential form ω , then the cohomology class $[\Omega_I]$ varies holomorphically with $\beta + i\gamma$ and is independent of α . In particular, the classes $[\omega_J]$ and $[\omega_K]$ are independent of α .

Analog In Other Complex Structures

Of course, there is a similar story in complex structure J . The quotient $\nu_J^{-1}(0)/\mathcal{G}_{p,\mathbb{C}}$ can be given a holomorphic interpretation as the moduli space of stable parabolic (or filtered) local systems. A parabolic local system is a $G_{\mathbb{C}}$ -valued flat connection on $C \setminus p$ (C with the point p omitted), with a constraint on the monodromy around p . We call this monodromy V . If the element $U = \exp(-2\pi(\alpha - i\gamma))$ of $G_{\mathbb{C}}$ is regular (which means that the subgroup of $G_{\mathbb{C}}$ that commutes with it is precisely the torus \mathbb{T}), the constraint on the monodromy is that V must be conjugate to U , as we would expect from eqn. (2.12). Otherwise, one needs a more careful description, which we postpone to section 3.3. The condition for a parabolic local system to be stable depends on β in general, but is independent of β if $\gamma + i\alpha$ is generic [24].

The moduli space $\widehat{\mathcal{M}}_H$ of stable parabolic local systems is a complex symplectic manifold, with holomorphic symplectic form $\Omega_J = \omega_K + i\omega_I$. We now make the same argument as in complex structure I . The fact that \mathcal{M}_H , viewed as a complex manifold in complex structure J , is the same as $\widehat{\mathcal{M}}_H$, makes obvious certain claims made in the table in section 2.2. In particular, the complex structure J on \mathcal{M}_H varies holomorphically with $\gamma + i\alpha$ and is independent of β . Similarly, the cohomology class $[\Omega_J]$ is holomorphic in $\gamma + i\alpha$, and independent of β . Therefore $[\omega_K]$ and $[\omega_I]$ are independent of β .

We can apply the same reasoning, of course, in complex structure K , to show that the complex structure K and symplectic structure Ω_K of \mathcal{M}_H are holomorphic in $\alpha + i\beta$ and independent of γ . In particular, therefore, the cohomology classes $[\omega_I]$ and $[\omega_J]$ are independent of γ . This completes the justification of the claims made in the table in section 2.2.

Linearity of The Cohomology Classes

We can get considerably farther using the fact that the same reasoning applies for any of the complex structures that make up the hyper-Kähler structure of \mathcal{M}_H . These are complex structures of the form $\widehat{I} = pI + qJ + rK$, with real parameters p, q , and r obeying $p^2 + q^2 + r^2 = 1$. All statements in the table of section 2.2 have analogs for any \widehat{I} . Indeed, these statements remain valid if we make an $SO(3)$ rotation of the space spanned by the complex structures I, J, K , along with the same rotation of the spaces spanned by the three symplectic structures ω_I, ω_J , and ω_K and by the three variables α, β , and γ . This

$SO(3)$ is not a symmetry of \mathcal{M}_H , but it is a symmetry of the reasoning we have used in deducing the statements in the table.

This makes possible some simple inferences about the cohomology classes $[\omega_I]$, $[\omega_J]$, and $[\omega_K]$. Let us write x_i , $i = 1, 2, 3$, for those cohomology classes. Thus the x_i take values in the vector space $H^2(\mathcal{M}_H, \mathbb{R})$. They are functions of α, β , and γ , which we will denote as y_i , $i = 1, 2, 3$. Then what we have established so far is that

$$(3.5) \quad \frac{\partial x_i}{\partial y_j} = 0, \quad i \neq j.$$

However, by making the same argument in a generic complex structure \widehat{I} , we learn that (3.5) also holds after making an $SO(3)$ rotation on $\vec{x} = (x_1, x_2, x_3)$ along with $\vec{y} = (y_1, y_2, y_3)$. This implies that

$$(3.6) \quad \frac{\partial x_i}{\partial y_j} = \delta_{ij} v$$

for some function v (which takes values in $H^2(\mathcal{M}_H, \mathbb{R})$).¹⁶ By differentiating again, we learn from (3.6) that

$$(3.7) \quad \delta_{ij} \frac{\partial v}{\partial x_k} = \delta_{ik} \frac{\partial v}{\partial x_j}.$$

But this implies (by considering the case $i = j$, $i \neq k$) that $\partial v / \partial x_k = 0$ for all k , so that v is constant.

Since v is a constant, (3.6) implies that the x_i are linear functions of the y_j . In particular, $[\omega_I]$, which we already know to depend only on α , is actually a *linear* function of α . Thus

$$(3.8) \quad [\omega_I] = a + \text{Tr } \alpha h$$

with some constants a and h . (a takes values in $H^2(\mathcal{M}_H, \mathbb{R})$, and h in $H^2(\mathcal{M}_H, \mathbb{R}) \otimes \mathfrak{t}$.) Similarly, we already know that $[\omega_J]$ only depends on β and vanishes at $\beta = 0$; we can now deduce from (3.6) that

$$(3.9) \quad [\omega_J] = \text{Tr } \beta h$$

with the same h as in (3.8). And by the same token,

$$(3.10) \quad [\omega_K] = \text{Tr } \gamma h.$$

We will describe a and h in sections 3.5 and 3.6.

\mathcal{M}_H As A Symplectic Variety

As we have just seen, the cohomology class $[\omega_K]$ is independent of α and β . If \mathcal{M}_H were compact, it would follow that \mathcal{M}_H as a real symplectic variety with symplectic structure ω_K is independent of α and β . Since \mathcal{M}_H is noncompact, this conclusion does not follow just

¹⁶This is equivalent to the statement that for any complex structure $\widehat{I} = pI + qJ + rK$, the cohomology class of the corresponding holomorphic $(2, 0)$ -form $\Omega_{\widehat{I}}$ (which is obtained from Ω_I by a suitable $SO(3)$ rotation) is independent of the Kahler form (which is a multiple of $p\alpha + q\beta + r\gamma$).

from constancy of the cohomology class of the symplectic form, but it does follow by considering more carefully the facts we used to prove this constancy.

To change α without changing the real symplectic variety $(\mathcal{M}_H, \omega_K)$, we simply view \mathcal{M}_H in complex structure I . Because of the interpretation via parabolic Higgs bundles, this complex structure and the corresponding holomorphic two-form $\Omega_I = \omega_J + i\omega_K$ are independent of α , as long as certain singularities are avoided. So we can vary α keeping fixed the real symplectic variety $(\mathcal{M}_H, \omega_K)$.

Similarly, to vary β without changing the real symplectic variety $(\mathcal{M}_H, \omega_K)$, we view \mathcal{M}_H in complex structure J . Because of the interpretation via filtered local systems, this complex structure and the corresponding holomorphic two-form $\Omega_J = \omega_K + i\omega_I$ are independent of β , as long as certain singularities are avoided. So we can vary β keeping fixed the real symplectic variety $(\mathcal{M}_H, \omega_K)$.

The singularities that have to be avoided in this process are described in section 3.6 and are of real codimension at least two. So we conclude that the real symplectic variety $(\mathcal{M}_H, \omega_K)$ is independent of α and β , an important result for applications to the geometric Langlands program.

An important point is that the isomorphism given by this argument is not canonical but depends on the path by which α and β are varied. The reason is simply that we keep one structure fixed in varying α and a different structure fixed in varying β . If we vary both α and β , no structure except ω_K is held fixed, and \mathcal{M}_H varies by a symplectomorphism. In varying α and β around a closed loop that avoids the singularities, one will in general get a symplectomorphism of \mathcal{M}_H that is “topologically” non-trivial, that is, it cannot be deformed to the identity by a family of symplectomorphisms. We will develop this idea systematically in sections 3.7 and 4.5 to get an action of the affine braid group on branes on \mathcal{M}_H .

\mathbb{C}^* Action

The $SO(3)$ group that rotates the space of complex or symplectic structures of \mathcal{M}_H is not a symmetry of \mathcal{M}_H . However, a subgroup of it is a symmetry, just as for ordinary Higgs bundles without parabolic structure, where this is described in [21], pp. 107-8.

Let us first consider the case that $\beta = \gamma = 0$. Let the group $\mathcal{U}_1 \cong U(1)$ act on (A, ϕ) by leaving A invariant and transforming $\varphi \rightarrow \lambda\varphi$, with $|\lambda| = 1$. Since $\phi = \varphi + \bar{\varphi}$ (where φ and $\bar{\varphi}$ are of type $(1, 0)$ and $(0, 1)$), this determines the transformation of ϕ :

$$(3.11) \quad \phi \rightarrow \lambda\varphi + \bar{\lambda}\bar{\varphi}.$$

The action of \mathcal{U}_1 leaves invariant the characterization (3.13) of the singularity (as long as $\beta = \gamma = 0$). It leaves invariant the moment map μ_I ,

while rotating the $\mu_J - \mu_K$ plane. So it gives a manifest symmetry of the hyper-Kähler moment map construction, and a group of symmetries of the hyper-Kähler metric on \mathcal{M}_H . In fact, \mathcal{U}_1 is an $SO(2)$ subgroup of the $SO(3)$ that rotates the three complex structures. It acts on the family of complex structures $\widehat{I} = pI + qJ + rK$ by leaving p fixed and rotating the $q - r$ plane. Alternatively, if we parametrize the family of complex structures by a complex variable w , setting

$$(3.12) \quad I_w = \frac{1 - \bar{w}w}{1 + \bar{w}w}I + \frac{i(w - \bar{w})}{1 + \bar{w}w}J + \frac{w + \bar{w}}{1 + \bar{w}w}K,$$

then \mathbb{C}^* acts on the parameter w by $w \rightarrow \lambda^{-1}w$.

If we relax the condition $|\lambda| = 1$, we no longer get a symmetry of the hyper-Kähler metric of \mathcal{M}_H . However, exactly as in [21], we get a group $\mathcal{U} \cong \mathbb{C}^*$ that acts on \mathcal{M}_H preserving the complex structure I and transforming the family I_w by $w \rightarrow \lambda^{-1}w$, just as for $|\lambda| = 1$. The fixed points are $w = 0$, $I_w = I$, and $w = \infty$, $I_w = -I$. All other complex structures I_w are equivalent under the action of \mathcal{U} .

All this carries over to the case $\beta, \gamma \neq 0$, except that the parameters β and γ must be transformed by $(\beta + i\gamma) \rightarrow \lambda(\beta + i\gamma)$. This follows for $|\lambda| = 1$ by observing that the transformation (3.11) leaves fixed the singularity (3.13) if β and γ are transformed as claimed.

3.3. The Non-Regular Case. Now we will describe what happens to some of the above statements when the pair (α, γ) or the pair (β, γ) is non-regular. This may help the reader understand the constructions of [22], and is useful background in some of the applications to the geometric Langlands program.

Complex Structure J

We begin with complex structure J , in which a solution of Hitchin's equations corresponds to a parabolic local system. Also, for simplicity, we consider first the basic case that the Levi subgroup used to define our surface operator is $\mathbb{L} = \mathbb{T}$.

We consider a solution of Hitchin's equations with an isolated singularity at a point $p \in C$. We pick coordinates near p as in section 2.1 (so p is defined by $z = 0$ where $z = x_1 + ix_2 = re^{i\theta}$), and we assume a solution that behaves near $r = 0$ as

$$(3.13) \quad \begin{aligned} A &= \alpha d\theta + \dots \\ \phi &= \beta \frac{dr}{r} - \gamma d\theta + \dots, \end{aligned}$$

where the ellipses refer to terms that are less singular than $1/r$ as $r \rightarrow 0$. Hitchin's equations ensure that the $G_{\mathbb{C}}$ -valued connection $\mathcal{A} = A + i\phi$ is flat.

For $b > 0$ (but small enough so that the coordinates r, θ are defined for $r \leq b$), let C_b be the circle $r = b$, and let V_b be the monodromy

around C_b of the flat connection \mathcal{A} . The V_b have two basic properties: (1) The conjugacy class of V_b is independent of b . This is so because the connection is flat. Since we only care about the monodromy up to conjugacy, we pick any one of the V_b and call it V , the monodromy around the singularity. (2) If we set $U = \exp(-2\pi(\alpha - i\gamma))$, then

$$(3.14) \quad \lim_{b \rightarrow 0} V_b = U.$$

This is so because in the limit of $b \rightarrow 0$, we can compute the holonomy just from the most singular terms in A and ϕ , and if we do so then the result is U .

If U is regular, the two properties imply that V is conjugate to U . For $SL(2, \mathbb{C})$, we can prove this as follows. The two properties imply that $\text{Tr } U = \text{Tr } V$ (where the trace is taken in the two-dimensional representation); if U is regular, this implies that U and V are conjugate. Similar reasoning holds for any G , with the trace replaced by a full set of invariant functions. If U is not regular, U and V need not be conjugate. For example, for $G = SL(2, \mathbb{C})$, let us consider the non-regular element $U = 1$, which corresponds to $\alpha = \gamma = 0$. For $U = 1$, we can satisfy the two conditions with

$$(3.15) \quad V_b = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}.$$

The V_b for $b > 0$ are all conjugate, and $\lim_{b \rightarrow 0} V_b = 1$. An element V of this form is called unipotent (this means simply that $V - 1$ is nilpotent).

So it is possible for a solution of Hitchin's equations with $\alpha = \gamma = 0$, $\beta \neq 0$ to have monodromy that is unipotent but not equal to 1. Not only is this possible, but it is the generic behavior, simply because the condition for V to be unipotent is one complex condition (which one can formulate as $\text{Tr } V = 2$), while for V to equal 1 is three complex conditions. Comparing these dimensions, one might think that if $\alpha = \gamma = 0$, then those Higgs bundles for which V is actually 1 would be a family of complex codimension 2. A more careful analysis shows that this is correct if $\beta = 0$ (in which case the Higgs bundles with $V = 1$ are a locus of A_1 orbifold singularities), but that for $\beta \neq 0$ the locus with $V = 1$ is "blown up," and is of complex codimension 1. The statements about the generic behavior, the singularity, and the blowup should become clearer below, especially in section 3.6. The blowup is described by specifying what Simpson [22] calls a "filtration" of the local system.

Let \mathfrak{C} be the conjugacy class in $G_{\mathbb{C}}$ containing U . There is, for any G and any choice of \mathfrak{C} , a finite set of conjugacy classes \mathfrak{C}_λ , $\lambda = 1, \dots, s$, with the property that a family of elements $V_b \in \mathfrak{C}_\lambda$ can have a limit in \mathfrak{C} for $b \rightarrow 0$. Differently put, \mathfrak{C} is in the closure of \mathfrak{C}_λ . We call the \mathfrak{C}_λ the conjugacy classes that are affiliated to \mathfrak{C} . For example, if U is regular, the only affiliated conjugacy class is \mathfrak{C} itself. At the other

extreme, if $U = 1$, the affiliated conjugacy classes are precisely the ones that parametrize unipotent elements of G . For G of large rank, there are many such classes (given for $SL(N, \mathbb{C})$ by block triangular matrices with 1's on the diagonal and blocks of different sizes).

In any event, there is always a unique affiliated conjugacy class \mathfrak{C}^* of maximal dimension, in fact of dimension $\dim G - r$, where r is the rank of G . It is called a regular conjugacy class, because it parametrizes elements that are regular, that is they commute with only an r -dimensional subgroup of $G_{\mathbb{C}}$.

The class \mathfrak{C} may not be regular, but it has another distinguishing property. Among the affiliated conjugacy classes, \mathfrak{C} is the unique one that is “semi-simple,” that is, it parametrizes group elements that are semi-simple (they can be conjugated to the maximal torus). \mathfrak{C} is certainly semi-simple, since it contains $U = \exp(-2\pi(\alpha - i\gamma))$, which is an element of \mathbb{T} .

But if U is not regular, then $\mathfrak{C} \neq \mathfrak{C}^*$, and elements of the regular conjugacy class affiliated to U are not semi-simple. For example, if $G = SU(N)$ and $U = 1$, then the regular conjugacy class affiliated to U contains the “principal unipotent element”

$$(3.16) \quad V = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & \dots & 0 \\ & & & \vdots & \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

with 1's on and just above the main diagonal and zeroes elsewhere. For any α and γ , the generic Higgs bundle gives a local system whose monodromy is in the regular conjugacy class associated to U .

Complex Structure I

Now we consider complex structure I , in which a solution of Hitchin's equations is a parabolic Higgs bundle.

Here it will be helpful to begin by considering in detail the example of $SL(2, \mathbb{C})$. We take $\beta = \gamma = 0$ (this being the only non-regular choice of β and γ for $SL(2, \mathbb{C})$), and as we want to assume that α is regular, we take

$$(3.17) \quad \alpha = iy \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

with $0 < y < 1/2$. (The reason for the factor of i is that as α takes values in the real Lie algebra of $SU(2)$, it is anti-hermitian in a unitary representation.) The limiting form of the solution of Hitchin's equations

for $r \rightarrow 0$ is therefore

$$(3.18) \quad \begin{aligned} A &= iy d\theta \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ \phi &= 0. \end{aligned}$$

We want to see what sort of pole φ , defined as the $(1,0)$ part of ϕ , can acquire when we replace (3.18) with a more general solution of Hitchin's equations that has the same asymptotic behavior as $r \rightarrow 0$.

At first sight, it might appear that no pole at all is possible. By definition, we want to perturb the limiting solution (3.18) by terms that are less singular than $1/r$. So it may seem that φ will not be sufficiently singular to have a pole. However, to decide whether φ has a pole, we need to trivialize the holomorphic structure of the bundle E near the singular point p . It turns out that once this is done, φ can have a pole.

When we expand Hitchin's equations around the solution (3.18), the linearized equations have a solution

$$(3.19) \quad \phi = \epsilon \frac{dz}{z} (\bar{z}z)^y \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

with ϵ a small parameter. This solution is less singular than $1/r$, so including this perturbation is compatible with the asymptotic behavior (3.18). (Of course, to get a real solution for ϕ , one must subtract the hermitian conjugate solution. The $(1,0)$ part of ϕ must be upper triangular, since otherwise an analogous solution is more singular than $1/r$ at $r = 0$.)

To trivialize the holomorphic bundle E near $r = 0$, we write down the appropriate $\bar{\partial}_A$ operator that defines the holomorphic structure of E :

$$(3.20) \quad \bar{\partial}_A = d\bar{z} \left(\frac{\partial}{\partial \bar{z}} + A_{\bar{z}} \right) = d\bar{z} \left(\frac{\partial}{\partial \bar{z}} - \frac{y}{2\bar{z}} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right).$$

We can write this as

$$(3.21) \quad \bar{\partial}_A = f \bar{\partial} f^{-1},$$

where $\bar{\partial} = d\bar{z} \partial / \partial \bar{z}$ is the standard $\bar{\partial}$ operator, and

$$(3.22) \quad f = \begin{pmatrix} (\bar{z}z)^{y/2} & 0 \\ 0 & (\bar{z}z)^{-y/2} \end{pmatrix}.$$

So if $\bar{\partial}_A \varphi = 0$ (which is part of Hitchin's equations), then $\bar{\partial}(f^{-1} \varphi f) = 0$. But

$$(3.23) \quad f^{-1} \varphi f = \epsilon \frac{dz}{z} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

So the conclusion is that φ can have a pole, relative to the trivialization of the bundle E , but the residue of this pole is strictly upper triangular (in a basis in which $-i\alpha$ is diagonal with decreasing eigenvalues along the diagonal).

In this discussion, we started with $\beta = \gamma = 0$, so $\sigma = \frac{1}{2}(\beta + i\gamma)$ also vanishes. The residue of the pole in φ turned out to be

$$(3.24) \quad \tau = \epsilon \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

The conjugacy class of τ is independent of ϵ , and the limit of τ for $\epsilon \rightarrow 0$ is σ . These properties imply that all invariant polynomials take the same value for σ and τ ; for $SL(2, \mathbb{C})$, this simply means that $\text{Tr } \sigma^2 = \text{Tr } \tau^2 = 0$.

In general, for any G , let \mathfrak{c} be the orbit or conjugacy class in the Lie algebra $\mathfrak{g}_{\mathbb{C}}$ that contains $\sigma = \frac{1}{2}(\beta + i\gamma)$. This conjugacy class parametrizes semi-simple elements of $\mathfrak{g}_{\mathbb{C}}$ (that is, elements that can be conjugated to a maximal torus) since σ itself is semi-simple. We say that a conjugacy class \mathfrak{c}_{λ} is affiliated to \mathfrak{c} if a sequence of elements of \mathfrak{c}_{λ} can converge to an element of \mathfrak{c} , or in other words if \mathfrak{c} is in the closure of \mathfrak{c}_{λ} . In general, there are finitely many conjugacy classes affiliated to \mathfrak{c} . The residue τ of the pole of the Higgs field always takes values in an affiliated conjugacy class. This is true by reasoning similar to what we have explained in the above example.

If σ is regular as well as semi-simple, then \mathfrak{c} itself is the only affiliated conjugacy class, and in particular τ is conjugate to σ . At the opposite extreme, if $\sigma = 0$, then the conjugacy classes affiliated to \mathfrak{c} are precisely the classes of nilpotent elements of $\mathfrak{g}_{\mathbb{C}}$.

For every \mathfrak{c} , there is a unique affiliated conjugacy class \mathfrak{c}^* of maximal dimension, in fact dimension $\dim(G) - r$. It parametrizes regular elements of $\mathfrak{g}_{\mathbb{C}}$, that is, elements that commute with only an r -dimensional subgroup of $G_{\mathbb{C}}$. For example, if $\sigma = 0$ and $G = SU(N)$, then the affiliated regular conjugacy class contains the element $v = V - 1$, where V was defined in (3.16). For any σ , the generic parabolic Higgs bundle has a pole whose residue τ is in the regular affiliated conjugacy class \mathfrak{c}^* .

We can summarize much of this by saying that conjugacy classes in the Lie algebra $\mathfrak{g}_{\mathbb{C}}$ behave in many relevant respects like conjugacy classes in the group $G_{\mathbb{C}}$.

Reformulation

Going back to the $SU(2)$ example, our result about the polar behavior of φ for the case $\beta = \gamma = 0$ can be described as follows. For $G = SU(2)$, the only possible Levi subgroup is $\mathbb{L} = \mathbb{T}$, which is the case considered in the above discussion. The choice of α determines a parabolic subgroup \mathcal{P} and a unipotent radical \mathcal{N} . In the above example, \mathcal{P} is the group of upper triangular matrices and \mathcal{N} is the group whose

Lie algebra \mathfrak{n} consists of strictly upper triangular matrices. The result (3.23) says that the polar residue of φ takes values in \mathfrak{n} . The mechanism by which this came about is simply that in order for ϕ (in the differential geometric description) to be less singular than $1/r$, its polar residue τ (in the holomorphic description) must obey $-i[\alpha, \tau] = \lambda\tau$, $\lambda > 0$. The analog for G of higher rank, still assuming $\beta = \gamma = 0$, is that τ must be a linear combination of elements of \mathfrak{g} that obey this condition. So in other words, τ takes values in \mathfrak{n} . This holds for any choice of Levi subgroup \mathbb{L} and any \mathbb{L} -regular α .

Going back to $SU(2)$, and without changing α , let us perturb β and γ to be nonzero, say $\frac{1}{2}(\beta + i\gamma) = \text{diag}(q, -q)$ for some $q \in \mathbb{C}$. What happens in this case? We must set $\phi = \beta(dr/r) - \gamma d\theta$ plus terms that are less singular at $r = 0$. One possibility is to have $\phi = \beta(dr/r) - \gamma d\theta$ exactly. This corresponds to $\varphi = (dz/z)\text{diag}(q, -q)$. But just as in (3.19), we can make an upper triangular deformation¹⁷ of φ , so the general possibility for the polar part of φ is

$$(3.25) \quad \varphi \sim \frac{dz}{z} \begin{pmatrix} q & * \\ 0 & -q \end{pmatrix}.$$

The upper right element denoted $*$ does not affect the conjugacy class of the residue of the pole if $q \neq 0$. Note that if φ and $\tilde{\varphi}$ are two Higgs fields with a pole of this kind, then $\varphi - \tilde{\varphi}$ has a pole with \mathfrak{n} -valued residue:

$$(3.26) \quad \varphi - \tilde{\varphi} \sim \frac{dz}{z} \begin{pmatrix} 0 & * \\ 0 & 0 \end{pmatrix}.$$

By similar reasoning, this is so not just for $SL(2, \mathbb{C})$, but for any gauge group G , with any choice of Levi subgroup \mathbb{L} and any \mathbb{L} -regular α . The general statement is that the polar residue of φ takes values in \mathfrak{p} , the Lie algebra of \mathcal{P} , and is equal to $\sigma = \frac{1}{2}(\beta + i\gamma)$ modulo an element of \mathfrak{n} . The last statement can be informally summarized by saying that the “eigenvalues” of the polar residue coincide with those of σ .

More On Complex Structure J

There is an asymmetry in our discussion of the non-regular behavior in complex structures J and I . In complex structure J , we reasoned somewhat abstractly about the closures of conjugacy classes, but in complex structure I , we analyzed the behavior of perturbations of Hitchin’s equations. Of course, by analyzing Hitchin’s equations, we can be more explicit about what happens in complex structure J . This will also enable us to get more information.

¹⁷For $q \neq 0$, the details of the solution are more complicated, and it is necessary to also modify A . The appropriate solution is described by Nahm’s equations; see section 3.8. It remains true, as at $q = 0$, that a lower-triangular modification of φ is more singular than $1/r$.

We carry out the discussion for any gauge group G and Levi subgroup \mathbb{L} . To begin with, we take $\alpha = \gamma = 0$, but we take β to be generic or in other words \mathbb{L} -regular. This corresponds to a solution of Hitchin's equations with

$$(3.27) \quad A = 0, \quad \phi = \beta \frac{dr}{r}.$$

We perturb this to

$$(3.28) \quad \begin{aligned} A &= a(r) d\theta, \\ \phi &= \beta \frac{dr}{r} + c(r) d\theta, \end{aligned}$$

where a and c must vanish at $r = 0$ (since the deviation from the limiting solution (3.27) must be less singular than $1/r$) and we will work to first order in a and c . The resulting equations can be written

$$(3.29) \quad r \frac{d}{dr}(a + ic) = [-i\beta, a + ic].$$

For $a + ic$ to vanish at $r = 0$, it must take values in \mathfrak{n} , the subspace of \mathfrak{g} spanned by vectors ψ with $-i[\beta, \psi] = \lambda\psi$, $\lambda > 0$. Hence, the monodromy of the flat connection $\mathcal{A} = A + i\phi$, which in this approximation is $U = \exp(-2\pi(a + ic))$, takes values in \mathcal{N} , the unipotent radical of the parabolic subgroup determined by β .

Every element of \mathcal{N} is a unipotent element of $G_{\mathbb{C}}$. In the theory of semi-simple Lie groups, it is shown that the generic element of \mathcal{N} lies in a unipotent conjugacy class in $G_{\mathbb{C}}$ called the Richardson class $\mathfrak{C}_{\mathbb{L}}$. From the definition, it seems that the Richardson class depends on \mathcal{P} , and thus β , but it can be shown that the Richardson class is actually determined by \mathbb{L} . (In general, distinct \mathbb{L} 's can lead to the same Richardson class.) The monodromy U found in the last paragraph is unconstrained except for taking values in \mathcal{N} , so generically it takes values in the conjugacy class $\mathfrak{C}_{\mathbb{L}}$.

For example, if $\mathbb{L} = \mathbb{T}$, then $\mathfrak{C}_{\mathbb{L}}$ is the regular unipotent conjugacy class described for $SL(N, \mathbb{C})$ in (3.16). An $N \times N$ matrix U such that $U - 1$ is strictly upper triangular is generically in this conjugacy class. To give another example, for $SL(3, \mathbb{C})$, if \mathbb{L} consists of matrices of the form

$$(3.30) \quad \begin{pmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & * \end{pmatrix}$$

and α is such that the unipotent radical \mathcal{N} consists of matrices

$$(3.31) \quad \begin{pmatrix} 1 & 0 & * \\ 0 & 1 & * \\ 0 & 0 & 1 \end{pmatrix},$$

then a generic element of \mathcal{N} is conjugate under \mathbb{L} to

$$(3.32) \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

This is a representative of the Richardson conjugacy class $\mathfrak{C}_{\mathbb{L}}$. In general, for $SL(N, \mathbb{C})$, every unipotent conjugacy class is a Richardson class, but this is not true for other groups.

This discussion can be generalized to the case that $\alpha, \gamma \neq 0$. If $U = \exp(-2\pi(\alpha - i\gamma))$ is \mathbb{L} -regular, then the monodromy V is conjugate to U . In general, it takes the form $V = UN$, where N takes values in the unipotent radical \mathcal{N} . Equivalently, the monodromy lies in a conjugacy class in \mathcal{P} whose closure includes U . Moreover, up to conjugacy of V , one can assume that U and N commute. Generically, N is then simply a general element of \mathcal{N} that commutes with U .

If N is of this form, then the orbit in $G_{\mathbb{C}}$ of the element $V = UN$ of \mathcal{P} has the same dimension as $G_{\mathbb{C}}/\mathbb{L}_{\mathbb{C}}$, which is the orbit in $G_{\mathbb{C}}$ of a generic \mathbb{L} -regular element of $\mathbb{T}_{\mathbb{C}}$. We call elements of \mathcal{P} that have this property \mathbb{L} -regular, so in particular, for any U , the generic monodromy $V = UN$ of the local system in complex structure J is \mathbb{L} -regular.

For instance, Richardson orbits are \mathbb{L} -regular. In the above $SL(3, \mathbb{C})$ example, the dimension of $G_{\mathbb{C}}/\mathbb{L}_{\mathbb{C}}$ is 4, which is also the dimension of the Richardson orbit described in eqn. (3.32). If a given unipotent orbit is the Richardson orbit of several different Levi subgroups \mathbb{L}_i , then it is \mathbb{L}_i -regular for each i .

3.4. Parabolic Bundles. At this point, we should perhaps explain a notion that is usually taken as the starting point in the mathematical theory, but that we have hidden so far. This is the notion of a parabolic bundle (as opposed to a parabolic Higgs bundle). All statements have obvious analogs with parabolic structure at several points, but for simplicity we consider mainly the case of one point. Until further notice, we consider only gauge theory, without the Higgs field.

For motivation, we return to the $SU(2)$ example of section 3.3. A holomorphic section s of the bundle $\text{ad}(E)$ is an $\text{ad}(E)$ -valued function annihilated by the $\bar{\partial}_A$ operator. Given the explicit form (3.20) of this operator, this means that near $z = 0$,

$$(3.33) \quad s = \begin{pmatrix} u & v(\bar{z}z)^y \\ w(\bar{z}z)^{-y} & -u \end{pmatrix},$$

where u, v , and w are ordinary holomorphic functions. If we want $|s|$ to be bounded for $z \rightarrow 0$, we require that $w(0) = 0$. Hence at $z = 0$,

$\tilde{s} = f^{-1}sf$ takes the form

$$(3.34) \quad \tilde{s} = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix}.$$

From a holomorphic point of view, what is happening is that the choice of α determines in the fiber of $\text{ad}(E)$ at p a Borel subalgebra \mathfrak{b} , spanned by vectors $\psi \in \mathfrak{g}$ with $-i[\alpha, \psi] = \lambda\psi$, $\lambda \geq 0$. Eqn. (3.34) says that $\tilde{s}(p)$ takes values in \mathfrak{b} . In this form, the result holds for any G : if α is regular, then $\tilde{s}(p)$ takes values in the Borel subgroup determined by α . More generally, if we pick a Levi subgroup \mathbb{L} and α is \mathbb{L} -regular, then by the same sort of reasoning, $\tilde{s}(p)$ takes values in the Lie algebra \mathfrak{p} of the parabolic subgroup \mathcal{P} determined by the pair (\mathbb{L}, α) .

A choice of parabolic structure for a G -bundle $E \rightarrow C$ at a point p is simply a reduction of the structure group of E at p to a parabolic subgroup \mathcal{P} . What we have just seen is that a bundle with a singularity that in differential geometry is described by $A = \alpha d\theta + \dots$ near a point $p \in C$, for \mathbb{L} -regular α , corresponds in complex geometry to a bundle with a choice of parabolic structure at p .

A theorem of Mehta and Seshadri [50] (which generalizes a theorem of Narashimhan and Seshadri [51] in the absence of parabolic structure) puts this in a systematic framework. This theorem establishes a one-to-one correspondence between stable parabolic $G_{\mathbb{C}}$ -bundles and flat G -bundles with the familiar singularity $A = \alpha d\theta + \dots$. On the left hand side of this correspondence, one considers a holomorphic $G_{\mathbb{C}}$ -bundle with a reduction of its structure group at a point $p \in C$ to a parabolic subgroup \mathcal{P} . We let \mathbb{L} be the Levi subgroup of \mathcal{P} , and we pick an α such that \mathcal{P} is determined in the usual way by the pair (\mathbb{L}, α) . We assume that α is generic enough so that the subgroup of G that commutes with $U = \exp(-2\pi\alpha)$ is precisely \mathbb{L} . We say that such an α is strictly \mathbb{L} -regular. For each such α , there is a natural notion of stability for bundles with parabolic structure of type \mathcal{P} .¹⁸ We will not describe the stability condition here, though it is fundamental in the mathematical theory. On the right hand side of the correspondence, one considers flat G -bundles on $C \setminus p$ with monodromy around p conjugate to $U = \exp(-2\pi\alpha)$. Equivalently, one considers solutions of the familiar equation

$$(3.35) \quad F = 2\pi\alpha\delta_p$$

modulo gauge transformations that take values in \mathbb{L} at the point p . We denote as $\mathcal{M}(\alpha; p)$ the moduli space of such flat bundles with monodromy. The theorem of Mehta and Seshadri is that the moduli space of stable parabolic $G_{\mathbb{C}}$ -bundles is the same as $\mathcal{M}(\alpha; p)$. The analogous

¹⁸Moreover, up to equivalence, this notion is invariant under shifts of α by a lattice vector. The equivalence in question involves a Hecke modification of the bundle.

theorem [51] in the absence of parabolic structure says that the moduli space \mathcal{M} of stable $G_{\mathbb{C}}$ -bundles over C is the same as the moduli space of flat G -bundles.

Both results are natural from the point of view of the symplectic quotient of the space of gauge fields by the group of gauge transformations [52]. The space of connections (or connections with a singularity $A = \alpha d\theta + \dots$) is a symplectic manifold with symplectic form

$$(3.36) \quad \omega = -\frac{1}{4\pi} \int_C \text{Tr } \delta A \wedge \delta A$$

and moment map

$$(3.37) \quad \mu = -\frac{1}{2\pi} \int_C \text{Tr } \epsilon F.$$

The symplectic quotient of the space of connections (or connections with singularity) by the appropriate group of gauge transformations is \mathcal{M} (or $\mathcal{M}(\alpha; p)$). By reinterpreting the symplectic quotient as a quotient by the group $\mathcal{G}_{\mathbb{C}}$ of complex-valued gauge transformations, these spaces can be alternatively interpreted as moduli spaces of stable bundles, or stable bundles with parabolic structure. The reasoning is similar to what we described in detail for Higgs bundles in sections 3.1, 3.2.

The notation $\mathcal{M}(\alpha; p)$ for the moduli space of flat bundles on $C \setminus p$ with monodromy $U = \exp(-2\pi\alpha)$ around p is slightly misleading, because $\mathcal{M}(\alpha; p)$ does not vary smoothly with α . Its dimension depends on the subgroup of G that commutes with U . $\mathcal{M}(\alpha; p)$ varies smoothly with α only if α is constrained to be strictly \mathbb{L} -regular for some fixed \mathbb{L} . To emphasize this, we consider \mathbb{L} as part of the definition and denote this space as $\mathcal{M}_{\mathbb{L}}(\alpha; p)$. For given \mathbb{L} , the space of strictly \mathbb{L} -regular α 's has distinct connected components (which in general are associated with non-isomorphic parabolic subgroups, as we learned in section (2.6)). When we do not write \mathbb{L} explicitly, it will mean that we are taking $\mathbb{L} = \mathbb{T}$.

Parabolic Higgs Bundles

Now we include the Higgs field and consider the analogous concept for Higgs bundles.

In the mathematical theory, the concept of a parabolic Higgs bundle (E, φ) is usually defined as follows. E is a parabolic bundle in the above sense, with a reduction of the structure group to \mathcal{P} at the point p . And the differential φ has a pole at p with a residue that is required to take values in the corresponding Lie algebra \mathfrak{p} ; moreover, this polar residue has the same ‘‘eigenvalues’’ as $\sigma = \frac{1}{2}(\beta + i\gamma)$, and in fact, it equals σ modulo \mathfrak{n} . (See (3.25) for a concrete illustration of this.)

If α and $\beta + i\gamma$ are regular, it is not necessary to make explicit the concept of parabolic structure. For generic (α, β, γ) , it is enough to give the bundle E together with the differential φ , which is required

to have a pole at p with residue conjugate to $\frac{1}{2}(\beta + i\gamma)$. Then α is determined in terms of φ , since it must commute with β and γ and its conjugacy class is known.¹⁹ So the parabolic structure is determined, and for generic parameters, we can define a parabolic Higgs bundle without ever explaining what it means for a bundle to have parabolic structure. That is essentially what we have done in our initial approach to the subject.

3.5. Topology Of $\mathcal{M}_{\mathbb{L}}(\alpha; p)$. We next use some of these ideas to get a rough understanding of the topology of $\mathcal{M}_{\mathbb{L}}(\alpha; p)$. We aim to give a first orientation to the topology of these spaces for readers who have never encountered them before. And we aim to develop the necessary background for certain results about \mathcal{M}_H described in sections 3.6 and 3.7. Hopefully, the very incomplete explanations we give will suffice for these particular goals.

Suppose we are given a particular $G_{\mathbb{C}}$ -bundle $E \rightarrow C$, and a parabolic subgroup \mathcal{P} of $G_{\mathbb{C}}$, and we wish to pick parabolic structure of type \mathcal{P} at a point $p \in C$. We have to pick at p a subgroup of $G_{\mathbb{C}}$ that is conjugate to \mathcal{P} . The space of all such subgroups is isomorphic to $G_{\mathbb{C}}/\mathcal{P}$. This suggests that $\mathcal{M}_{\mathbb{L}}(\alpha; p)$ should be a fiber bundle with fiber $G_{\mathbb{C}}/\mathcal{P}$ over \mathcal{M} , the moduli space of stable G -bundles (without parabolic structure):

$$(3.38) \quad \begin{array}{ccc} G_{\mathbb{C}}/\mathcal{P} & \rightarrow & \mathcal{M}_{\mathbb{L}}(\alpha; p) \\ & & \downarrow \\ & & \mathcal{M}. \end{array}$$

To the extent that it is valid (which we discuss shortly), this fibration elucidates the complex structure of $\mathcal{M}_{\mathbb{L}}(\alpha; p)$; the fiber and base are both complex manifolds, and the fibration is holomorphic.

In addition, $\mathcal{M}_{\mathbb{L}}(\alpha; p)$ has a natural symplectic structure, which is conveniently understood from its interpretation as the moduli space of flat G -bundles with monodromy. The symplectic form is $\omega = -\frac{1}{4\pi} \int_C \text{Tr } \delta A \wedge \delta A$. The complex and symplectic structures of $\mathcal{M}(\alpha; p)$ combine to a Kahler structure. To see the symplectic structure of $\mathcal{M}(\alpha; p)$, a variant of (3.38) is more helpful. A basic fact in the theory of complex Lie groups is that the quotient $G_{\mathbb{C}}/\mathcal{P}$ is the same as G/\mathbb{L} , where $\mathbb{L} = G \cap \mathcal{P}$ is a Levi subgroup of G that is a maximal compact subgroup of \mathcal{P} . So instead of (3.38) we can exhibit $\mathcal{M}_{\mathbb{L}}(\alpha; p)$

¹⁹The action of the Weyl group introduces no ambiguity, since it acts diagonally on the triple (α, β, γ) .

as a fibration of symplectic manifolds:

$$(3.39) \quad \begin{array}{ccc} G/\mathbb{L} & \rightarrow & \mathcal{M}_{\mathbb{L}}(\alpha; p) \\ & & \downarrow \\ & & \mathcal{M}. \end{array}$$

Here the base and fiber are symplectic and the fibration will be used below to describe the symplectic structure of $\mathcal{M}_{\mathbb{L}}(\alpha; p)$.

In fact, the fibrations (3.38) and (3.39) are valid precisely to the extent that we can assume that every point in \mathcal{M} is represented by a stable (and not just semi-stable) bundle. Otherwise, it is possible to have a stable parabolic bundle (E, \mathcal{P}) , where the underlying bundle E is not stable. In that case, the fibration breaks down, since it is not possible to construct the moduli space $\mathcal{M}_{\mathbb{L}}(\alpha; p)$ of stable parabolic bundles by first picking the bundle E and then endowing it with all possible parabolic structures. The fibrations (3.38) and (3.39) do hold away from singularities of \mathcal{M} and give a good first approximation to the topology of $\mathcal{M}_{\mathbb{L}}(\alpha; p)$.

There is one important and widely studied case in which all semi-stable bundles are stable, and therefore the fibrations (3.41) and (3.42) are precisely valid. This occurs if $G = PSU(N)$ and E is a bundle whose characteristic class in $H^2(C; \mathbb{Z}_N)$ is of order N . In general, the codimension at which the fibrations breaks down increases when the genus g_C of C or the rank r of G is increased. For example, the real codimension exceeds 2 if $g_C > 2$ or $g_C = 2$, $r > 1$.

If the singularities do not play an important role, then we can use (3.38) or (3.39) to describe the second cohomology group of $\mathcal{M}_{\mathbb{L}}(\alpha; p)$. (We do this because it will eventually help us understand the symplectic structure of \mathcal{M}_H and give a concrete illustration of the action of the affine Weyl group on its cohomology.) We will do this mainly assuming that G is simply-connected. In this case, the Leray spectral sequence for the cohomology of $\mathcal{M}_{\mathbb{L}}(\alpha; p)$ begins with

$$(3.40) \quad \begin{array}{c|cccc} 2 & H^2(G/\mathbb{L}; \mathbb{Z}) & 0 & * & * \\ 1 & 0 & 0 & 0 & 0 \\ 0 & \mathbb{Z} & 0 & H^2(\mathcal{M}; \mathbb{Z}) & * \\ \hline & 0 & 1 & 2 & 3 \end{array}$$

(We have plotted the q^{th} cohomology of \mathcal{M} with values in the p^{th} cohomology of G/\mathbb{T} , with $q = 0, 1, 2, 3$ running horizontally and $p = 0, 1, 2$ running vertically. The precise form of the groups labeled $*$ will not be important.) In dimension two, as we will explain later, the differentials in the spectral sequence vanish if G is simply-connected, so the spectral sequence for the fibration reduces to an exact sequence

$$(3.41) \quad 0 \rightarrow H^2(\mathcal{M}; \mathbb{Z}) \rightarrow H^2(\mathcal{M}_{\mathbb{L}}(\alpha; p); \mathbb{Z}) \rightarrow H^2(G/\mathbb{L}; \mathbb{Z}) \rightarrow 0.$$

Moreover $H^2(\mathcal{M}; \mathbb{Z}) \cong \mathbb{Z}$ for simply-connected G , generated [53], [54] by the first Chern class of a line bundle \mathfrak{L} that we will loosely call the determinant line bundle (for $G = SU(N)$, \mathfrak{L} can be defined as the determinant line bundle of a Dirac operator). So the exact sequence becomes

$$(3.42) \quad 0 \rightarrow \mathbb{Z} \rightarrow H^2(\mathcal{M}_{\mathbb{L}}(\alpha; p); \mathbb{Z}) \rightarrow H^2(G/\mathbb{L}; \mathbb{Z}) \rightarrow 0.$$

For simply-connected G , the spectral sequence for the fibration

$$(3.43) \quad \begin{array}{ccc} \mathbb{L} & \rightarrow & G \\ & & \downarrow \\ & & G/\mathbb{L} \end{array}$$

gives $H^2(G/\mathbb{L}; \mathbb{Z}) = H^1(\mathbb{L}; \mathbb{Z})$. For $\mathbb{L} = \mathbb{T}$, this gives

$$(3.44) \quad H^2(G/\mathbb{T}; \mathbb{Z}) = H^1(\mathbb{T}; \mathbb{Z}) = \Lambda_{\text{wt}},$$

with Λ_{wt} the weight lattice of G . This particular result (with the same lattice Λ_{wt}) holds whether G is simply-connected or not, since if we replace G by a finite cover of itself, then G/\mathbb{T} is unchanged (the cover extends \mathbb{T} in the same way, and cancels out of the quotient G/\mathbb{T}).

By virtue of (3.44), (3.42) is equivalent, for $\mathbb{L} = \mathbb{T}$, to

$$(3.45) \quad 0 \rightarrow \mathbb{Z} \rightarrow H^2(\mathcal{M}(\alpha; p); \mathbb{Z}) \rightarrow \Lambda_{\text{wt}} \rightarrow 0.$$

For more general \mathbb{L} , a similar reasoning gives instead

$$(3.46) \quad 0 \rightarrow \mathbb{Z} \rightarrow H^2(\mathcal{M}_{\mathbb{L}}(\alpha; p); \mathbb{Z}) \rightarrow \Lambda_{\text{wt}, \mathbb{L}} \rightarrow 0,$$

where $\Lambda_{\text{wt}, \mathbb{L}}$ is the sublattice of Λ_{wt} that is invariant under the Weyl group of \mathbb{L} .

Like any exact sequence of lattices, (3.45) can be split to give

$$(3.47) \quad H^2(\mathcal{M}(\alpha; p); \mathbb{Z}) = \mathbb{Z} \oplus \Lambda_{\text{wt}}.$$

This result is the theorem stated (in terms of the Picard group) in section (1.1) of [54]. The theorem is stated there for the case of parabolic structure at several points p_1, \dots, p_s . In this case, one has a fibration like that considered above, with a copy of G/\mathbb{T} at each of the points p_i , so a similar analysis gives

$$(3.48) \quad H^2(\mathcal{M}(\alpha_1, p_1; \dots; \alpha_s, p_s); \mathbb{Z}) = \mathbb{Z} \oplus (\oplus_{i=1}^s \Lambda_{\text{wt}, i}),$$

with s copies of the weight lattices.

A splitting of lattices such as (3.47) is in general non-canonical, but in this case there is a canonical splitting. One way to see this is to start with the universal bundle $E_{\text{ad}} \rightarrow \mathcal{M} \times C$ in the adjoint representation. Upon restriction to $\mathcal{M} \times p$, this gives a G_{ad} -bundle $E_{\text{ad}, p} \rightarrow \mathcal{M}$. When pulled back to $\mathcal{M}(\alpha; p) \rightarrow \mathcal{M}$, the structure group of this bundle reduces to \mathbb{T} , so it splits as a sum of line bundles. The first Chern classes of these line bundles generate rationally the summand Λ_{wt} in (3.47). This

gives the splitting, which is also evident in the interpretation we give below via the affine weight lattice.

In our derivation of (3.47), we have made use of the approximate fibration (3.38). Actually, the singularities that we have neglected are not important for the second cohomology except at special values of α at which a vanishing cycle collapses and the second Betti number of $\mathcal{M}(\alpha; p)$ is smaller. For a description of the values of α at which this happens for $G = SU(2)$, see [55]. Alternatively, if one considers the cohomology of the “stack” of parabolic bundles (rather than the moduli space $\mathcal{M}(\alpha; p)$ of stable parabolic bundles), no jumping occurs. This is actually the right thing for gauge theory, since the starting point is the space of all gauge fields, which is a differential-geometric analog of what in algebraic geometry is the stack of all bundles or parabolic bundles. (In sigma models, branes supported on a vanishing cycle do not disappear when the vanishing cycle collapses; they simply become branes supported on the resulting singularity.) In down to earth terms, the drop that occurs in the second cohomology of $\mathcal{M}_{\mathbb{L}}(\alpha; p)$ at certain values of α is inessential for our applications, for the following reasons. When we describe the cohomology classes of the symplectic forms ω_I , ω_J , and ω_K of \mathcal{M}_H , the jumping just means that certain periods must vanish at certain values of α , as will be manifest in the formula we give. Alternatively, when we describe the action of the affine Weyl group on the cohomology of \mathcal{M}_H , we will avoid the bad values of the parameters.

The above description of the second cohomology of $\mathcal{M}(\alpha; p)$ will make it possible to usefully describe the cohomology class of the symplectic form of this space. The symplectic form

$$(3.49) \quad \omega = -\frac{1}{4\pi} \int_C \text{Tr } \delta A \wedge \delta A$$

of $\mathcal{M}(\alpha; p)$ takes values in $H^2(\mathcal{M}(\alpha; p); \mathbb{R}) = H^2(\mathcal{M}(\alpha; p); \mathbb{Z}) \otimes \mathbb{R} = \mathbb{R} \oplus \mathfrak{t}^\vee$. Its cohomology class is described in Theorem 3.2 and Proposition 3.7 of [56]; for a proof for $SU(2)$ using gluing arguments, see [57], and for a more general argument based on realizing $\mathcal{M}(\alpha; p)$ as a symplectic quotient with α as a parameter in the moment map, see [58]. We will explain the formula after a few preliminaries.

The Affine Weyl Group

We recall that, up to equivalence, α takes values in \mathbb{T}/\mathcal{W} , where \mathcal{W} is the Weyl group. Alternatively, we can lift α to \mathfrak{t} , but then there is an equivalence in transforming α by elements of the cocharacter lattice Λ_{cochar} . Since we have assumed that G is simply-connected, the cocharacter lattice is the same as the coroot lattice Λ_{cort} . The combined group

$$(3.50) \quad \mathcal{W}_{\text{aff}} = \Lambda_{\text{cort}} \rtimes \mathcal{W}$$

of lattice shifts and Weyl transformations is known as the affine Weyl group. It is the Weyl group of the Kac-Moody or affine Lie algebra of G . So we can think of α as taking values in $\mathfrak{t}/\mathcal{W}_{\text{aff}}$.

\mathfrak{t} can be usefully divided as follows into fundamental domains for the action of \mathcal{W}_{aff} . On certain hyperplanes in \mathfrak{t} , $U = \exp(-2\pi\alpha)$ is non-regular (it commutes with a nonabelian subgroup of G). Each such hyperplane is a locus of fixed points of some element of \mathcal{W}_{aff} . These hyperplanes divide \mathfrak{t} into fundamental domains for the action of the affine Weyl group.

For example, in the case of $G = SU(N)$, generalizing (3.17), we can think of α as a diagonal matrix $i \text{diag}(y_1, \dots, y_N)$, with $\sum_a y_a = 0$. \mathcal{W} acts by permutations, and Λ_{cort} acts by integer shifts preserving the vanishing of the sum of the y_a . The condition for $U = \exp(-2\pi\alpha)$ to be non-regular is that $n = y_a - y_b$ is an integer for some a and b , in which case α is invariant under the affine Weyl transformation $y_a \rightarrow y_b + n$, $y_b \rightarrow y_a - n$. The hyperplanes $n = y_a - y_b$ divide \mathfrak{t} into fundamental domains of \mathcal{W}_{aff} .

The ordinary Weyl group has, once we pick a set of positive roots, a distinguished fundamental domain in \mathfrak{t} called the positive Weyl chamber. It is the region in which $\langle \alpha, w \rangle > 0$ for every fundamental weight w . For $G = SU(N)$, this chamber is described by $y_1 \geq y_2 \geq \dots \geq y_N$. Upon intersecting it with the above-mentioned hyperplanes, the positive Weyl chamber decomposes as a union of infinitely many fundamental domains for the affine Weyl group. There is a distinguished one, which we will call D , whose closure contains $\alpha = 0$. For $G = SU(N)$, D is characterized by $y_1 - y_N \leq 1$ (or $y_1 - y_N < 1$ if one wishes U to be regular), generalizing the condition $y < 1/2$ in (3.17).

Now we can explain the above-mentioned formula for the cohomology class of the symplectic form ω of $\mathcal{M}(\alpha; p)$. We write the formula for α in the distinguished affine Weyl chamber of \mathfrak{t} , as just described. Then the cohomology class $[\omega/2\pi]$ is

$$(3.51) \quad \left[\frac{\omega}{2\pi} \right] = e \oplus (-\alpha^*).$$

Here we use the fact that $[\omega/2\pi]$ takes values in $\mathbb{R} \oplus \mathfrak{t}^\vee$. In (3.51), $e \in H^2(\mathcal{M}; \mathbb{R}) \cong \mathbb{R}$ is the pullback to $\mathcal{M}(\alpha; p)$ of the first Chern class of the determinant line bundle $\mathfrak{L} \rightarrow \mathcal{M}$ (equivalently, the pullback of the cohomology class of the symplectic form $\omega/2\pi$ of \mathcal{M}). And in the second summand on the right, $\alpha^* \in \mathfrak{t}^\vee$ is (as in section 2.4) the image of $\alpha \in \mathfrak{t}$ under the map from \mathfrak{t} to \mathfrak{t}^\vee that comes from the quadratic form $-\text{Tr}$.

Rather than summarize here the arguments of [57], [58] leading to this formula, we give a brief explanation using four-dimensional gauge theory. To determine the cohomology class of ω , we need to compute periods $\int_\Sigma \omega$, for $\Sigma \subset \mathcal{M}(\alpha; p)$ a closed two-manifold. Given a choice of Σ , we set M to be the four-manifold $M = \Sigma \times C$, and make our usual

construction on M . Let $\Sigma_p = \Sigma \times p$. Each point in Σ determines a flat G -bundle over $C \setminus p$ with monodromy around $p \in C$; these fit together to a G -bundle $E \rightarrow M \setminus \Sigma_p$ with a monodromy around Σ_p . (E may only exist as a G_{ad} -bundle, but this does not affect the following derivation.) The relation between the symplectic form of $\mathcal{M}(\alpha; p)$ and gauge theory in four dimensions is

$$(3.52) \quad \int_{\Sigma} \frac{\omega}{2\pi} = \int_M \frac{\text{Tr } F \wedge F}{8\pi^2}.$$

This is shown exactly like the corresponding statement without parabolic structure; see eqn. (4.18) of [8]. Now using (2.41) (and observing that $\Sigma_p \cap \Sigma_p = 0$) we have

$$(3.53) \quad \int_{\Sigma} \frac{\omega}{2\pi} = \int_{\Sigma_p} \text{Tr } \alpha \frac{F}{2\pi} = -\langle \alpha^*, \mathfrak{m} \rangle \text{ mod } \mathbb{Z},$$

where $\mathfrak{m} \in H^2(\Sigma_p, \pi_1(\mathbb{T})) \cong \Lambda_{\text{cort}}$ is the cohomology class of $F/2\pi$, and $\langle \cdot, \cdot \rangle$ is the natural pairing between \mathfrak{t} and \mathfrak{t}^{\vee} . This is equivalent to (3.51). We also see that the homology cycles with which the symplectic form can naturally be paired are labeled by Λ_{cort} , the dual of Λ_{wt} .

This description of the symplectic form of $\mathcal{M}(\alpha; p)$ has an obvious similarity to our claims in section 3.2 about Higgs bundles. For Higgs bundles, we expect the cohomology classes such as $[\omega_I]$ to vary linearly, as claimed in (3.8), while for ordinary bundles we have the linearity seen in (3.51). We explain the relation between these results in section 3.6.

However, in the case of Higgs bundles, the argument leading to (3.8) is valid for all α , while the analogous statement (3.51) for bundles holds only for α in a fundamental affine Weyl chamber. For bundles, there is no obvious way to continue the formula (3.51) beyond the fundamental affine Weyl chamber, since on the boundary of \mathbb{D} , the manifold $\mathcal{M}(\alpha; p)$ collapses to a manifold of lower dimension, as we will see in section 3.5. For Higgs bundles, as we describe in sections 3.6 and 3.7, we can take $\beta, \gamma \neq 0$ and smoothly continue beyond the boundaries of the affine Weyl chamber.

In this discussion, we have implicitly taken the Levi subgroup to be $\mathbb{L} = \mathbb{T}$. However, the same result holds for any \mathbb{L} , with the same derivation; one merely has to restrict α to be \mathbb{L} -invariant.

The Affine Weight Lattice

Let us reconsider the description (3.47) of the second cohomology of $\mathcal{M}(\alpha; p)$ for the case $\mathbb{L} = \mathbb{T}$:

$$(3.54) \quad H^2(\mathcal{M}(\alpha; p); \mathbb{Z}) = \mathbb{Z} \oplus \Lambda_{\text{wt}}.$$

We can describe this by saying that $H^2(\mathcal{M}(\alpha; p); \mathbb{Z})$ is the affine weight lattice of G , that is, the weight lattice of the affine Lie algebra or Kac-Moody algebra or centrally extended loop group associated to G . This description is natural in the existing mathematical theory [54], where

(3.54) is obtained via the theory of loop groups. We will sketch the idea in a physical language in the context of Chern-Simons gauge theory in 2+1 dimensions [59]. Let k be a positive integer, and let $w \in \Lambda_{wt}$ be the highest weight of an integrable representation of the Kac-Moody algebra of G at level k . Using the quadratic form $-\text{Tr}$ to identify $\mathfrak{t}^\vee = {}^L\mathfrak{t}$ with \mathfrak{t} , $w \in {}^L\mathfrak{t}$ maps to an element $w^* \in \mathfrak{t}$. We consider Chern-Simons gauge theory on C with a single marked point p labeled by a representation of G with highest weight w . The classical phase space is $\mathcal{M}(\alpha; p)$ with $\alpha = w^*/k$. Quantization is carried out by taking the global sections of a line bundle \mathfrak{L}^k whose first Chern class is represented in de Rham cohomology by $k\omega/2\pi$ where $\omega = -\frac{1}{4\pi} \int_C \text{Tr} \delta A \wedge \delta A$. (\mathfrak{L}^k is not really the k^{th} power of a line bundle unless $w = 0$, but we write it as \mathfrak{L}^k because of the factor of k in the rational first Chern class.) Consequently, the class $k\omega/2\pi$ in de Rham cohomology lifts to an element in the lattice $H^2(\mathcal{M}(\alpha; p); \mathbb{Z})$. Accordingly, this lattice must contain a point corresponding to the pair (k, w) , that is, it must contain the affine weight lattice of G . These points have distinct images in de Rham cohomology, in view of the formula (3.51) for the cohomology class of $[\omega/2\pi]$. At this point, it is also clear that the differentials in the spectral sequence (3.40) for $H^2(\mathcal{M}(\alpha; p); \mathbb{Z})$ do vanish (as we assumed in our above discussion), or that cohomology group could not contain the affine weight lattice.

Since $H^2(\mathcal{M}(\alpha; p); \mathbb{Z})$ is the affine weight lattice of G , it admits a natural action of the affine Weyl group of G , although this action has no evident meaning in terms of the geometry of $\mathcal{M}(\alpha; p)$. In section 3.7, we will explain more conceptually why $H^2(\mathcal{M}(\alpha; p); \mathbb{Z})$ admits this action of the affine Weyl group. We will also describe the action more precisely.

What happens if we relax the assumption that G is simply-connected? The argument via Kac-Moody algebras shows for any G , not necessarily simply-connected, that the lattice $H^2(\mathcal{M}(\alpha; p; G); \mathbb{Z})$ modulo torsion contains the affine character lattice of G as a sublattice. By the affine character lattice of a simple but perhaps not simply-connected group G , we mean

$$(3.55) \quad \Lambda_{\text{aff char}} = \mathbb{Z} \oplus \Lambda_{\text{char}}(G),$$

where \mathbb{Z} classifies central extensions²⁰ of the loop group of G . We suspect that for generic α , $H^2(\mathcal{M}(\alpha; p; G); \mathbb{Z}) \bmod \text{torsion}$ is always precisely this lattice:

$$(3.56) \quad H^2(\mathcal{M}(\alpha; p)) = \mathbb{Z} \oplus \Lambda_{\text{char}}(G).$$

²⁰Some care is needed here. Let \overline{G} be the universal cover of G . Related to the fact that instanton number is \mathbb{Z} -valued for \overline{G} but not for G , not every central extension of the loop group of \overline{G} corresponds to one for G . The summand \mathbb{Z} in (3.55) is naturally understood as a proper subgroup of the summand \mathbb{Z} in (3.54).

At any rate, we will show in section 3.7 that a natural symmetry group of this lattice acts on $H^2(\mathcal{M}(\alpha, p; G); \mathbb{Z})$. It will turn out that, if we restrict ourselves to a connected component of $\mathcal{M}(\alpha, p; G)$, then the group that acts on the cohomology is the same affine Weyl group whether G is simply-connected or not.

Alternative Point Of View

The approximate fibration (3.39) can also be seen from a purely topological or symplectic point of view, without mentioning parabolic structure (for example, see the brief summary in [56]). Pick on the Riemann surface C a standard set of A - and B -cycles. Let A_i and B_i be the monodromies of a flat G -bundle around these cycles. They obey

$$(3.57) \quad \prod_{i=1}^g [A_i, B_i] = 1,$$

where $[A, B] = ABA^{-1}B^{-1}$. The moduli space \mathcal{M} of flat bundles is the space of solutions of this equation, modulo conjugation. Pick an element U of G that is close to the identity and replace (3.57) by

$$(3.58) \quad \prod_{i=1}^g [A_i, B_i] = U.$$

A solution of this equation describes a flat bundle over $C \setminus p$ with monodromy U around p . As long as we do not encounter singularities of \mathcal{M} , the solution spaces of (3.57) and (3.58) are topologically the same, for U sufficiently close to 1. Now suppose that we want to specify only the conjugacy class of U , which we assume to be semi-simple and, again, sufficiently close to 1. Then the conjugacy class of U contains an element $\exp(-2\pi\alpha)$, where α is an element of \mathfrak{t} close to the origin. Let \mathbb{L} be the subgroup of G that commutes with α or equivalently with U ; then \mathbb{L} is a Levi subgroup. The possible choices of U in its conjugacy class are parametrized by a copy of G/\mathbb{L} , and so the solution space of (3.58) is a G/\mathbb{L} -bundle over the solution space of (3.57). This remains so after dividing both spaces by conjugation by G , so we arrive again at the fibration (3.39):

$$(3.59) \quad \begin{array}{ccc} G/\mathbb{L} & \rightarrow & \mathcal{M}_{\mathbb{L}}(\alpha; p) \\ & & \downarrow \\ & & \mathcal{M}. \end{array}$$

Suppose that we want to delete points p_1, \dots, p_s from C , and denote as U_1, \dots, U_s the monodromies about these points. The analog of (3.57) is

$$(3.60) \quad \prod_{i=1}^g [A_i, B_i] = \prod_{a=1}^s U_a.$$

If the U_a are close enough to the identity and commute with Levi subgroups \mathbb{L}_a , a similar reasoning to the above leads to a fibration with a factor of G/\mathbb{L}_a for each puncture (and this for instance leads to the description of the second cohomology of $\mathcal{M}(\alpha_1, p_1; \dots; \alpha_s, p_s)$ in eqn. (3.48)). However, we want to consider another issue aimed at later applications. Let y_a , $a = 1, \dots, s$ take values in the center of G and suppose that

$$(3.61) \quad \prod_{a=1}^s y_a = 1.$$

Then eqn. (3.60) is completely invariant under

$$(3.62) \quad U_a \rightarrow y_a U_a, \quad a = 1, \dots, s.$$

If U_a is conjugate to $\exp(-2\pi\alpha_a)$ for some $\alpha_a \in \mathfrak{t}$, and $y_a = \exp(-2\pi u_a)$ for some $u_a \in \Lambda_{\text{cowt}}$, then the transformation (3.62) amounts to

$$(3.63) \quad \alpha_a \rightarrow \alpha_a + u_a.$$

And (3.61) is equivalent to

$$(3.64) \quad \sum_a u_a \in \Lambda_{\text{cochar}} \subset \Lambda_{\text{cowt}}.$$

This is an illustration of the situation that was described in eqns. (2.6) and (2.7). The shifts $\alpha_a \rightarrow \alpha_a + u_a$ individually would shift the characteristic class ξ of the G_{ad} bundle derived from E , and the condition (3.61) or (3.64) ensures that globally ξ is actually unchanged.

Let us return now for simplicity to the case of one puncture. If $U = \exp(-2\pi\alpha)$ is regular, which we can achieve by placing α in the interior of the distinguished affine Weyl chamber \mathbb{D} , then the fibration (3.59) takes the form

$$(3.65) \quad \begin{array}{ccc} G/\mathbb{T} & \rightarrow & \mathcal{M}(\alpha; p) \\ & & \downarrow \\ & & \mathcal{M}. \end{array}$$

Now suppose that α approaches a boundary point $\bar{\alpha}$ of \mathbb{D} . Then U ceases to be regular, and its orbit under conjugation is a copy of G/\mathbb{L} for some \mathbb{L} , rather than G/\mathbb{T} . So now the fibration takes the form in (3.59):

$$(3.66) \quad \begin{array}{ccc} G/\mathbb{L} & \rightarrow & \mathcal{M}_{\mathbb{L}}(\bar{\alpha}; p) \\ & & \downarrow \\ & & \mathcal{M}. \end{array}$$

What happens as $\alpha \rightarrow \bar{\alpha}$? G/\mathbb{T} is fibered over G/\mathbb{L} with fiber \mathbb{L}/\mathbb{T} . As $\alpha \rightarrow \bar{\alpha}$, the orbit of U collapses from a copy of G/\mathbb{T} to a copy of G/\mathbb{L}

and we get a fibration

$$(3.67) \quad \begin{array}{ccc} \mathbb{L}/\mathbb{T} & \rightarrow & \mathcal{M}(\alpha; p) \\ & & \downarrow \\ & & \mathcal{M}_{\mathbb{L}}(\bar{\alpha}; p). \end{array}$$

The fibers \mathbb{L}/\mathbb{T} are symplectic manifolds, and they collapse to points as $\alpha \rightarrow \bar{\alpha}$. So this is what happens to $\mathcal{M}(\alpha; p)$ as α approaches a non-regular value $\bar{\alpha}$: it collapses to a variety of lower dimension, with vanishing cycles \mathbb{L}/\mathbb{T} .

Generically, a boundary point $\bar{\alpha}$ is contained in only one of the hyperplanes that mark the boundary of \mathbb{D} , and $\mathbb{L}/\mathbb{T} = SU(2)/U(1) = \mathbb{C}\mathbb{P}^1 = S^2$. For example, for $G = SU(N)$, we have $\alpha = i \operatorname{diag}(y_1, y_2, \dots, y_N)$, with $y_1 \geq y_2 \geq y_3 \dots \geq y_N$, $y_1 - y_N \leq 1$, and $\sum y_a = 1$. At a generic boundary point, precisely one of the inequalities is an equality, and then $\mathbb{L} = U(2) \times U(1)^{N-3}$, and $\mathbb{L}/\mathbb{T} = U(2)/U(1)^2 = \mathbb{C}\mathbb{P}^1 = S^2$. So as α approaches a generic boundary point of \mathbb{D} , $\mathcal{M}(\alpha; p)$ is fibered by vanishing two-spheres, that is, two-spheres that shrink to points. In general, the rank of the semi-simple part of \mathbb{L} is the number of boundary hyperplanes that contain $\bar{\alpha}$.

More generally, we can consider a pair of Levi subgroups \mathbb{L}_1 and \mathbb{L}_2 , with \mathbb{L}_1 a proper subgroup of \mathbb{L}_2 but not necessarily equal to \mathbb{T} . The same sort of reasoning as above applies. If $U = \exp(-2\pi\alpha_1)$, where α_1 is \mathbb{L}_1 -regular, we get a fibration

$$(3.68) \quad \begin{array}{ccc} G/\mathbb{L}_1 & \rightarrow & \mathcal{M}_{\mathbb{L}_1}(\alpha_1; p) \\ & & \downarrow \\ & & \mathcal{M}. \end{array}$$

If α_2 is \mathbb{L}_2 -regular, then for $U = \exp(-2\pi\alpha_2)$, the fibration looks like

$$(3.69) \quad \begin{array}{ccc} G/\mathbb{L}_2 & \rightarrow & \mathcal{M}_{\mathbb{L}_2}(\alpha_2; p) \\ & & \downarrow \\ & & \mathcal{M}. \end{array}$$

As α_1 approaches α_2 through \mathbb{L}_1 -regular values, the orbit G/\mathbb{L}_1 of U degenerates to G/\mathbb{L}_2 . G/\mathbb{L}_1 maps to G/\mathbb{L}_2 with fiber $\mathbb{L}_2/\mathbb{L}_1$, and for α_1 approaching α_2 we get, away from singularities, a fibration

$$(3.70) \quad \begin{array}{ccc} \mathbb{L}_2/\mathbb{L}_1 & \rightarrow & \mathcal{M}_{\mathbb{L}_1}(\alpha_1; p) \\ & & \downarrow \\ & & \mathcal{M}_{\mathbb{L}_2}(\alpha_2; p). \end{array}$$

The fibers $\mathbb{L}_2/\mathbb{L}_1$ are vanishing cycles for $\alpha_1 \rightarrow \alpha_2$.

As one might expect, there is also a natural description of the limiting behavior in terms of complex geometry. The pair (\mathbb{L}_1, α_1) determines

a parabolic subgroup \mathcal{P}_1 , and the pair (\mathbb{L}_2, α_2) determines a parabolic subgroup \mathcal{P}_2 that contains \mathcal{P}_1 . From a holomorphic point of view, according to the theorem of Mehta and Seshadri, $\mathcal{M}_{\mathbb{L}_1}(\alpha_1; p)$ or $\mathcal{M}_{\mathbb{L}_2}(\alpha_2; p)$ parametrizes bundles with parabolic structure at p of type \mathcal{P}_1 or \mathcal{P}_2 , respectively. After reducing the structure group of a bundle to \mathcal{P}_2 , the ways of further reducing it to the subgroup \mathcal{P}_1 are parameterized by $\mathcal{P}_2/\mathcal{P}_1$. So from a holomorphic point of view, we get a fibration

$$(3.71) \quad \begin{array}{ccc} \mathcal{P}_2/\mathcal{P}_1 & \rightarrow & \mathcal{M}_{\mathbb{L}_1}(\alpha_1; p) \\ & & \downarrow \\ & & \mathcal{M}_{\mathbb{L}_2}(\alpha_2; p). \end{array}$$

This is in accord with (3.70), since $\mathcal{P}_2/\mathcal{P}_1 = \mathbb{L}_2/\mathbb{L}_1$. This fibration is Proposition 3.4 in [60].

3.6. Topology Of \mathcal{M}_H . In section 3.5, we explored the topology of the moduli space $\mathcal{M}(\alpha; p)$ of bundles with parabolic structure. Here we will consider the analogous questions for Higgs bundles.

If we simply set $\phi = 0$, then Hitchin's equations reduce to the equations $F = 0$ for a flat unitary G -bundle. The moduli space of such flat bundles is isomorphic, by a theorem of Narasimhan and Seshadri mentioned above [51], to the moduli space of stable G -bundles over C , which we call $\mathcal{M}(G, C)$, or \mathcal{M} when the context is clear.

The moduli space \mathcal{M}_H , viewed in complex structure I , parametrizes stable pairs (E, φ) . We can define a "foliation" of \mathcal{M}_H by forgetting φ and remembering only the holomorphic type of E . For a generic stable (or semistable) pair (E, φ) , E is a stable (or semistable) bundle, and the foliation gives a meromorphic map $\psi : \mathcal{M}_H \rightarrow \mathcal{M}$ (in [8], this was called Hitchin's second fibration). The fiber of this map is a linear space parametrized by $\varphi \in H^0(C, K \otimes \text{ad}(E))$, which (if E is stable) is the cotangent space to \mathcal{M} at the point defined by E . Moreover, the map ψ has a natural section (holomorphic in complex structure I), because we can embed \mathcal{M} in \mathcal{M}_H as the space of solutions of Hitchin's equations with $\varphi = 0$. So birationally in complex structure I , \mathcal{M}_H is the cotangent bundle $T^*\mathcal{M}$:

$$(3.72) \quad \mathcal{M}_H \cong T^*\mathcal{M}.$$

More specifically, \mathcal{M}_H contains $T^*\mathcal{M}$ as a dense open set. Somewhat like the fibrations considered in section 3.5, this gives a useful first approximation to the topology of \mathcal{M}_H if the genus g_C of C and the rank r of G are large enough. For example, it is good enough for discussing the second cohomology of \mathcal{M}_H if $g_C \geq 2$.

This has an analog for ramified Higgs bundles if the parameter α is \mathbb{L} -regular. Given a point in $\mathcal{M}_{H, \mathbb{L}}(\alpha, \beta, \gamma; p)$ associated with a stable or semistable ramified Higgs bundle (E, φ) , we can forget φ and simply think of E as a $G_{\mathbb{C}}$ -bundle with the parabolic structure determined by

α . Generically, this parabolic bundle is stable, and thus defines a point in $\mathcal{M}_{\mathbb{L}}(\alpha; p)$. So we get a meromorphic map $\psi : \mathcal{M}_{H, \mathbb{L}}(\alpha, \beta, \gamma; p) \rightarrow \mathcal{M}_{\mathbb{L}}(\alpha; p)$. If $\beta = \gamma = 0$, the story is exactly as it was above. The fiber of the map ψ is a space of Higgs fields with a nilpotent pole that takes values in \mathfrak{n} , the Lie algebra of the unipotent radical of the parabolic subgroup \mathcal{P} determined by (\mathbb{L}, α) . This is precisely the cotangent space to $\mathcal{M}_{\mathbb{L}}(\alpha; p)$. Moreover, $\mathcal{M}_{\mathbb{L}}(\alpha; p)$ can be embedded holomorphically in $\mathcal{M}_{H, \mathbb{L}}(\alpha, 0, 0; p)$ as the space of solutions of Hitchin's equations with $\varphi = 0$. We can think of this as the zero section of the cotangent bundle. So $\mathcal{M}_{H, \mathbb{L}}(\alpha, 0, 0; p)$ contains the cotangent bundle to $\mathcal{M}_{\mathbb{L}}(\alpha; p)$ as a dense open subspace

$$(3.73) \quad \mathcal{M}_{H, \mathbb{L}}(\alpha, 0, 0; p) \cong T^* \mathcal{M}_{\mathbb{L}}(\alpha; p),$$

and the two are birational in complex structure I .

For $\beta, \gamma \neq 0$, this requires some modification. We can still forget φ and thus define the meromorphic map

$$(3.74) \quad \psi : \mathcal{M}_{H, \mathbb{L}}(\alpha, \beta, \gamma; p) \rightarrow \mathcal{M}_{\mathbb{L}}(\alpha; p).$$

However, we cannot set φ to zero, since its polar part has eigenvalues determined by β and γ . So the map ψ has no holomorphic section. Hence, $\mathcal{M}_{H, \mathbb{L}}(\alpha, \beta, \gamma; p)$ is not birational to a vector bundle over $\mathcal{M}_{\mathbb{L}}(\alpha; p)$, but to an ‘‘affine bundle.’’ This means that the fibers of ψ are copies of \mathbb{C}^N for some N , but the structure group of the fibration is a group of affine transformations $x \rightarrow ax + b$ (not just linear transformations $x \rightarrow ax$). Related to this, the polar part of φ is not nilpotent, and so φ does not represent a cotangent vector to $\mathcal{M}_{H, \mathbb{L}}(\alpha; p)$. But if φ and $\tilde{\varphi}$ are two Higgs fields (with the same bundle E) then, as we saw in eqn. (3.26), their difference $\varphi - \tilde{\varphi}$ has a polar part valued in \mathfrak{n} and hence represents a cotangent vector to $\mathcal{M}_H(\alpha; p)$. The upshot is that $\mathcal{M}_H(\alpha, \beta, \gamma; p)$, for general β and γ , and regular α , contains a dense open set that is an ‘‘affine deformation’’ of the cotangent bundle of $\mathcal{M}(\alpha; p)$.

In making this statement, we require α to be \mathbb{L} -regular, since otherwise $\mathcal{M}(\alpha; p)$ collapses to a manifold of lower dimension. However, as long as β and γ are generic, the topology of $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ is independent of α whether α is regular or not [24], so we get a rough description of the topology for any α and generic β, γ . The topology does change in real codimension three when the triple (α, β, γ) is non-regular, as we explain presently.

An immediate application of the relation between \mathcal{M}_H and \mathcal{M} is that we can describe the second cohomology of \mathcal{M}_H and determine the unknown constants in the formula (3.8) for the cohomology class of the symplectic form ω_I . First of all, the second cohomology of \mathcal{M}_H is isomorphic to that of \mathcal{M} , because \mathcal{M}_H , being an affine bundle, is contractible to \mathcal{M} (away from a codimension that is too high to affect the second cohomology, barring special cases of small genus and rank

that we will not consider). So

$$(3.75) \quad H^2(\mathcal{M}_H; \mathbb{Z}) = \mathbb{Z} \oplus \Lambda_{wt}$$

is the affine weight lattice of G , just like $H^2(\mathcal{M}; \mathbb{Z})$. Also, according to eqn. (3.4), $\omega_I = \psi^*(\omega)$, where ω is the usual symplectic form of \mathcal{M} , whose cohomology class was described in eqn. (3.51). So we can simply borrow the result of eqn. (3.51) and write

$$(3.76) \quad \left[\frac{\omega_I}{2\pi} \right] = e \oplus (-\alpha^*).$$

(This formula actually holds for every \mathbb{L} ; it makes sense because α^* always takes values in the \mathbb{L} -invariant part of $\Lambda_{wt} \otimes_{\mathbb{Z}} \mathbb{R}$.) From (3.9) and (3.10), we have therefore

$$(3.77) \quad \left[\frac{\omega_J}{2\pi} \right] = 0 \oplus (-\beta^*), \quad \left[\frac{\omega_K}{2\pi} \right] = 0 \oplus (-\gamma^*).$$

After some discussion of the singularities of $\mathcal{M}_H(\alpha, \beta, \gamma; p)$, we will be able to draw some interesting conclusions from these formulas.

Singularity At Special Values Of (α, β, γ)

Even without ramification, the moduli space \mathcal{M} of $G_{\mathbb{C}}$ bundles and the moduli space \mathcal{M}_H of Higgs bundles can have singularities at points that correspond to reducible bundles or Higgs bundles. Such singularities involve the global behavior on C , and occur in high codimension if the genus of C is large.

Parabolic structure at a point $p \in C$ introduces a new kind of singularity, which depends only on the parameters characterizing the ramification. These parameters are $\alpha \in \mathbb{T}$ in the case of bundles, or the trio $(\alpha, \beta, \gamma) \in \mathbb{T} \times \mathfrak{t} \times \mathfrak{t}$ in the case of Higgs bundles. These singularities depend only on the local behavior at p and their codimension is independent of the genus of C . They will be much more prominent in applications to the geometric Langlands program than the singularities that are global in nature.

For bundles, we described this kind of singularity in section 3.5. The result is described in eqns. (3.67) and (3.71). Suppose that α approaches a value $\bar{\alpha}$ that is not \mathbb{L} -regular, but is \mathbb{L}' -regular, where \mathbb{L}' is a Levi group that contains \mathbb{L} as a proper subgroup. And let the corresponding parabolic subgroups be \mathcal{P} and \mathcal{P}' . Then as $\alpha \rightarrow \bar{\alpha}$, $\mathcal{M}_{\mathbb{L}}(\alpha; p)$ is fibered by vanishing cycles of the form \mathbb{L}'/\mathbb{L} (or \mathcal{P}'/\mathcal{P}). The fibers collapse everywhere, so the codimension of the singularity is zero. The picture looks like

$$(3.78) \quad \begin{array}{ccc} \mathcal{P}'/\mathcal{P} & \rightarrow & \mathcal{M}_{\mathbb{L}}(\alpha; p) \\ & & \downarrow \\ & & \mathcal{M}_{\mathbb{L}'}(\bar{\alpha}; p), \end{array}$$

with the fibers collapsing as $\alpha \rightarrow \bar{\alpha}$. This is a local singularity, in the sense that it has only to do with the value of α and the behavior near the point p . It has nothing to do with a global problem such as finding a semistable bundle or a reducible flat connection.

Now what does this imply for ramified Higgs bundles? First let us see what happens if β and γ have the most special possible values, namely 0. But we begin with an \mathbb{L} -regular value of α . (See [61] for a mathematical discussion of these issues, justifying many statements below from a different point of view.) As we have discussed, $\mathcal{M}_{H,\mathbb{L}}(\alpha, 0, 0; p)$ is generically the cotangent bundle of $\mathcal{M}_{\mathbb{L}}(\alpha; p)$. Likewise, if \mathbb{L}' is a Levi subgroup that properly contains \mathbb{L} and $\bar{\alpha}$ is \mathbb{L}' -regular, then $\mathcal{M}_{H,\mathbb{L}'}(\bar{\alpha}, 0, 0; p)$ is generically the cotangent bundle of $\mathcal{M}_{\mathbb{L}'}(\bar{\alpha}; p)$. To go from parabolic bundles to ramified Higgs bundles, we just replace everything in (3.78) with its cotangent bundle. So for α near $\bar{\alpha}$, $\mathcal{M}_{H,\mathbb{L}}(\alpha, 0, 0; p)$ is generically fibered over $\mathcal{M}_{H,\mathbb{L}'}(\bar{\alpha}, 0, 0; p)$:

$$(3.79) \quad \begin{array}{ccc} T^*(\mathcal{P}'/\mathcal{P}) = T^*(\mathbb{L}'/\mathbb{L}) & \rightarrow & \mathcal{M}_{H,\mathbb{L}}(\alpha, 0, 0; p) \\ & & \downarrow \\ & & \mathcal{M}_{H,\mathbb{L}'}(\bar{\alpha}, 0, 0; p). \end{array}$$

We have used the fact that $\mathcal{P}'/\mathcal{P} = \mathbb{L}'/\mathbb{L}$. When α approaches $\bar{\alpha}$, \mathbb{L}'/\mathbb{L} becomes a “vanishing cycle” and collapses to a point. The codimension of \mathbb{L}'/\mathbb{L} inside its cotangent bundle is equal to the dimension of \mathbb{L}'/\mathbb{L} . So this is the codimension of the vanishing cycle that $\mathcal{M}_{H,\mathbb{L}}(\alpha, 0, 0; p)$ acquires as α approaches a nonregular value $\bar{\alpha}$. At $\alpha = \bar{\alpha}$, the vanishing cycle collapses to a point, and \mathcal{M}_H acquires a singularity of whose codimension is twice as great. We call this a local singularity, since it only depends on the local behavior near p (which is determined by the parameters (α, β, γ)) and not on solving any global problem. For example, if $\mathbb{L} = \mathbb{T}$ and if $\bar{\alpha}$ is a generic nonregular value, then $\mathbb{L}'/\mathbb{L} = \mathbb{C}\mathbb{P}^1 = S^2$, as we have seen in section 3.5. So the vanishing cycle in this case has real codimension two.

Because $\mathcal{M}_{H,\mathbb{L}}(\alpha, 0, 0; p)$ is hyper-Kähler, the geometry near the vanishing cycle is highly constrained. The highly curved geometry near an almost vanishing cycle (for α near $\bar{\alpha}$) must itself be hyper-Kähler, in order for it to be possible for $\mathcal{M}_{H,\mathbb{L}}(\alpha, 0, 0; p)$ to be hyper-Kähler. It is possible to see explicitly how this happens. In fact, a family of hyper-Kähler metrics on $T^*(\mathbb{L}'/\mathbb{L})$ can be constructed [34] using Nahm’s or Hitchin’s equations. This is part of the construction of hyper-Kähler metrics on coadjoint orbits of the complex Lie group $\mathbb{L}'_{\mathbb{C}}$; we review this construction in section 3.8.

Rather than use this full machinery, we will consider in some detail the case that $\mathbb{L} = \mathbb{T}$, and $\bar{\alpha}$ is a generic non-regular value, so that

$\mathbb{L}'/\mathbb{L} = \mathbb{C}\mathbb{P}^1$. In this case, the relevant family of hyper-Kähler metrics on $T^*(\mathbb{L}'/\mathbb{L}) = T^*(\mathbb{C}\mathbb{P}^1)$ is a more familiar family of metrics first constructed by Calabi and by Eguchi and Hansen.

As the area of $\mathbb{C}\mathbb{P}^1$ converges to zero, $T^*\mathbb{C}\mathbb{P}^1$ converges, in the Calabi-Eguchi-Hansen metric, to $\mathbb{R}^4/\mathbb{Z}_2 = \mathbb{C}^2/\mathbb{Z}_2$, that is, to an A_1 singularity. Hence, in the limit $\alpha = \bar{\alpha}$, $\mathcal{M}_{H,\mathbb{L}}(\alpha, 0, 0; p)$ (which is the same as $\mathcal{M}_H(\alpha, 0, 0; p)$, since we have set $\mathbb{L} = \mathbb{T}$) has a family of A_1 singularities. The singular locus is precisely the hyper-Kähler manifold $\mathcal{M}_{H,\mathbb{L}'}(\bar{\alpha}, 0, 0; p)$, embedded inside $\mathcal{M}_H(\bar{\alpha}, 0, 0; p)$. This follows from the fibration (3.79), since when the fiber $T^*(\mathbb{L}'/\mathbb{L})$ degenerates to $\mathbb{C}^2/\mathbb{Z}_2$, which has an isolated singularity, the singular locus of $\mathcal{M}_H(\alpha, 0, 0; p)$ becomes a section of the fibration or in other words a copy of $\mathcal{M}_{H,\mathbb{L}'}(\bar{\alpha}, 0, 0; p)$. We will also explain below a slightly different approach to this result (see the discussion of eqn. (3.82)). If α is close to but not equal to $\bar{\alpha}$, or $\alpha = \bar{\alpha}$ but β and γ are not quite zero, then the local singularity is deformed or resolved, and the behavior near the singularity is described by the Calabi-Eguchi-Hansen metric. These assertions will hopefully become clear below and in section 3.8.

It will help to recall a few facts about the $\mathbb{R}^4/\mathbb{Z}_2$ singularity. If we single out one complex structure (which in our application corresponds to the complex structure I of \mathcal{M}_H), then $\mathbb{R}^4/\mathbb{Z}_2$ can be described as the complex singularity $a_1^2 + a_2^2 + a_3^2 = 0$. It can be deformed to the smooth complex manifold

$$(3.80) \quad a_1^2 + a_2^2 + a_3^2 = \epsilon.$$

It also can be resolved to make the cotangent bundle $T^*\mathbb{C}\mathbb{P}^1$, with the exceptional cycle having an area r . Moreover, the deformation and resolution can be made simultaneously. The Calabi-Eguchi-Hansen metric depends on the real parameter r as well as the complex parameter ϵ . The A_1 singularity appears precisely if $r = \epsilon = 0$; otherwise the manifold is smooth. The picture can be described very naturally by constructing $\mathbb{R}^4/\mathbb{Z}_2$ and its smooth deformations via hyper-Kähler quotients of a Euclidean space [62]. Alternatively, it is a special case of the construction based on Nahm's equations that we review in section 3.8.

Now let us apply this to \mathcal{M}_H . In complex structure I , α is a Kähler parameter, as asserted in the table in section 2.2. So when α is varied, \mathcal{M}_H can change only by a birational transformation, and this only when a vanishing cycle appears (as occurs when the triple (α, β, γ) becomes nonregular). Hence varying α will give us the blowup parameter r of the A_1 singularity. To see the complex parameter ϵ , we must vary β and γ .

Analysis For $G = SU(2)$

To see what happens in varying β and γ , we will, to keep things simple, take $G = SU(2)$. A point in \mathcal{M}_H corresponds in complex structure I to a Higgs bundle (E, φ) . φ takes values in $V = H^0(C, K_C \otimes \text{ad}(E) \otimes \mathcal{O}(p))$; thus, it is a section of $K_C \otimes \text{ad}(E)$ with a possible pole at p . The space V has dimension $3g_C$. φ is constrained to obey

$$(3.81) \quad \text{Tr } \varphi^2 = \text{Tr } \sigma^2 (dz/z)^2 + \dots,$$

where $\sigma = \frac{1}{2}(\beta + i\gamma)$.

Pick a basis b_i , $i = 1, 2, 3$, of the fiber of $\text{ad}(E)$ at p , normalized so that $\text{Tr } b_i b_j = \delta_{ij}$. For $i = 1, 2, 3$, pick an element $\varphi_i \in V$ whose polar part is $\varphi_i \sim b_i (dz/z)$. And complete the φ_i to a basis $\varphi_1, \dots, \varphi_{3g_C}$ of V , such that the φ_i , $i > 3$, have no pole at p . Now introduce complex parameters a_i , $i = 1, \dots, 3g$, and expand φ as $\varphi = \sum_{i=1}^{3g_C} a_i \varphi_i$. $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ is parametrized by the choice of a bundle E ($3g_C - 3$ parameters) and the coefficients a_i ($3g_C$ parameters), subject to the equation (3.81), so its dimension is $6g_C - 4$. The equation (3.81) tells us that

$$(3.82) \quad \sum_{i=1}^3 a_i^2 = \text{Tr } \sigma^2.$$

This equation describes the deformation of the A_1 singularity, with $\text{Tr } \sigma^2$ playing the role of the parameter ϵ in (3.80).

At $\sigma = 0$, we potentially recover the A_1 singularity, but now we must remember the parameter α that controls the Kahler structure. For generic α , we get the resolution of the A_1 singularity. The reason that this occurs is simply that when the polar part of φ is nonzero, it (plus the specification of the conjugacy class of α) fixes what α must be. But when φ has no pole, there is a family of possible choices of α that (since the conjugacy class of α is fixed) is parametrized by $\mathbb{C}\mathbb{P}^1$. Since the locus at which φ has no pole is precisely where the A_1 singularity would be, taking into account the choice of α replaces the A_1 singularity in each transverse slice with a copy of $\mathbb{C}\mathbb{P}^1$. This operation is the resolution of the singularity. If, however, $\alpha = 0$, then \mathcal{M}_H does develop a family of A_1 singularities.

If we set $\alpha = \sigma = 0$, then the locus of A_1 singularities is given by $a_1 = a_2 = a_3 = 0$. This condition ensures that φ has no pole, so that (E, φ) is an ordinary Higgs bundle, without ramified structure. Thus, the locus of A_1 singularities in this example is simply $\mathcal{M}_H \subset \mathcal{M}_H(0, 0, 0; p)$.

This is a special case of the general description of local singularities of $\mathcal{M}_{H, \mathbb{L}}(\alpha, \beta, \gamma; p)$. If we set (α, β, γ) to a triple $(\bar{\alpha}, \bar{\beta}, \bar{\gamma})$ that is not \mathbb{L} -regular but is \mathbb{L}' -regular, where \mathbb{L}' properly contains \mathbb{L} , then

$\mathcal{M}_{H,\mathbb{L}}(\bar{\alpha}, \bar{\beta}, \bar{\gamma}; p)$ contains $\mathcal{M}_{H,\mathbb{L}'}(\bar{\alpha}, \bar{\beta}, \bar{\gamma}; p)$ as a locus of local singularities. For given \mathbb{L} , all possible \mathbb{L}' that properly include \mathbb{L} can occur, including $\mathbb{L}' = G$.

3.7. Action Of The Affine Weyl Group. To keep things simple, we begin this section with the assumption that G is simply-connected and the Levi group is $\mathbb{L} = \mathbb{T}$.

The parameters (α, β, γ) in general take values in $\mathbb{T} \times \mathfrak{t} \times \mathfrak{t}$, modulo the action of the Weyl group \mathcal{W} . Equivalently, we can take a slightly different point of view and think of (α, β, γ) as taking values in $\mathfrak{t} \times \mathfrak{t} \times \mathfrak{t}$ modulo the action of the affine Weyl group \mathcal{W}_{aff} . We recall that \mathcal{W}_{aff} is an extension of \mathcal{W} by the coroot lattice of G :

$$(3.83) \quad 0 \rightarrow \Lambda_{\text{cort}} \rightarrow \mathcal{W}_{\text{aff}} \rightarrow \mathcal{W} \rightarrow 1.$$

We let \mathcal{W}_{aff} act on the trio (α, β, γ) by acting on α in the natural fashion, while acting on β and γ via the quotient \mathcal{W} . With this action of \mathcal{W}_{aff} , we have

$$(3.84) \quad (\mathfrak{t} \times \mathfrak{t} \times \mathfrak{t})/\mathcal{W}_{\text{aff}} = (\mathbb{T} \times \mathfrak{t} \times \mathfrak{t})/\mathcal{W},$$

simply because $\mathbb{T} = \mathfrak{t}/\Lambda_{\text{cort}}$ for simply-connected G . We call an element of a space acted on by \mathcal{W} regular if it is not left invariant by any element of \mathcal{W} except the identity, and likewise for an element of a space acted on by \mathcal{W}_{aff} . $(\alpha, \beta, \gamma) \in \mathfrak{t}^3$ is regular for the action of \mathcal{W}_{aff} if and only if its projection to $\mathbb{T} \times \mathfrak{t} \times \mathfrak{t}$ is regular for the action of \mathcal{W} .

Now before dividing by \mathcal{W}_{aff} , we would like to omit from $\mathfrak{t} \times \mathfrak{t} \times \mathfrak{t}$ the points on which $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ develops a singularity. We analyzed in section 3.6 the local singularities, which depend only on the behavior near p . These occur precisely when the triple $(\alpha, \beta, \gamma) \in \mathfrak{t} \times \mathfrak{t} \times \mathfrak{t}$ is non-regular for the action of \mathcal{W}_{aff} (in the discussion in section 3.6, the more natural criterion is the equivalent one that the projection to $\mathbb{T} \times \mathfrak{t} \times \mathfrak{t}$ is nonregular in the usual sense). This happens in real codimension three, because being invariant under some element $x \in \mathcal{W}_{\text{aff}}$ that has a nonzero image in \mathcal{W} places a non-trivial condition on each of α , β , and γ separately. (And an element of Λ_{cort} acts freely on α and hence has no fixed points at all.)

There also are global singularities, which arise at values of (α, β, γ) at which there are reducible solutions of Hitchin's equations. Even after we have suitably adjusted the triple (α, β, γ) so that such solutions exist, they occur on \mathcal{M}_H only in high codimension if the genus of C is large. (This contrasts with local singularities, which generically are A_1 singularities, of real codimension four, as we saw in section 3.6.) The values of (α, β, γ) at which these singularities occur can be described precisely [24]. They arise on a discrete set of affine linear spaces in \mathfrak{t}^3 of real codimension three. (Actually, the analysis shows that for the case of precisely one parabolic point, there are no global singularities for a regular triple (α, β, γ) , but for two or more parabolic points, there can

be global singularities with each triple being regular.) The basic reason that the global singularities are in real codimension three is the same as for the local singularities: the hyper-Kähler nature of \mathcal{M}_H . Viewing \mathcal{M}_H as a complex manifold in one of its complex structures, to obtain a singularity one must always adjust at least one complex parameter that controls the complex structure and one real parameter that controls the Kähler metric, making three real parameters in all.

Let us omit from \mathfrak{t}^3 all of the codimension three affine linear spaces on which a local or global singularity occurs. Call what remains X . X is connected and simply-connected, since \mathfrak{t}^3 is a linear space, and what we have omitted is of codimension three. The topology of $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ does not change when we vary the parameters without meeting a singularity. So it does not change if we vary the parameters in X . Since X is connected, the varieties $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ are independent of α, β , and γ topologically, as long as we restrict ourselves to $(\alpha, \beta, \gamma) \in X$. This statement is one of the main results of [24].

We can learn more by observing that the group \mathcal{W}_{aff} acts freely on X , since all triples $(\alpha, \beta, \gamma) \in X$ are regular. So $X/\mathcal{W}_{\text{aff}}$ is a smooth manifold, with fundamental group \mathcal{W}_{aff} . The cohomology of $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ varies as the fiber of a flat bundle over $X/\mathcal{W}_{\text{aff}}$. Taking the monodromy of the flat bundle, we get an action of \mathcal{W}_{aff} on the cohomology of $\mathcal{M}_H(\alpha, \beta, \gamma; p)$. Similarly, \mathcal{W}_{aff} acts on, for example, the K -theory of this space.

This result is somewhat analogous to the Springer representations of the Weyl group [26], which are also naturally understood, as suggested in [31], by varying the parameters of hyper-Kähler metrics on coadjoint orbits and their Slodowy slices [35]. (We review the framework for this in section 3.8.) The Springer representations have been generalized to an action of the affine Hecke algebra on equivariant K -theory [27], [28] and more recently to an action of the affine braid group on certain derived categories [29] of sheaves on the Springer resolution. See the book by Chriss and Ginzburg [30] for an exposition of some of these results. The affine Weyl group action on cohomology or K -theory of \mathcal{M}_H is enriched to an affine braid group action on the categories of A -branes or B -branes, as we will see in section 4.5. As for whether one can see the affine Hecke algebra in the context of \mathcal{M}_H , to attempt to do so, we would set $\beta = \gamma = 0$, whereupon \mathcal{M}_H admits an action of \mathbb{C}^* , and one can define its equivariant cohomology or K -theory. These may well admit an action of the affine Hecke algebra, which would improve the analogy between the “global” problem involving \mathcal{M}_H and the “local” problem involving complex coadjoint orbits.

Example

The affine Weyl group acts on the cohomology of \mathcal{M}_H in all dimensions. However, we can describe this action explicitly if we restrict to

the two-dimensional cohomology, which we described in eqn. (3.75):

$$(3.85) \quad H^2(\mathcal{M}_H(\alpha, \beta, \gamma; p); \mathbb{Z}) = \mathbb{Z} \oplus \Lambda_{\text{wt}}.$$

The right hand side is the affine weight lattice of G , and so admits a natural action of \mathcal{W}_{aff} . To justify the obvious guess that \mathcal{W}_{aff} actually does act on $H^2(\mathcal{M}_H)$ in this natural way, we use the result (3.76) for the cohomology class of the symplectic form ω_I :

$$(3.86) \quad \left[\frac{\omega_I}{2\pi} \right] = e \oplus (-\alpha^*).$$

The cohomology class of ω_I must be invariant under the combined action of \mathcal{W}_{aff} on α and on $H^2(\mathcal{M}_H(\alpha, \beta, \gamma; p); \mathbb{Z})$. This uniquely determines the action of \mathcal{W}_{aff} on $H^2(\mathcal{M}_H(\alpha, \beta, \gamma; p); \mathbb{Z})$ to be its natural action on the affine weight lattice. This means that the subgroup $\mathcal{W} \subset \mathcal{W}_{\text{aff}}$ acts trivially on \mathbb{Z} and in the usual fashion on Λ_{wt} . And $\mathfrak{m} \in \Lambda_{\text{cort}}$ acts by

$$(3.87) \quad e \rightarrow e \oplus \mathfrak{m}^*,$$

while acting trivially on Λ_{wt} . Clearly, the right hand side of (3.86) is invariant under this transformation together with $\alpha \rightarrow \alpha + \mathfrak{m}$.

Several Ramification Points

We can readily generalize this to the case of several ramification points p_1, \dots, p_s . Associated with each such point is a triple $(\alpha_i, \beta_i, \gamma_i) \in \mathfrak{t}^3$, with its own action of \mathcal{W}_{aff} . Thus a group $(\mathcal{W}_{\text{aff}})^s$ acts on the collection of s triples, taking values in $(\mathfrak{t}^3)^s = \mathfrak{t}^{3s}$.

The corresponding moduli space $\mathcal{M}_H(\alpha_1, \beta_1, \gamma_1, p_1; \dots; \alpha_s, \beta_s, \gamma_s, p_s)$ of ramified Higgs bundles is a smooth manifold if the triples $(\alpha_i, \beta_i, \gamma_i)$ take values in a suitable parameter space \mathcal{X}_s . This space is obtained from \mathfrak{t}^{3s} by omitting certain affine linear spaces of codimension three. To avoid local singularities, one must require that each triple $(\alpha_i, \beta_i, \gamma_i)$ is separately regular. To avoid global singularities, one must omit certain additional affine linear spaces described in [24].

The group $(\mathcal{W}_{\text{aff}})^s$ acts freely on \mathcal{X}_s , since in defining \mathcal{X}_s , we require each triple to be separately regular. Hence, by the same logic as before, we get an action of $(\mathcal{W}_{\text{aff}})^s$ on the cohomology, K -theory, etc., of \mathcal{M}_H .

As before, we can describe this action explicitly if we specialize to the two-dimensional cohomology, which was described in (3.48):

$$(3.88) \quad H^2(\mathcal{M}(\alpha_1, p_1; \dots; \alpha_s, p_s); \mathbb{Z}) = \mathbb{Z} \oplus (\oplus_{i=1}^s \Lambda_{\text{wt}, i}).$$

s copies of the ordinary Weyl group act on the s copies of Λ_{wt} . And the lattices act by a generalization of (3.87),

$$(3.89) \quad e \rightarrow e \oplus (\oplus_{i=1}^s \mathfrak{m}_i^*).$$

That this is the right action of $(\mathcal{W}_{\text{aff}})^s$ actually follows from the analog of (3.86), which is $[\omega_I/2\pi] = e \oplus (\oplus_i (-\alpha_i^*))$.

One might be puzzled by these results, since a slightly larger group could act on the lattice (3.88). Instead of shifting e by $\oplus_{i=1}^s \mathfrak{m}_i^*$ for a

collection of coroots m_i , why not simply shift it by $\bigoplus_{i=1}^s w_i$ for an arbitrary set of weights $w_1, \dots, w_s \in \Lambda_{\text{wt}}$, without worrying about whether w_i is of the form m_i^* for a coroot m_i ? In fact, in the case of several ramification points, a group smaller than this but larger than we have so far described does act naturally on the cohomology of $\mathcal{M}_H(\alpha_1, \dots, p_s)$. According to (3.63) and (3.64), there is a symmetry $\alpha_i \rightarrow \alpha_i + u_i$ for any family of coweights u_i , $i = 1, \dots, n$, such that $\sum_{i=1}^s u_i$ is a coroot. (We demonstrated the symmetry for parabolic bundles, but the same reasoning applies for ramified Higgs bundles.) So the cohomology of $\mathcal{M}_H(\alpha_1, \dots, p_s)$ must admit the action of the group $\Lambda^* \rtimes (\mathcal{W})^s$, where Λ^* is the sublattice of $\bigoplus_{i=1}^n \Lambda_{\text{cowt},i}$ consisting of elements $u_1 \oplus \dots \oplus u_s$, with all $u_i \in \Lambda_{\text{cowt}}$ and $\sum_i u_i \in \Lambda_{\text{cort}}$. Since Λ^* has $\bigoplus_{i=1}^s \Lambda_{\text{cort},i}$, whose action we have already described, as a sublattice of finite index, its action must be given by the same formula:

$$(3.90) \quad e \rightarrow e \oplus \left(\bigoplus_{i=1}^s u_i^* \right).$$

This makes sense because the map $u \rightarrow u^*$ does map Λ_{cowt} to Λ_{wt} , as explained in Appendix A.

Non-Simply-Connected G

What happens if G is not simply-connected? α now takes values in $\mathbb{T} = \mathfrak{t}/\Lambda_{\text{cochar}}$, and it might seem that the group that would act on cohomology or K -theory of \mathcal{M}_H would be now $\Lambda_{\text{cochar}} \rtimes \mathcal{W}$.

Whether this is correct depends on precisely what one means. When G is not simply-connected, \mathcal{M}_H has $\#\pi_1(G)$ components, labeled by the value of the characteristic class $\xi(E)$ that measures the obstruction to lifting E to a bundle with simply-connected structure group \overline{G} . The action of Λ_{cochar} permutes the components, as discussed in eqn. (2.6). The subgroup that acts on the cohomology of one given component is Λ_{cort} . Hence if we restrict our attention to one fixed component, the group that acts on the cohomology, for the case of one parabolic point, is $\mathcal{W}_{\text{aff}} = \Lambda_{\text{cort}} \rtimes \mathcal{W}$, just as if G is simply-connected. Similarly, with s parabolic points, the group that acts on the cohomology of a single component of \mathcal{M}_H is $\Lambda^* \rtimes (\mathcal{W})^s$, whether G is simply-connected or not. But the larger group acts if one wants to include transformations that permute the components.

Ramification Of Type \mathbb{L}

We can consider in a similar fashion a point p endowed with a singularity labeled by an arbitrary Levi subgroup \mathbb{L} . The parameters (α, β, γ) labeling such a point are invariant under \mathbb{L} , and a local singularity is avoided precisely if this triple is \mathbb{L} -regular. The group that naturally acts is the subgroup of \mathcal{W}_{aff} that commutes with \mathbb{L} . Let us call this group $\mathcal{W}_{\text{aff},\mathbb{L}}$.

The same reasoning as above shows that $\mathcal{W}_{\text{aff},\mathbb{L}}$ acts on the cohomology (or K -theory, etc.) of $\mathcal{M}_{H,\mathbb{L}}$. In the extreme case that $\mathbb{L} = G$, this statement becomes trivial, as $\mathcal{W}_{\text{aff},\mathbb{L}}$ is then the trivial group.

Generalization To Four-Dimensional Gauge Theory And Sigma Models

In applications to the geometric Langlands program, we really care not about the variety \mathcal{M}_H , but about a two-dimensional sigma model in which the target space is \mathcal{M}_H . This sigma model arises as a low energy approximation to a four-dimensional gauge theory. More relevant than whether the classical variety is smooth is whether the sigma model is smooth. We call the sigma model smooth when its spectrum, correlation functions, etc., vary smoothly with the parameters.

When \mathcal{M}_H is smooth, the sigma model is certainly smooth. However, the sigma model may remain smooth even when classically \mathcal{M}_H has a singularity. Typically, in two-dimensional sigma models with $(4, 4)$ supersymmetry, to get a singularity of the quantum theory, one must adjust four parameters, not just three. Three parameters control the classical geometry, and the fourth controls a theta-angle of the sigma model.

As we learned in section 2.3, in addition to the parameters $(\alpha, \beta, \gamma) \in \mathbb{T} \times \mathfrak{t} \times \mathfrak{t}$, the gauge theory depends on another parameter, a theta-angle $\eta \in {}^L\mathbb{T}$. So the parameters labeling a point with ramified structure are really a quartet²¹ $(\alpha, \beta, \gamma, \eta) \in \mathbb{T} \times \mathfrak{t} \times \mathfrak{t} \times {}^L\mathbb{T}$, with an equivalence under the action of the Weyl group on all four variables. This makes sense because G and ${}^L G$ have the same Weyl group!

By a local singularity of the sigma model with target \mathcal{M}_H , we mean a singularity whose position only depends on the parameters $(\alpha, \beta, \gamma, \eta)$ labeling a single ramification point. Global singularities are those whose positions depend on the parameters of two or more points.

We recall that a point in a space acted on by \mathcal{W} is regular if no non-trivial element of \mathcal{W} leaves the point fixed. To keep this discussion simple, we begin with the case of one ramification point, and use the fact that in this case \mathcal{M}_H is singular only for non-regular triples $(\alpha, \beta, \gamma) \in \mathbb{T} \times \mathfrak{t} \times \mathfrak{t}$. It follows that the sigma model is smooth if (α, β, γ) is regular. However, the gauge theory also has the S -duality transformation $\tau \rightarrow -1/\tau$ which exchanges α and η , as we discussed in section 2.4. The smoothness of the sigma model must be invariant under this transformation, so we learn that the sigma model is smooth if the triple (η, β, γ) is regular.

More generally, we have an infinite discrete duality group Γ acting on α, η as in (2.53). If G is simply-laced or we restrict to an index two subgroup of Γ , then β, γ transform by multiplication by a positive real

²¹If we wish, we can lift α and η to be \mathfrak{t} -valued. Then the group by which we must divide is not the affine Weyl group but an extension of \mathcal{W} by the product of a pair of lattices, which act by shifting α and η , as we discuss in section 4.5.

number (note the discussion following (2.24)). Such a rescaling does not affect the question of whether a triple (α, β, γ) is regular. For example, for $G = E_8$, the duality group is $SL(2, \mathbb{Z})$, and contains a transformation that maps α to $m\alpha + n\eta$, for any relatively prime integers m, n . So the sigma model is smooth if $(m\alpha + n\eta, \beta, \gamma)$ is regular. The fact that this is so for all pairs m, n implies that actually, the sigma model is smooth if the quartet $(\alpha, \beta, \gamma, \eta)$ is regular. We would still reach the same conclusion if we replace $SL(2, \mathbb{Z})$ by a congruence subgroup. For any G , the duality group contains a congruence subgroup of $SL(2, \mathbb{Z})$ that acts as just described, so it is always the case that the sigma model is smooth if the quartet $(\alpha, \beta, \gamma, \eta)$ is regular.

For the case of ramification at several points p_i , we can similarly consider global singularities, which we can define to be simply singularities whose positions depend upon the parameters $(\alpha_i, \beta_i, \gamma_i, \eta_i)$ associated with more than one point. The duality symmetry can now be used to show that the conditions found in [24] can be extended by an additional condition involving the η 's, so the global singularities now occur in real codimension four.

3.8. Nahm's Equations And Local Singularity Of \mathcal{M}_H . Gauge theory and Nahm's equations can be used to obtain hyper-Kähler metrics on coadjoint orbits of complex Lie groups. See [34], [35] for the original constructions, [36], [37] for generalizations to arbitrary orbits, and [63], [31] for reviews and further references. Our interest in these metrics is that they give the behavior of $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ near a local singularity.

There are several routes to the construction of these hyper-Kähler metrics. For us, it is most convenient to consider Hitchin's equations on a punctured disc C , defined as the region of the complex z -plane with $|z| \leq 1$, and $z \neq 0$. We write as usual $z = re^{i\theta}$. On C , we consider solutions of Hitchin's equations that are invariant under rotations of the disc and have the familiar singularity near $r = 0$:

$$(3.91) \quad \begin{aligned} A &= \alpha d\theta + \dots \\ \phi &= \beta \frac{dr}{r} - \gamma d\theta + \dots \end{aligned}$$

We suppose that the triple (α, β, γ) is regular, that is, that the subgroup of G that commutes with this triple is precisely the torus \mathbb{T} . We let $\mathcal{G}_{C,p}$ be the group of rotation-invariant gauge transformations $g : C \rightarrow G$ that equal 1 for $|z| = 1$ and take values in \mathbb{T} at $z = 0$. The space of rotation-invariant solutions of Hitchin's equations, with the boundary condition (3.91), and modulo the action of $\mathcal{G}_{C,p}$, is a hyper-Kähler manifold that we will call $\mathcal{Q}(\alpha, \beta, \gamma)$.

In fact, it can be constructed as the hyper-Kähler quotient by $\mathcal{G}_{C,p}$ of the space of rotation-invariant pairs (A, ϕ) . Such a pair is given in

general by

$$(3.92) \quad \begin{aligned} A &= a(r) d\theta + h(r) \frac{dr}{r} \\ \phi &= b(r) \frac{dr}{r} - c(r) d\theta \end{aligned}$$

with \mathfrak{g} -valued functions a, b, c , and h . The functions a, b , and c are related in a fairly obvious way to the usual parameters α, β , and γ , but rotational symmetry allows a fourth function h . Though h can be gauged away, it is more convenient not to do so for the moment. The space of solutions of Hitchin's equations that are of this form has a hyper-Kähler structure that is the usual one appropriate to Hitchin's equations, specialized to this case. One way to describe it is to think of the functions (h, a, b, c) as giving a map from the open unit interval to the quaternions $\mathbb{H} \cong \mathbb{R}^4$, tensored with \mathfrak{g} . The hyper-Kähler structure comes from the hyper-Kähler structure on \mathbb{H} . Concretely, in one complex structure, which we will call I , the complex variables are $h - ia$ and $b + ic$. The others can be obtained by applying an $SO(3)$ rotation to the triple (a, b, c) .

If we set $s = -\ln r$ and $D/Ds = d/ds + [h, \cdot]$, then Hitchin's equations become

$$(3.93) \quad \begin{aligned} \frac{Da}{Ds} &= [b, c] \\ \frac{Db}{Ds} &= [c, a] \\ \frac{Dc}{Ds} &= [a, b]. \end{aligned}$$

These become Nahm's equations [64] if we set h to zero, which we can do locally by a gauge transformation.

To elucidate the nature of the moduli space $\mathcal{Q}(\alpha, \beta, \gamma)$ of solutions of these equations, first note that a linear combination of two of the equations gives

$$(3.94) \quad \frac{d}{ds}(b + ic) = -[h - ia, b + ic].$$

This implies that the conjugacy class of $b + ic$ in $\mathfrak{g}_{\mathbb{C}}$ is independent of s . We also have the boundary condition $\lim_{s \rightarrow \infty} (b + ic) = \beta + i\gamma$. If $\sigma = \frac{1}{2}(\beta + i\gamma)$ is regular, this implies that $\tau = \frac{1}{2}(b + ic)$ is everywhere in the conjugacy class \mathfrak{c} that contains σ . In particular,

$$(3.95) \quad \tau = \frac{1}{2}(b(0) + ic(0))$$

is contained in this conjugacy class. τ is also gauge-invariant (since we only allow gauge transformations that equal 1 at $s = 0$). So by mapping a solution of Nahm's equations to the corresponding value of τ , we get a map $\Phi : \mathcal{Q} \rightarrow \mathfrak{c}$ that is holomorphic in complex structure I . By

interpreting the remaining part of Nahm's equations as a moment map condition, it is shown in [34] that this map is an isomorphism.

By definition, \mathfrak{c} is the orbit under conjugation of the vector $\sigma \in \mathfrak{g}_{\mathbb{C}}$ (or equally well in the dual space $\mathfrak{g}_{\mathbb{C}}^{\vee}$, since the quadratic form $-\text{Tr}$ gives a $G_{\mathbb{C}}$ -invariant identification between $\mathfrak{g}_{\mathbb{C}}$ and $\mathfrak{g}_{\mathbb{C}}^{\vee}$). It is known as an adjoint (or co-adjoint) orbit. For regular σ , the subgroup in $G_{\mathbb{C}}$ that leaves σ invariant is precisely $\mathbb{T}_{\mathbb{C}}$, the complexification of the torus \mathbb{T} . Hence \mathfrak{c} is isomorphic to $G_{\mathbb{C}}/\mathbb{T}_{\mathbb{C}}$.

If σ is not regular, then τ is in one of the affiliated orbits \mathfrak{c}_{λ} , described in section 3.3. All possibilities occur, but for a generic point in \mathcal{Q} , τ takes values in the affiliated orbit of maximal dimension, which is the regular orbit \mathfrak{c}^* . To get some insight about what happens when σ is non-regular, let us go to the extreme case $\sigma = 0$. Then the equations and boundary conditions enable us to find solutions in which b and c identically vanish. The equations collapse to $Da/Ds = 0$, so the conjugacy class of a is independent of s and equal to that of α . This conjugacy class must be regular, since we have assumed that (α, β, γ) is regular and we have set $\beta = \gamma = 0$. We can now reason somewhat as before. $a(0)$ is gauge-invariant and is conjugate to α . Moreover, modulo the gauge group, $a(0)$ is the only invariant of a solution with $b = c = 0$. Finally, $a(0)$ can be any element of the orbit of $\alpha \in \mathfrak{g}$. That orbit is (for regular α) a copy of G/\mathbb{T} , which is a Kahler manifold, known as the flag manifold. To include b and c , we note that from a holomorphic point of view in complex structure I , $b+ic$ is characterized by the linear equation (3.94). Thus \mathcal{Q} is a holomorphic vector bundle over G/\mathbb{T} . For \mathcal{Q} to be a complex symplectic manifold (and actually hyper-Kahler), this bundle must be the cotangent bundle.

So when $\beta = \gamma = 0$ but α is generic, \mathcal{Q} in complex structure I is the cotangent bundle of G/\mathbb{T} :

$$(3.96) \quad \mathcal{Q} \cong T^*(G/\mathbb{T}).$$

If we deform to $\beta, \gamma \neq 0$, the cotangent bundle is deformed to an affine bundle over G/\mathbb{T} . When $\sigma = \frac{1}{2}(\beta + i\gamma)$ is regular, the affine bundle is isomorphic to the $G_{\mathbb{C}}$ -orbit of $\sigma \in \mathfrak{g}_{\mathbb{C}}$.

Analog Of Type \mathbb{L}

This construction can be repeated with \mathbb{T} replaced by any Levi subgroup \mathbb{L} of G .

We require the triple (α, β, γ) to be \mathbb{L} -regular (that is, it commutes precisely with \mathbb{L}), and we modify the definition of $\mathcal{G}_{C,p}$ so that it comprises gauge transformations that at p take values in \mathbb{L} . We write $\mathcal{Q}_{\mathbb{L}}(\alpha, \beta, \gamma)$ for the space of solutions of the equations (3.93), with boundary conditions set by α, β , and γ , modulo the action of $\mathcal{G}_{C,p}$.

The same reasoning as above shows that if $\sigma = \frac{1}{2}(\beta + i\gamma)$ is \mathbb{L} -regular, then

$$(3.97) \quad \mathcal{Q}_{\mathbb{L}} \cong G_{\mathbb{C}}/\mathbb{L}_{\mathbb{C}}$$

is the orbit of $\sigma \in \mathfrak{g}_{\mathbb{C}}$ under the adjoint action of $G_{\mathbb{C}}$. At the other extreme, if $\sigma = 0$ but α is generic,

$$(3.98) \quad \mathcal{Q}_{\mathbb{L}} \cong T^*(G/\mathbb{L}).$$

In general, $\mathcal{Q}_{\mathbb{L}}$ is an affine deformation of $T^*(G/\mathbb{L})$. If we replace G by a general semi-simple Lie group \mathbb{L}' containing \mathbb{L} as a Levi subgroup, the same construction based on gauge theory with gauge group \mathbb{L}' gives a family of hyper-Kähler metrics on

$$(3.99) \quad \mathcal{Q}_{\mathbb{L}',\mathbb{L}} \cong T^*(\mathbb{L}'/\mathbb{L})$$

and its affine deformations.

Local Model Of \mathcal{M}_H

Many properties that we have described for Hitchin's moduli space \mathcal{M}_H have local analogs involving the hyper-Kähler metrics \mathcal{Q} .

For example, the cohomology classes of the symplectic forms of \mathcal{M}_H are linear functions of the parameters α, β , and γ . A similar linearity holds for \mathcal{Q} , as stated in Theorem 2.6 of [34].

To give another example, just as the affine Weyl group acts on the cohomology of \mathcal{M}_H , the ordinary Weyl group similarly acts on the cohomology of \mathcal{Q} . The framework for proving this is described in section 5 of [31], using the action of the Weyl group on the parameters (α, β, γ) and the fact that the singularities are in codimension three – in other words, the same facts that we used in section 3.7 to construct an affine Weyl group action on cohomology of \mathcal{M}_H . Actually, what is considered in [31] is a somewhat larger class of hyper-Kähler varieties constructed in [35] and involving Slodowy slices. (These more general varieties are constructed by solving the same equations as above but with different asymptotic behavior at $s = 0$.) The Weyl group representations that arise for these varieties are known as the Springer representations. They can be understood geometrically in a relatively elementary construction [65] without hyper-Kähler metrics; perhaps this has also an analog for \mathcal{M}_H .

The basic reason that the hyper-Kähler manifolds \mathcal{Q} give local models for many properties of \mathcal{M}_H is that they do in fact describe the behavior of the moduli space of ramified Higgs bundles near a local singularity. For example, as in (3.79), the local behavior of $\mathcal{M}_{H,\mathbb{L}}(\alpha, 0, 0; p)$ when α approaches a value $\bar{\alpha}$ that is not \mathbb{L} -regular is modeled by $T^*(\mathbb{L}'/\mathbb{L})$ for some Levi group \mathbb{L}' that properly contains \mathbb{L} . To get a local model of the situation, we need a suitable family of hyper-Kähler metrics on $T^*(\mathbb{L}'/\mathbb{L})$ (and its affine deformations with $\beta, \gamma \neq 0$). This is what we get from the construction summarized above, as noted in (3.99), if we

take the gauge group to be \mathbb{L}' . There is a simple rationale for using \mathbb{L}' gauge theory rather than G gauge theory to construct the local model: the boundary conditions at the singularity, defined in this case by the trio $(\bar{\alpha}, 0, 0)$, are invariant only under \mathbb{L}' , not G , so a theory with \mathbb{L}' as the gauge symmetry suffices for describing the singularity.

Detailed Analysis For $SU(2)$

To understand more fully why the varieties \mathcal{Q} give a good description of local singularities of \mathcal{M}_H , we will examine more closely the behavior of \mathcal{Q} near a non-regular point. To keep things simple, we concentrate on $G = SU(2)$ for illustration.

The only nonregular value of the triple (α, β, γ) for $G = SU(2)$ is $\alpha = \beta = \gamma = 0$. At $\alpha = \beta = \gamma = 0$, \mathcal{Q} describes solutions of Hitchin's equations on the disc $|z| \leq 1$ that are rotation-invariant and less singular than $1/r$ at $r = 0$.

Equivalently, taking the gauge $h = 0$, we need solutions of Nahm's equations

$$(3.100) \quad \begin{aligned} \frac{da}{ds} &= [b, c] \\ \frac{db}{ds} &= [c, a] \\ \frac{dc}{ds} &= [a, b], \end{aligned}$$

on the half-line $[0, \infty)$, with $a, b, c \rightarrow 0$ for $s \rightarrow \infty$.

Obviously, one such solution is $a = b = c = 0$. Another simple solution, which is the starting point in [35], is

$$(3.101) \quad \begin{aligned} a &= -\frac{1}{s}t_1 \\ b &= -\frac{1}{s}t_2 \\ c &= -\frac{1}{s}t_3, \end{aligned}$$

where t_1, t_2 , and t_3 are fixed elements of $\mathfrak{sl}(2, \mathbb{C})$ obeying $[t_1, t_2] = t_3$, and cyclic permutations thereof.

In complex structure I , this solution describes a Higgs bundle (E, φ) in which φ has a pole with nilpotent residue. (Notice that $b + ic = -s^{-1}(t_2 + it_3)$ is in fact nilpotent for all s .) In complex structure J , it describes a flat bundle with monodromy around the point $r = 0$ that is unipotent but not equal to 1. These results are what one might expect from section 3.3.

A slight generalization is to introduce a positive constant f and take

$$(3.102) \quad \begin{aligned} a &= -\frac{1}{s+f^{-1}}t_1 \\ b &= -\frac{1}{s+f^{-1}}t_2 \\ c &= -\frac{1}{s+f^{-1}}t_3, \end{aligned}$$

We have parametrized the solutions in this particular way, because (3.102) actually has a limit for $f \rightarrow 0$, namely the trivial solution $a = b = c = 0$. So f takes values in $\mathbb{R}_{\geq 0} = [0, \infty)$. In our application, we will be concerned with small f . We can also pick an element $R \in SO(3)$ and generalize (3.102) to

$$(3.103) \quad \begin{aligned} a &= -\frac{1}{s+f^{-1}}Rt_1R^{-1} \\ b &= -\frac{1}{s+f^{-1}}Rt_2R^{-1} \\ c &= -\frac{1}{s+f^{-1}}Rt_3R^{-1}. \end{aligned}$$

The parameter space of this family is thus $\mathbb{R}_{\geq 0} \times SO(3) = \mathbb{R}^4/\mathbb{Z}_2$, where \mathbb{Z}_2 acts by a reflection on all four coordinates of \mathbb{R}^4 . This is the A_1 singularity. In particular, the natural metric on this family \mathcal{Q} of solutions of Nahm's or Hitchin's equations, obtained by integrating the L^2 norm of the variation of the fields (A, ϕ) over the disc $r \leq 1$, is the flat metric on $\mathbb{R}^4/\mathbb{Z}_2$; all modes are square-integrable, thanks to the factor of $1/s = 1/(-\ln r)$, and $f = 0$ is at finite distance. For nonzero α, β, γ , we get instead the Eguchi-Hansen metric, describing the deformation and resolution of the A_1 singularity.

Now we can explain the basic reason that Nahm's equations and Kronheimer's construction give a good model for local singularities of \mathcal{M}_H . The singularity of \mathcal{Q} occurs at $f = 0$ and corresponds to the trivial solution $a = b = c = 0$. A point in \mathcal{Q} near the singularity corresponds to a solution with very small f . When Nahm's equations are embedded in Hitchin's equations, there is an extra factor of $1/r$, visible in eqn. (3.92). This means that for small f , all fields are very small except for $r \lesssim f$. This continues to be the case if we perturb α, β, γ to be nonzero but of order f . Because of this, we can start with an arbitrary Higgs bundle (E, φ) without ramification, and "glue in" the above family of solutions, to get a family of ramified Higgs bundles. We use the old solution for $r \gg f$, and the exact solution (3.103) for $r \lesssim f$. The family of solutions obtained this way acquires an A_1 singularity if one sets $f = 0$, whereupon the "new" solutions reduce to the old ones.

Of course, the gluing operation does not give an exact solution. The set of fields obtained by gluing must be modified to get an exact

solution, but the requisite modification is small if f is small. This is somewhat analogous to the construction [66] of Yang-Mills instanton solutions on a four-manifold by gluing in an exact solution from \mathbb{R}^4 that has its support mainly on a very small region in \mathbb{R}^4 . However, in contrast to instanton moduli space, the deformation theory of Higgs bundles is unobstructed (as long as we keep away from reducible Higgs bundles) so there is no analog of the topological conditions described in [66] that can potentially obstruct the process of deforming the glued fields to an exact solution.

Thus we get a more precise way to see what was argued in section 3.6: for $\alpha, \beta, \gamma \rightarrow 0$, $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ develops an A_1 singularity, the singular locus being precisely \mathcal{M}_H , the moduli space of unramified Higgs bundles. This specific result is of course special to $G = SU(2)$. For G of higher rank, as we have argued, a similar construction leads in general to more complicated singularities whose resolution is $T^*(\mathbb{L}'/\mathbb{L})$ for various \mathbb{L}' and \mathbb{L} .

3.9. The Hitchin Fibration. Now we describe the Hitchin fibration and complete integrability in the context of Higgs bundles with ramification. All of these matters have been understood in the literature in much more detail; for example, see [67], [68].

We begin with the example of $SL(2, \mathbb{C})$, from section 3.3. Up to conjugacy, the local behavior of φ near a parabolic point with $\sigma \neq 0$ is

$$(3.104) \quad \varphi = \frac{dz}{z} \sigma (1 + \mathcal{O}(z)).$$

This implies that

$$(3.105) \quad \text{Tr } \varphi^2 = \text{Tr } \sigma^2 \left(\frac{dz}{z} \right)^2 (1 + \mathcal{O}(z)).$$

This statement is actually true uniformly for all σ , zero or not. Indeed, for $\sigma = 0$, the polar part of φ is nilpotent, but in a general solution φ also has regular terms. After trivializing the holomorphic structure near $z = 0$, we have up to conjugacy

$$(3.106) \quad \varphi = \frac{dz}{z} \left(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + az + bz^2 + \dots \right),$$

with $a, b \in \mathfrak{sl}(2, \mathbb{C})$. Hence the quadratic differential $\text{Tr } \varphi^2$ may have a simple pole at $z = 0$, but no double pole, showing that (3.105) also holds for $\sigma = 0$.

The reason that this is important is that the quadratic differential $\text{Tr } \varphi^2$ is the key to the complete integrability of \mathcal{M}_H as a complex manifold in complex structure I . In the absence of ramification, \mathcal{M}_H has (complex) dimension equal, for $G_{\mathbb{C}} = SL(2, \mathbb{C})$, to $6g_C - 6$, where g_C is the genus of C . To establish complete integrability, one requires $3g_C - 3$

commuting Hamiltonians. These are precisely the components of $\text{Tr } \varphi^2$, which takes values in the space of quadratic differentials on C . The dimension of that space is $3g_C - 3$. (For a very brief explanation of complete integrability of Hitchin systems, see section 4.3 of [8].)

Let us carry out the analogous computation in the presence of ramification. For simplicity in the exposition, we suppose that there is just one ramified point p . A Higgs bundle is a pair (E, φ) , where E is an $SL(2, \mathbb{C})$ -bundle over C and $\varphi \in H^0(C, K_C \otimes \text{ad}(E) \otimes \mathcal{O}(p))$. Here we include the factor of $\mathcal{O}(p)$ (the bundle whose sections are functions that may have a simple pole at p), since φ is allowed to have a pole at p . The number of parameters required to specify the bundle E is $3g_C - 3$. By Riemann-Roch, the dimension of $H^0(C, K_C \otimes \text{ad}(E) \otimes \mathcal{O}(p))$ is $3g_C$. (Indeed, $H^0(C, K_C \otimes \text{ad}(E))$ is the cotangent space to the moduli space of stable bundles and has dimension $3g_C - 3$; tensoring with $\mathcal{O}(p)$ adds 3 to the dimension, since $K_C \otimes \text{ad}(E)$ has rank 3.) However, φ obeys the one constraint (3.106). So the choice of φ depends on $3g_C - 1$ parameters, and the dimension of $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ is $(3g_C - 3) + (3g_C - 1) = 6g_C - 4$. So we need $3g_C - 2$ commuting Hamiltonians to establish complete integrability.

These are precisely the components of $\text{Tr } \varphi^2$. In general, $\text{Tr } \varphi^2$ is a quadratic differential with a double pole. The space of such quadratic differentials has dimension $3g_C - 1$, but one parameter is determined by (3.105), so $\text{Tr } \varphi^2$ lives in a $3g_C - 2$ dimensional space. This gives the $3g_C - 2$ parameters needed for complete integrability of $\mathcal{M}_H(\alpha, \beta, \gamma; p)$.

All generalities about the Hitchin fibration of \mathcal{M}_H , complete integrability, etc., have natural analogs for Higgs bundles with ramification. We write \mathbf{B} for the space of quadratic differentials with double pole at p obeying (3.105). The Hitchin fibration is a map $\pi : \mathcal{M}_H(\alpha, \beta, \gamma; p) \rightarrow \mathbf{B}$ which is holomorphic in complex structure I . It is defined by mapping a Higgs bundle (E, φ) to the point in \mathbf{B} defined by $\text{Tr } \varphi^2$. The functions on \mathbf{B} are the commuting Hamiltonians, and the fibers are complex Lagrangian submanifolds (that is, they are Lagrangian from the point of view of the holomorphic two-form Ω_I). The generic fiber is an abelian variety. Moreover, all this remains true if one has ramification at several points, not just one.

For any G , the analog of this is as follows. Let r be the rank of G . The ring of invariant polynomials on the Lie algebra \mathfrak{g} is freely generated by r polynomials \mathcal{P}_i , which are homogeneous of degree d_i for some integers d_i . These integers obey

$$(3.107) \quad \sum_{i=1}^r (2d_i - 1) = \dim(G).$$

Instead of the quadratic differential $\text{Tr } \varphi^2$, we consider its analogs $\mathcal{P}_i(\varphi)$, which are holomorphic sections of $K_C^{d_i}$. They obey

$$(3.108) \quad \mathcal{P}_i(\varphi) = \mathcal{P}_i(\sigma) \left(\frac{dz}{z} \right)^{d_i} (1 + \mathcal{O}(z)).$$

These conditions hold for all σ , and, for $\mathbb{L} = \mathbb{T}$, the differentials $\mathcal{P}_i(\varphi)$ otherwise have arbitrary poles of degree d_i at $z = 0$. For other \mathbb{L} , there are additional conditions on the poles. We will concentrate here on the case $\mathbb{L} = \mathbb{T}$. \mathbf{B} is defined by saying that a point in \mathbf{B} is a collection of d_i -differentials, $i = 1, \dots, r$, that are holomorphic away from ramification points and behave like (3.108) near such a point. The dimension of \mathbf{B} (for the case of one ramified point) is $(g_C - 1)\dim(G) + \sum_{i=1}^r (d_i - 1)$, as one can prove with the help of (3.107).

The dimension of \mathcal{M}_H is computed as we did for $SU(2)$. The choice of a G -bundle E depends on $(g_C - 1)\dim(G)$ parameters. The dimension of $H^0(C, K_C \otimes \text{ad}(E) \otimes \mathcal{O}(p))$ is $g_C \dim(G)$, but φ is subject to r constraints (3.108), so it depends on $g_C \dim(G) - r$ parameters. The dimension of $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ is thus

$$(3.109) \quad \begin{aligned} \dim(\mathcal{M}_H(\alpha, \beta, \gamma; p)) &= (2g_C - 1) \dim(G) - r \\ &= (2g_C - 2) \dim(G) + \sum_{i=1}^r (2d_i - 2) \\ &= \dim(\mathcal{M}_H) + \dim(G) - r. \end{aligned}$$

The number of commuting Hamiltonians needed to establish complete integrability is therefore $(g_C - 1)\dim(G) + \sum_{i=1}^r (d_i - 1)$. This is precisely the dimension of \mathbf{B} . The Hitchin fibration $\pi : \mathcal{M}_H(\alpha, \beta, \gamma; p) \rightarrow \mathbf{B}$, defined by mapping a pair (E, φ) to the point in \mathbf{B} defined by $(\mathcal{P}_1(\varphi), \dots, \mathcal{P}_r(\varphi))$, has all the key properties that we described for $G = SL(2, \mathbb{C})$.

The formula (3.109) for the (complex) dimension of $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ is, of course, compatible with the existence, in a suitable limit, of the generic fibration (3.79). For a more general Levi subgroup \mathbb{L} , there is a natural modification of this, with a suitable extension of the constraints (3.108).

4. Geometric Langlands With Tame Ramification

4.1. Review Of Unramified Case. The basic steps to get from $\mathcal{N} = 4$ Yang-Mills theory to geometric Langlands are described in [8], beginning in section 3. We briefly review them, since the same procedure applies in the presence of ramification.

One first makes a certain topological twist of the $\mathcal{N} = 4$ theory, the GL twist. The twisting depends on a complex parameter t , and leads to a family of four-dimensional topological field theories parametrized by $\mathbb{C}\mathbb{P}^1$. It is convenient to think of this $\mathbb{C}\mathbb{P}^1$ as the complex Ψ plane

plus a point at infinity. Ψ is a certain combination of t and the gauge coupling parameter $\tau = \theta/2\pi + 4\pi i/e^2$. A duality transformation of the $\mathcal{N} = 4$ theory that acts on τ by $\tau \rightarrow (a\tau + b)/(c\tau + d)$ acts likewise on Ψ by $\Psi \rightarrow (a\Psi + b)/(c\Psi + d)$. In particular, the duality operation $S : \tau \rightarrow -1/n_{\mathfrak{g}}\tau$ maps $\Psi = \infty$ to $\Psi = 0$. The usual form of the geometric Langlands program involves these two values of Ψ and the duality between them.

The next step [19], [20] is to consider the theory on a four-manifold $M = \Sigma \times C$, where C is the Riemann surface on which one wishes to study the geometric Langlands program. In an appropriate limit of Σ much larger than C , or in the topological field theory (in which distances are irrelevant), the compactified theory can be described in terms of a sigma model on Σ in which the target is the moduli space $\mathcal{M}_H(G, C)$ of Higgs bundles on C .

At $\Psi = \infty$, as explained in section 5 of [8], the resulting two-dimensional theory is the B -model of \mathcal{M}_H in complex structure J . It varies holomorphically in $\alpha + i\gamma$, and is locally independent of β and η . At $\Psi = 0$, it is the A -model with symplectic structure

$$(4.1) \quad \omega = (\text{Im } \tau)\omega_K,$$

where the factor of $\text{Im } \tau = 4\pi/e^2$ comes from the normalization of the kinetic energy. We call these models the B -model of type J and the A -model of type K . Each of these models is independent of the four-dimensional gauge coupling e^2 , and hence can be studied at weak coupling.

The next step is to let Σ have a boundary, labeled by a brane \mathcal{B} . S -duality automatically exchanges B -branes of type J with A -branes of type K . This is the basic geometric Langlands duality. To get the usual formulation of geometric Langlands duality, one must incorporate Wilson and 't Hooft operators and the duality between them (sections 6,8,9, and 10 of [8]), and one must reinterpret the A -model of type K in terms of \mathcal{D} -modules on $\mathcal{M}(G, C)$ (section 11).

4.2. Sigma Model With Ramification. We now begin to describe an analogous program in the presence of ramified structure at a point $p \in C$. (To keep the notation simple, we consider the case of one ramification point, except when it is essential to allow several.) We consider gauge theory on $M = \Sigma \times C$, with gauge group ${}^L G$, in the presence of a “surface operator” supported on $\Sigma_p = \Sigma \times p$. For tame ramification, which is the subject of this paper, we take the surface operator to be of the familiar type, labeled by the parameters $({}^L\alpha, {}^L\beta, {}^L\gamma, {}^L\eta) \in {}^L\mathbb{T} \times \mathfrak{t}^\vee \times \mathfrak{t}^\vee \times \mathbb{T}$. ${}^L\mathbb{T}$ and \mathbb{T} are exchanged, of course, relative to most of our previous discussion, since we take the gauge group to be ${}^L G$. Until further notice, we take the Levi subgroup to be $\mathbb{L} = \mathbb{T}$.

The duality transformation $S : \tau \rightarrow -1/n_{\mathfrak{g}}\tau$ maps the ${}^L G$ theory with coupling ${}^L\tau$ to a theory with gauge group G and coupling $\tau = -1/n_{\mathfrak{g}}{}^L\tau$. To get the basic geometric Langlands duality, it is convenient to take ${}^L\tau$ to be imaginary and to take the twisting parameter of the ${}^L G$ theory to be $t = i$. This gives the B -model at $\Psi = \infty$, and is mapped by S to a model with $t = 1$ and imaginary τ . The latter gives a convenient description of the A -model at $\Psi = 0$.

A surface operator with parameters $({}^L\alpha, {}^L\beta, {}^L\gamma, {}^L\eta)$ maps to a surface operator of the same type but with different parameters $(\alpha, \beta, \gamma, \eta)$. The relation among the parameters was analyzed in section 2.4:

$$(4.2) \quad \begin{aligned} (\alpha, \eta) &= ({}^L\eta, -{}^L\alpha) \\ (\beta, \gamma) &= (\text{Im } {}^L\tau) ({}^L\beta^*, {}^L\gamma^*). \end{aligned}$$

The second formula is eqn. (2.24), and can be inverted as in (2.23):

$$(4.3) \quad ({}^L\beta, {}^L\gamma) = (\text{Im } \tau)(\beta^*, \gamma^*).$$

Our goal is to adapt the steps that were summarized in section 4.1. The four-dimensional theory on $M = \Sigma \times C$ reduces in this situation to a two-dimensional sigma model with target the moduli space of ramified Higgs bundles. Just as in the absence of ramification, this sigma model has $(4, 4)$ supersymmetry, since the target space is hyper-Kähler.

The surface parameters (α, β, γ) enter the classical geometry of the moduli space. The remaining surface parameter η was introduced in section 2.3 by analogy with theta-angles of two-dimensional gauge theory. Indeed, for G of rank r (and barring some exceptional cases of small r and small g_C), $\mathcal{M}_H(\alpha, \beta, \gamma; p)$ has second Betti number $r + 1$, so the sigma model has room for $r + 1$ theta-angles. One such angle descends from the theta-angle θ of the underlying four-dimensional gauge theory, by precisely the reasoning described in eqn. (4.18) of [8]. The other r theta-angles of the effective two-dimensional theory are derived from η . As we see in (4.2), S -duality exchanges theta-angles and geometrical parameters.

Discrete Electric and Magnetic Fluxes

The sigma model additionally depends on discrete electric and magnetic fluxes (see section 7 of [8], where a somewhat different point of view is taken). We describe them in some detail, since otherwise it is impossible to give a completely precise description of the action of electric-magnetic duality.

The bundle $E \rightarrow M = \Sigma \times C$ has a characteristic class $\xi(E) \in H^2(M, \pi_1(G))$. We write \mathbf{m}_0 for the restriction of this class to $q \times C$, for q a generic point in Σ . \mathbf{m}_0 takes values in $H^2(C, \pi_1(G)) \cong \pi_1(G)$, and \mathcal{M}_H has components labeled by \mathbf{m}_0 .

Discrete electric flux needs more explanation. We introduce a slight twist into G gauge theory to allow bundles $E \rightarrow M = \Sigma \times C$ with

structure group G_{ad} but with the property that, when restricted to $q \times C$ (for q a point in Σ), E can be lifted to a G -bundle. Thus, locally along Σ but globally along C , the theory has gauge group G , but globally along Σ there may be a “twist” that prevents lifting E to a G -bundle. A bundle of this type, restricted to $\Sigma \times r$ for generic $r \in C$, has a characteristic class $\mathbf{a}_0 \in H^2(\Sigma, \pi_1(G_{\text{ad}}))$.

Now in performing the path integral, we introduce a discrete theta-angle \mathbf{e}_0 and include in the path integral a phase factor

$$(4.4) \quad \exp(2\pi i \langle \mathbf{e}_0, \mathbf{a}_0 \rangle).$$

Thus, \mathbf{e}_0 (though written additively) is a character of $\pi_1(G_{\text{ad}})$. However, in G gauge theory, it is natural to consider only \mathbf{e}_0 that annihilate $\pi_1(G) \subset \pi_1(G_{\text{ad}})$ (so that \mathbf{e}_0 is only sensitive to the “twist”). Thus \mathbf{e}_0 is a character of $\pi_1(G_{\text{ad}})/\pi_1(G)$ (this is the same as the center of G , $\mathcal{Z}(G)$). Equivalently,

$$(4.5) \quad \mathbf{e}_0 \in \Lambda_{\text{char}}/\Lambda_{\text{rt}}.$$

If we restrict \mathbf{e}_0 in this way, we can project \mathbf{a}_0 to

$$(4.6) \quad \pi_1(G_{\text{ad}})/\pi_1(G) = \Lambda_{\text{cowt}}/\Lambda_{\text{cochar}}.$$

The duality transformation $S : \tau \rightarrow -1/n_{\mathfrak{g}}\tau$ exchanges \mathbf{e}_0 and \mathbf{m}_0 . This is possible because $\pi_1({}^L G) = \pi_1(G_{\text{ad}})/\pi_1(G)$, and similarly with ${}^L G$ and G exchanged.

Ramification does not substantially change the definition of \mathbf{m}_0 and \mathbf{e}_0 . However, their role is qualitatively different in the presence of ramification. As we observed in discussing eqn. (2.6), \mathbf{m}_0 can be changed by shifting α . Dually, as we will now explain, \mathbf{e}_0 can be changed by shifting η . The effect of η in the path integral is a factor

$$(4.7) \quad \exp(2\pi i \langle \eta, \mathbf{m} \rangle),$$

where \mathbf{m} is the characteristic class of the \mathbb{T} -bundle obtained by restricting E to $\Sigma \times p$, for $p \in C$. In standard G gauge theory, \mathbf{m} takes values in Λ_{cochar} , but the “twist” means that in the present context \mathbf{m} takes values in Λ_{cowt} . Indeed, we have

$$(4.8) \quad \mathbf{m} \cong \mathbf{a}_0 \text{ mod } \Lambda_{\text{cochar}}.$$

This is a general fact about the two-dimensional characteristic class of a G_{ad} -bundle whose structure group reduces to the torus \mathbb{T}_{ad} .

Because \mathbf{m} does not necessarily take values in Λ_{cochar} , the usual symmetry $\eta \rightarrow \eta + v$ for $v \in \Lambda_{\text{char}}$ does not necessarily hold. Rather, comparing (4.4) to (4.7) and using (4.8), we find that $\eta \rightarrow \eta + v$ is equivalent to $\mathbf{e}_0 \rightarrow \mathbf{e}_0 + \bar{v}$, where \bar{v} is the image of v in $\Lambda_{\text{char}}/\Lambda_{\text{rt}}$. The shifts of η that do not change \mathbf{e}_0 are by Λ_{rt} .

This mirrors the corresponding statement for α , which is that the shifts that do not change \mathbf{m}_0 are precisely those by Λ_{cort} . The parallelism remains if there are several ramification points p_1, \dots, p_s . As

in the discussion of eqn. (3.90), the shifts of $(\alpha_1, \dots, \alpha_s)$ that do not change \mathbf{m}_0 are $\alpha_i \rightarrow \alpha_i + u_i$, with $u_i \in \Lambda_{\text{cowt}}$ and $\sum_i u_i \in \Lambda_{\text{cort}}$. Likewise, the shifts in (η_1, \dots, η_s) that do not change \mathbf{e}_0 are $\eta_i \rightarrow \eta_i + v_i$, where $v_i \in \Lambda_{\text{wt}}$ and $\sum_i v_i \in \Lambda_{\text{rt}}$.

4.3. Branes. Now we specialize to the topological field theories that arise at $\Psi = \infty$ with gauge group ${}^L G$ or at $\Psi = 0$ with gauge group G .

At $\Psi = \infty$, we get the B -model in complex structure J . In this complex structure, $\mathcal{M}_H({}^L\alpha, {}^L\beta, {}^L\gamma; p)$ parametrizes flat ${}^L G_{\mathbb{C}}$ -bundles $E \rightarrow C \setminus p$ whose monodromy V around the point p obeys a condition that depends on ${}^L\alpha$ and ${}^L\gamma$.

As explained most fully in section 3.3, the condition is that the orbit of V under conjugation contains in its closure the element $U = \exp(-2\pi({}^L\alpha - i{}^L\gamma))$. If U is regular, this means that V is conjugate to U . If, at the other extreme, $U = 1$, this means that V is unipotent. At any rate, for any $V \in {}^L G_{\mathbb{C}}$, there is a choice of ${}^L\alpha$ and ${}^L\gamma$, unique up to a Weyl transformation, such that a flat connection with monodromy V represents a point in $\mathcal{M}_H({}^L\alpha, {}^L\beta, {}^L\gamma; p)$. We simply choose ${}^L\alpha$ and ${}^L\gamma$ so that U is contained in the closure of the orbit of V .

So if we want to use the B -model in complex structure J to say something about a flat bundle with a given monodromy, then ${}^L\alpha$ and ${}^L\gamma$ are uniquely determined. But ${}^L\beta$ and ${}^L\eta$ are arbitrary, since from the point of view of complex structure J , they are Kahler parameters. The B -model in complex structure J , of course, varies holomorphically with ${}^L\gamma + i{}^L\alpha$, but it is locally independent of ${}^L\beta$ and ${}^L\eta$. Globally, when we vary ${}^L\beta$ and ${}^L\eta$, the B -model has monodromies, which we will study in section 4.5.

In the geometric Langlands program with tame ramification, we begin with a flat ${}^L G_{\mathbb{C}}$ -bundle $E \rightarrow C \setminus p$ with monodromy V . Roughly speaking, at the right value of ${}^L\alpha$ and ${}^L\gamma$, E determines a zerobrane \mathcal{B}_E on $\mathcal{M}_H({}^L\alpha, {}^L\beta, {}^L\gamma; p)$. (When V is non-regular, this statement requires some elaboration, which we provide in section 4.6.)

Now we apply the S -duality transformation $S: \Psi \rightarrow -1/n_{\mathfrak{g}}\Psi$. The gauge group is transformed from ${}^L G$ to G . Ψ is mapped from ∞ to 0; the resulting model at $\Psi = 0$ is the A -model with symplectic structure $\omega = (\text{Im } \tau)\omega_K$. The parameters $(\alpha, \beta, \gamma, \eta)$ of the model with gauge group G are expressed in (4.2) in terms of the ${}^L G$ parameters $({}^L\alpha, {}^L\beta, {}^L\gamma, {}^L\eta)$. In particular, ${}^L\gamma + i{}^L\alpha$, on which the model at $\Psi = \infty$ depends holomorphically, is equal according to (4.3) to $(\text{Im } \tau)\gamma^* - i\eta$. So we expect the A -model at $\Psi = 0$ to vary holomorphically in that variable. Indeed, $(\text{Im } \tau)\gamma^* - i\eta$ is the complexified Kahler class from the standpoint of complex structure K . (The symplectic form $\omega = (\text{Im } \tau)\omega_K$ has cohomology class proportional to $(\text{Im } \tau)\gamma^*$, in view of (3.77), while η supplies the imaginary part of the complexified Kahler class.) Likewise,

since the model at $\Psi = \infty$ with gauge group ${}^L G$ is locally independent of ${}^L \beta$ and ${}^L \eta$, we expect the dual model at $\Psi = 0$ with gauge group G to be locally independent of α and β . This is in accord with the fact that, as a symplectic variety with symplectic structure ω_K , \mathcal{M}_H is independent of α and β . The A -model depends only on this symplectic variety with its complexified Kahler form $(\text{Im } \tau)\gamma^* - i\eta$. On the other hand, the complex structure K , which is irrelevant in the A -model, varies holomorphically in $\alpha + i\beta$.

The duality transformation S maps branes in the B -model of type J to branes in the A -model of type K . So in particular, the zerobrane \mathcal{B}_E that is determined by a ramified flat bundle E is mapped to an A -brane $\widehat{\mathcal{B}}_E$. We can see quite concretely what sort of A -brane this will be. With or without ramification, \mathcal{M}_H admits the Hitchin fibration $\pi : \mathcal{M}_H \rightarrow \mathbf{B}$, as we described in section 3.9. The generic fibers of the Hitchin fibration are complex tori, holomorphic in complex structure I . S -duality acts via T -duality on the fibers of the Hitchin fibration, according to the same reasoning as in sections 5.4 and 5.5 of [8]. (This duality has been studied mathematically from several points of view, including applications to the geometric Langlands program [69]–[71].) T -duality on the fibers of a torus fibration maps a zero-brane to a brane wrapped on a fiber, and endowed with a flat Chan-Paton line bundle \mathcal{L} . So just as in the absence of ramification, the S -dual of a zerobrane is a brane of this type. Such a brane is called in [8] a brane of type \mathbf{F} .

Now we can make a preliminary statement of the geometric Langlands duality. In stating it, we think of \mathcal{M}_H not just as a classical space but as defining a quantum sigma model, so we will include η when we list its parameters. Also, in making the statement, we make the gauge groups explicit, recalling that electric-magnetic duality exchanges G and ${}^L G$.

Our first statement of geometric Langlands duality is that for every zerobrane in the sigma model with target $\mathcal{M}_H({}^L \alpha, {}^L \beta, {}^L \gamma, {}^L \eta, p; {}^L G)$, there is a corresponding brane of type \mathbf{F} in the sigma model with target $\mathcal{M}_H(\alpha, \beta, \gamma, \eta, p; G)$; as usual, the parameters are related by (4.2). More generally, for every B -brane of type J (that is, every brane of the B -model in complex structure J) in the sigma model with target $\mathcal{M}_H({}^L \alpha, {}^L \beta, {}^L \gamma, {}^L \eta, p; {}^L G)$, there is a naturally corresponding A -brane of type K in the sigma model with target $\mathcal{M}_H(\alpha, \beta, \gamma, \eta, p; G)$. This correspondence extends to a natural correspondence between all of the structures of the B -model on one side and all of the structures of the A -model on the other side.

The analogs of Wilson and 't Hooft operators are described in section 4.5, and some details will be clarified in section 4.6. But the most pressing problem is that the geometric Langlands correspondence is usually stated with B -branes of type J on the left hand side (just as in the

last paragraph), but with \mathcal{D} -modules of an appropriate sort, rather than A -branes of type K , on the right hand side. To reconcile the two points of view, we will follow the approach of section 11 of [8], which the reader may want to consult.

4.4. Twisted \mathcal{D} -Modules. In our formulation, the right hand side of the geometric Langlands correspondence involves a brane in the topological field theory at $\Psi = 0$. As in [8], a convenient way to get to $\Psi = 0$ is to take the twisting parameter t to equal 1, and to take the four-dimensional theta-angle $\theta = 2\pi \operatorname{Re} \tau$ to vanish. The branes at $\Psi = 0$ are branes of the A -model of type K .

The usual statement of the geometric Langlands duality involves \mathcal{D} -modules on the moduli space $\mathcal{M}(\alpha, p; G)$ of parabolic bundles. To reconcile the two formulations, we must associate to every A -brane of type K on $\mathcal{M}_H(\alpha, \beta, \gamma, \eta, p; G)$ a \mathcal{D} -module of a suitable type on $\mathcal{M}(\alpha, p; G)$. When confusion is unlikely, to make the notation less clumsy, we will generally omit to specify the gauge group G .

As explained in section 3.6, precisely if $\beta = \gamma = 0$ and α is regular, $\mathcal{M}_H(\alpha, \beta, \gamma, \eta; p)$ is birational in complex structure I to the cotangent bundle of $\mathcal{M}(\alpha; p)$. More precisely, for sufficiently large g_C , away from high codimension, \mathcal{M}_H contains the cotangent bundle as a dense open set:

$$(4.9) \quad T^*\mathcal{M}(\alpha, p; G) \subset \mathcal{M}_H(\alpha, 0, 0, \eta; p).$$

Setting β to zero and assuming that α is regular are not severe restrictions, since the A -model of type K is independent of these parameters. But the A -model does depend on γ , so we will want to restore the γ dependence later.

Also, taking α to be regular means choosing an affine Weyl chamber that contains α . This in a sense reduces the symmetry of the model. So one could argue that the description of the duality in terms of \mathcal{D} -modules, which depends on this choice, is less natural than the description in terms of mirror symmetry between an A -model and a B -model.²² However, the description by \mathcal{D} -modules is motivated by an analogy with number theory, and we want to see how it comes about.

Let X be any hyper-Kähler manifold with complex structures I, J, K and symplectic forms $\omega_I, \omega_J, \omega_K$ such that $[\omega_J] = 0$ (we recall that $[\omega]$ denotes the cohomology class of a closed form ω). Pick a positive real number $\operatorname{Im} \tau$, and consider the A -model on X with symplectic structure $\omega = (\operatorname{Im} \tau) \omega_K$. There is as explained in section 11.1 of [8] a distinguished space-filling A -brane. Its Chan-Paton line bundle is a

²²One could argue the same based on the fact that the mirror symmetry preserves the full supersymmetry of branes on both sides of the duality, while the description in terms of \mathcal{D} -modules does not. For an explanation of this point, see the introduction to section 5.

trivial line bundle endowed with a connection whose curvature form is $F = (\text{Im } \tau) \omega_J$. This formula ensures that $(\omega)^{-1}F$ is a complex structure, indeed the one that we have called I . This is the condition for a coisotropic brane in the sense of [72]. We call this brane the c.c. or canonical coisotropic brane, and denote it as $\mathcal{B}_{\text{c.c.}}$. For $\mathcal{B}_{\text{c.c.}}$ to exist, the cohomology class $[\omega_J]$ must vanish, so that it is possible for a trivial line bundle to have a connection form $F = (\text{Im } \tau) \omega_J$. Indeed, $[\omega_J] = 0$ precisely if $\beta = 0$, as was explained most directly at the end of section 3.1. At any rate, we have already set $\beta = 0$ to ensure the relationship of \mathcal{M}_H to $T^*\mathcal{M}$.

For any brane \mathcal{B} , the $(\mathcal{B}, \mathcal{B})$ strings form a ring. To all orders in sigma model perturbation theory, one can “localize” the $(\mathcal{B}, \mathcal{B})$ strings, by considering wavefunctions that are regular in an open set $\mathcal{U} \subset X$. The open strings that are regular in \mathcal{U} form a ring, and by letting \mathcal{U} vary, one gets, to all orders in perturbation theory, a sheaf of rings over X . In general, this construction of a sheaf of rings is only valid to all orders in perturbation theory, not as an exact statement. However, as explained in [8] (and as is certainly known in the mathematical literature; for a much deeper analysis, see [73]), one can under certain conditions go beyond perturbation theory in a very special way in the case of the canonical coisotropic brane $\mathcal{B}_{\text{c.c.}}$ of a hyper-Kähler manifold X . The requisite conditions are that X should contain as a dense open set a cotangent bundle $T^*\mathcal{M}$ (for some Kähler manifold \mathcal{M}), where this identification is holomorphic in complex structure I and ω_K is the imaginary part of the natural holomorphic two-form of the cotangent bundle. In this case, one can associate a ring of $(\mathcal{B}_{\text{c.c.}}, \mathcal{B}_{\text{c.c.}})$ strings to each open set in X that is of the form $T^*\mathcal{U}$ for \mathcal{U} an open set in \mathcal{M} . This association gives a sheaf of rings over \mathcal{M} . This sheaf of rings is precisely the sheaf of differential operators on \mathcal{M} acting on sections of some line bundle²³ \mathcal{L} . We write $\mathcal{D}_{\mathcal{L}}$ for this sheaf of rings.

\mathcal{L} can be identified in a particularly simple way if the sigma model of X has time-reversal symmetry, which is the case if its theta-angles vanish. If $K_{\mathcal{M}}$ denotes the canonical line bundle of \mathcal{M} , then time-reversal symmetry implies that $\mathcal{L} = K_{\mathcal{M}}^{1/2}$ and the sheaf of rings is $\mathcal{D}_{K_{\mathcal{M}}^{1/2}}$.

To apply this to our problem, we want the theta-angles of the sigma model with target $\mathcal{M}_H(\alpha, \beta, \gamma, \eta; p)$ to vanish. We have already set the four-dimensional parameter θ to zero; the remaining theta-angles of the sigma model are precisely η . So we achieve the time-reversal symmetry

²³More generally, as reviewed in section 11.1 of [8], \mathcal{L} may be a complex power of a line bundle, or a tensor product of such complex powers. The sheaf of rings $\mathcal{D}_{\mathcal{L}}$ is invariant under twisting \mathcal{L} by a flat line bundle, and as a result makes sense in this greater generality [74]. For the same reason, in what follows it does not matter if $K_{\mathcal{M}}^{1/2}$ exists or is unique as a line bundle. It actually does exist but is not always unique.

by setting $\eta = 0$. We have already simplified the problem by setting $\gamma = 0$, so our combined condition is equivalent to the vanishing of the complexified Kahler class $(\text{Im } \tau)\gamma^* - i\eta$. The condition $\eta = 0$ is equivalent, of course, to ${}^L\alpha = 0$.

The fact that the sheaf of $(\mathcal{B}_{\text{c.c.}}, \mathcal{B}_{\text{c.c.}})$ strings coincides with the sheaf of rings $\mathcal{D}_{K_{\mathcal{M}}^{1/2}}$ makes possible a very general construction. Let \mathcal{B}' be any A -brane on X of type K . To \mathcal{B}' , we can associate the sheaf of $(\mathcal{B}_{\text{c.c.}}, \mathcal{B}')$ strings. It is automatically a sheaf of modules for the $(\mathcal{B}_{\text{c.c.}}, \mathcal{B}_{\text{c.c.}})$ strings, that is, for $\mathcal{D}_{K_{\mathcal{M}}^{1/2}}$. So we get a natural way to associate a $\mathcal{D}_{K_{\mathcal{M}}^{1/2}}$ -module to every A -brane of type K .

Combining this with what we learned from S -duality, we see that a B -brane of type J on $\mathcal{M}_H(0, {}^L\beta, 0, {}^L\eta, p; {}^L G)$ is naturally associated to a sheaf of modules for the sheaf of rings $\mathcal{D}_{K_{\mathcal{M}}^{1/2}}$ over $\mathcal{M}(\alpha, p; G)$ (with as usual $\alpha = {}^L\eta$). This is essentially²⁴ the usual statement of the geometric Langlands program with tame ramification, for the case ${}^L\alpha = {}^L\gamma = 0$, or in other words for flat ${}^L G$ bundles with unipotent monodromy.

As in [8], the fact that the $(\mathcal{B}_{\text{c.c.}}, \mathcal{B}_{\text{c.c.}})$ strings are the differential operators on $K_{\mathcal{M}}^{1/2}$ (rather than on some more general line bundle) can be seen more explicitly, without relying on time-reversal symmetry. For this, we make use of another important brane on \mathcal{M}_H : the brane \mathcal{B}' supported at $\varphi = 0$, that is, on the zero section of the cotangent bundle $T^*\mathcal{M}$, and endowed with a trivial Chan-Paton line bundle. Quantization of the $(\mathcal{B}_{\text{c.c.}}, \mathcal{B}')$ strings shows that they are sections of $K_{\mathcal{M}}^{1/2}$. Since they also furnish a sheaf of modules for the $(\mathcal{B}_{\text{c.c.}}, \mathcal{B}_{\text{c.c.}})$ strings, the latter are the differential operators acting on sections of $K_{\mathcal{M}}^{1/2}$.

Restoring The Parameters

The case $\gamma = \eta = 0$ is the case that $U = \exp(-2\pi({}^L\alpha - i{}^L\gamma))$ is equal to 1. Being limited to this case would mean describing geometric Langlands only for the case of unipotent monodromy. To get beyond this case, we must restore the dependence on γ and η .

As we will see, incorporating either γ or η has the effect of replacing differential operators that act on $K_{\mathcal{M}}^{1/2}$ by differential operators that act on \mathcal{S} , where loosely speaking \mathcal{S} is a more general line bundle – more precisely, it is a tensor product of complex powers of line bundles. The sheaf of differential operators acting on such an \mathcal{S} makes sense, though \mathcal{S} itself cannot be defined as a line bundle. We aim to (i) justify the claim that for any γ and η , the A -branes on \mathcal{M}_H of type K are naturally

²⁴The usual statement involves the sheaf of rings \mathcal{D} , that is the sheaf of differential operators acting on functions. This sheaf of rings is Morita-equivalent to $\mathcal{D}_{K_{\mathcal{M}}^{1/2}}$. There are some subtleties that we will not consider here involving the dependence of this Morita-equivalence on a choice of spin structure on C .

associated with modules for some sheaf of rings of this type, and (ii) identify \mathcal{S} .

Let us first justify the claim (i). Let us consider the effect of having $\gamma \neq 0$, with η still vanishing. The result of this, as explained in section 3.6, is that \mathcal{M}_H is no longer birational to $T^*\mathcal{M}$. Rather, it contains as a dense open set an affine deformation of $T^*\mathcal{M}$ (giving us a situation similar to the case of $\theta \neq 0$, as analyzed in section 11.3 of [8]). By itself, replacing the cotangent bundle with an affine deformation of it does not modify the definition of the c.c. brane $\mathcal{B}_{c.c.}$ (as long as we keep $\beta = 0$; see below). Nor does it modify the argument that the sheaf of $(\mathcal{B}_{c.c.}, \mathcal{B}_{c.c.})$ strings is the sheaf of differential operators acting on some “line bundle” \mathcal{S} . However, it does spoil the use of time-reversal symmetry to show that $\mathcal{S} = K_{\mathcal{M}}^{1/2}$. Indeed, the definition of the relevant time-reversal operation requires that \mathcal{M}_H should have a symmetry $\varphi \rightarrow -\varphi$ (acting holomorphically in complex structure I); this symmetry is absent when $T^*\mathcal{M}$ is deformed to an affine bundle. The alternative argument for the identification $\mathcal{S} \cong K_{\mathcal{M}}^{1/2}$ uses the brane \mathcal{B}' supported on $\mathcal{M} \subset T^*\mathcal{M}$, which is precisely the zero-section of the cotangent bundle. This argument also fails for $\gamma \neq 0$, since the affine deformation of $T^*\mathcal{M}$ admits no holomorphic section.

Now let us consider the effect of having $\eta \neq 0$, with γ zero or nonzero. Because of holomorphy in $(\text{Im } \tau)\gamma^* - i\eta$, the effect of η must be similar to that of γ , but it is interesting to see how this comes about. Having $\eta \neq 0$ causes the B -field of the sigma model, which we simply call B , to be nonzero. As a result, some care is needed in defining the canonical coisotropic brane. Let F denote the curvature of the Chan-Paton bundle \mathcal{L} of this brane. In the presence of a B -field, the condition for a coisotropic brane is that $\omega^{-1}(F + B)$ should be an integrable complex structure, which we will take to be I . We achieve this by taking $F + B = (\text{Im } \tau)\omega_J$. We will still assume that \mathcal{L} is topologically trivial. This requires at the level of cohomology classes that $[F] = 0$, so $[B] = (\text{Im } \tau)[\omega_J]$. Since $[\omega_J/2\pi] = -\beta^*$, as we learned in (3.77), and $[B/2\pi] = \eta$, the c.c. brane can be defined only for $\beta^* = -(\text{Im } \tau)^{-1}\eta$. So for nonzero η , we must have $\beta \neq 0$, and hence again \mathcal{M}_H is birational to an affine deformation of $T^*\mathcal{M}$, not to $T^*\mathcal{M}$ itself.

We assumed here that \mathcal{L} is topologically trivial, but we can also construct branes with non-trivial \mathcal{L} . As we explain below, the results that can be obtained this way are Morita-equivalent to what we can learn from the case that \mathcal{L} is topologically trivial.

As before, once the brane $\mathcal{B}_{c.c.}$ is defined, the standard arguments show that the sheaf of $(\mathcal{B}_{c.c.}, \mathcal{B}_{c.c.})$ strings is the sheaf of differential operators acting on some “line bundle” $\mathcal{S} \rightarrow \mathcal{M}$. However, as in the case of deforming γ , we cannot argue that $\mathcal{S} = K_{\mathcal{M}}^{1/2}$. We cannot use

time-reversal invariance to prove that $\mathcal{S} = K_{\mathcal{M}}^{1/2}$, because having $\eta \neq 0$ violates time-reversal symmetry. What happens if we try to argue using the brane \mathcal{B}' is described later.

To determine \mathcal{S} , we proceed as follows. We already know that at $\gamma = \eta = 0$, $\mathcal{S} = K_{\mathcal{M}}^{1/2}$. So we write $\mathcal{S} = K_{\mathcal{M}}^{1/2} \otimes \mathcal{S}'$, where \mathcal{S}' is trivial for $\gamma = \eta = 0$. As a tensor product of complex powers of line bundles over \mathcal{M} , \mathcal{S}' has a first Chern class $w \in H^2(\mathcal{M}, \mathbb{C})$. (Concretely, if $\mathcal{S}' = \otimes_i \mathcal{L}_i^{g_i}$ with ordinary line bundles \mathcal{L}_i and $g_i \in \mathbb{C}$, then $w = \sum_i g_i c_1(\mathcal{L}_i)$.) We formally write $\mathcal{S}' = \mathcal{L}^w$ to express the statement that the first Chern class of \mathcal{S}' is w . We wish to determine w .

Actually, w is linear in the Kahler parameter $z = (\text{Im } \tau)\gamma^* - i\eta$. This follows by precisely the reasoning that was used in section 11.3 of [8] to show that an analogous exponent $f(\theta)$ depends linearly on the four-dimensional parameter θ . The idea is to calculate for weak coupling, taking $e^2 = 4\pi/\text{Im } \tau$ to be small with z fixed. Then w can be computed in perturbation theory in $z/\text{Im } \tau$. Concretely, perturbation theory is used to compute the cocycle that enters in an explicit description of the ring structure of the sheaf of $(\mathcal{B}_{c.c.}, \mathcal{B}_{c.c.})$ strings. The structure of perturbation theory is such that only a linear term in z can appear. For more detail, see the discussion of eqn. (11.39) in [8].

It remains to determine exactly which linear function of $(\text{Im } \tau)\gamma^* - i\eta$ is equal to w . w takes values in the second cohomology of $\mathcal{M}(\alpha, p)$ with complex coefficients. According to (3.47) and (3.56), this is $(\mathbb{Z} \oplus \Lambda_{\text{char}}) \otimes \mathbb{C}$. As $\eta + i(\text{Im } \tau)\gamma^*$ takes values in $\Lambda_{\text{char}} \otimes \mathbb{C}$, there is a naturally defined linear function

$$(4.10) \quad w(\gamma, \eta) = 0 \oplus (-\eta - i(\text{Im } \tau)\gamma^*),$$

and we claim that this is the right result.

By analogy with the treatment of dependence on Ψ or θ in section 11.3 of [8], we will justify this result by showing that it holds when $\gamma = 0$ and η is equal to a lattice vector $v \in \Lambda_{\text{char}}$. The key step is to define the brane \mathcal{B}' that is, roughly speaking, supported on the zero section of the cotangent bundle $\mathcal{M} \subset T^*\mathcal{M} \subset \mathcal{M}_H$. Once we find this brane, the $(\mathcal{B}_{c.c.}, \mathcal{B}')$ strings give a natural sheaf of modules for the sheaf of rings provided by the $(\mathcal{B}_{c.c.}, \mathcal{B}_{c.c.})$ strings. By identifying this module, we can identify the sheaf of rings.

In carrying out this program, we run into an important detail. To define the brane $\mathcal{B}_{c.c.}$ at $\eta \neq 0$, assuming that its Chan-Paton bundle \mathcal{L} is trivial, we need to set $\beta^* = -(\text{Im } \tau)^{-1}\eta$, as we explained above. This means that from a holomorphic point of view in complex structure I , \mathcal{M}_H is not birational to $T^*\mathcal{M}$, but to an affine deformation thereof, and hence we do not have a holomorphic embedding $\mathcal{M} \subset \mathcal{M}_H$. This may appear to obstruct the definition of the brane \mathcal{B}' , which is supported on \mathcal{M} .

However, for our present purposes, we are not interested in what happens holomorphically in complex structure I . Rather, our concern is the A -model of type K , in which the target space is \mathcal{M}_H understood as a symplectic variety with symplectic structure a multiple of ω_K . As we explained in section 3.2, as a symplectic variety with symplectic structure ω_K , \mathcal{M}_H is naturally independent of β . The β -independence is established by thinking of \mathcal{M}_H as a moduli space of flat bundles in complex structure J .

For $\beta \neq 0$, we do not have an embedding $\mathcal{M} \subset \mathcal{M}_H$ that is holomorphic in complex structure I , but we do have such an embedding that is Lagrangian with respect to the symplectic structure ω_K . To get such an embedding, we start at $\beta = 0$ with the usual embedding $\mathcal{M} \subset T^*\mathcal{M} \subset \mathcal{M}_H$, holomorphic in complex structure I . Defined this way, \mathcal{M} is a complex Lagrangian submanifold in complex structure I – that is, it is holomorphic in complex structure I and Lagrangian for the holomorphic symplectic form $\Omega_I = \omega_J + i\omega_K$. In particular, \mathcal{M} is Lagrangian with respect to ω_K .

Now we change our point of view and think of \mathcal{M}_H as a complex symplectic manifold with complex structure J . From this point of view, \mathcal{M}_H is canonically independent of β . (We encounter no singularities in varying β , since we are taking α regular in order to aim for an answer involving \mathcal{D} -modules on $\mathcal{M}(\alpha, p)$.) When β is varied, we do not change the holomorphic symplectic structure $\Omega_J = \omega_K + i\omega_I$. So in particular, we vary β keeping fixed the symplectic structure ω_K . So the submanifold $\mathcal{M} \subset \mathcal{M}_H$ found in the last paragraph, if understood merely as a Lagrangian submanifold of type K , is naturally defined for any β . We take it to be literally independent of β , if β is varied keeping fixed the ramified flat bundle determined (in complex structure J) by a point in \mathcal{M}_H .

To get an A -brane \mathcal{B}' supported on \mathcal{M} , we need a unitary line bundle $\mathcal{L}' \rightarrow \mathcal{M}$ whose curvature F obeys $F + B|_{\mathcal{M}} = 0$. \mathcal{L}' will be the Chan-Paton line bundle of the brane \mathcal{B}' . Here B is the B -field which (since we assume that $\theta = 0$) is determined by η . In fact, B is a closed two-form on \mathcal{M}_H whose cohomology class is $0 \oplus \eta \subset (\mathbb{Z} \oplus \Lambda_{\text{char}}) \otimes_{\mathbb{Z}} \mathbb{R} = H^2(\mathcal{M}(\alpha, p); \mathbb{R})$. For generic η , this cohomology class is not a lattice vector, and hence no suitable line bundle \mathcal{L}' or brane \mathcal{B}' exists. However, if $\eta = v$ for some $v \in \Lambda_{\text{char}}$, then we can take ²⁵ $\mathcal{L}' = \mathcal{L}^{-v}$, that is, \mathcal{L}' is a line bundle with first Chern class $0 \oplus (-v)$.

Having thus defined an A -brane \mathcal{B}' of type K , we now consider the $(\mathcal{B}_{c.c.}, \mathcal{B}')$ strings, which will furnish a sheaf of modules for the sheaf of rings derived from the $(\mathcal{B}_{c.c.}, \mathcal{B}_{c.c.})$ strings. In quantizing the $(\mathcal{B}_{c.c.}, \mathcal{B}')$

²⁵The minus sign in the exponent of the following formula arises as follows. There is no minus sign in the relation between B and η : $[B/2\pi] = \eta$. So as $[F + B] = 0$, we have $[F/2\pi] = -\eta$. Hence if $\eta = v$, the first Chern class of the Chan-Paton line bundle is $-v$.

strings, the only thing that is new, compared to the case $\eta = 0$, is that the Chan-Paton line bundle of the brane \mathcal{B}' , instead of being trivial, is now $\mathcal{L}' = \mathcal{L}^{-v}$. This means that the sheaf of $(\mathcal{B}_{c.c.}, \mathcal{B}')$ strings, instead of being the sheaf of sections of $K_{\mathcal{M}}^{1/2}$, is now the sheaf of sections of $K_{\mathcal{M}}^{1/2} \otimes \mathcal{L}^{-v}$. This shows that (4.10) is valid for $\gamma = 0$, $\eta \in \Lambda_{\text{char}}$, and hence for all γ, η .

We can now restate the geometric Langlands duality for tamely ramified flat bundles in terms of \mathcal{D} -modules rather than mirror symmetry. On the left hand side of the geometric Langlands correspondence, we consider a ramified flat bundle $E \rightarrow C$ with structure group ${}^L G$ and monodromy whose orbit contains in its closure the semisimple element $U = \exp(-2\pi(L\alpha - iL\gamma))$. On the right hand side, the dual of such a ramified flat bundle is a sheaf of modules for the sheaf of differential operators on $\mathcal{M}(\alpha, p; G)$ twisted by $K_{\mathcal{M}}^{1/2} \otimes \mathcal{L}^{-(\eta + i(\text{Im } \tau)\gamma^*)}$, which is the same as $K_{\mathcal{M}}^{1/2} \otimes \mathcal{L}^{L\alpha - iL\gamma}$. Thus, the ‘‘exponent’’ of the line bundle is the same as the ‘‘eigenvalue’’ of the monodromy. S -duality establishes a natural correspondence between objects of these two kinds. (For a flat bundle with non-regular monodromy, this statement needs some clarification, which we defer to section 4.6.) This version of the geometric Langlands duality has been conjectured mathematically. See the survey in section 9.4 of [7], and additional references in the introduction.

Informal Explanation

The explanation that we have just given is the most precise one that we know. However, some readers may prefer an alternative explanation that we will just present informally.

Consider any complex variety \mathcal{M} and let $X = T^*\mathcal{M}$, endowed with its natural holomorphic symplectic form Ω_X . Quantization of X leads to differential operators on \mathcal{M} twisted by $K_{\mathcal{M}}^{1/2}$. Now let us replace X by a variety Y that is an affine deformation of the cotangent bundle, meaning that there is a holomorphic fibration $\pi : Y \rightarrow \mathcal{M}$, whose fibers are those of the cotangent bundle, but π admits no holomorphic section. Locally, one can pick a holomorphic section and identify Y with X ; globally, the obstruction to this is determined by an element $w \in H^1(\mathcal{M}, T^*\mathcal{M})$. Let us furthermore require that Y admits a holomorphic symplectic form Ω_Y that locally (once we pick a local holomorphic section of π , giving a local identification of Y with X) coincides with Ω_X . This is equivalent to saying that w can be represented by a complex-valued $(1, 1)$ -form on \mathcal{M} that is annihilated by both the $\bar{\partial}$ and ∂ operators and hence in particular by $d = \partial + \bar{\partial}$. This is automatically so if \mathcal{M} is a compact Kahler manifold. In this situation, it is natural to claim that quantization of Y with symplectic form Ω_Y leads to differential operators on \mathcal{M} twisted by $K_{\mathcal{M}}^{1/2} \otimes \mathcal{L}^w$, where \mathcal{L}^w is symbolically a ‘‘line bundle’’ with first Chern class w .

A convenient way to characterize the twisting parameter w is as follows. The holomorphic symplectic form Ω_X has vanishing cohomology class. Indeed, if q^α are local coordinates on \mathcal{M} and p_α are the canonical momenta, then $\Omega_X = \sum_\alpha dp_\alpha \wedge dq^\alpha = d(\sum_\alpha p_\alpha dq^\alpha)$. Here the one-form $\lambda = \sum_\alpha p_\alpha dq^\alpha$ is globally defined, so $[\Omega_X] = 0$. By contrast, it is a standard result that the cohomology class of Ω_Y is equal to the class in $H^2(\mathcal{M}, \mathbb{C})$ of the closed $(1, 1)$ -form w (or more precisely, the pullback of this class to Y). We will simply write w for this cohomology class.

Now in our problem, \mathcal{M}_H in complex structure I is (away from a set of high codimension) an affine deformation of the cotangent bundle of \mathcal{M} , with holomorphic symplectic form $\Omega = (\text{Im } \tau)\Omega_I = (\text{Im } \tau)(\omega_J + i\omega_K)$. According to (3.77), the cohomology class of Ω is $w = -(\text{Im } \tau)(\beta^* + i\gamma^*)$. To construct the c.c. brane that is needed to relate the category of A -branes to \mathcal{D} -modules, we had to take $\beta^* = -(\text{Im } \tau)^{-1}\eta$, so the twisting parameter is actually $w = \eta - i(\text{Im } \tau)\gamma^*$, as found above.

The reason that we consider this explanation heuristic, even if perhaps more understandable for some readers, is that we are not entitled to arbitrarily postulate what properties the c.c. brane should have. This should be deduced as part of the standard framework of quantum field theory. The arguments given above, together with those in section 11 of [8], are an attempt to do this.

Symmetry Of Lattice Shift

One might now ask why have we not, in the above derivation, seen the symmetry of shifting η by a lattice vector, say $\eta \rightarrow \eta + v_0$ for $v_0 \in \Lambda_{\text{char}}$. In fact, we lost this symmetry by assuming that the Chan-Paton bundle \mathcal{L} of the c.c. brane was trivial. The lattice shift $\eta \rightarrow \eta + v_0$ acts on all branes by tensoring their Chan-Paton bundle with \mathcal{L}^{-v_0} . To restore the symmetry, we simply add additional coisotropic branes $\mathcal{B}_{c.c.}^{v_0}$ constructed from $\mathcal{B}_{c.c.}$ by shifting η by v_0 and tensoring the Chan-Paton bundle with \mathcal{L}^{-v_0} . We do not need to introduce any more \mathcal{B}' branes. In fact, the full set of \mathcal{B}' branes constructed above already has the shift symmetry, as these branes were defined for all possible lattice vectors v and all β .

From a mathematical point of view, what we learn by contemplating the shift symmetry is something called Morita-equivalence. If \mathcal{L}^{v_0} is an honest line bundle (rather than a complex power of line bundles), then the sheaf of differential operators acting on sections of $K_{\mathcal{M}}^{1/2} \otimes \mathcal{L}^w \otimes \mathcal{L}^{v_0}$ is Morita-equivalent to the sheaf of differential operators acting on sections of $K_{\mathcal{M}}^{1/2} \otimes \mathcal{L}^w$. The Morita-equivalence is established by considering a ‘‘bimodule’’ which in the present context is the sheaf of $(\mathcal{B}_{c.c.}, \mathcal{B}_{c.c.}^{v_0})$ strings.

Analog of Type \mathbb{L}

Everything that we have said has a direct analog for an arbitrary Levi subgroup \mathbb{L} , as opposed to the case $\mathbb{L} = \mathbb{T}$ that we have considered so far. We write $\mathcal{W}_{\mathbb{L}}$ for the Weyl group of \mathbb{L} .

For general \mathbb{L} , the parameters $(\alpha, \beta, \gamma, \eta)$ are restricted to be $\mathcal{W}_{\mathbb{L}}$ -invariant. With this understood, the statement of the duality as a relation between a B -model at $\Psi = \infty$ and an A -model at $\Psi = 0$ requires no essential modification. The duality establishes a natural correspondence between B -branes of $\mathcal{M}_{H, \mathbb{L}}({}^L\alpha, {}^L\beta, {}^L\gamma, {}^L\eta; p, {}^L G)$ and A -branes of $\mathcal{M}_{H, \mathbb{L}}(\alpha, \beta, \gamma, \eta; p, G)$, with the usual relation among the parameters.

To express the duality in terms of \mathcal{D} -modules, we must assume that α is \mathbb{L} -regular. Then we denote as \mathcal{P} the parabolic subgroup determined by the pair (\mathbb{L}, α) , and we introduce $\mathcal{M}_{\mathbb{L}}(\alpha; p, G)$, the moduli space of G -bundles over C with parabolic structure of type \mathcal{P} at the point p . For $\beta = \gamma = 0$, the Higgs bundle moduli space²⁶ $\mathcal{M}_{H, \mathbb{L}}(\alpha, \beta, \gamma; p, G)$ contains $T^*\mathcal{M}_{\mathbb{L}}(\alpha; p, G)$ as a dense open subspace, as we observed in (3.73). For $\beta, \gamma \neq 0$, $\mathcal{M}_{H, \mathbb{L}}(\alpha, \beta, \gamma; p, G)$ contains as a dense open subspace an affine deformation of the cotangent bundle.

These geometrical facts and the existence of a canonical coisotropic brane can be used exactly as above to get a natural map from an A -brane of $\mathcal{M}_{H, \mathbb{L}}(\alpha, \beta, \gamma, \eta; p, G)$ to a \mathcal{D} -module. All the key statements and arguments are the same. The only changes are that the statements now refer to $\mathcal{M}_{H, \mathbb{L}}$ and $\mathcal{M}_{\mathbb{L}}$ (rather than \mathcal{M}_H and \mathcal{M}), and the parameters are $\mathcal{W}_{\mathbb{L}}$ -invariant.

In one sense, the statement for general \mathbb{L} is slightly less natural than the statement for $\mathbb{L} = \mathbb{T}$. Different choices of \mathbb{L} -regular α (which correspond of course to different choices of ${}^L\eta$) lead to different choices of \mathcal{P} . They therefore lead to \mathcal{D} -modules over moduli spaces of bundles with parabolic structures of different types. The parabolic structure is of type \mathcal{P} , where \mathcal{P} can be any parabolic subgroup of $G_{\mathbb{C}}$ that contains \mathbb{L} as a Levi subgroup. These different spaces, however, are birationally equivalent.

4.5. Line Operators And Monodromies. In the absence of ramification, line operators, supported at a point $q \in C$ (times a line in Σ that runs along its boundary), act in a natural way on branes. This is explained in section 6.4 of [8].

At $\Psi = \infty$, the natural line operators are the Wilson operators, which are classified by a choice of representation of the gauge group. We take the gauge group at $\Psi = \infty$ to be ${}^L G$, so a Wilson operator is labeled by a representation of that group. In general, Wilson operators can change the discrete electric flux \mathbf{e}_0 . The Wilson operators that leave fixed \mathbf{e}_0 are those that are derived from representations of ${}^L G_{\text{ad}}$, the adjoint form of ${}^L G$.

²⁶We omit η here as it does not enter the classical geometry.

At $\Psi = 0$, the natural line operators are 't Hooft operators, and the gauge group is G . The 't Hooft operators of G gauge theory are constructed using singular BPS monopoles and again are classified by representations of ${}^L G$. In general, 't Hooft operators change the discrete magnetic flux \mathbf{m}_0 (into which \mathbf{e}_0 transforms under the duality). The 't Hooft operators that leave \mathbf{m}_0 fixed are those associated with representations of ${}^L G_{\text{ad}}$.

So at either $\Psi = \infty$ or $\Psi = 0$, the line operators that act at a given point $q \in C$, in the absence of ramification, correspond to representations of ${}^L G$, or of ${}^L G_{\text{ad}}$ if we want only operators that leave the discrete fluxes fixed. The composition of these Wilson or 't Hooft operators corresponds to the tensor product of representations. A central statement of the geometric Langlands program is that the duality between ${}^L G$ and G maps a Wilson operator labeled by a representation of ${}^L G$ to an 't Hooft operator labeled by the same representation.

Now suppose that $p \in C$ is a ramification point. We want to determine what structure at p replaces the action of the Wilson or 't Hooft line operators at a generic point. We will show that the answer to this question can be described in terms of the monodromies in the space of ramification parameters.

Duality-Symmetric Monodromies

We begin by repeating the reasoning of section 3.7 in a duality symmetric way. For simplicity, we start with the case that the Levi subgroup is $\mathbb{L} = \mathbb{T}$.

For gauge group G , the definition of ramified structure at a point p depends on a choice of parameters $(\alpha, \beta, \gamma, \eta) \in \mathfrak{t}^3 \times \mathfrak{t}^\vee$, modulo the action of a certain group of equivalences. This group is generated by (i) the translations of α by the lattice Λ_{cochar} ; (ii) the translations of η by the dual lattice Λ_{char} ; and (iii) the Weyl group \mathcal{W} . So the group that acts on the parameters is an extension

$$(4.11) \quad \widehat{\mathcal{V}} = (\Lambda_{\text{cochar}} \oplus \Lambda_{\text{char}}) \rtimes \mathcal{W}.$$

The quotient $(\mathfrak{t}^3 \times \mathfrak{t}^\vee)/\widehat{\mathcal{V}}$ is the same as $(\mathbb{T} \times \mathfrak{t} \times \mathfrak{t} \times {}^L \mathbb{T})/\mathcal{W}$.

If we want to restrict to translations of α and η that do not shift the discrete fluxes \mathbf{m}_0 and \mathbf{e}_0 , we should replace Λ_{cochar} by Λ_{cort} and Λ_{char} by Λ_{rt} . Then we get a smaller group

$$(4.12) \quad \mathcal{V} = (\Lambda_{\text{cort}} \oplus \Lambda_{\text{rt}}) \rtimes \mathcal{W}.$$

\mathcal{V} is a sort of duality-symmetric version of the affine Weyl group \mathcal{W}_{aff} , since the affine Weyl group of G is $\Lambda_{\text{cort}} \rtimes \mathcal{W}$, and the affine Weyl group of ${}^L G$ is $\Lambda_{\text{rt}} \rtimes \mathcal{W}$. The relationship between \mathcal{V} and $\widehat{\mathcal{V}}$ is extremely simple. The quotient $\Lambda_{\text{cochar}}/\Lambda_{\text{cort}}$ is $\pi_1(G)$, and likewise $\Lambda_{\text{char}}/\Lambda_{\text{rt}} = \pi({}^L G)$. The Weyl group acts trivially on each of these, so we have a group

extension

$$(4.13) \quad 1 \rightarrow \mathcal{V} \rightarrow \widehat{\mathcal{V}} \rightarrow \pi_1(G) \times \pi_1({}^L G) \rightarrow 1.$$

We say that a point $(\alpha, \beta, \gamma, \eta) \in \mathfrak{t}^3 \times \mathfrak{t}^\vee$ is \mathcal{V} -regular if it is not invariant under any non-trivial element of \mathcal{V} . According to section 3.6, the points that are not regular in this sense are the points at which the sigma model with target $\mathcal{M}_H(\alpha, \beta, \gamma, \eta; p)$ has a local singularity. We let Y denote the complement of the non- \mathcal{V} -regular points in $\mathfrak{t}^3 \times \mathfrak{t}^\vee$. (With two or more ramification points, we would also omit the locus of global singularities in defining Y ; this has real codimension four and will not affect the argument.) The group \mathcal{V} acts freely on Y so the quotient Y/\mathcal{V} is a smooth manifold.

We have therefore a family of smooth $(4, 4)$ sigma models parametrized by the manifold Y/\mathcal{V} (as well as other data, such as the gauge coupling parameter τ). We can now make an argument that is a sort of quantum version of the reasoning in section 3.7. A basic fact about branes is that every brane has a charge or K -theory class, taking values roughly speaking in the complex K -theory of the target space \mathcal{M}_H . $K(\mathcal{M}_H)$ varies with $(\alpha, \beta, \gamma, \eta)$ as the fiber of a flat bundle over Y/\mathcal{V} . Taking monodromies, we get an action of \mathcal{V} on $K(\mathcal{M}_H)$.

To be more precise about this, we must recall that \mathcal{M}_H has distinct topological components labeled by \mathfrak{m}_0 . Also, for each \mathfrak{e}_0 , one defines a space of twisted branes, whose charge takes values in a twisted version²⁷ of K -theory that we might call $K_{\mathfrak{e}_0}(\mathcal{M}_H)$. Let $K_{\mathfrak{e}_0}(\mathcal{M}_H; \mathfrak{m}_0)$ be the \mathfrak{e}_0 -twisted K -theory of the component of \mathcal{M}_H defined by \mathfrak{m}_0 . The above argument shows that \mathcal{V} acts on $K_{\mathfrak{e}_0}(\mathcal{M}_H; \mathfrak{m}_0)$ for each choice of \mathfrak{e}_0 and \mathfrak{m}_0 . The larger group $\widehat{\mathcal{V}}$ also acts, by a slight extension of this reasoning; its action, of course, permutes the possible values of \mathfrak{e}_0 and \mathfrak{m}_0 . In particular, as an abelian group, $K_{\mathfrak{e}_0}(\mathcal{M}_H; \mathfrak{m}_0)$ is independent of \mathfrak{e}_0 and \mathfrak{m}_0 up to isomorphism.

We can understand in a relatively concrete way how \mathcal{V} (or its extension $\widehat{\mathcal{V}}$) acts on $K(\mathcal{M}_H)$. The action of the subgroup $\mathcal{W}_{\text{aff}} = \Lambda_{\text{cort}} \rtimes \mathcal{W}$ was already described in section 3.7.

On the other hand, the shift $\eta \rightarrow \eta + v$, for $v \in \Lambda_{\text{rt}}$, is a shift of the world-sheet B -field that changes its periods by integer multiples of 2π . Such a shift acts on a brane \mathcal{B} by tensoring the Chan-Paton vector bundle of \mathcal{B} by the line bundle \mathcal{L}^v whose first Chern class is v . The associated action on K -theory is thus simply the tensor product with \mathcal{L}^v , for $v \in \Lambda_{\text{rt}}$. Since this action is clear classically, the fact that \mathcal{V} acts on $K(\mathcal{M}_H)$ does not by itself tell us much beyond the action of \mathcal{W}_{aff} , which we already know from section 3.7.

²⁷This is the K -theory of twisted vector bundles, which are described in the present context in section 7.1 of [8]. The idea is that \mathfrak{e}_0 defines a flat gerbe over \mathcal{M}_H , and $K_{\mathfrak{e}_0}(\mathcal{M}_H)$ is the twisted K -theory defined relative to this gerbe.

What does appear to give more information is, however, the action of $S : \tau \rightarrow -1/n_{\mathfrak{g}}\tau$. This exchanges $K(\mathcal{M}_H({}^L G, C))$ with $K(\mathcal{M}_H(G, C))$ while exchanging Λ_{rt} with Λ_{cort} (and \mathfrak{e}_0 with \mathfrak{m}_0). So it exchanges the action of Λ_{rt} via twisting by a line bundle with the action of Λ_{cort} via monodromy. This symmetry of the K -theory of \mathcal{M}_H is not obvious classically.

We will briefly indicate a simple example of how this works. The identity element of the multiplicative structure of $K(\mathcal{M}_H)$ is the K -theory class of a brane \mathcal{B} whose support is all of \mathcal{M}_H and whose Chan-Paton line bundle is trivial. This brane varies continuously when we vary α , β , and γ , so its K -theory class is invariant under the action of the affine Weyl group $\mathcal{W}_{\text{aff}} = \Lambda_{\text{cort}} \rtimes \mathcal{W}$. So the duality operation S must transform it into a brane $\tilde{\mathcal{B}}$ that is invariant under the action of Λ_{rt} . In other words, the K -theory class of $\tilde{\mathcal{B}}$ must be invariant under the operation of tensoring the Chan-Paton bundle of $\tilde{\mathcal{B}}$ by a line bundle of the form \mathcal{L}^v . This statement means that \mathcal{L}^v must be (topologically) trivial when restricted to the support of $\tilde{\mathcal{B}}$. Actually, $\tilde{\mathcal{B}}$ is a brane supported on a section of the Hitchin fibration, so its support is isomorphic to the base \mathbf{B} of this fibration. As this is a contractible space (isomorphic to \mathbb{C}^N for some integer N), any line bundle \mathcal{L}^v is indeed trivial when restricted to the support of $\tilde{\mathcal{B}}$.

Monodromies Of The A-Model

What we have said so far is hopefully interesting, but may not seem to get us very close to finding the analog of Wilson and 't Hooft operators.

To get farther, we must discuss, not branes in general, but specific kinds of branes. For example, let us discuss the branes of the A -model of type K , with gauge group G . This is the relevant model at $\Psi = 0$. The definition of this model depends on the parameters γ and η . So to study branes of the A -model, we should keep γ and η fixed. On the other hand, the A -model is locally independent of α and β . So the group of monodromies that we obtain by varying α and β will be a symmetry group of the A -model. This group will depend on γ and η .

The process of varying α and β while keeping \mathcal{M}_H fixed as a symplectic variety with symplectic structure ω_K was discussed in section 3.2. From this discussion, we know that when one goes around a loop in $\alpha - \beta$ space, avoiding singularities and keeping γ fixed, \mathcal{M}_H changes by a symplectomorphism. Thus, the group of symmetries of the A -model that we are about to analyze is simply a group of classical symplectomorphisms (or rather a group of components of the group of symplectomorphisms). This contrasts with the dual monodromies of the B -model, which involve varying the quantum parameter η , and so only make sense quantum mechanically.

Now let us discuss the action of \mathcal{V} . Since Λ_{rt} acts only on η , which is held fixed in studying the A -model and its monodromies, it will play no interesting role. We may as well divide by Λ_{rt} and replace the pair γ, η by $Y = \exp(-2\pi(\eta - i\gamma))$. Y takes values in, roughly speaking, the complex maximal torus ${}^L\mathbb{T}_{\mathbb{C}}$ of ${}^L G_{\mathbb{C}}$. To be more precise, since we have only divided by Λ_{rt} , not Λ_{char} , Y takes values in the maximal torus of ${}^L\overline{G}_{\mathbb{C}}$, the simply-connected cover of ${}^L G_{\mathbb{C}}$. Since the A -model depends on Y , we are not interested in monodromies that involve varying Y . So the relevant part of the Weyl group is the subgroup that leaves Y fixed. We call this \mathcal{W}_Y .

The subgroup of \mathcal{V} that acts only on α and β is an extension of \mathcal{W}_Y by Λ_{cort} :

$$(4.14) \quad \mathcal{V}_Y = \Lambda_{\text{cort}} \rtimes \mathcal{W}_Y.$$

So if $Y = 1$, \mathcal{V}_Y is just the affine Weyl group. We say that a pair $(\alpha, \beta) \in \mathfrak{t} \times \mathfrak{t}$ is \mathcal{V}_Y -regular if it is not left fixed by any element of \mathcal{V}_Y other than the identity. The pairs that are not \mathcal{V}_Y -regular correspond to non-regular quadruples $(\alpha, \beta, \gamma, \eta)$, and therefore to local singularities of \mathcal{M}_H . We define Z_Y to be the space of \mathcal{V}_Y -regular pairs. The group \mathcal{V}_Y acts freely on Z_Y , so we get a family of smooth sigma models parametrized by Z_Y/\mathcal{V}_Y . If we pass from the sigma model to the associated A -model of type K , the family becomes locally constant, since this A -model is locally independent of the parameters α and β . So the fundamental group $B_Y = \pi_1(Z_Y/\mathcal{V}_Y)$ acts as a group of automorphisms of the A -model. (B loosely stands for “braid,” as will be clearer in a moment.)

This reasoning is similar to what we presented earlier in discussing the actions of both \mathcal{W}_{aff} and \mathcal{V} . However, there are two notable differences:

(A) The pairs (α, β) that are not \mathcal{V}_Y -regular are in general of real codimension two, simply because we are considering only two variables α and β . (By contrast, nonregular triples (α, β, γ) or quadruples $(\alpha, \beta, \gamma, \eta)$ are of real codimension three or four, respectively.) As a result, the space Z_Y of \mathcal{V}_Y -regular pairs is not necessarily simply-connected, and the fundamental group B_Y of the quotient Z_Y/\mathcal{V}_Y is in general not equal to \mathcal{V}_Y . We do, however, get an exact sequence

$$(4.15) \quad 1 \rightarrow \pi_1(Z_Y) \rightarrow B_Y \rightarrow \mathcal{V}_Y \rightarrow 1.$$

(B) Related to this, in the case of two or more ramification points, the global singularities are of real codimension two and might play a role in a complete treatment. They can be avoided if one considers only branes supported for Higgs bundles (E, φ) such that the underlying bundle E , forgetting its parabolic structure, is stable. In this paper, we will not be concerned with the global singularities.

A useful alternative description is to divide first by the action of Λ_{cort} on α . Modulo this action, the pair (α, β) combine to the element

$T = \exp(-2\pi(\alpha + i\beta)) \in \overline{\mathbb{T}}_{\mathbb{C}}$, the maximal torus of the simply-connected cover $\overline{G}_{\mathbb{C}}$ of $G_{\mathbb{C}}$. We say that T is \mathcal{W}_Y -regular if it is not invariant under any non-trivial element of \mathcal{W}_Y . We write Q_Y for the space of \mathcal{W}_Y -regular elements of $\overline{\mathbb{T}}_{\mathbb{C}}$. Then \mathcal{W}_Y acts freely on Q_Y , and $B_Y = \pi_1(Q_Y/\mathcal{W}_Y)$.

As usual, to get some understanding of the group B_Y , it is helpful to consider the cases that Y is regular or $Y = 1$. Other cases are intermediate between these two. In the extreme cases, we have:

(1) If Y is regular, then $\mathcal{W}_Y = 1$, and $Q_Y = \overline{\mathbb{T}}_{\mathbb{C}}$. Then $B_Y = \pi_1(Q_Y/\mathcal{W}_Y) = \pi_1(Q_Y) = \Lambda_{\text{cort}}$.

(2) Alternatively, if $Y = 1$, then $\mathcal{W}_Y = \mathcal{W}$. So Q_Y is the space of regular points in $\overline{\mathbb{T}}_{\mathbb{C}}$, and $Q_Y/\mathcal{W}_Y = Q_Y/\mathcal{W}$ is the space of regular conjugacy classes in $\overline{G}_{\mathbb{C}}$. The fundamental group of this space is called the affine braid group²⁸ of ${}^L G$. We denote it as $B_{\text{aff}}({}^L G)$. So the monodromy group for $Y = 1$ is $B_1 = B_{\text{aff}}({}^L G)$. Now, for $Y = 1$, $\mathcal{V}_Y = \Lambda_{\text{cort}} \rtimes \mathcal{W}$ is the affine Weyl group of ${}^L G$. (4.15) in this case gives an exact sequence

$$(4.16) \quad 1 \rightarrow \pi_1(Z) \rightarrow B_{\text{aff}}({}^L G) \rightarrow \mathcal{W}_{\text{aff}}({}^L G) \rightarrow 1,$$

where $Z = Z_1$ is the space of regular pairs $(\alpha, \beta) \in \overline{\mathbb{T}} \times \mathfrak{t}$, or equivalently \mathcal{W}_{aff} -regular pairs $(\alpha, \beta) \in \mathfrak{t} \times \mathfrak{t}$.

The affine braid group was introduced in [29] in the context of, roughly speaking, a local version of the present problem. We have discussed the relationship in section 3.8 and return to it in section 6.

The group B_Y that we have defined is the group of monodromy transformations of the ramification parameters of the A -model that leave fixed \mathfrak{m}_0 . We can relax this condition and consider an extended group $\widehat{B}_Y = B_Y \rtimes \mathcal{Z}(G)$ that includes transformations that shift \mathfrak{m}_0 . For regular Y , $\widehat{B}_Y = \Lambda_{\text{cochar}}$, and for $Y = 1$, it is an extension

$$(4.17) \quad \widehat{B}_{\text{aff}}({}^L G) = B_{\text{aff}}({}^L G) \rtimes \mathcal{Z}(G).$$

We have carried out this analysis at $\Psi = 0$. Obviously we could make a similar analysis for the B -model with complex structure J , which corresponds to $\Psi = \infty$. This exchanges the roles of α and η along with G and ${}^L G$. The net effect is that the group of monodromies acting on the branes of the A -model of type K with gauge group G is the same as the group of monodromies of the B -model of type J with gauge group ${}^L G$. (However, as we noted above, the monodromy group of the A -model acts by classical symplectomorphisms, while that of the B -model is highly non-classical.) As we will now discuss, this generalizes the correspondence between 't Hooft operators in one case and Wilson operators in the other case.

²⁸The motivation for the name ‘‘affine braid group’’ is as follows. The ordinary braid group on n letters is the fundamental group of the space of unordered n -plets of distinct points in \mathbb{C} . If we replace \mathbb{C} by \mathbb{C}^* , we get the affine braid group of $GL(n, \mathbb{C})$, since an unordered n -plet in \mathbb{C}^* is equivalent to a regular semisimple conjugacy class of $GL(n, \mathbb{C})$.

Relation To Line Operators

We have expressed all this in terms of a group \mathcal{V}_Y that acts on the B -branes at $\Psi = \infty$ or A -branes at $\Psi = 0$. More generally, since we can take the direct sum of two branes or multiply a brane by a positive integer, positive integer linear combinations of elements of \mathcal{V}_Y can act on branes. Such linear combinations form what we will call the group semiring of \mathcal{V}_Y .

We will here mainly consider the case of regular Y . Then $\mathcal{V}_Y = \Lambda_{\text{cort}}$, and $\widehat{\mathcal{V}}_Y = \Lambda_{\text{cochar}}$. Let \mathcal{B} be a brane with Chan-Paton bundle \mathcal{U} . We found above that Λ_{cort} and Λ_{cochar} act on \mathcal{B} by $\mathcal{U} \rightarrow \mathcal{L}^v \otimes \mathcal{U}$, for $v \in \Lambda_{\text{cort}}$ or $v \in \Lambda_{\text{cochar}}$. So the group semiring acts by tensor product with a direct sum of such line bundles

$$(4.18) \quad \mathcal{U} \rightarrow \left(\bigoplus_{i=1}^w \mathcal{L}^{v_i} \right) \otimes \mathcal{U}.$$

Now we are going to show, for the case of regular Y , exactly how this structure is related to the line operators that exist in the absence of ramification. We express this in the language of the B -model. (For background, see sections 7 and 8 of [8].) Let ${}^L R$ be a representation of ${}^L G$. Let $(\mathcal{E}, \widehat{\varphi})$ be the universal Higgs bundle over $\mathcal{M}_H({}^L G, C) \times C$, and let \mathcal{E}_{LR} be the associated bundle²⁹ in the representation ${}^L R$. Let $\mathcal{E}_{LR}|_q$ denote the restriction of \mathcal{E}_{LR} to $\mathcal{M}_H \times q$ for q a point in C . Now consider a Wilson operator $W_q({}^L R)$ at the point q in the representation ${}^L R$. Its action on a brane \mathcal{B} is as follows. If \mathcal{U} is the Chan-Paton vector bundle of a brane \mathcal{B} , then the brane obtained by acting on \mathcal{B} with $W_q({}^L R)$ has Chan-Paton bundle $\mathcal{E}_{LR}|_q \otimes \mathcal{U}$.

Now let q approach a ramification point p for which Y is regular. At p , the structure group of the ${}^L G$ -bundle E is reduced to the maximal torus ${}^L \mathbb{T}$, and accordingly the fiber at p of the ${}^L G$ bundle $E \rightarrow C$ splits as a direct sum of subspaces corresponding to representations of ${}^L \mathbb{T}$. Accordingly, the vector bundle $\mathcal{E}_{LR}|_q$ breaks up, for $q = p$, as a direct sum of line bundles \mathcal{L}^v for $v \in \Lambda_{\text{char}}$. The tensor product with such a sum of line bundles is an example of the action of an element of the group semiring, as described in (4.18).

So we have shown that, at least for regular Y , what we get at a generic point q from Wilson or 't Hooft operators is, in the limit that q approaches a ramification point p , a special case of what we get from monodromies of the B - or A -model. The exact sequence (4.15) shows that it is tricky to generalize this to non-regular Y . The Wilson line operator $W({}^L R)$ gives, similarly to what we have seen above, an element of the group semiring of $\mathcal{V}_Y = \Lambda_{\text{char}}({}^L G) \rtimes \mathcal{W}_Y$. But there is no unique way to lift this to the group semiring of B_Y . In fact, the “direction” in

²⁹If the center of ${}^L G$ acts nontrivially in the representation ${}^L R$, then \mathcal{E}_{LR} is a twisted vector bundle rather than an ordinary one, and the corresponding Wilson operator shifts \mathbf{e}_0 .

which q approaches p (on the Riemann surface C) affects the limit of the Wilson operator, as has been analyzed in another language [75]

The explanation we have given may seem rather abstract. In section 5, we re-examine these questions more directly in gauge theory. We will get a fairly satisfactory description for the A -model, and some insight for the B -model.

Generalization For Any \mathbb{L}

All this has a natural generalization to any Levi subgroup \mathbb{L} of G . All we have to do is to include \mathbb{L} in all statements.

To begin with, of course, we require the parameters $(\alpha, \beta, \gamma, \eta)$ to be invariant under the Weyl group $\mathcal{W}_{\mathbb{L}}$ of \mathbb{L} . Focusing for simplicity on the A -model, to describe its monodromies we must keep fixed $Y = \exp(-2\pi(\eta - i\gamma))$. We denote as $\mathcal{W}_{Y, \mathbb{L}}$ the subgroup of the Weyl group of G that normalizes $\mathcal{W}_{\mathbb{L}}$ and commutes with Y . This is the subgroup that acts on (α, β) , preserving their \mathbb{L} -invariance and keeping Y fixed. In shifting α , we should restrict ourselves to the $\mathcal{W}_{\mathbb{L}}$ -invariant sublattice $\Lambda_{\text{cort}}^{\mathbb{L}} \subset \Lambda_{\text{cort}}$. So the analog of (4.14) is that the group that acts only on α and β is now

$$(4.19) \quad \mathcal{V}_{Y, \mathbb{L}} = \Lambda_{\text{cort}}^{\mathbb{L}} \rtimes \mathcal{W}_{Y, \mathbb{L}}.$$

We denote as $Z_{Y, \mathbb{L}}$ the space of \mathbb{L} -invariant pairs $(\alpha, \beta) \in \mathfrak{t} \times \mathfrak{t}$ that are not left fixed by any element of $\mathcal{V}_{Y, \mathbb{L}}$. The group $\mathcal{V}_{Y, \mathbb{L}}$ acts freely on $Z_{Y, \mathbb{L}}$, so we get a family of smooth A -models parametrized by the quotient $Z_{Y, \mathbb{L}}/\mathcal{V}_{Y, \mathbb{L}}$. This family is locally constant, so the fundamental group $B_{Y, \mathbb{L}} = \pi_1(Z_{Y, \mathbb{L}}/\mathcal{V}_{Y, \mathbb{L}})$ acts as a group of automorphisms of the A -model. This statement has an immediate analog for the B -model.

4.6. Representations And Branes. In the absence of ramification, an irreducible flat ${}^L G$ -bundle $E \rightarrow C$ corresponds to a smooth point $x_E \in \mathcal{M}_H$. The B -model of \mathcal{M}_H is independent of the Kahler metric, which is controlled by the gauge coupling parameter $\tau = \theta/2\pi + 4\pi i/e^2$. By taking $\text{Im } \tau$ large, we can go to a region in which the B -model can be treated semiclassically. Then there is a zerobrane \mathcal{B}_E supported at the smooth point $x_E \in \mathcal{M}_H$. It is an eigenbrane for the action of the Wilson line operators, and is a primary object of study in the geometric Langlands program.

The role here of irreducibility of E is that it keeps us away from the singularities of \mathcal{M}_H . If E is reducible (but semistable), it still corresponds to a point $x_E \in \mathcal{M}_H$, but a singular point. We can still consider branes that are supported at x_E , but the theory of such branes is more complicated. There is not, in general, a unique, canonically determined, brane associated with E . Instead, we can define a space or “category” of branes supported at x_E ; it can be argued that this space is closed under the action of the Wilson operators. This complication is a geometrical analog of a phenomenon that is known in number theory.

We want to discuss the analogous question for the ramified case, that is, for a flat bundle over $C \setminus p$ (or more generally over $C \setminus \{p_1, \dots, p_s\}$ for some finite set $\{p_1, \dots, p_s\}$), with monodromy around p . Let E be such a bundle, and let V denote its monodromy around p . We assume that E is irreducible, as otherwise we would meet a global singularity just as in the absence of ramification. Our goal here is to consider the role of local singularities that depend only on the conjugacy class of V .

For any V , we pick ${}^L\alpha$ and ${}^L\gamma$ such that $U = \exp(-2\pi({}^L\alpha - i{}^L\gamma))$ is in the closure of the orbit of V . After making some choice of ${}^L\beta$, we want to associate a brane on $\mathcal{M}_H({}^L\alpha, {}^L\beta, {}^L\gamma; p)$ with the given flat bundle $E \rightarrow C \setminus p$. If V (and therefore U) is regular and semisimple, then there is no problem. $\mathcal{M}_H({}^L\alpha, {}^L\beta, {}^L\gamma; p)$ is smooth (for any ${}^L\beta$), and we can treat it semi-classically by taking $\text{Im } \tau$ large. The flat bundle E determines a zerobrane \mathcal{B}_E , for any choices of ${}^L\beta$ and ${}^L\eta$.

Since U is semi-simple, the monodromies studied in section 4.5 are abelian and act by tensoring the Chan-Paton bundle of a brane with a line bundle \mathcal{L}^v . But \mathcal{L}^v is trivial when restricted to a point. So \mathcal{B}_E is an eigenbrane for the abelian group of monodromies. It likewise is an eigenbrane for Wilson line operators acting at a generic point $q \in C$. This follows by the same reasoning as in the absence of ramification; see section 8.2 of [8]. As an eigenbrane for all relevant operations, \mathcal{B}_E is a good analog of a zerobrane supported at a smooth point in the absence of ramification.

So in short, for regular semi-simple V , we are in essentially the same situation as in the absence of ramification. Now let us suppose that V ceases to be semi-simple, but is still regular. An example to keep in mind is that V might be a regular unipotent element of ${}^L G_{\mathbb{C}}$, for example, the element

$$(4.20) \quad V = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & \dots & 0 \\ & & & \vdots & \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

of $SL(N, \mathbb{C})$. If V is not semi-simple, then the pair $({}^L\alpha, {}^L\gamma)$ is non-regular; for example, ${}^L\alpha = {}^L\gamma = 0$ if V is unipotent. Then $\mathcal{M}_H({}^L\alpha, {}^L\beta, \gamma; p)$ may, depending on ${}^L\beta$, have a local singularity as described in sections 3.6 and 3.8. This will happen precisely if the triple $({}^L\alpha, {}^L\beta, {}^L\gamma)$ is nonregular.

Even if $\mathcal{M}_H({}^L\alpha, {}^L\beta, {}^L\gamma; p)$ has a local singularity, the point x_E is away from this singularity if V is regular. This means that there is no problem in defining the brane \mathcal{B}_E . Since it is defined in a natural and unique way, it is also an eigenbrane for all of the monodromy operations. So

again, for regular V , we are in essentially the same situation as in the absence of ramification.

A problem does occur if V is non-regular. Then, if we take ${}^L\beta = 0$, the point x_E is contained in the locus of local singularities. This gives us a problem in defining the brane \mathcal{B}_E . We can resolve the singularity by taking ${}^L\beta \neq 0$ and generic enough so that the triple $({}^L\alpha, {}^L\beta, {}^L\gamma)$ is regular. ${}^L\beta$ is a Kahler parameter in complex structure J , and taking sufficiently generic ${}^L\beta$ has the effect of blowing up the locus of local singularities. (This gives us a global version of what locally is the Springer resolution of the nilpotent cone.) This makes $\mathcal{M}_H({}^L\alpha, {}^L\beta, {}^L\gamma)$ non-singular, but the blow-up replaces the point x_E by a variety Υ_E of positive dimension. The best we can do is to associate to E the whole family or “category” of branes supported on Υ_E . The monodromy group acts on this category, possibly in a way that is in some sense irreducible.

Alternatively, we can keep ${}^L\beta = 0$ (or sufficiently special to avoid blowing up the point x_E), and make the sigma model smooth by taking ${}^L\eta$ to be sufficiently generic. Such a choice of ${}^L\eta$ may enable us to define a zerobrane $\mathcal{B}_{E, {}^L\eta}$. But the possible choices of sufficiently generic ${}^L\eta$ are divided into “chambers.” We suspect that if it is possible to use a choice of ${}^L\eta$ to define a zerobrane $\mathcal{B}_{E, {}^L\eta}$, then this zerobrane depends on the “chamber” containing ${}^L\eta$.

For example, in the case ${}^L\alpha = {}^L\gamma = 0$ of unipotent monodromy, if we also keep ${}^L\beta = 0$ to avoid resolving singularities, then the condition on ${}^L\eta$ to avoid a singularity of the sigma model is that ${}^L\eta$ must be regular. But the space of regular ${}^L\eta$'s is not connected, as the non-regular ${}^L\eta$'s divide ${}^L\mathbb{T}$ into affine Weyl chambers. If we try to pass from one chamber to another by varying ${}^L\beta$ so as to go around the singularities that separate between the different chambers, we will run into the monodromies that were used to define the group \mathcal{V}_U .

The result is that for non-regular V , we can define a family or “category” of branes on Higgs bundle moduli space, associated with the flat bundle $E \rightarrow C \setminus p$, and supported at x_E or its blowup. This category is acted on by the nonabelian group \mathcal{V}_U . All branes in this category are eigenbranes for Wilson operators acting at points away from p . But there is no apparent way to pick a particular brane in this family. It might be that in some sense the group \mathcal{V}_U acts irreducibly on this family.

Applying the duality transformation $S : \tau \rightarrow -1/n_{\mathfrak{g}}\tau$, we get the same sort of picture in the A -model of gauge group G : a family of A -branes that are acted on by \mathcal{V}_U , and are eigenbranes for 't Hooft/Hecke operators acting away from p . Finally, using the c.c. brane, we can map this family to a family of twisted \mathcal{D} -modules over $\mathcal{M}(\alpha, p; G)$, related in the same way to the Hecke operators and the group \mathcal{V}_U .

Accordingly, for non-regular V , the geometric Langlands program can be expressed as a duality that maps a family of B -branes, acted

on by \mathcal{V}_U , to a family of A -branes or \mathcal{D} -modules, acted on by the same group.

Searching For The Canonical Zerobrane

We have just seen that the duality statement that one can deduce using the standard surface operator associated with the maximal torus $\mathbb{T} \subset G$ becomes more involved when the ramified flat bundle E has a non-regular monodromy V . Although this complication will probably surprise most physicists who have gotten this far, if any, it will come as no surprise to geometers, since it is expected based on an analogy with number theory.

We can, however, ask the following question. Given a flat bundle $E \rightarrow C \setminus p$ with non-regular monodromy V , can we find another duality statement involving this flat bundle that involves a canonical zerobrane rather than a whole category? It is not clear that there is anything wrong if the answer is “no,” but if the answer is “yes,” we would like to find out.

For some conjugacy classes of V , we can answer this question simply by using the surface operator based on a general Levi subgroup \mathbb{L} , rather than the generic surface operator that we have used so far. In the presence of a surface operator of type \mathbb{L} , a solution of Hitchin’s equations describes in complex structure J a flat bundle $E \rightarrow C \setminus p$ whose possible monodromy V was analyzed at the end of section 3.3. This monodromy is conjugate to an element of \mathcal{P} , the parabolic subgroup that is determined by the pair $(\mathbb{L}, {}^L\alpha)$. Generically the monodromy is \mathbb{L} -regular (in a sense described in section 3.3). Precisely if V is \mathbb{L} -regular, the ramified flat bundle E corresponds to a smooth point on $\mathcal{M}_{H,\mathbb{L}}$, and hence to a canonical zerobrane \mathcal{B}_E in the sigma model of target $\mathcal{M}_{H,\mathbb{L}}$. This canonical zerobrane is an eigenbrane for all of the relevant Wilson operators and monodromies by the same arguments as above. Applying to it S -duality, we get a brane of the A -model with symplectic structure ω_K (and gauge group G rather than ${}^L G$), and then using the relation between A -branes and \mathcal{D} -modules, we associate to E a \mathcal{D} -module on the moduli space $\mathcal{M}_{\mathbb{L}}(\alpha; p, G)$ of parabolic bundles.

So in short, if the monodromy is \mathbb{L} -regular for some \mathbb{L} , we can reduce to a situation as simple as the unramified case by considering the theory with a surface operator of type \mathbb{L} . For instance, referring back to the Levi subgroup described in eqn. (3.30) with $G = SL(3, \mathbb{C})$, typical examples of \mathbb{L} -regular conjugacy classes are the semi-simple conjugacy class containing the element $\text{diag}(\lambda, \lambda, \lambda^{-2})$ with $\lambda^3 \neq 1$, and the Richardson conjugacy class with a representative given in eqn. (3.32).

For $G = SL(N, \mathbb{C})$, every conjugacy class is \mathbb{L} -regular for some \mathbb{L} . Given $V \in SL(N, \mathbb{C})$, we simply let $U \in \mathbb{T}$ be contained in the closure of the orbit of V , and we let \mathbb{L} be the subgroup of $SL(N, \mathbb{C})$ that commutes with U . However, for other groups, this is not the case. For example,

semi-simple Lie groups other than $SL(N, \mathbb{C})$ contain rigid noncentral orbits. (An orbit is called rigid if it cannot be deformed to any nearby orbit, usually because all nearby orbits have greater dimension.) Such an orbit is not \mathbb{L} -regular for any \mathbb{L} .

We suspect that it may be possible to define additional supersymmetric surface operators in $\mathcal{N} = 4$ super Yang-Mills theory with the following properties. For every ramified flat bundle $E \rightarrow C \setminus p$, we hope to be able to pick a surface operator such that, in the presence of this surface operator, E corresponds to a canonical zerobrane \mathcal{B}_E in the B -model at $\Psi = \infty$. Moreover, this surface operator should have an S -dual (which would be a surface operator of a roughly similar type). S -duality applied to the zerobrane \mathcal{B}_E would give an A -brane in the dual model at $\Psi = 0$. However, the new surface operators will not be related to parabolic subgroups, so in the presence of such an operator, we do not expect \mathcal{M}_H to be related in the usual fashion to a cotangent bundle. Consequently, it would not be possible to relate the dual A -brane to a \mathcal{D} -module, and the duality will have to be expressed as a mirror symmetry or S -duality between the B -model of $\mathcal{M}_H(LG)$ and an A -model of $\mathcal{M}_H(G)$, rather than a relation between coherent sheaves (or B -branes) on \mathcal{M}_H and \mathcal{D} -modules on some other space. The lack of an interpretation via \mathcal{D} -modules might mean that this more general duality, if it exists, will not be relevant to number theory, but it might still give an elegant application of S -duality and mirror symmetry.

5. Line Operators And Ramification

In section 4.5, we described the analog of Wilson and 't Hooft operators at ramification points. The description may have seemed rather abstract, and the answer – especially in the case of unipotent monodromy – is surprisingly complicated. By contrast, the Wilson and 't Hooft operators that act at a generic point on C are usually defined using quite different methods of gauge theory.

It is unsatisfying to describe the two cases with completely different methods, so in the present section we will attempt to give a gauge theory definition in the ramified case. This has another advantage: it will help us understand the full supersymmetry of these operators.

The topological field theory that arises at $\Psi = \infty$, after compactifying to two dimensions, is the B -model with target \mathcal{M}_H in complex structure J . This supersymmetry is preserved by the monodromies considered in section 4.5. However, the relevant Wilson operators at an unramified point preserve a greater supersymmetry. They have supersymmetry of type (B, B, B) ; that is, they preserve the supersymmetry of the B -model in any of the complex structures that make up the hyper-Kähler structure of \mathcal{M}_H .

Similarly, at $\Psi = 0$, the relevant topological field theory is the A -model of \mathcal{M}_H with symplectic structure ω_K . The appropriate supersymmetry is preserved by the monodromies studied in section 4.5. However, the relevant 't Hooft operators actually preserve supersymmetry of type (B, A, A) ; that is, they preserve the supersymmetry of the B -model of type I and of the A -model of type J or K .

We will aim to use gauge theory to show that the appropriate operators acting at a ramification point really have the same supersymmetry as the Wilson and 't Hooft operators at a generic point. We will be able to achieve this for the A -model, and partially for the B -model.

5.1. General Framework. For simplicity, we will work on a four-manifold $M = \Sigma \times C$. As usual, C is the Riemann surface on which we study the geometric Langlands program. We take $\Sigma = \mathbb{R} \times I$, where \mathbb{R} parametrizes the “time,” and I is a closed interval. We thus can also write $M = \mathbb{R} \times W$, where $W = I \times C$ is a three-manifold. Branes are chosen to define boundary conditions on the boundary of W .

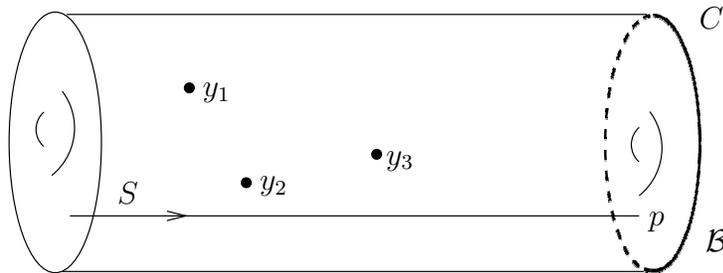


FIGURE 2. A time zero slice of a time-independent configuration. The support D of a surface operator intersects the time zero slice on the line S . The supports L_i of several line operators intersect the time zero slice at points y_i , $i = 1, 2, 3$.

For simplicity, we take our line and surface operators to have time-independent support. Therefore, we can describe any configuration of line and surface operators by indicating what is happening in W . A time-independent surface operator is supported on a two-manifold $D = \mathbb{R} \times S$, for S a curve in W . For applications to the ramified case of geometric Langlands, we take $S = I \times p$ for p a point in C . We also include line operators with time-independent support $L_i = \mathbb{R} \times y_i$ for some points $y_i \in W$. We will denote these line operators simply as L_i . This configuration is sketched in fig. 2.

The line operators can be understood as operators acting on branes, as explained in detail in section 6.4 of [8]. For example, using the topological invariance, we can move the points y_i to approach the right

boundary in fig. 2, which is labeled by the brane \mathcal{B} . As y_i approaches the boundary, the corresponding line operator L_i acts on \mathcal{B} to give a new brane \mathcal{B}_i which we describe symbolically as $L_i \cdot \mathcal{B}$. By contrast, because $S = I \times p$ ends on the boundary of W , the surface operator with support on $D = \mathbb{R} \times S$ must be included as part of the definition of what we mean by a brane; it does not really give in this situation an operator acting on branes.³⁰

Topological invariance means that there is a natural flat connection such that a line operator supported at a point $y_i \in W$ is naturally equivalent to one supported at a nearby point y'_i . This implies that line operators commute with each other, since the space of configurations of distinct points in a small open set in a three-manifold is connected and simply-connected. The connectedness means that we can move two line operators past each other without any singularity, and the simple connectivity means that there is (up to homotopy) no ambiguity about how to do this.

Line Operators Supported On A Surface

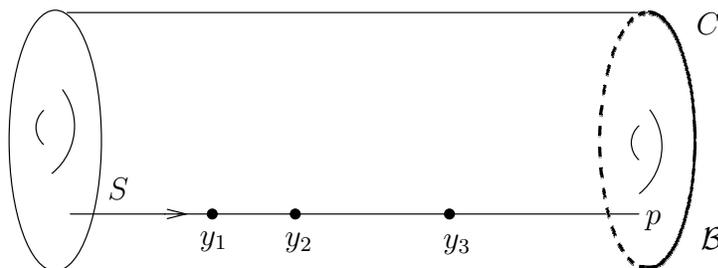


FIGURE 3. Line operators L_i supported on a surface D . Sketched is the time zero slice of a time-independent situation. At time zero, D is represented by the indicated line S , and the line operators are represented by points $y_i \in S$.

Now, however, consider the case that some of the points y_i are actually contained in the curve S . In the situation just considered, this will occur if we take $y_i = u_i \times p$ where $u_i \in I$ and p is the ramification point in C (fig. 3). In this case, of course, the line $L_i = \mathbb{R} \times y_i$ is contained in the surface $D = \mathbb{R} \times S$, so what we have is a line operator supported on a surface. Of course, what kinds of line operator are possible on the support of a given surface operator is one question that we will have to address.

³⁰A line or surface operator whose support is of finite extent in the time direction will give an operator acting on branes.

Line operators supported on a surface are quite different from generic line operators in four-space, because, granted suitable orientations, they have a natural ordering. This is shown in fig. 3, in which the ordering is $y_1 < y_2 < y_3$. A line operator with support $\mathbb{R} \times y_i$ gives us an operator that acts on branes, since we can move the points y_i to the right, whereupon the corresponding line operators act on the brane \mathcal{B} . However, in general, we should not expect line operators supported on branes to commute with each other, since they have a definite ordering and there is no way to move them past one another without meeting a singularity.

Thus, at a generic point $q \in C$, we get a commutative algebra \mathcal{S}_q of Wilson or 't Hooft operators that act on branes. But at a ramification point p , we get instead in general a noncommutative algebra $\widehat{\mathcal{S}}_p$.

Although $\widehat{\mathcal{S}}_p$ is in general noncommutative, it commutes with the algebra \mathcal{S}_q of Wilson or 't Hooft operators supported at a generic point $q \in C$. This is so simply because, as long as q is not a ramification point, a line operator that acts at q can be freely moved to the left or right, without encountering a singularity.

Of course, a noncommutative algebra may contain a large center. We can map \mathcal{S}_q to the center of $\widehat{\mathcal{S}}_p$ by taking the limit as q approaches p . In a topological field theory, this limit will exist since the distance between q and p is anyway irrelevant. However, generically there will be a monodromy as q is circled around p , and if so, the limit of an element of \mathcal{S}_q as $q \rightarrow p$ depends on the direction from which the limit is taken. This is obvious for the B -model, as we note in section 5.2, and has been analyzed in [75] for the A -model (in a language very different from that of this paper).

So the usual algebra of Wilson or 't Hooft operators can be mapped to the center of the noncommutative algebra that acts at a ramification point, but not quite in a canonical way.

Relation To Monodromies

Now let us discuss how this formulation is related to the description by monodromies in section 4.5. In that approach, one considers operations on branes that preserve only one topological supersymmetry – the symmetry of the B -model in complex structure J or of the A -model with symplectic structure ω_K . For simplicity, we will just express the following argument in terms of the B -model. For the A -model, the story is the same with α and η exchanged.

In the B -model in complex structure J , the dependence of the Lagrangian on the Kahler parameters (β, η) is entirely of the form $\int_M d^4x \sqrt{g} \{Q, V\}$, where Q is the appropriate topological supercharge and V is suitably chosen. The dependence on the parameters (α, γ) which control the complex structure cannot be put in this form.

With an appropriate choice of V , we can let the pair (β, η) be non-constant functions along the support D of a surface operator. For our

purposes, we take these functions to be time-independent, but to vary as one moves from left to right along the curve S of fig. 3. We must keep away from values of the pair (β, η) at which the quantum theory would become singular. Letting (β, η) vary along S , keeping away from the locus of singularities, gives us back exactly the monodromies that were described in section 4.5.

Now we can see the relation between monodromies and line operators. In describing a monodromy, we can let the pair (β, η) be constant outside of a small interval $I_0 \in S$. In a topological field theory, the size of the interval I_0 does not matter. If we think of this interval as being essentially pointlike, the monodromy is just a particular way to describe a line operator.

However, the description of line operators by monodromies has a drawback if we wish to see the full topological supersymmetry of the B -model at $\Psi = \infty$. We recall that this symmetry is of type (B, B, B) . To preserve B -type supersymmetry in complex structure J , we must keep α and γ fixed (as we did above), since this complex structure depends on those parameters. Likewise, to preserve B -type supersymmetry in complex structure I , we must keep β fixed, as complex structure I depends on β . So altogether, to preserve the full topological supersymmetry of type (B, B, B) , we can only vary η , and at best we will only be able to see abelian monodromies that correspond to shifting η by a lattice vector.

The most interesting nonabelian singularities are expected if $\alpha = \beta = \gamma = 0$. In this case, to avoid a singularity, η must be regular. Thus, η is confined to the interior of an affine Weyl chamber. With this restriction, we cannot observe any monodromies at all.

There is a similar problem, of course, in the A -model at $\Psi = 0$ if we wish to see the full topological supersymmetry of type (B, A, A) . If we want to preserve the full symmetry of the problem, monodromies are not an adequate framework. Our goal in the rest of this section is to see how much better we can do by thinking in terms of line operators rather than monodromies – that is, by defining discrete operations analogous to Wilson and 't Hooft operators in the unramified case.

5.2. The B -Model. The easy case to discuss is the B -model, but this is also the case in which we will not be able to get a fully satisfactory answer.

In the unramified case, with gauge group ${}^L G$, the important operators that act on branes are Wilson operators in a representation ${}^L R$ of ${}^L G$. They are defined by parallel transport using the complex-valued connection $\mathcal{A} = A + i\phi$. If L is a closed loop in the four-manifold M , we define the Wilson operator

$$(5.1) \quad W({}^L R; L) = \text{Tr } P \exp \left(- \int_L \mathcal{A} \right),$$

where $P \exp(-\int_L \mathcal{A})$ is the holonomy of the connection \mathcal{A} along the line L , and the trace is taken in the representation ${}^L R$. If instead L is an open line (connecting boundaries or ends of the four-manifold M), we define $W({}^L R; L)$ as this holonomy, regarded as an operator acting between initial and final states in the representation ${}^L R$.

For the geometric Langlands program in the unramified case, we take $M = \mathbb{R} \times I \times C$, where \mathbb{R} parametrizes the “time,” I is an interval, and C is a Riemann surface. To get a Wilson operator that can act on branes, we take $L = \mathbb{R} \times u \times q$, with chosen points $u \in I$ and $q \in C$.

If L is contained in the support D of a surface operator, as in fig. 3, we can carry out much the same construction. There is one important difference. On the support of a surface operator, the group ${}^L G$ is reduced to its maximal torus ${}^L \mathbb{T}$. Hence, if L is contained in D , we can define Wilson operators associated with a choice of representation of ${}^L \mathbb{T}$. This is more general than a choice of representation of ${}^L G$, since a representation of ${}^L G$ can be decomposed as a direct sum of representations of ${}^L \mathbb{T}$, but in general not the other way around.

So in general, at a ramification point the representation ring of ${}^L \mathbb{T}$ acts on branes of the model at $\Psi = \infty$, extending the action of representations of ${}^L G$ that occurs at a generic point. What happens is simply that a Wilson operator at a generic point q , associated with a representation of ${}^L G$, splits up, in the limit that q approaches p , as a sum of Wilson operators associated with representations of ${}^L \mathbb{T}$.

Because the ramified flat bundle of the B -model has a monodromy around the support of a surface operator, there is a subtlety in the limit $q \rightarrow p$ in this situation. The limit depends on the path via which one takes q to approach p .

Interpretation

The representation ring of ${}^L \mathbb{T}$ is the coweight lattice Λ_{cowt} . This contains a sublattice Λ_{cort} consisting of Wilson operators that do not change the discrete electric flux \mathbf{e}_0 .

We similarly encountered in section 4.5 an action on branes of the lattice Λ_{cort} or Λ_{cowt} . However, the explanation given there was somewhat different: in that approach, the lattice acts by shifts of η , that is, by 2π shifts of the theta-angles that are defined on the impurity surface.

It is, however, a classic result [76] that a Wilson operator in abelian gauge theory in two dimensions causes a shift in the theta-angle by an integer multiple of 2π . Let us recall how this comes about. We consider a gauge theory with gauge group $H = U(1)$ on a two-manifold D . We let L be a one-manifold on D of a suitable type. The simplest case to consider is that L is a closed one-manifold that is the boundary of a region $R \subset D$. Let A be an abelian gauge field³¹ on D with curvature

³¹We consider A as a connection on a principal $U(1)$ bundle, represented locally by a real one-form.

F . Then for the holonomy of A around L , we have the identity

$$(5.2) \quad \exp\left(i \oint_L A\right) = \exp\left(i \int_R F\right).$$

This says that the theta-angle jumps by 2π in crossing the line on which a charge 1 Wilson operator is supported. With more care in the analysis, one can reach the same conclusion even if the one-manifold L is not closed.

Discussion

So for the case of regular semi-simple monodromy, the lattice action on branes that we have found from gauge theory is all that we expect to see. However, according to the analysis of section 4.5, a larger noncommutative group should act on branes in case the pair (α, γ) is non-regular. Unfortunately, as we will now explain, it seems difficult at $\Psi = \infty$ to see this larger symmetry by semi-classical gauge theory methods. Thus we will not really be able with these methods to do better than we did with monodromies.

The monodromies that we want to see arise by varying β and η , and are absent if β is constrained so that the triple (α, β, γ) is always regular. So the existence of these monodromies depends on what happens when that triple becomes non-regular. At this point, the classical moduli space \mathcal{M}_H gets a local singularity, in the language of section 3.6.

The monodromies still make sense, because the quantum theory remains smooth if η is generic. However, it is hard to use this fact about the quantum theory in a semi-classical gauge theory construction of line operators. Any such construction begins with a classical construction which is then implemented quantum mechanically. Operators whose definition depends on quantum properties of the theory are difficult to see using gauge theory methods.

The situation is different in the A -model, because the roles of α and η are exchanged. The relevant monodromies are found by varying α and β , with γ and η fixed. Even if $\gamma = \eta = 0$, one can see all the monodromies in the region in which the pair (α, β) is regular, which means that the classical geometry of \mathcal{M}_H is smooth.

This suggests that in the A -model, we might be more successful in using gauge theory methods to describe the full set of operators that act on branes and all the supersymmetry that they preserve. That will be our next goal.

5.3. The A -Model. In the absence of ramification, the natural operators acting on branes in the A -model are 't Hooft operators, defined by prescribing a singularity that the fields are required to have. So in the presence of ramification, we will look for 't Hooft-like operators. To keep things simple, we will take the Levi subgroup defining our surface operator to be $\mathbb{L} = \mathbb{T}$.

It is convenient to get to $\Psi = 0$, that is, the A -model with symplectic structure K , by taking the twisting parameter of the underlying four-dimensional theory to be $t = 1$ and the four-dimensional theta-angle to vanish. On a general four-manifold M , the conditions to preserve the topological symmetry of the A -model were described in eqns. (9.1), (9.2) of [8]. If we set $\mathcal{A} = A + i\phi$, and write $\mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A}$ for the curvature of \mathcal{A} , and set $d_A = d + A$, then the conditions are

$$(5.3) \quad \begin{aligned} \mathcal{F} + i \star \bar{\mathcal{F}} &= 0 \\ d_A \star \phi &= 0. \end{aligned}$$

We specialize to $M = \mathbb{R} \times W$, with \mathbb{R} parametrized by a “time” coordinate s . We restrict the equations (5.3) to the time-independent case. This is motivated by the fact that the usual ’t Hooft operators are defined by a time-independent singularity (and by arguments given in the introductory part of section 9 of [8], which are also relevant in the presence of ramification). In a time-independent situation, the four-dimensional connection reduces to $A = A' + A_0 ds$ with A' a connection on a G -bundle $E \rightarrow W$ and A_0 an $\text{ad}(E)$ -valued zero-form. Similarly we write $\phi = \phi' + \phi_0 ds$, where ϕ' and ϕ_0 are $\text{ad}(E)$ -valued forms on W . We set $\mathcal{A}' = A' + i\phi'$, $\mathcal{F}' = d\mathcal{A}' + \mathcal{A}' \wedge \mathcal{A}'$, and $\Phi_0 = \phi_0 - iA_0$. Since the whole subsequent discussion will occur in three dimensions, we omit the primes. The first equation in (5.3) becomes

$$(5.4) \quad \mathcal{F} = \star \bar{\mathcal{D}} \Phi_0,$$

where $\bar{\mathcal{D}} = d + \bar{A}$ and of course \star is now the three-dimensional Hodge \star operator. The second equation becomes

$$(5.5) \quad D \star \phi = \frac{i}{2} [\bar{\Phi}_0, \Phi_0].$$

Scaling Symmetry

An ’t Hooft-like operator is defined using a singular solution of these equations. For example, ordinary ’t Hooft operators are defined by using a singular solution of these equations that has a singularity in codimension three, that is, at an isolated point $r \in W$. The construction was described in section 6.2 of [8], and will be generalized below. In the present paper, our main topic has been surface operators, which arise by specifying a codimension two singularity – the familiar singularity with parameters (α, β, γ) .

We are interested in a more complicated situation in which a curve S of codimension two singularities contains an isolated point y at which the singular behavior is “worse.” Thus, near a generic point of S , we see the usual codimension two singularity, but a new singularity occurs at the point $y \in S$. *A priori*, to the left of y , the line S is labeled by parameters (α, β, γ) , and to the right it is labeled by parameters $(\alpha', \beta', \gamma')$. If we are to get a surface operator that preserves the full supersymmetry of

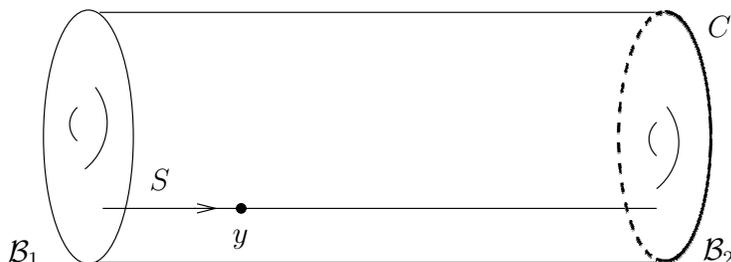


FIGURE 4. A line operator supported on a surface is represented in this time zero slice by a point y on a line S . In the A -model, the fields have a prescribed singularity along S . Near y , the singular behavior is different from what it is near a generic point on S . The generic singular behavior along S may in turn be different on the two sides of y .

type (B, A, A) , we must have $(\beta, \gamma) = (\beta', \gamma')$, because β and γ are physical parameters in the B -model of type I , and cannot jump if we are to preserve the supersymmetry of this model. However, there is no reason to consider only the case that $\alpha = \alpha'$. The topological field theories of type (B, A, A) are locally independent of α , so we should expect to get results that are essentially independent of α and α' .

To simplify things, we will make an assumption about the nature of the singularities that we are looking for. We will assume that they are scale-invariant, like the singularities used to define ordinary 't Hooft operators. Indeed, since we are only interested in the local singular behavior, we can replace W by \mathbb{R}^3 , with y as the origin in \mathbb{R}^3 . We choose Euclidean coordinates x^1, x^2, x^3 on \mathbb{R}^3 , and we take the line S of fig. 5 to be the x^1 axis, defined by $x^2 = x^3 = 0$. \mathbb{R}^3 admits a scaling symmetry $\vec{x} \rightarrow \lambda \vec{x}$, for real positive λ . The equations of interest are invariant under this scaling symmetry (if we take Φ_0 to scale with dimension 1), and the required singularities along S away from the point y are compatible with scale invariance. So it makes sense to look for a scale-invariant singular solution, and this will suffice for our purposes here.

If we remove the point y from \mathbb{R}^3 and divide by scaling, we get $(\mathbb{R}^3 \setminus y) / \mathbb{R}_+^* = S^2 \cong \mathbb{CP}^1$. We will call this quotient V . The interesting solutions can be largely described in terms of data on V . The fields on V will be singular at two points sketched in fig. 5, namely the points p and p' at which V meets the line S .

The Case $\phi = 0$

As we learned in section 4.5, the operators acting on branes are most interesting if $\beta = \gamma = 0$. In this case, it is possible to have $\phi = 0$ (since

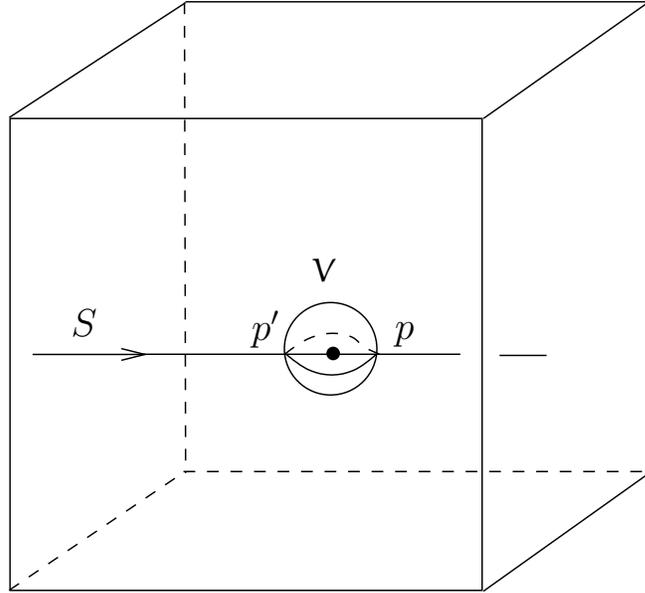


FIGURE 5. To describe the singularity locally, we replace the three-manifold W by \mathbb{R}^3 , and the line S by a straight line $\mathbb{R} \subset \mathbb{R}^3$. The point y becomes the origin in \mathbb{R}^3 , marked here by the black dot. V is a unit sphere surrounding the origin. It intersects the line S at two points, p and p' .

ϕ has no singularity along a generic point of S). It turns out that the 't Hooft-like operators that we are seeking can be defined using a singular solution with $\phi = 0$, and also $A_0 = 0$.

This leads to a drastic simplification in the equations (5.4) and (5.5). They reduce to the Bogomolny equations

$$(5.6) \quad F = \star d_A \Phi_0,$$

where now $F = dA + A \wedge A$ and $\Phi_0 = \phi_0$ take values in the real Lie algebra \mathfrak{g} of G . We will make an assumption that is stronger than scale-invariance: we assume that A and Φ_0 are pullbacks from $V = S^2$. We will think of V as the unit sphere in \mathbb{R}^3 , defined by $r = 1$ where $r = |\vec{x}|$. The equations on S^2 come out to be

$$(5.7) \quad \begin{aligned} F &= -\star \Phi_0 \\ d_A \Phi_0 &= 0, \end{aligned}$$

where now these are equations on the two-manifold V and \star is meant in the two-dimensional sense. For example, to get the first equation, we

use the fact that the dependence of Φ_0 on r is precisely a factor of $1/r$ (since Φ_0 scales with weight 1), so that at $r = 1$ we have $\partial\Phi_0/\partial r = -\Phi_0$. The second equation is obtained using the fact that F vanishes when contracted with $\partial/\partial r$ (because we assume the solution to be a pullback from \mathbf{V}).

Obviously, the two equations combine to

$$(5.8) \quad d_A \star F = 0.$$

These are the two-dimensional Yang-Mills equations, that is, they are the Euler-Lagrange equations that can be derived from the Yang-Mills action

$$(5.9) \quad \mathbf{I} = -\frac{1}{2} \int_{\mathbf{V}} \text{Tr } F \wedge \star F.$$

Another interpretation of these equations is as follows. Given a Riemann surface \mathbf{V} , every connection A on a G -bundle $E \rightarrow \mathbf{V}$ endows E with a holomorphic structure, defined by the $\bar{\partial}$ operator $\bar{\partial}_A = d\bar{z}(\partial_{\bar{z}} + A_{\bar{z}})$. Turning this around, we can begin with a holomorphic $G_{\mathbb{C}}$ -bundle $E \rightarrow \mathbf{V}$ and ask what sort of unitary connections E admits that are compatible with its holomorphic structure (in the sense that the holomorphic structure induced by the connection agrees with the given holomorphic structure of E). Stable and semistable bundles are precisely the ones that admit a flat unitary connection compatible with their holomorphic structure [51]. More generally, if we endow \mathbf{V} with a Kahler metric and thus a \star operator (so that the Yang-Mills equations and action are defined), then [52] every $G_{\mathbb{C}}$ -bundle $E \rightarrow \mathbf{V}$, stable or not, admits a unitary connection, unique up to gauge transformation, that obeys the Yang-Mills equations and is compatible with the holomorphic structure of E . Thus, in particular, solutions of the Yang-Mills equations up to gauge transformation are in natural correspondence with equivalence classes of holomorphic $G_{\mathbb{C}}$ -bundles.

This result has a natural interpretation in terms of the gradient flow of the Yang-Mills action [52], but we will not need this here. More important for us is the generalization of the above result to allow parabolic structure [77]. We will state this theorem for any number of ramification points with arbitrary choices of parabolic subgroups. We remove from the Riemann surface \mathbf{V} a finite set of points p_1, \dots, p_s . We label each of the p_i by a parabolic weight $\alpha_i \in \mathfrak{t}$, and an associated conjugacy class $U_i = \exp(-2\pi\alpha_i)$. We assume α_i to be generic enough that for each i , the Levi subgroup $\mathbb{L}_i \subset G$ that commutes with α_i is the same as the subgroup that commutes with U_i ; we write \mathcal{P}_i for the parabolic subgroup determined by the pair (\mathbb{L}, α_i) . We consider a connection A on the G -bundle $E \rightarrow \mathbf{V} \setminus \{p_1, \dots, p_s\}$ that obeys the Yang-Mills equations and whose holonomy around the point p_i is conjugate to U_i . The claim is that such connections, up to gauge transformation, are in one-to-one correspondence with holomorphic $G_{\mathbb{C}}$ -bundles $E \rightarrow \mathbf{V}$ with parabolic

structure (of type \mathcal{P}_i) at the points p_i . A parabolic bundle is stable or semi-stable if and only if the corresponding solution of the Yang-Mills equations is flat; otherwise, it is unstable.

In applying this theorem, since we are taking our surface operators to be of type \mathbb{T} , we are for the moment interested in the case that the \mathbb{L}_i are all conjugate to the maximal torus \mathbb{T} . Hence, until the end of this section, where we briefly discuss surface operators of general type \mathbb{L} , the phrase “parabolic structure” means a reduction of the structure group to a Borel subgroup \mathcal{B} .

Parabolic Bundles For $SL(2, \mathbb{C})$

For constructing 't Hooft-like operators, we are interested in $G_{\mathbb{C}}$ -bundles over $V = \mathbb{C}\mathbb{P}^1$ with parabolic structure at two points, p and p' . The parabolic weights at the two points are α and $-\alpha'$. The reason for the minus sign is that in fig. 5, the two-sphere V intersects the line S with opposite orientation at the points p and p' .

Parabolic bundles in this situation can be described very explicitly. To explain the key ideas, we take $G_{\mathbb{C}} = SL(2, \mathbb{C})$, and carry out the analysis both holomorphically and using the relation to the Yang-Mills equations.

Looking at things from a holomorphic point of view, the $SL(2, \mathbb{C})$ bundle $E \rightarrow V$ (regarded as a rank two vector bundle of trivial determinant) must be $\mathcal{O}(m) \oplus \mathcal{O}(-m)$ for some integer m , which we may as well assume to be non-negative. In addition, we must endow E with parabolic structure at the two points $p, p' \in V$. Parabolic structure for a rank two bundle E at a given point on a Riemann surface is simply a choice of a one-dimensional subspace of the fiber of E at that point. We denote these fibers as $E|_p$ and $E|_{p'}$. So the parabolic structures at the points p and p' are given by choices of subspaces $E_p \subset E|_p$ and $E_{p'} \subset E|_{p'}$. The isomorphism classes of such parabolic bundles can be classified as follows:

(1) Suppose first that $m = 0$. Then the bundle E is trivial, and the group $G_{\mathbb{C}} \cong SL(2, \mathbb{C})$ acts on it. We say that the parabolic structures at p and p' agree if there is a non-zero global section of E whose restriction to p takes values in E_p and whose restriction to p' takes values in $E_{p'}$. Otherwise we say they disagree. Up to isomorphism, that is, modulo the action of $SL(2, \mathbb{C})$, there are precisely two choices: the parabolic structures at p and p' may agree or disagree.

(2) Now suppose that $m > 0$. The bundle $E = \mathcal{O}(m) \oplus \mathcal{O}(-m)$ has a unique sub-bundle isomorphic to $\mathcal{O}(m)$ (characterized by the fact that any global holomorphic section of E is actually a section of this sub-bundle). By contrast, the embedding of $\mathcal{O}(-m)$ in E and hence the splitting $E = \mathcal{O}(m) \oplus \mathcal{O}(-m)$ is not canonical. Up to an automorphism of E , the only invariant information in the parabolic structure is whether the subspaces E_p and $E_{p'}$ coincide or do not coincide with the fibers of

$\mathcal{O}(m)$ at p and p' . So for given m , there are $2 \times 2 = 4$ choices. We organize the four possibilities as follows. We label the parabolic bundle E by the integer m if $E_p = \mathcal{O}(m)|_p$, and otherwise we label it by $-m$. And we say that the parabolic structure at p' agrees with that at p if E_p and $E_{p'}$ both coincide with, or both differ from, the relevant fiber of $\mathcal{O}(m)$. Otherwise, we say that the parabolic structures disagree.

Now let us look at things from the point of view of the Yang-Mills equations. In this case, we need to pick parabolic weights at the points p and p' . (The weights did not enter the holomorphic description, since we were not concerned with the question of which parabolic bundles were stable.) We take

$$(5.10) \quad \begin{aligned} \alpha &= iy \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ \alpha' &= iy' \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \end{aligned}$$

with $0 < y, y' < 1/2$. It is also convenient to take initially $y \neq y'$.

The Yang-Mills equations (5.8) tell us that the curvature is covariantly constant, so we can pick a gauge in which the curvature is actually constant

$$(5.11) \quad F = i \begin{pmatrix} f & 0 \\ 0 & -f \end{pmatrix},$$

with a real constant f that we may as well take to be nonnegative. We will see momentarily that $f \neq 0$ if $y \neq y'$. Since F is covariantly constant, it commutes with the holonomies U and U' around the points p and p' , and therefore (assuming $f \neq 0$) these holonomies are diagonal matrices. Since the holonomies must be conjugate to $\exp(-2\pi\alpha)$ and $\exp(-2\pi(-\alpha'))$, respectively, it follows that $U = \exp(-2\pi\epsilon\alpha)$, $U' = \exp(2\pi\epsilon'\alpha')$, with $\epsilon, \epsilon' = \pm 1$. Finally, a bundle E with the connection and curvature that we have just described exists if and only if

$$(5.12) \quad \exp\left(\int_V F\right) = UU'.$$

This condition is equivalent to

$$(5.13) \quad \int_V \frac{f}{2\pi} + \epsilon y - \epsilon' y' = m',$$

for some integer m' .

The Weyl group acts by changing the sign of F , y , y' , and m' . So by a Weyl transformation, we can fix $\epsilon = 1$, after which the solutions of the Yang-Mills equations are labeled by an integer m' and the variable

$\epsilon' \in \{\pm 1\}$. The classification we have given of the solutions of the Yang-Mills equations is also valid if $y = y'$, but in that case one needs a little more care in treating the case $\epsilon' = 1$, $m' = 0$.

Thus, the classification of Yang-Mills solutions agrees with the classification of parabolic bundles. In this comparison, m' corresponds to m , while $\epsilon' = 1$ (or -1) corresponds to the case that the two parabolic structures agree (or disagree).

Each such solution or bundle enables us to define an 't Hooft-like operator. Thus, for every pair $(m, \epsilon') \in \mathbb{Z} \times \mathbb{Z}_2$, we get an 't Hooft-like operator $T_{m, \epsilon'}$.

These operators do preserve the full topological supersymmetry of type (B, A, A) . To see this, we first note that, since they are derived from solutions of the requisite equations (5.4) and (5.5), they preserve the topological symmetry of the A -model of type K (the one with symplectic structure ω_K). As these solutions are invariant under $\varphi \rightarrow i\varphi$, which rotates the symplectic structure ω_K into ω_J , they also preserve the topological symmetry of the A -model of type J . Linear combinations of these two supercharges generate the full topological supersymmetry of type (B, A, A) .

Action On Branes

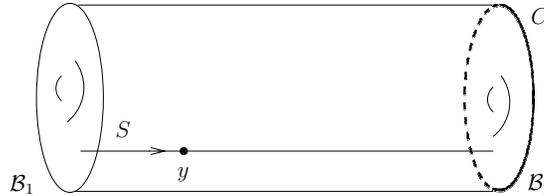


FIGURE 6. Action of a line operator, here inserted at the point y , on branes. The fields (A, ϕ) determine a ramified Higgs bundle $E_u \rightarrow C_u = \{u\} \times C$ in complex structure I for every $u \in S$. In a solution of the supersymmetric equations, the holomorphic type of E_u is constant except when C_u crosses the point y , at which point it jumps.

In broad outline, the action of one of these 't Hooft-like operators on branes is analogous to the action of ordinary 't Hooft operators in the unramified case, as described in [8]. We will hence be rather brief.

In fig. 6, we consider fields (A, ϕ) on a G -bundle E over the three-manifold $W = I \times C$ with the usual singularity along the line S . For any $u \in I$, the restriction of the bundle E and fields A, ϕ to $C_u = \{u\} \times C$ give a ramified Higgs bundle $E_u \rightarrow C_u$. If the Bogomolny equations (5.6) are obeyed, the holomorphic type of E_u is independent of u , except in crossing the position of a line operator. If one crosses a line operator

supported at a generic point in C (times a point in I), the bundle E_u undergoes an ordinary Hecke modification, and if as in fig. 6 one crosses a line operator supported on S , one gets a more general Hecke-like modification that can involve the parabolic structure. The justification of these statements is as in [8].

Now suppose that we want to see how the line operator L_y at the point y in fig. 6 acts on, say, the brane \mathcal{B}_1 on the left of the figure. What we have just described means that the operator L_y determines a “correspondence” $\mathcal{Q} \subset \mathcal{M}_H \times \mathcal{M}_H$. \mathcal{Q} parametrizes pairs (\tilde{E}_u, E_u) , where \tilde{E}_u can be obtained from E_u in “jumping” over the line operator L_y . \mathcal{Q} can be regarded as a brane in $\mathcal{M}_H \times \mathcal{M}_H$ of type (B, A, A) , since it is obtained by solving equations that preserve this symmetry. We are mainly interested in it as an A -brane of type K .

In fact, most parabolic bundles over $V \cong \mathbb{CP}^1$ that may be used to define the ‘t Hooft-like operator L_y are unstable. When this is the case, we really want to define \mathcal{Q} using a compactification of the space of pairs (\tilde{E}_u, E_u) by including Hecke-like modifications defined using a less unstable parabolic bundle (for $G_{\mathbb{C}} = SL(2, \mathbb{C})$ this simply means that $|m|$ is smaller). This is just as in the unramified case. One would expect that in principle use of such a compactification could be justified based on the underlying four-dimensional gauge theory.

Now let π_i , $i = 1, 2$, be the two projections $\pi_i : \mathcal{M}_H \times \mathcal{M}_H \rightarrow \mathcal{M}_H$. Given a brane \mathcal{B}_1 over \mathcal{M}_H , we “pull it back” to $\mathcal{M}_H \times \mathcal{M}_H$ via π_1^* , tensor it with \mathcal{Q} , and “push forward” to \mathcal{M}_H via p_2 , to get the new brane $L_y \cdot \mathcal{B}_1$. Thus $L_y \cdot \mathcal{B}_1 = \pi_{2*}(p_1^*(\mathcal{B}_1) \otimes \mathcal{Q})$. If \mathcal{B}_1 is a brane of type (B, A, A) , these operations can be carried out using complex geometry in complex structure I , but more generally, if \mathcal{B}_1 is simply an A -brane of type K , one must use the analogs of these operations in the A -model, that is, in Floer cohomology.

Here is a possibly more down-to-earth explanation. To construct the physical Hilbert space in the situation of fig. 6, with branes defining the boundary conditions at the ends and the operator L_y in the interior, one proceeds as follows. One considers the compactified moduli space $\overline{\mathcal{N}}$ of supersymmetric configurations, with boundary conditions determined at the two ends by the branes \mathcal{B}_1 , \mathcal{B}_2 , and L_y in the interior. The physical Hilbert space is the cohomology of $\overline{\mathcal{N}}$. To define the action of L_y on the branes, the idea is that the cohomology of $\overline{\mathcal{N}}$ is the same as it would be if the singularity due to L_y were omitted and \mathcal{B}_1 replaced by some other brane $L_y \cdot \mathcal{B}_1$, which will not depend on \mathcal{B}_2 . The Chan-Paton sheaf of $L_y \cdot \mathcal{B}_1$ has a fiber at a point $E_u \in \mathcal{M}_H$ that is equal to the cohomology of the space $\Xi(E_u)$ obtained as follows: $\Xi(E_u)$ parametrizes Higgs bundles \tilde{E}_u such that the pair (\tilde{E}_u, E_u) represents a point in \mathcal{Q} . (In this oversimplified explanation, we have ignored the singularities of \mathcal{Q} .) This recipe can be justified by thinking through how one would

compute the physical Hilbert spaces with or without insertion of the operator L_y .

Affine Weyl And Braid Groups

Now let us return to the example $G_{\mathbb{C}} = SL(2, \mathbb{C})$, in which we classified the 't Hooft-like operators by pairs $(m, \epsilon') \in \mathbb{Z} \times \mathbb{Z}_2$. The affine Weyl group \mathcal{W}_{aff} of $SL(2, \mathbb{C})$, which was introduced in eqn. (3.50), is an extension

$$(5.14) \quad 0 \rightarrow \mathbb{Z} \rightarrow \mathcal{W}_{\text{aff}} \rightarrow \mathbb{Z}_2 \rightarrow 0,$$

and elements of this group are fairly naturally labeled by pairs (m, ϵ') , just like the 't Hooft-like operators. Actually, because \mathbb{Z}_2 acts on \mathbb{Z} by reversing the sign, there is a subtlety about the sign of m , just as there was for the 't Hooft operators. The correspondence of 't Hooft operators with elements of \mathcal{W}_{aff} is more natural than the correspondence of either one with pairs (m, ϵ') .

This has an analog for any simply-connected G : bundles on \mathbb{CP}^1 with parabolic structure at precisely two points correspond naturally to elements of \mathcal{W}_{aff} . This is fairly easy to see in the Yang-Mills approach, by generalizing the arguments we gave for $SL(2, \mathbb{C})$, and is left to the reader. It is also a standard result in the holomorphic approach.³²

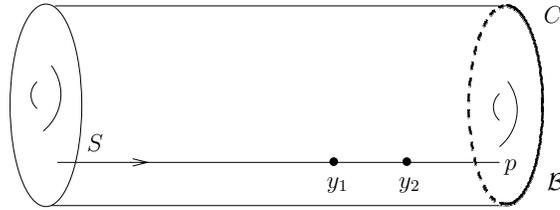


FIGURE 7. Two line operators living on the support of a surface operator can be composed by simply moving the lines together. In the static case depicted here, this is done by taking the limit as $y_1 \rightarrow y_2$. In topological field theory, line operators generate an associative algebra that acts on branes.

The 't Hooft-like operators that we have defined can be composed in a natural fashion by moving two lines together, as sketched in fig. 7. One might expect that these operators would compose according to the multiplication law of the affine Weyl group. However, according to [29] as well as our arguments in section 4.5, one expects that the group that acts on branes of the A -model at a ramification point $p \in C$ should be

³²It is equivalent to the statement that bundles on \mathbb{CP}^1 with parabolic structure at two points correspond to orbits in the action of the Iwahori group on the affine flag manifold.

the affine braid group, not the affine Weyl group. As in (4.16), the two groups are related by an exact sequence

$$(5.15) \quad 1 \rightarrow \pi_1(\mathcal{Z}) \rightarrow B_{\text{aff}} \rightarrow \mathcal{W}_{\text{aff}} \rightarrow 1,$$

where \mathcal{Z} is the space of pairs $(\alpha, \beta) \in \mathfrak{t} \times \mathfrak{t}$ on which \mathcal{W}_{aff} acts freely. This is a non-trivial extension, so \mathcal{W}_{aff} can be lifted to B_{aff} , but not in a way that respects the group law. Therefore, we do not expect the operators $T_{m, \epsilon'}$ acting on branes to respect the relations of the affine Weyl group.

We can be slightly more specific. For G of rank r , \mathcal{W}_{aff} is generated by elementary reflections T_i , $i = 1, \dots, r+1$, one for each node of the Dynkin diagram. They obey $T_i^2 = 1$ and certain additional relations. For example, for simply-laced G of rank greater than 1, one has

$$(5.16) \quad T_i T_j T_i = T_j T_i T_j$$

if the vertices i and j are connected by an edge, and

$$(5.17) \quad T_i T_j = T_j T_i$$

otherwise. To get the affine braid group, one omits the relation $T_i^2 = 1$, but keeps the others. A word of shortest length in the affine Weyl group can be lifted in a fairly natural way to the “same” word in the affine braid group, but this does not give a group homomorphism.

We will not try here to describe via gauge theory methods the operations on branes that correspond to $\pi_1(\mathcal{Z})$. In examples, it appears that $\pi_1(\mathcal{Z})$ acts by tensoring the Chan-Paton bundle of a brane with certain vector bundles with a non-trivial action of the ghost number.

Incorporation Of β And γ

So far we have taken the Higgs field ϕ to vanish, which only makes sense if $\beta = \gamma = 0$. If $(\beta, \gamma) \neq 0$, then ϕ has prescribed singularities, and cannot vanish.

It is surprisingly straightforward to incorporate β and γ in the analysis, for the following reason. The solutions of the Yang-Mills equations that we have used to describe the operators $T_{m, \epsilon'}$ are actually all abelian, that is, the structure group reduces from G to its maximal torus \mathbb{T} . This reflects the fact that the fundamental group of the twice-punctured two-sphere is abelian.

This being so, we can incorporate β and γ by simply adding an explicit abelian solution for ϕ . Given $(\beta, \gamma) \in \mathfrak{t}$, we define

$$(5.18) \quad \varphi = \frac{1}{2}(\beta + i\gamma) \frac{dz}{z},$$

where $z = x^2 + ix^3$, with x^2 and x^3 being the normal coordinates to the line $S \subset W = \mathbb{R}^3$. (φ is not a pullback from V .) Then we obey the equations (5.4) and (5.5) by setting $\phi = \varphi + \bar{\varphi}$ (still with $A_0 = 0$). The singularity of φ is independent of x^1 , so as expected this construction

is only possible given that β and γ , unlike α , do not jump in crossing the support of a line operator.

The line operators obtained this way do preserve topological symmetry of type (B, A, A) . In adapting the argument that we gave at $\varphi = 0$, care is needed only at one point. The transformation $\varphi \rightarrow i\varphi$ is not a symmetry of the solution just described, but rather maps this solution into a solution of the same form with different (β, γ) . This is enough to justify the argument.

Including ϕ in this way, even though it may seem like a trivial operation, actually places an interesting restriction on the part of the solution involving the gauge field. (For an analogous statement without ramification, see section 10.5 of [8].) With $\beta = \gamma = 0$, the parabolic structures at the points p and p' are related by a relative Weyl transformation. For example, in the Yang-Mills construction, this happened because a knowledge of the conjugacy classes of the monodromies U_1 and U_2 , plus the fact that they commute with the curvature F , fixes them only up to Weyl transformations. However, the Weyl group acts diagonally on the triples (α, β, γ) and (α', β, γ) . So since β and γ are constant, we cannot act on α and α' with arbitrary Weyl transformations. We are restricted to Weyl transformations that act trivially on β and γ . Quantum mechanically, if we include η , which likewise cannot jump because it is a physical parameter of the A -model of type K , the Weyl transformation must commute with (β, γ, η) . Hence, a surface operator of this type is determined by the characteristic class \mathfrak{m} and a Weyl transformation that commutes with (β, γ, η) . This is the result that one would expect from section 4.5.³³

In particular, if the triple (β, γ, η) is regular, there is no freedom at all to make a Weyl transformation. The 't Hooft-like operators that act at a ramification point are hence classified entirely by the characteristic class $\mathfrak{m} \in \Lambda_{\text{cort}}$ (or Λ_{cowl} if one wishes to include operators that change the characteristic class of the bundle). This agrees with the analysis in section 4.5, where we found that in the case of a ramification point with regular semi-simple monodromy, what acts on branes is precisely this lattice.

Analog Of Type \mathbb{L}

As usual, we can generalize all this to the case of a surface operator based on an arbitrary Levi subgroup \mathbb{L} . The details are fairly obvious and are left to the reader. We will just note that 't Hooft-like operators supported on a surface of type \mathbb{L} with $\beta = \gamma = 0$ are still classified by $G_{\mathbb{C}}$ -bundles over $V \cong \mathbb{C}\mathbb{P}^1$ with parabolic structure at the two points p

³³The analysis there gave a slightly different answer (the Weyl transformation had to commute with (γ, η)) because, as only the A -model of type K was considered, β was free to vary. Here we are constructing 't Hooft-like operators that preserve the full (B, A, A) symmetry.

and p' . Now, however, the parabolic structures at p and p' are respectively of type \mathcal{P} and \mathcal{P}' , where \mathcal{P} and \mathcal{P}' are the parabolic subgroups of $G_{\mathbb{C}}$ determined by the pairs (\mathbb{L}, α) and $(\mathbb{L}, -\alpha')$. Bundles with parabolic structure of this type can be easily classified using the Yang-Mills equations and correspond to elements of the group called $\mathcal{W}_{\text{aff}, \mathbb{L}}$ in section 3.7. This is the group that acts on the cohomology of $\mathcal{M}_{H, \mathbb{L}}$ via monodromies in the space of ramification parameters. The composition of such 't Hooft-like operators is expected to give the braid-like group described in section 4.5. Including β and γ has again the effect that the Weyl group element that enters in comparing the two parabolic structures must commute with β and γ .

6. Local Models And Realizations By String Theory

6.1. Overview. The concluding section of this paper is devoted to alternative approaches to a few of the topics in this paper. We first provide an overview.

Instead of describing a surface operator by simply postulating that the fields have the familiar singularity

$$(6.1) \quad \begin{aligned} A &= \alpha d\theta \\ \phi &= \beta \frac{dr}{r} - \gamma d\theta \end{aligned}$$

along a codimension two surface D , it is tempting to describe such an operator by introducing new degrees of freedom supported on D and coupling them to the gauge fields of the $\mathcal{N} = 4$ super Yang-Mills theory in bulk. For example, hypermultiplets or vector multiplets with $(4, 4)$ supersymmetry in two dimensions can be coupled to $\mathcal{N} = 4$ super Yang-Mills theory in bulk. This will give a theory somewhat like the one that we have described “by hand” in the present paper.

The physics literature of course contains many analyses of conformal field theory coupled to defects of various kinds, that is, fields supported on a submanifold of positive codimension. The Kondo model in condensed matter physics is a classic example. Closer to our present concerns, hypermultiplets localized in codimension one have been coupled to $\mathcal{N} = 4$ super Yang-Mills theory via brane constructions, for example in [78], [79], and hypermultiplets of codimension two, the case most relevant to us, have been described in [18].

From section 3.8, we actually can see what sort of impurity model would reproduce some properties of the surface operators that we have considered. One of the important themes in the present paper is the “local singularity” of the moduli space \mathcal{M}_H of Higgs bundles when the triple (α, β, γ) becomes non-regular. The hyper-Kähler resolution of such a local singularity is conveniently described by hyper-Kähler metrics [34] on certain complex coadjoint orbits, or equivalently (in a different complex structure) hyper-Kähler metrics on $X = T^*(G/\mathbb{L})$, for various Levi

subgroups $\mathbb{L} \subset G$. Therefore, in fact, some topics explored in this paper can be expressed in terms of a combined system consisting of $\mathcal{N} = 4$ super Yang-Mills theory on a four-dimensional spacetime together with a supersymmetric sigma model supported on a two-dimensional submanifold of spacetime. We call this sigma model the local model.

Despite the usefulness of this approach, it has (as far as we know) a fundamental limitation. The theory developed in the present paper depends on the parameters³⁴ $(\alpha, \beta, \gamma, \eta) \in \mathbb{T} \times \mathfrak{t} \times \mathfrak{t} \times {}^L\mathbb{T}$. The sigma model depends on somewhat similar parameters $(\alpha, \beta, \gamma, \eta)$, where (α, β, γ) are geometrical parameters that enter in the hyper-Kähler metric on X , and η incorporates the theta-angles of the sigma model. The parameter η takes values in ${}^L\mathbb{T}$, just as in the gauge theory (see (3.44)), and likewise β and γ are \mathfrak{t} -valued in each case. But α , which is \mathbb{T} -valued in the full geometry of \mathcal{M}_H , is \mathfrak{t} -valued (and related to β and γ by an $SO(3)$ symmetry) in the hyper-Kähler geometry of $T^*(G/\mathbb{L})$, as we have reviewed in section 3.8. Consequently, the parameters of the full model and the local model do not quite match.

This leads to no contradiction with the claim that the local model captures the behavior near a local singularity, since for that purpose, the behavior when α becomes large is not important. However, it does apparently mean that the local model cannot be S -dual. Since S -duality exchanges α and η , it hardly can hold in the sigma model of target $T^*(G/\mathbb{L})$, given that in this model η is an angular variable but α takes values in a linear space. Hence, we cannot expect to maintain S -duality if we couple the $\mathcal{N} = 4$ theory to this local model.

Although this may seem surprising, another point of view perhaps makes the conclusion more natural. The spaces $T^*(G/\mathbb{L})$ are not torus fibrations and do not admit mirror symmetry, at least not in a G -invariant way. (A mirror symmetry of the sigma model that is not G -invariant would not help much when the sigma model is coupled to four-dimensional gauge theory with gauge group G .) So expecting to maintain S -duality when coupling $\mathcal{N} = 4$ super Yang-Mills theory to the local model may be unrealistic.

Nonetheless, as we have stressed, we do believe that the local model is useful for studying the behavior near a local singularity of \mathcal{M}_H , that is, near a non-regular triple (α, β, γ) . For example, consider the action of the affine braid group B_{aff} on the branes of the B -model at $\Psi = \infty$. In this action, α and γ are held fixed (and the most interesting case is that they are near zero), and one studies monodromies in β and η . Whether α is \mathfrak{t} -valued or \mathbb{T} -valued is unimportant, and B_{aff} acts on the B -branes of the local model for the same reason that it acts on the branes on \mathcal{M}_H . Indeed, the analysis of the B_{aff} action in [29] is equivalent, from

³⁴For brevity, we describe here the parameters for the case that the Levi subgroup is $\mathbb{L} = \mathbb{T}$. In general, all parameters are required to be \mathbb{L} -invariant.

a physical point of view, to studying the monodromy action on the B -branes of the local model. This seems to be a good local model for the B_{aff} action on B -branes of \mathcal{M}_H .

By contrast, the local model is not a good model for studying the action of the B_{aff} on the A -branes at $\Psi = 0$. This involves monodromy action on the pair (α, β) , and this monodromy action is lost if α takes values in \mathfrak{t} rather than \mathbb{T} . To study the monodromy action on A -branes of \mathcal{M}_H , it is essential, as far as we know, to use the full sigma model of target \mathcal{M}_H (or the full four-dimensional gauge theory). There is a slightly ironical role reversal here. The monodromy action of the A -model involves classical symplectomorphisms, while that of the B -model is highly nonclassical; but the monodromy action of the B -model can be studied in the local model, while that of the A -model cannot.

Contents Of This Section

Now we can give a brief overview of the contents of this section.

As we have discussed, despite its limitations, the local model does have its uses. The question arises of whether there is some physical approach to the local model by which we can learn more than we have gleaned already. Though we do not have a good answer to this question for general gauge groups, there is a useful description of the local model in the case of $G_{\mathbb{C}} = SL(2, \mathbb{C})$. This will be the topic of section 6.2.

The analysis in section 6.2 will reveal the importance of the fact that the instanton action (for certain supersymmetric instantons of the local model) vanishes when the triple (α, β, γ) becomes nonregular. This fact is actually inherited from a similar fact in the four-dimensional gauge theory, as we explain in section 6.3.

Finally, it is not very satisfactory to merely say that the local model lacks duality symmetry. What can we do instead that is similar but does preserve the duality symmetry? In fact, an answer to this question has already been given in string theory [18] in an interesting special case ($G_{\mathbb{C}} = SL(N, \mathbb{C})$ with a minimal coadjoint orbit). We explain this in section 6.4.

6.2. Linear Sigma Model For $G_{\mathbb{C}} = SL(2, \mathbb{C})$. To use the local model – the supersymmetric sigma model with target $T^*(G/\mathbb{L})$ – to learn something that is not obvious from four-dimensional gauge theory, we need a different way of studying it. We do not have a general approach that works uniformly for all groups, but we will explain an interesting approach for $G = SU(2)$. Here we rely upon the fact that the only relevant choice of \mathbb{L} is $\mathbb{L} = U(1)$, and that $T^*(SU(2)/U(1)) = T^*\mathbb{C}\mathbb{P}^1$ is the resolution of an A_1 singularity. As a result, it can be constructed as a hyper-Kähler quotient of a finite-dimensional linear space [62]. This has many applications in brane physics [80] and is relevant for our purposes because it can be used to construct a linear sigma model

[81] that gives a simple way to understand the vacuum structure of the supersymmetric sigma model with target $T^*\mathbb{C}\mathbb{P}^1$.

Complex coadjoint orbits of all the classical Lie groups can similarly be constructed as hyper-Kähler quotients of finite-dimensional linear spaces [82]. These more general constructions could be taken as the starting point in constructing a linear sigma model relevant to surface operators for any classical group. However, the case of $SU(2)$ has an important advantage, which is that the action of the Weyl group is transparent in the linear sigma model. This is not the case for G of higher rank, which is why we will in this analysis consider only $SU(2)$.

To get the A_1 singularity as a hyper-Kähler quotient, we start with $\mathbb{H}^2 \cong \mathbb{R}^8$, regarded as a flat hyper-Kähler manifold. It admits the action of $Sp(2)$, which in our notation is the group of 2×2 unitary matrices of quaternions. Two kinds of quaternionic matrices will play an important role. $Sp(2)$ contains the group $SU(2)$ consisting of quaternion multiples of the identity

$$(6.2) \quad \begin{pmatrix} v & 0 \\ 0 & v \end{pmatrix},$$

where v is a quaternion of modulus 1, and also the group $O(2)$ of unitary quaternion matrices whose entries are actually all real:

$$(6.3) \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

These two groups commute with each other, but the global structure is not a product, because $SU(2)$ and $O(2)$ contain the common central element -1 . So the global structure is

$$(6.4) \quad (SU(2) \times O(2)) / \mathbb{Z}_2 \subset Sp(2).$$

The connected component of $O(2)$ is $U(1)$. Let $\vec{\mu} = (\mu_1, \mu_2, \mu_3)$ be the hyper-Kähler moment map for the action of this $U(1)$. We introduce real parameters (α, β, γ) , and define $X(\alpha, \beta, \gamma)$ to be the hyper-Kähler quotient $\mathbb{H}^2 // U(1)$ taken at $\vec{\mu} = (\alpha, \beta, \gamma)$. In other words, $X(\alpha, \beta, \gamma)$ is defined by setting $\vec{\mu} = (\alpha, \beta, \gamma)$ and dividing by $U(1)$. For $\alpha = \beta = \gamma = 0$, this gives the A_1 singularity $\mathbb{R}^4 / \mathbb{Z}_2$; in general, it gives the hyper-Kähler resolution of the A_1 singularity.

From a physical point of view, we construct a supersymmetric linear sigma model on a two-manifold D in which the fields are a $U(1)$ vector multiplet and a pair of hypermultiplets H_1, H_2 (whose bosonic components parametrize \mathbb{H}^2 , acted on by $U(1)$ as above). The parameters (α, β, γ) are the D -terms or Fayet-Iliopoulos parameters of the vector multiplet. The moduli space of classical vacua is precisely the hyper-Kähler quotient $X(\alpha, \beta, \gamma)$ described in the last paragraph, with one

caveat: if $\alpha = \beta = \gamma = 0$, this moduli space has a second branch, as described in [81].

The linear sigma model has an $SO(3)$ symmetry that rotates the parameters (α, β, γ) . This arises as follows. If we think of an element of \mathbb{H}^2 as a column of quaternions $\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$, then the action of $Sp(2)$ on the left commutes with a right action of $Sp(1)$. This $Sp(1)$ acts by rotating the complex structures of the hyper-Kähler manifold \mathbb{H}^2 , and likewise rotating the parameters (α, β, γ) .

In addition to (α, β, γ) , the quantum theory also depends on the theta-angle of the $U(1)$ vector multiplet, which as usual we will call η . Clearly, η is angle-valued. It appears in a factor in the path integral of the form

$$(6.5) \quad \exp\left(i\eta \int_D F\right).$$

Here F is the $U(1)$ curvature, which we represent by a real two-form. This means, in the usual form of the model, that there is a symmetry

$$(6.6) \quad \eta \rightarrow \eta + 1.$$

As we explain later, there is a certain sense in which the basic symmetry is $\eta \rightarrow \eta + 2$.

The parameters $(\alpha, \beta, \gamma, \eta)$ take values in $K_0 = \mathbb{R}^3 \times S^1$. But this is not quite the natural parameter space of the model. We must recall from (6.3) that the $U(1)$ symmetry of the model can be extended to $O(2)$. Let w be an element of the disconnected conjugacy class of $O(2)$. As we will see momentarily, w will play the role of the non-trivial element of the Weyl group of $SU(2)$. (This fact does not have a good analog for similar models for coadjoint orbits of groups of higher rank. That is why we limit ourselves here to $SU(2)$.) w acts by an outer automorphism of the gauge group $U(1)$ of the linear sigma model, and acts on the coupling parameters by changing all signs:

$$(6.7) \quad w : (\alpha, \beta, \gamma, \eta) \rightarrow (-\alpha, -\beta, -\gamma, -\eta).$$

This is the action that in the rest of the present paper has always come from the Weyl group of $SU(2)$. The parameter space of the model is thus really $K = K_0/\mathbb{Z}_2$. At its fixed points, which will be described momentarily, w becomes a symmetry of the linear sigma model.

The question of where in parameter space the model becomes singular was addressed in [81]. The answer is that it becomes singular precisely at the point $P_0 : \alpha = \beta = \gamma = \eta = 0$, when the vacuum becomes unnormalizable because of the role of the “second branch” of classical vacua.

Obviously, P_0 is one of the two fixed points for the action of \mathbb{Z}_2 on K_0 . The second is $P_1 : \alpha = \beta = \gamma = 0, \eta = 1/2$. What happens

there? From the point of view of classical geometry, one does not see the parameter η . The hyper-Kähler quotient at $\alpha = \beta = \gamma = 0$ is the orbifold singularity $\mathbb{R}^4/\mathbb{Z}_2$. There is a soluble conformal field theory associated with this orbifold [83]. One may guess that for some value of η , the linear sigma model will reduce to this orbifold theory at low energies. This in fact occurs [84] at $\eta = 1/2$, or more exactly at $P_1 : (\alpha, \beta, \gamma, \eta) = (0, 0, 0, 1/2)$. The theory at this point is a perfectly smooth, well-behaved quantum field theory, and in fact a simple and exactly soluble one. Since P_1 is a fixed point of w , the orbifold theory has w as a symmetry. In fact, w is the “quantum symmetry” of the orbifold (which acts as -1 for strings in the twisted sector and $+1$ on untwisted strings).

Now let us focus on the A -model or B -model derived from this theory (upon picking one of the symplectic or complex structures), and the associated monodromies as the parameters are varied. Actually, as was explained in the overview of this section, it is not very interesting to consider the A -model, because in the present local model α takes values in a linear space, preventing the existence of interesting monodromies. So we will concentrate on the B -model in one of the complex structures. Because of the $SO(3)$ symmetry that rotates the complex structures, it does not matter which one we pick. To agree with the terminology in the rest of this paper, we will consider the B -model in complex structure J .

The monodromies of this model come as in section 4.5 by varying β and η while keeping α and γ fixed. If the pair (α, γ) is nonzero, then the relevant parameters are β and η modulo the symmetry $\eta \rightarrow \eta + 1$. We do not get anything new from the “Weyl transformation” w , since it does not leave (α, γ) fixed. Thus, in this situation, the appropriate parameter space is simply the quotient of the $\beta - \eta$ plane by $\eta \rightarrow \eta + 1$, or $\mathbb{R}^2/\mathbb{Z} = \mathbb{R} \times S^1$. The monodromy group is simply \mathbb{Z} . This answer is familiar from section 4.5.

As usual, the most interesting case is $(\alpha, \gamma) = (0, 0)$. Here after dividing by \mathbb{Z} to get $\mathbb{R} \times S^1 \cong \mathbb{C}^*$, we must still divide by $w : (\beta, \eta) \rightarrow (-\beta, -\eta)$. w acts on $\lambda = \exp(2\pi(\beta + i\eta))$ by $\lambda \rightarrow \lambda^{-1}$. There are two fixed points, at $\lambda = 1$ and $\lambda = -1$. The fixed point at $\lambda = 1$ is the point P_0 at which the theory becomes singular. We omit this point from the parameter space and call what remains $\tilde{\mathbb{C}}^*$. We still have the action of \mathbb{Z}_2 on $\tilde{\mathbb{C}}^*$, with a fixed point at $P_1 : \lambda = -1$. We do not want to remove this fixed point, since the quantum field theory is well-defined there. Rather, we think of the parameter space on which the quantum field theory depends as an orbifold $\tilde{\mathbb{C}}^*/\mathbb{Z}_2$, and the relevant monodromy group is the fundamental group in the orbifold sense, $\pi_1^{\text{orb}}(\tilde{\mathbb{C}}^*/\mathbb{Z}_2)$. This group is generated by A , the monodromy around $\lambda = 1$, and T , the monodromy

around $\lambda = -1$, with the sole relation being $T^2 = 1$, expressing the fact that $\lambda = -1$ is a \mathbb{Z}_2 orbifold point.

This answer for the monodromies seen in the local model is essentially equivalent to the answer obtained in section 4.5, but some explanation is required. First of all, let us consider the maximal torus \mathbb{T} of the group $SO(3)$. It consists of elements of the following form:

$$(6.8) \quad U = \begin{pmatrix} \cos(2\pi\eta) & \sin(2\pi\eta) & 0 \\ -\sin(2\pi\eta) & \cos(2\pi\eta) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

η is real if we want the compact group $SO(3)$, or should be replaced by its complexification $\eta - i\beta$ in the case of the complex Lie group $SO(3, \mathbb{C}) = SL(2, \mathbb{C})/\mathbb{Z}_2$. In that case, we can write

$$(6.9) \quad U = \begin{pmatrix} \frac{1}{2}(\lambda + \lambda^{-1}) & \frac{i}{2}(\lambda - \lambda^{-1}) & 0 \\ -\frac{i}{2}(\lambda - \lambda^{-1}) & \frac{1}{2}(\lambda + \lambda^{-1}) & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

with $\lambda = \exp(2\pi i(\eta - i\beta)) \in \mathbb{C}^*$. We call an element $U \in \mathbb{T}$ very regular if the subgroup of $SO(3)$ that commutes with U is precisely \mathbb{T} , and regular if the connected component of this group is equal to \mathbb{T} . This criterion does not depend on whether we work in $SO(3)$ or $SO(3, \mathbb{C})$. The element $U = 1$ corresponding to $\lambda = 1$ is nonregular. By contrast, the element U that corresponds to $\lambda = -1$ is regular but it is not very regular, since it commutes with the element

$$(6.10) \quad w = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

which acts by $\lambda \rightarrow \lambda^{-1}$, and, in fact, generates the Weyl group of $SO(3)$.

We get a regular conjugacy class for any $\lambda \in \tilde{\mathbb{C}}^*$. However, the conjugacy classes in $SO(3)$ that correspond to λ and λ^{-1} are conjugate by the action of w . So the moduli space of regular conjugacy classes in $SO(3)$ is the orbifold $\tilde{\mathbb{C}}^*/\mathbb{Z}_2$, and, as we have seen, its orbifold fundamental group $\pi_1^{\text{orb}}(\tilde{\mathbb{C}}^*/\mathbb{Z}_2)$ is also the monodromy group of our B -model for the case $(\alpha, \gamma) = (0, 0)$.

In section 4.5, it was convenient to consider the subgroup of the monodromy group that consists of transformations that act trivially on the discrete electric flux \mathbf{e}_0 . So let us identify that subgroup in the local model. As we explained in section 4.1, to define \mathbf{e}_0 in $SU(2)$ gauge theory, one allows a “twist” so that the gauge bundle $E \rightarrow M = \Sigma \times C$ has structure group $SO(3) = SU(2)_{\text{ad}}$, but lifts to an $SU(2)$ bundle if restricted to $q \times C$ for q a point in Σ . In the local model, we do not necessarily have such a lifting to $SU(2)$, since the support of the local

model is not $q \times C$ but is $D = \Sigma \times p$ for some point $p \in C$. In defining \mathbf{e}_0 , we include $SO(3)$ -bundles $E \rightarrow D$ with $\int_D w_2(E) \neq 0$.

What does this mean in the local model? The group that acts on the hypermultiplets of the local model is not $SO(3)$ but $SU(2)$, so at first sight one might think it is impossible to twist the local model by an $SO(3)$ bundle E with non-zero $w_2(E)$. However, the local model has a gauge group $U(1)$, and by restricting (6.4) to the connected component, we see that the global form of the symmetry group is not $SU(2) \times U(1)$ but $(SU(2) \times U(1))/\mathbb{Z}_2$. An $SO(3)$ bundle $E \rightarrow D$ with $w_2(E) \neq 0$ cannot be lifted to an $SU(2)$ bundle, but it can be lifted to a bundle with structure group $(SU(2) \times U(1))/\mathbb{Z}_2$. When we do this, the $U(1)$ curvature F obeys

$$(6.11) \quad \int_D \frac{F}{2\pi} = \frac{1}{2} \int_D w_2(E) \pmod{\mathbb{Z}}.$$

Looking back to (6.5), we see that this means that the effect of $\eta \rightarrow \eta + 1$ is to multiply the integrand of the path integral by

$$(6.12) \quad (-1)^{\int_D w_2(E)}.$$

This operation is equivalent to shifting \mathbf{e}_0 .

The result is that the fundamental group $\pi_1^{\text{orb}}(\tilde{C}^*)$ is the full monodromy group of the model, including transformations that shift \mathbf{e}_0 . If we want to identify the subgroup of monodromies that keep \mathbf{e}_0 fixed, which will facilitate the comparison to the result of section 4.5, we should consider the pair (β, η) subject to the symmetries $w : (\beta, \eta) \rightarrow (-\beta, -\eta)$ and

$$(6.13) \quad \eta \rightarrow \eta + 2.$$

These are the symmetries that keep \mathbf{e}_0 fixed. Identifying $\eta \pmod{2}$ is equivalent to lifting U to an element of the maximal torus of $SU(2)$, rather than $SO(3)$. As such it is conjugate to

$$(6.14) \quad \hat{U} = \begin{pmatrix} \exp(i\pi(\eta - i\beta)) & 0 \\ 0 & \exp(-i\pi(\eta - i\beta)) \end{pmatrix}.$$

The Weyl group still acts on the pair (β, η) with two fixed points, but there is an essential difference. The fixed points are now $P_0 : (\beta, \eta) = (0, 0)$ and $P'_1 : (\beta, \eta) = (0, 1)$. The essential difference is that the two fixed points of the Weyl group are now both points at which the sigma model is singular (since its singularities are invariant under $\eta \rightarrow \eta + 1$). In contrast, before lifting to $SU(2)$, one fixed point was a singularity of the sigma model and one was an orbifold point.

The non-regular values of \hat{U} are 1 and -1 , which correspond precisely to the fixed points P_0 and P'_1 . So when we omit the points at which the local model is singular, and divide by the Weyl group and by $\eta \rightarrow \eta + 2$, we get precisely the moduli space of regular conjugacy classes

in $SL(2, \mathbb{C})$. Its fundamental group is known as the affine braid group of $SL(2, \mathbb{C})$, and denoted $B_{\text{aff}}(SL(2, \mathbb{C}))$. This group is freely generated by elements A and B that we can regard as the monodromies around the points 1 and -1 . They obey no relations at all.

The full monodromy group of the local model, including transformations that change \mathbf{e}_0 , is an extension:

$$(6.15) \quad 1 \rightarrow B_{\text{aff}} \rightarrow \pi_1^{\text{orb}}(\tilde{\mathbb{C}}^*) \rightarrow \mathbb{Z}_2 \rightarrow 1.$$

Here B_{aff} is extended by an outer automorphism of order 2 that acts by $\eta \rightarrow \eta + 1$ and exchanges the points $\hat{U} = 1$ and $\hat{U} = -1$, which are equivalent in $SO(3)$. We call this automorphism T ; it obeys $T^2 = 1$, $TAT = B$, $TBT = A$. Since B is the same as TAT , the extension $\pi_1^{\text{orb}}(\tilde{\mathbb{C}}^*)$ is generated by A and T with the sole relation $T^2 = 1$. This is the result we gave earlier for the full monodromy group.

All of this is in accord with the analysis of section 4.5. The subgroup of the monodromy group that acts trivially on \mathbf{e}_0 is the affine braid group, and the full monodromy group is an extension of the affine braid group by the center of the simply-connected form of the gauge group G . In the present example, $G = SL(2, \mathbb{C})$, and the center is \mathbb{Z}_2 .

To tie up a loose end, we should perhaps mention that $SL(2, \mathbb{C})$ is an exception to the description of the affine braid group given in eqns. (5.16) and (5.17). The reason for the exception is that in $SL(N, \mathbb{C})$ for $N > 2$, the adjacent nodes of the affine Dynkin diagram correspond to vectors at an angle $2\pi/3$, but this is not so for $SL(2, \mathbb{C})$ (where the angle is π). To describe what happens for $SL(2, \mathbb{C})$, we go back to the definition of the affine Weyl group as an extension $\Lambda_{\text{cort}} \rtimes \mathcal{W}$, with \mathcal{W} the Weyl group. For $SU(2)$, $\Lambda_{\text{cort}} \cong \mathbb{Z}$, so a vector in Λ_{cort} is just an integer n . The Weyl group is \mathbb{Z}_2 , generated by $A : n \rightarrow -n$. The affine Weyl group is generated by A together with $R : n \rightarrow n + 1$. They obey $A^2 = 1$ and $ARA = R^{-1}$. Equivalently, \mathcal{W}_{aff} is generated by A and $B = AR$ with $A^2 = B^2 = 1$ and no other relations. A and B are the reflections corresponding to the two nodes of the extended Dynkin diagram. To get the affine braid group of $SL(2, \mathbb{C})$, we just drop the conditions $A^2 = B^2 = 1$, so $B_{\text{aff}}(SL(2, \mathbb{C}))$ is simply a free group with the two generators A and B .

6.3. Instantons And The Local Singularity. In this discussion, the fact that the local model is singular at $\alpha = \beta = \gamma = \eta = 0$ (and non-singular elsewhere) played a crucial role.

One explanation of this is given in [81], using the fact that a second branch of classical vacua becomes relevant precisely when these parameters all vanish. There is, however, another standard explanation of the significance of having $\alpha = \beta = \gamma = \eta = 0$. When $(\alpha, \beta, \gamma) \neq 0$, the A_1 singularity undergoes a hyper-Kähler resolution to produce a smooth manifold with the topology of $T^*\mathbb{C}\mathbb{P}^1$. The zero section of this

cotangent bundle is holomorphic in the complex structure

$$(6.16) \quad \mathcal{I} = \frac{(\alpha I + \beta J + \gamma K)}{\sqrt{\alpha^2 + \beta^2 + \gamma^2}}.$$

It is antiholomorphic in the opposite complex structure. Its area A is proportional to $\sqrt{\alpha^2 + \beta^2 + \gamma^2}$. A holomorphic or antiholomorphic map of D to $\mathbb{C}\mathbb{P}^1 \subset T^*\mathbb{C}\mathbb{P}^1$ is holomorphic in complex structure \mathcal{I} or $-\mathcal{I}$. We interpret it as an instanton of the sigma model. In other complex structures, there are no compact holomorphic or anti-holomorphic curves. An instanton of the sigma model that is defined by a degree d holomorphic mapping of D to the zero section of $T^*\mathbb{C}\mathbb{P}^1$ makes a contribution to the path integral that is proportional to q^d , where $q = \exp(-kA + 2\pi i\eta)$. (Here k is a constant that depends on the gauge coupling of the linear sigma model.) Upon summing over d , the instanton series has a pole at $q = 1$, or in other words when $(\alpha, \beta, \gamma, \eta)$ all vanish. This pole reflects the singularity of the sigma model.

What has just been summarized is a standard analysis in two dimensions. We want to consider here how this story looks in four dimensions. The twisted $\mathcal{N} = 4$ super Yang-Mills theory that underlies the present discussion depends on a parameter t . The analog of the instanton equation in four dimensions, according to section 3.2 of [8], is the condition $\mathcal{V}^+(t) = \mathcal{V}^-(t) = \mathcal{V}^0 = 0$, where

$$(6.17) \quad \begin{aligned} \mathcal{V}^+(t) &= (F - \phi \wedge \phi + td_A\phi)^+ \\ \mathcal{V}^-(t) &= (F - \phi \wedge \phi - t^{-1}d_A\phi)^- \\ \mathcal{V}^0 &= D \star \phi. \end{aligned}$$

Moreover, according to eqn. (5.28) (or eqns. (5.7) and (5.11)) of [8], if we specialize to $M = \Sigma \times C$, then the four-dimensional topological field theory reduces on Σ to a two-dimensional A -model with target space $\mathcal{M}_H(G, C)$, and symplectic structure a multiple of

$$(6.18) \quad \omega_t = \frac{1-t^2}{1+t^2}\omega_I - \frac{2t}{1+t^2}\omega_K.$$

The instantons in this A -model are holomorphic curves for the corresponding complex structure

$$(6.19) \quad I_t = \frac{1-t^2}{1+t^2}I - \frac{2t}{1+t^2}K.$$

Now let us incorporate a surface operator supported on $D = \Sigma_p = \Sigma \times p$, for p a point in C . We endow the surface operator with parameters (α, β, γ) . From the point of view of the local model, we expect holomorphic instantons only in the complex structure \mathcal{I} . The local model should be adequate at least for describing those instantons whose action goes to zero as $(\alpha, \beta, \gamma) \rightarrow 0$. We will call these the instantons with small action. Four-dimensional instantons of small action should correspond

to instantons of the sigma model with complex structure I_t . The sigma model only has instantons in complex structure \mathcal{I} . Therefore, we expect the four-dimensional gauge theory to have instantons – or at least instantons of small action – only if $I_t = \mathcal{I}$. The condition to have $I_t = \mathcal{I}$, in view of (6.16) and (6.19), is that

$$(6.20) \quad \frac{\alpha}{\gamma} = \frac{t - t^{-1}}{2}, \quad \beta = 0.$$

We want to derive these conditions directly in four dimensions.

First of all, from a four-dimensional point of view, \mathcal{V}^0 cannot vanish if $\beta \neq 0$. Indeed, having $\beta \neq 0$ causes $d_A \star \phi$ to have a delta function along D , so it cannot vanish. This is one of the desired results; it remains to derive the first equation in (6.20). We will do this by generalizing the vanishing arguments of section 3.3 of [8].

What in that reference is called Vanishing Theorem 1 asserts that in any solution of the equations (6.17) (in the absence of surface operators) the Pontryagin number vanishes. This follows from the following identity:

$$(6.21) \quad \int_M \text{Tr} (\mathcal{V}^+(t) \wedge \mathcal{V}^+(-t^{-1}) + \mathcal{V}^-(t) \wedge \mathcal{V}^-(-t^{-1})) = \int_M \text{Tr} F \wedge F.$$

The derivation of this identity depends on integration by parts. The left hand side vanishes in a solution of the equations, and of course the right hand side is a multiple of the Pontryagin number. So the identity implies vanishing of the Pontryagin number.

A few things are different in the presence of a surface operator supported on $D \subset M$. As in the derivation of (2.42), we interpret the integral in (6.21) as an integral over the complement of D in M . With that understood, the Pontryagin number is not simply a multiple of $\int_M \text{Tr} F \wedge F$, but has the extra terms indicated in (2.42). Second, in the integration by parts that is needed to derive (6.21), one runs into a term $(t - t^{-1}) \int_M d(\text{Tr} \phi \wedge F)$. For $\gamma \neq 0$, despite being the integral of an exact form, this does not vanish, even for compact M . Instead, there is a sort of surface contribution localized on D , of the form $-2\pi(t - t^{-1}) \int_D \text{Tr} \gamma F$. With this understood, the generalization of (6.21) is

$$(6.22) \quad \int_M \text{Tr} (\mathcal{V}^+(t) \wedge \mathcal{V}^+(-t^{-1}) + \mathcal{V}^-(t) \wedge \mathcal{V}^-(-t^{-1})) \\ = -8\pi^2 \mathbf{N} + 4\pi^2 (D \cap D) \text{Tr} \alpha^2 + 4\pi \int_D \text{Tr} \left(\alpha - \frac{t - t^{-1}}{2} \gamma \right) F.$$

The left hand side vanishes in any solution of the supersymmetric equations. So obviously, if such a solution is to exist for small (α, β, γ) , the integer \mathbf{N} must vanish. In addition, in our application with $M =$

$\Sigma \times C$ and $D = \Sigma \times p$, we have $D \cap D = 0$. So all terms on the right vanish except the last one, which hence must also vanish. The G -bundle $E \rightarrow M$, when restricted to D , has structure group \mathbb{T} , and is characterized topologically by the characteristic class \mathfrak{m} of its curvature F . For $G = SU(2)$, as assumed in the derivation of (6.20), \mathfrak{m} is equivalent to the instanton number of the sigma model (the degree of the map $D \rightarrow \mathbb{C}\mathbb{P}^1$), and the singularity of the sigma model at (α, β, γ) approaching zero is supposed to come from instantons with $\mathfrak{m} \neq 0$ and small action. But (6.22) shows us that if $N = D \cap D = 0$, then an instanton with $\mathfrak{m} \neq 0$ must have $\alpha - \frac{1}{2}(t - t^{-1})\gamma = 0$. This is the desired condition in eqn. (6.20), so we have explained from a four-dimensional point of view the fact that instantons or supersymmetric field configurations appear only when this condition is obeyed.

It is similarly possible to derive the same condition by generalizing Vanishing Theorem 2 of [8] to incorporate surface operators.

6.4. Some String Theory Constructions. The assertion that the local model does not preserve S -duality is a little perplexing. In order to understand this better, we will consider a string theory construction in which S -duality is manifest, and see how this construction fails to give the local model.

As in [18], we will consider a much studied situation in Type IIB superstring theory: a stack of N $D3$ -branes, giving a $U(N)$ gauge theory in four dimensions, that intersects in codimension two another $D3$ -brane which we will call $D3'$. Thus everything happens in a six-dimensional subspace of the ten-dimensional spacetime of the Type IIB theory. We will take this six-dimensional subspace to be a product $Z = \Sigma \times C \times \tilde{C}$ of three Riemann surfaces. The N $D3$ -branes are supported on $M = \Sigma \times C \times r$, with r a point in \tilde{C} , and the $D3'$ -brane is supported on $\tilde{M} = \Sigma \times p \times \tilde{C}$, with p a point in C . \tilde{C} may be either the complex plane \mathbb{C} or a compact Riemann surface. In our usual application, C is the Riemann surface on which we consider the geometric Langlands program, and Σ is the spacetime of the two-dimensional effective field theory that results from compactification on C . One can also replace $M = \Sigma \times C$ with a more general four-manifold, and presently this will be convenient.

The stack of N $D3$ -branes wrapped on M produces a $U(N)$ gauge theory on M with $\mathcal{N} = 4$ supersymmetry, and the $D3'$ -brane produces a $U(1)$ gauge theory on \tilde{M} also with $\mathcal{N} = 4$ supersymmetry. On the intersection $M \cap \tilde{M} = \Sigma \times p \times r$, one gets hypermultiplets in the N -dimensional representation of $U(N)$ and with charge 1 under $U(1)$. To try to make contact with the situation explored in this paper, one might hope that the hypermultiplets, together with the $U(1)$ gauge multiplet supported on \tilde{M} , would be equivalent at low energies to a sigma model

with target $\mathbb{H}^N // U(1)$, where $\mathbb{H}^N \cong \mathbb{C}^{2N}$ is parametrized by the hypermultiplets and $\mathbb{H}^N // U(1)$ is its hyper-Kähler quotient by $U(1)$.

This hyper-Kähler quotient depends on the constant values chosen for the three components of the hyper-Kähler moment map. Physically, the constants will have to arise from the Fayet-Iliopoulos D -terms of the $U(1)$ gauge theory on \widetilde{M} , which one would like to somehow derive from parameters of the string theory construction. For non-zero D -terms, the hyper-Kähler quotient is equivalent in one complex structure to $T^*\mathbb{C}\mathbb{P}^{N-1}$, and in a different complex structure it is equivalent to the orbit of an element

$$(6.23) \quad \delta = \begin{pmatrix} N-1 & & & \\ & -1 & & \\ & & \ddots & \\ & & & -1 \end{pmatrix}$$

of the Lie algebra $\mathfrak{gl}(N)$. What we get here is a special case of the local model discussed in section —reoverview, namely the case with $G = U(N)$, $\mathbb{L} = U(N-1)$, and so $T^*(G/\mathbb{L}) = T^*(U(N)/U(N-1)) = T^*\mathbb{C}\mathbb{P}^{N-1}$. The parameters (α, β, γ) must each, in this case, be real multiples of δ , with the proportionality determined by the Fayet-Iliopoulos parameters. Thus, each depends on only one real parameter, an assertion that will remain true in the Anti de Sitter construction explained below.

This looks like a plausible way to get from branes an example of the situation discussed in this paper. However, since the brane construction (like any configuration of $D3$ -branes in Type IIB superstring theory) has $SL(2, \mathbb{Z})$ S -duality, and the model consisting of the N hypermultiplets plus $U(1)$ gauge field does not, the brane configuration cannot really be equivalent to the coupling of $\mathcal{N} = 4$ super Yang-Mills theory to the two-dimensional system of hypermultiplets and gauge fields. What fails depends on the choice of \widetilde{C} .

If we take $\widetilde{C} = \mathbb{C}$, then we do not quite get the reduction we want, since the infinite area of \widetilde{C} means that one cannot treat the $D3'$ -brane in two-dimensional terms. More fundamentally, the dynamics of this problem in general involves brane recombination, in which one of the N $D3$ -branes combines with the $D3'$ -brane to make a smooth (but deformed) brane. Locally, if C is defined by a complex equation $x = 0$ and \widetilde{C} by $y = 0$, the recombination can be described by deforming an

intersection $xy = 0$ to a smooth curve³⁵

$$(6.24) \quad xy = \epsilon.$$

The low energy effective physics cannot then be described without taking into account the non-compactness of \tilde{C} , and does not lead to the sort of surface operator considered in this paper. On the plus side of the ledger, the noncompactness of \tilde{C} means that whatever we get from this construction depends on parameters (α, β, γ) that can be defined in terms of the behavior at infinity on \tilde{C} . For example, α is the holonomy at infinity on \tilde{C} of the $U(1)$ gauge field, and $\beta + i\gamma$ is the parameter ϵ of eqn. (6.24), which can also be measured at infinity (in terms of the way the brane is “bent”).

If we take \tilde{C} to be compact, we have the opposite problem. Compactness of \tilde{C} means that all fields on \tilde{C} other than zero modes can be eliminated in a low energy description. So we get some sort of description in terms of four-dimensional fields defined on $M = \Sigma \times C$ and impurity fields supported on $\Sigma \times p$. In particular, the $U(1)$ gauge field on the $D3'$ -brane can be eliminated at low energies by taking a hyper-Kähler quotient. But when \tilde{C} is compact, all variables describing the physics of the brane intersection and the fields on \tilde{C} are dynamical (rather than being “frozen” at infinity). So there is no apparent way to introduce parameters corresponding to $(\alpha, \beta, \gamma, \eta)$. It would be interesting, however, to know what kind of surface operator one does get from the case of compact \tilde{C} .

An Alternative

There is an alternative to this, explained in [18]. This alternative enables us to maintain the S -duality and to see the parameters $(\alpha, \beta, \gamma, \eta)$. It is not equivalent to the local model (it hardly can be, as the local model is not S -dual), but it is an elegant way to use branes and string theory to describe an example of the construction studied in this paper.

The alternative is merely to use the AdS/CFT duality [85] to replace the $D3$ -branes with a description involving the Type IIB geometry that these $D3$ -branes create. The $D3'$ -brane will remain as part of the Type IIB description, and this will give the desired surface operator.

To simplify things, we will take $M = S^4$, in which case the simplest relevant Type IIB spacetime is $AdS_5 \times S^5$. The simplest choice of the $D3'$ -brane is $AdS_3 \times S^1$, where $AdS_3 \times S^1$ is embedded in $AdS_5 \times S^5$ in the obvious way. The embedding of S^1 in S^5 is unstable topologically, but energetically it is stable and in fact leads to a supersymmetric membrane on AdS_5 . The conformal boundary of AdS_3 is $D \cong S^2$, embedded in

³⁵For small ϵ , in the space \mathbb{C}^2 parametrized by x and y , there is a supersymmetric disc, of area proportional to ϵ , whose boundary is on the curve $xy = \epsilon$. One can also consider a multiple covering of this disc. This gives a stringy model of the instantons of small action that were discussed in section 6.3.

$M \cong S^4$ in the obvious way (the S^1 factor of $AdS_3 \times S^1$ effectively shrinks to a point near the conformal boundary). D is, of course, a two-dimensional submanifold of M . By quantizing in the presence of a $D3'$ -brane whose asymptotic behavior is required to coincide with that of $AdS_3 \times S^1$, one defines in the $\mathcal{N} = 4$ super Yang-Mills theory on M a supersymmetric surface operator supported on D . This is precisely in parallel with the use [86], [87] of strings in $AdS_5 \times S^5$ to describe supersymmetric Wilson and 't Hooft operators on M . One can modify this to consider more general M or D , but this is not essential for what we will say.

The surface operator obtained this way has the familiar parameters $(\alpha, \beta, \gamma, \eta)$. To see them, we consider the $\mathcal{N} = 4$ super Yang-Mills theory, with gauge group $U(1)$, supported on the $D3'$ -brane. The world-volume of that brane, in our simple example, is $X = Y \times S^1$, with $Y = AdS_3$. And in general this gives the right asymptotic behavior, which is all that one needs in describing the parameters $(\alpha, \beta, \gamma, \eta)$. They arise as follows. The parameter α is the holonomy of the $U(1)$ gauge field around S^1 , measured near the conformal boundary of Y . The part of the $U(1)$ gauge field on X that is invariant under rotations around S^1 is a $U(1)$ gauge field on Y that can be dualized to give an angle-valued scalar field. The expectation value of this scalar field near the boundary of Y is η . Finally, to find β and γ , we use the ‘‘Higgs field’’ φ of the gauge multiplet on the $D3'$ -brane. In Poincaré coordinates, the metric of AdS_3 can be written as $\frac{1}{r^2}(dr^2 + dx^2 + dy^2)$, where $0 \leq r \leq \infty$, and the (x, y) -plane is the two-sphere D with a point at infinity removed. The metric of $X = Y \times S^1$ is thus $\frac{1}{r^2}(dr^2 + dx^2 + dy^2) + d\theta^2$ where θ is an angle. Upon setting $z = re^{i\theta}$, we introduce β and γ by requiring that φ should be asymptotic at $r = 0$ (that is, on the conformal boundary of Y) to $(\beta + i\gamma)(dz/2z)$. One can verify that the underlying S -duality of the $U(1)$ gauge theory, which we reviewed in section 2.4, acts in the expected fashion on these parameters.

Another Variant

For another variant of this construction, we begin with the $(0, 2)$ conformal field theory in six dimensions [81]. This theory on S^6 can be described [85] in terms of M -theory on $AdS_7 \times S^4$ (or more precisely on spacetimes with that asymptotic behavior) with N units of flux on S^4 . The $(0, 2)$ model has surface operators [88], [89], which can be defined by considering $M2$ -branes in $AdS_7 \times S^4$ whose world-volume is asymptotic at infinity to a two-dimensional surface $D \subset S^6$. One can replace S^6 by a more general six-manifold Z by replacing $AdS_7 \times S^4$ with spacetimes with the appropriate asymptotic behavior.

Now take the six-manifold on which the $(0, 2)$ model is formulated to be $Z = M \times T^2$ with M a four-manifold. The $(0, 2)$ model formulated on Z is equivalent at low energies to $\mathcal{N} = 4$ super Yang-Mills theory on

M , with gauge group $U(N)$. A surface operator on Z whose support is a circle in M times one of the two factors in $T^2 = S^1 \times S^1$ gives a Wilson or 't Hooft operator on M . (More generally, any choice of a geodesic circle in T^2 gives a mixed Wilson-'t Hooft operator.) However, a surface operator on Z whose support is a surface $D \subset M$ times a point $r \in T^2$ gives a surface operator on M , with α and η determined by the choice of r . S -duality comes from the geometrical action of $SL(2, \mathbb{Z})$ on T^2 . This construction is closely related to the previous one, since M -theory on T^2 is dual to Type IIB on a circle.

Appendix A. Review Of Duality

In this appendix, we will give a brief review of the relation between a Lie group G and its Langlands or GNO dual ${}^L G$. For general background, see chapter 3 of [90] or chapter 8 of [91]. For brevity, we describe only the case that G is simple, though the theory extends to any compact Lie group.

The root system Φ of G is a finite set of nonzero vectors in a fixed Euclidean space $E \cong \mathbb{R}^r$ (where r will be the rank of G). E will eventually be interpreted as \mathfrak{t}^\vee , the dual of the Lie algebra \mathfrak{t} of a maximal torus \mathbb{T} of G . We consider E to be endowed with a metric $(\ , \)$ which to begin with we consider to be defined only up to multiplication by a positive real scalar. Later, we will choose a particular metric.

Φ is required to obey certain axioms. The vectors in Φ generate a rank r lattice Λ_{rt} (eventually interpreted as the root lattice of G), which moreover has no decomposition as a direct sum of orthogonal sublattices. Additionally, for $\mu \in \Phi$, we require that the multiples of μ that are contained in Φ are precisely the vectors $\pm\mu$. This will ensure that there is an $\mathfrak{sl}(2)$ subalgebra of \mathfrak{g} with nonzero roots $\pm\mu$. We call this algebra $\mathfrak{sl}(2)_\mu$. We also ask that for any other root ν , the expression

$$(A.1) \quad \langle \nu, \mu \rangle = \frac{2(\nu, \mu)}{(\mu, \mu)}$$

is an integer. Note that the symbol $\langle \ , \ \rangle$, which is independent of the normalization of $(\ , \)$, is only linear in the first variable. The integrality of $\langle \nu, \mu \rangle$ is interpreted as the condition that each root ν has integer or half-integer weight with respect to $\mathfrak{sl}(2)_\mu$. Finally, we ask that the set of roots should be closed under the reflection with respect to μ , which is the operation

$$(A.2) \quad \nu \rightarrow \nu - \langle \nu, \mu \rangle \mu.$$

This operation is eventually interpreted as the Weyl transformation of the algebra $\mathfrak{sl}(2)_\mu$.

For any such root system Φ , one defines a dual root system Φ^\vee in the following way. For $\mu \in \Phi$, we define

$$(A.3) \quad \mu^\vee = \frac{2\mu}{(\mu, \mu)}.$$

Φ^\vee is defined to consist precisely of the vectors μ^\vee for $\mu \in \Phi$. A short computation shows that Φ^\vee obeys the same axioms as Φ . It will be interpreted as the root system of the dual group.

However, the definition of μ^\vee depends on the normalization of the metric $(\ , \)$, and therefore the vectors μ^\vee are not really naturally defined as vectors in \mathbb{E} . They actually are more naturally understood as vectors in the dual space \mathbb{E}^\vee . The reason for this claim is that for $\mu, \nu \in \Phi$, the pairing

$$(A.4) \quad \mu^\vee(\nu) = (\nu, \mu^\vee) = \frac{2(\nu, \mu)}{(\mu, \mu)} = \langle \nu, \mu \rangle$$

is independent of the normalization of the metric $(\ , \)$. Hence, while the original root system Φ lies in $\mathbb{E} \cong \mathfrak{t}^\vee$, the dual root system Φ^\vee lies in $\mathbb{E}^\vee \cong \mathfrak{t}$.

The Weyl group \mathcal{W} of Φ is the group generated by the reflections with respect to the roots μ , namely the operations $\nu \rightarrow \nu - \mu\langle \nu, \mu \rangle$. Similarly, the Weyl group \mathcal{W}^\vee of Φ^\vee is the group generated by the reflections with respect to the coroots μ^\vee , namely $\nu \rightarrow \nu - \mu^\vee\langle \nu, \mu^\vee \rangle$. However, as μ^\vee is a real multiple of μ , we have $\mu\langle \nu, \mu \rangle = \mu^\vee\langle \nu, \mu^\vee \rangle$, and the reflections with respect to μ or μ^\vee coincide. Hence the groups \mathcal{W} and \mathcal{W}^\vee likewise coincide. This leads to the fact that G and its dual group ${}^L G$ have the same Weyl group.

Just as the vectors $\mu \in \Phi$ generate a rank r lattice Λ_{rt} , the vectors $\mu^\vee \in \Phi^\vee$ generate a rank r lattice Λ_{cort} (the coroot lattice). We write Λ_{cowt} (the coweight lattice) for the dual to Λ_{rt} and Λ_{wt} (the weight lattice) for the dual of Λ_{cort} . Thus (momentarily regarding all these lattices as embedded in \mathbb{E} , using the metric $(\ , \)$), we have $\nu \in \Lambda_{\text{cowt}}$ if and only if $(\nu, \mu) \in \mathbb{Z}$ for all $\mu \in \Phi$, and $\nu \in \Lambda_{\text{wt}}$ if and only if $(\nu, \mu^\vee) \in \mathbb{Z}$ for all $\mu^\vee \in \Phi^\vee$. Integrality of the pairing (ν, μ^\vee) for $\nu \in \Lambda_{\text{rt}}$, $\mu^\vee \in \Lambda_{\text{cort}}$ means that we have inclusions

$$(A.5) \quad \begin{aligned} \Lambda_{\text{rt}} &\subset \Lambda_{\text{wt}} \\ \Lambda_{\text{cort}} &\subset \Lambda_{\text{cowt}}. \end{aligned}$$

These lattices all have the same rank r , so the inclusions are of finite index.

Now let Λ_{char} (the character lattice) be any lattice that is intermediate between Λ_{rt} and Λ_{wt} :

$$(A.6) \quad \Lambda_{\text{rt}} \subset \Lambda_{\text{char}} \subset \Lambda_{\text{wt}}.$$

The dual of Λ_{char} is a lattice Λ_{cochar} (the cocharacter lattice) that lies between the coroot and coweight lattices:

$$(A.7) \quad \Lambda_{\text{cort}} \subset \Lambda_{\text{cochar}} \subset \Lambda_{\text{cowt}}.$$

Of course, the three lattices in (A.7) are naturally regarded as embedded in E^\vee ; their embedding in E depends on the choice of metric $(\ , \)$.

The classification of compact simple Lie groups states that such groups correspond to character lattices that can arise in a construction of this type. For every choice of Λ_{char} , there is a compact Lie group G such that the highest weight of a representation of G is an element of Λ_{char} . In fact, Λ_{char} is the group of characters, that is of homomorphisms of the maximal torus \mathbb{T} of G to $U(1)$. Thus

$$(A.8) \quad \Lambda_{\text{char}} = \text{Hom}(\mathbb{T}, U(1)),$$

and dually,

$$(A.9) \quad \Lambda_{\text{cochar}} = \text{Hom}({}^L\mathbb{T}, U(1)) = \text{Hom}(U(1), \mathbb{T}).$$

If $\Lambda_{\text{char}} = \Lambda_{\text{wt}}$, then G is simply-connected. If $\Lambda_{\text{char}} = \Lambda_{\text{rt}}$, then G is of adjoint type.

More generally, let $\mathcal{Z}(G)$ denote the center of G . For any locally compact abelian group A , we write A^\vee for $\text{Hom}(A, U(1))$; according to Pontryagin duality, if $B = A^\vee$, then $A = B^\vee$. In an irreducible representation R of G , $\mathcal{Z}(G)$ acts centrally, or in other words via a homomorphism to $U(1)$. The highest weight of R is an element $w \in \Lambda_{\text{char}}$, and $\mathcal{Z}(G)$ acts trivially if $w \in \Lambda_{\text{rt}}$. The quotient $\Lambda_{\text{char}}/\Lambda_{\text{rt}}$ is $\text{Hom}(\mathcal{Z}(G), U(1))$:

$$(A.10) \quad \mathcal{Z}(G)^\vee = \Lambda_{\text{char}}/\Lambda_{\text{rt}}.$$

Using the duality between the inclusions in (A.6) and (A.7), this is equivalent to

$$(A.11) \quad \mathcal{Z}(G) = \Lambda_{\text{cowt}}/\Lambda_{\text{cochar}}.$$

Similarly, an element of the fundamental group $\pi_1(G)$, if lifted to the universal cover \overline{G} of G , determines a path from the identity to a central element ξ of \overline{G} . In a representation R of \overline{G} , ξ acts by multiplication by an element of $U(1)$. By this construction, R determines an element of $\pi_1(G)^\vee$, and this element vanishes if R is actually a representation of G , that is, if its highest weight is in Λ_{char} . This leads to a relation

$$(A.12) \quad \pi_1(G)^\vee = \Lambda_{\text{wt}}/\Lambda_{\text{char}}.$$

Again, the duality between (A.6) and (A.7) leads to an alternative version

$$(A.13) \quad \pi_1(G) = \Lambda_{\text{cochar}}/\Lambda_{\text{cort}}.$$

The Langlands or GNO dual group ${}^L G$ is related to the lattices Λ_{cort} , Λ_{cochar} , and Λ_{cowt} exactly as G is related to Λ_{rt} , Λ_{char} , and Λ_{wt} . All of

the above statements have obvious duals. For example, ${}^L G$ is simply-connected if $\Lambda_{\text{cochar}} = \Lambda_{\text{cowt}}$, and is of adjoint type if $\Lambda_{\text{cochar}} = \Lambda_{\text{cort}}$. And more generally,

$$(A.14) \quad \begin{aligned} \mathcal{Z}({}^L G)^\vee &= \Lambda_{\text{cochar}} / \Lambda_{\text{cort}} \\ \pi_1({}^L G)^\vee &= \Lambda_{\text{cowt}} / \Lambda_{\text{cochar}}. \end{aligned}$$

Comparing (A.14) to (A.11) and (A.13), we learn that

$$(A.15) \quad \begin{aligned} \pi_1({}^L G) &= \mathcal{Z}(G)^\vee \\ \mathcal{Z}({}^L G) &= \pi_1(G)^\vee. \end{aligned}$$

The whole construction is completely symmetric under exchange of E with E^\vee , the lattices in (A.6) with the dual lattices in (A.7), and the group G with the dual group ${}^L G$. In the theory of Lie algebras, the roots of G take values in \mathfrak{t}^\vee (the dual of the Lie algebra \mathfrak{t} of the maximal torus of G) and the coroots in \mathfrak{t} , so we identify E with \mathfrak{t}^\vee and E^\vee with \mathfrak{t} . The results (A.11) and (A.13) have more direct explanations using the fact that the lattices Λ_{cort} , Λ_{cochar} , and Λ_{cowt} are all naturally embedded in \mathfrak{t} .

Choice of Metric

At this point, it is useful to make a convenient choice for the metric $(\ , \)$ on E .

The group G is said to be simply-laced if all roots $\mu \in \Phi$ have the same length squared. This is so precisely if G is of type A , D , or E . In this case, it is convenient to pick a metric such that the roots actually have length squared equal to 2, that is $(\mu, \mu) = 2$ for all μ . Looking back to the definition of μ^\vee in eqn. (A.3), we see that this condition ensures that $\mu^\vee = \mu$ for all μ . This causes the construction summarized above to simplify. The distinction between E and E^\vee can be omitted without any loss of symmetry, and we then have $\Lambda_{\text{rt}} = \Lambda_{\text{cort}}$, $\Lambda_{\text{wt}} = \Lambda_{\text{cowt}}$. In this situation, the groups G and ${}^L G$ have the same Lie algebra, since their root lattices are the same, but they are not necessarily isomorphic, since the lattices Λ_{char} and Λ_{cochar} may differ.

Now let us consider the case that G is not simply-laced. In this case, it is shown via the classification of simple Lie algebras that there are precisely two values for the length squared of a root. We denote the ratio of these two values as $n_{\mathfrak{g}}$; it equals 2 for groups of type B , C , or F_4 , and 3 for G_2 . From $\mu^\vee = 2\mu/(\mu, \mu)$, we find

$$(A.16) \quad (\mu^\vee, \mu^\vee) = \frac{4}{(\mu, \mu)}.$$

So the function (μ^\vee, μ^\vee) again takes two values for $\mu^\vee \in \Phi^\vee$, and these values have the same ratio $n_{\mathfrak{g}}$. Moreover, from (A.16), we see that if μ is a short root, then μ^\vee is a long coroot, and if μ is a long root, then μ^\vee is a short coroot.

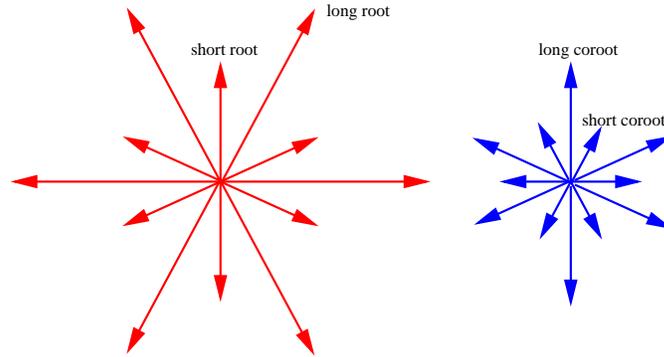


FIGURE 8. Shown here are the roots of the Lie group G_2 (left) and the coroots (right). The map $\mu \rightarrow \mu^*$ (defined using the metric in which short roots have length squared 2) maps a vector on the left to the “same” vector on the right. A short root maps to a long coroot, and a long root maps to 3 times a short coroot. To get an isomorphism between the root diagram and the coroot diagram, we can compose the map $\mu \rightarrow \mu^*$ with rotation by an angle $\pm\pi/6$ and multiplication by $1/\sqrt{3}$.

We obviously cannot now normalize the metric $(\ , \)$ on E so that all roots have length squared 2. However, it is convenient to normalize the metric so that short roots have length squared 2, and therefore long roots have length squared $2n_{\mathfrak{g}}$. When this choice is made, we have

$$(A.17) \quad \mu^\vee = \begin{cases} \mu & \text{if } \mu \text{ is a short root;} \\ \mu/n_{\mathfrak{g}} & \text{if } \mu \text{ is a long root.} \end{cases}$$

Reciprocally,

$$(A.18) \quad \mu = \begin{cases} \mu^\vee & \text{if } \mu \text{ is a short root;} \\ n_{\mathfrak{g}}\mu^\vee & \text{if } \mu \text{ is a long root.} \end{cases}$$

Once we pick the metric, we get a natural identification of E with E^\vee . This gives a map from E to E^\vee , mapping $\nu \in E$ to the linear form on E defined by $\nu(\mu) = (\nu, \mu)$. We denote this map as $\nu \rightarrow \nu^*$.

From (A.18), we can be more precise. If $\mu \in \Phi$ is a short root, then μ^* (which is the same as μ , but interpreted as an element of E^\vee) is equal to a long coroot (namely μ^\vee). But if μ is a long root, then μ^* is $n_{\mathfrak{g}}$ times the short coroot μ^\vee . This is illustrated in fig. 8 for the case of G_2 . In particular, μ^* is not always a coroot but is always an element of the coroot lattice, so the linear transformation $\mu \rightarrow \mu^*$ maps Λ_{rt} into Λ_{cort} .

Of course, there is also a dual of this construction. We can pick on E^\vee a metric in which a short coroot has length squared 2. This leads to a natural map $E^\vee \rightarrow E$, which we denote $\mu^\vee \rightarrow (\mu^\vee)^*$. By applying

the same reasoning, we see that this operation maps a short coroot to a long root, or a long coroot to $n_{\mathfrak{g}}$ times a short root. In particular, it maps Λ_{cort} into Λ_{rt} .

The composition of the two maps is multiplication by $n_{\mathfrak{g}}$, that is $(\mu^*)^* = n_{\mathfrak{g}}\mu$ for all μ . Indeed, a short root maps to a long coroot and thence to $n_{\mathfrak{g}}$ times the original short root, while a long root maps to $n_{\mathfrak{g}}$ times a short root, and thence to $n_{\mathfrak{g}}$ times the original long root.

Dynkin Diagrams

A collection of r roots $\mu_i \in \Phi$ is called a set of simple positive roots if every root μ has an expansion $\mu = \sum_i a_i \mu_i$, with coefficients a_i that are all non-negative or all non-positive. Such a set always exists and is unique up to a Weyl transformation.

If μ_i are a set of simple positive roots of Φ , then the dual roots $\mu_i^\vee \in \Phi^\vee$ are a set of simple positive roots of Φ^\vee . Indeed, the μ_i^\vee are positive multiples of the μ_i , and any dual root μ^\vee is similarly related to μ . So in an expansion $\mu^\vee = \sum_i b_i \mu_i^\vee$, the coefficients b_i have the same signs as the coefficients a_i in $\mu = \sum_i a_i \mu_i$.

The Dynkin diagram of G has a node for every simple positive root. The nodes are connected in a way that encodes the angles among the simple positive roots. When G is not simply-laced, the diagram is also commonly labeled with an arrow that points from long to short roots.

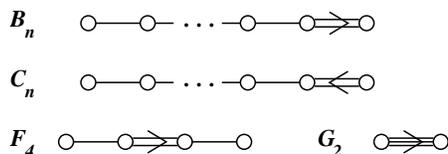


FIGURE 9. Dynkin diagrams of the non-simply-laced Lie groups. Duality has the effect of reversing the arrow that points from long to short roots. Thus, the groups B_n and C_n are exchanged by duality. However, F_4 and G_2 are self-dual, since for those groups, a reversal of the arrow is equivalent to exchanging the two ends of the diagram.

The angles between the simple positive roots μ_i of Φ are the same as the angles between the simple positive roots μ_i^\vee of Φ^\vee , since the μ_i^\vee are positive multiples of the μ_i . However, the transformation $\mu \rightarrow \mu^\vee$ maps short roots to long ones, and vice-versa, as we have seen above. So the root systems Φ and Φ^\vee have Dynkin diagrams that look just the same, except that, in the non-simply-laced case, the arrows point in opposite directions. This is illustrated in fig. 9.

A Levi subgroup \mathbb{L} of G has a Dynkin diagram obtained by simply removing some nodes from the Dynkin diagram of G . ${}^L G$ has a corresponding Levi subgroup ${}^L \mathbb{L}$ whose Dynkin diagram consists of the

“same” nodes. \mathbb{L} and ${}^L\mathbb{L}$ have the same Weyl group for the same reason that G and ${}^L G$ do.

Comparison To Gauge Theory

For simply-laced G , the metric in which the roots have length squared 2 is invariant under duality, since in this metric the roots and coroots coincide.

When G is not simply-laced, if we identify E with its dual using a metric in which the roots have length squared 2 or $2n_{\mathfrak{g}}$, then the coroots have length squared 2 or $2/n_{\mathfrak{g}}$, so this recipe is not invariant under duality. The factor of $n_{\mathfrak{g}}$ in the formula $(\mu^*)^* = n_{\mathfrak{g}}\mu$ reflects this discrepancy, as does the factor of $n_{\mathfrak{g}}$ in the S -duality transformation $\tau \rightarrow -1/n_{\mathfrak{g}}\tau$.

Of course, we could restore duality by simply dividing the metric by $\sqrt{n_{\mathfrak{g}}}$. However, this would lead to awkwardness in other formulas. Indeed, the metric on $\mathfrak{t} \cong E^\vee$ such that a short coroot has length squared 2 is widely used in gauge theory, for example in the present paper beginning in sections 2.4 and 2.5, because it leads to definitions of the instanton number (2.39) and theta-angle (2.38) that are uniformly valid for all G .

In addition, normalizing the metrics as we have done has the virtue that it leads to maps $\mu \rightarrow \mu^*$ and $\mu^\vee \rightarrow (\mu^\vee)^*$ that, although not isomorphisms of the relevant lattices, do map $\Lambda_{\mathfrak{rt}}$ into Λ_{cort} , and vice-versa. This is lost if we divide by $\sqrt{n_{\mathfrak{g}}}$.

Relation Of G and ${}^L G$

If G is simply-laced, then G and ${}^L G$ have the same Lie algebra, and this fact is manifest in the above construction, in the sense that the map $\mu \rightarrow \mu^*$ is an isomorphism from Φ to Φ^\vee .

If G is not simply-laced, the story is more complicated. For G of type B or C , the Lie algebras of G and ${}^L G$ are distinct, as is clear from the diagrams in fig. 9. For G of type G_2 or F_4 , the groups G and ${}^L G$ are actually isomorphic. This is obvious from the Dynkin diagrams; replacing Φ with Φ^\vee has the effect of reversing the arrow that points from long to short roots, but for G_2 and F_4 , this is equivalent to exchanging the two ends of the diagram, as one can see in fig. 9.

For G_2 and F_4 , the isomorphism between the group and the dual group does not simply come from the maps $\mu \rightarrow \mu^*$ and $\mu^\vee \rightarrow (\mu^\vee)^*$, which are not isomorphisms between Φ and Φ^\vee . To get such an isomorphism, one can compose the map $\mu \rightarrow \mu^*$ with a linear transformation $\mathfrak{R} : E^\vee \rightarrow E$. Modulo a Weyl transformation, \mathfrak{R} is determined by the way it must exchange the two ends of the Dynkin diagram. For G_2 , \mathfrak{R} can be chosen to be the composition of a rotation by an angle $\pm\pi/6$ and multiplication by $1/\sqrt{3}$, as one can see in fig. 8. For F_4 , it is the composition of a rotation and multiplication by $1/\sqrt{2}$.

In the physics literature [41], [42], S -duality is often defined so that, in a vacuum in which G is spontaneously broken to an abelian subgroup, the action of \mathfrak{R} is incorporated as part of the definition of S . At the expense of treating G_2 and F_4 as exceptional cases, this makes the self-duality of those groups manifest. We instead prefer, as for example in [71], to treat all groups uniformly, omitting the factor of \mathfrak{R} . In this way, formulas such as (2.22) can be written uniformly for all groups.

Appendix B. Index Of Notation

I. Gauge Theory

M	four-manifold
W	three-manifold
C	Riemann surface on which we do geometric Langlands
Σ	Riemann surface to which we compactify
G	compact gauge group, usually simple
\overline{G}	simply-connected cover of G
E	a G -bundle
A	the gauge field; a connection on E
d_A	the gauge-covariant exterior derivative $d + [A, \cdot]$
ϕ	$\text{ad}(E)$ -valued one-form; the Higgs field
φ	the $(1, 0)$ part of ϕ
\mathcal{A}	the complex-valued connection $A + i\phi$
V	monodromy of \mathcal{A}
U	semi-simple element in closure of orbit of V
τ	the gauge coupling parameter $\theta/2\pi + 4\pi i/e^2$
N	instanton number
m	magnetic flux; characteristic class of a \mathbb{T} -bundle
ξ	characteristic class that obstructs lifting a G -bundle E to a \overline{G} -bundle
D	support of a surface operator
L	support of a line operator
$(\alpha, \beta, \gamma, \eta)$	parameters of a surface operator
\mathbf{e}_0	discrete electric flux
\mathbf{m}_0	discrete magnetic flux
\mathcal{M}_H	moduli space of Higgs bundles
\mathcal{M}	moduli space of G -bundles
$\mathcal{M}_H(\alpha, \beta, \gamma)$	moduli space of ramified Higgs bundles
$\mathcal{M}(\alpha)$	moduli space of parabolic bundles
\mathbf{B}	base of the Hitchin fibration
\mathbf{F}	fiber of the Hitchin fibration
\mathcal{B}	a brane
\mathcal{L}	Chan-Paton line bundle of a brane
\mathcal{G}	the group of gauge transformations

\star	Hodge star operator
Γ	duality group

II. Hyper-Kahler Structure

X	generic hyper-Kahler manifold
I, J, K	complex structures on X
$\omega_I, \omega_J, \omega_K$	Kahler structures on X
$\Omega_I, \Omega_J, \Omega_K$	holomorphic two-forms on X
$\vec{\mu} = (\mu_I, \mu_J, \mu_K)$	hyper-Kahler moment map

III. Group Theory

G	compact Lie group, usually simple
\overline{G}	universal cover of G
G_{ad}	adjoint form of G
L_G	dual group
\mathbb{T}	maximal torus of G
\mathbb{L}	Levi subgroup of G
$G_{\mathbb{C}}, \mathbb{T}_{\mathbb{C}}, \text{etc.}$	complexifications
\mathcal{B}	Borel subgroup of $G_{\mathbb{C}}$
\mathcal{P}	parabolic subgroup of $G_{\mathbb{C}}$
\mathcal{N}	unipotent radical of \mathcal{P}
\mathfrak{g}	Lie algebra of G
\mathfrak{t}	Lie algebra of \mathbb{T}
\mathfrak{t}^{\vee}	dual of \mathfrak{t}
\mathfrak{b}	Lie algebra of \mathcal{B}
\mathfrak{p}	Lie algebra of \mathcal{P}
\mathfrak{n}	Lie algebra of \mathcal{N}
\mathcal{W}	Weyl group
\mathcal{W}_{aff}	affine Weyl group
D	affine Weyl chamber
B_{aff}	affine braid group
$\mathcal{Z}(G)$	center of G
$\pi_1(G)$	fundamental group of G
\mathfrak{c}	conjugacy class in $G_{\mathbb{C}}$
\mathfrak{c}	conjugacy class in $\mathfrak{g}_{\mathbb{C}}$
$\mathfrak{C}_{\mathbb{L}}$	Richardson conjugacy class associated to Levi subgroup \mathbb{L}

IV. Lattices

Λ_{rt}	root lattice
Λ_{cort}	coroot lattice
Λ_{wt}	weight lattice
Λ_{cowt}	coweight lattice
Λ_{char}	character lattice
Λ_{cochar}	cocharacter lattice

References

- [1] R. Langlands, *Problems In The Theory Of Automorphic Forms*, in ‘Lect. Notes in Math.’, **170** (Springer-Verlag, 1970), 18–61; *Where Stands Functoriality Today?* in ‘Representation Theory And Automorphic Forms’, Proc. Symp. Pure Math., **61** (American Mathematical Society, 1997), 457–471.
- [2] A. Beilinson and V. Drinfeld, *Quantization Of Hitchin’s Integrable System And Hecke Eigensheaves*, preprint (ca. 1995), <http://www.math.uchicago.edu/~arinkin/langlands/>.
- [3] E. Frenkel, *Affine Algebras, Langlands Duality, And Bethe Ansatz*, in ‘Proceedings of the XIth International Congress of Mathematical Physics’, Paris, 1994, ed. D. Iagolnitzer (International Press, 1995), 606–642, q-alg/9506003.
- [4] A. Beilinson and V. Drinfeld, *Chiral Algebras*, American Mathematical Society Colloquium Publications, **51**, (American Mathematical Society, 2004).
- [5] E. Frenkel and D. Ben-Zvi, *Vertex Algebras And Algebraic Curves*, Mathematical Surveys And Monographs, **88**, second edition (American Mathematical Society, 2004).
- [6] E. Frenkel and D. Gaitsgory, *Local Geometric Langlands Correspondence And Affine Kac-Moody Algebras*, math.RT/0508382.
- [7] E. Frenkel, *Ramifications Of The Geometric Langlands Program*, math.QA/0611294.
- [8] A. Kapustin and E. Witten, *Electric-Magnetic Duality And The Geometric Langlands Program*, hep-th/0604151.
- [9] E. Frenkel, *Lectures On The Langlands Program And Conformal Field Theory*, arXiv:hep-th/0512172.
- [10] R.M. Rohm, *Some Current Problems In Physics Beyond The Standard Model*, Princeton University Ph.D. Thesis (1985), unpublished.
- [11] P.B. Kronheimer and T.S. Mrowka, *Gauge Theory For Embedded Surfaces*, I, II, *Topology* **32** (1993), 773–826; **34** (1995) 37–97.
- [12] E. Witten, *Topological Sigma Models*, *Commun. Math. Phys.* **118** (1988) 411.
- [13] J. Preskill and L.M. Krauss, *Local Discrete Symmetry And Quantum Mechanical Hair*, *Nucl. Phys.* **B341** (1990) 50–100.
- [14] M.G. Alford, K.-M. Lee, J. March-Russell, and J. Preskill, *Quantum Field Theory Of Non-Abelian Strings And Vortices*, *Nucl. Phys.* **B384** (1992) 251–317, hep-th/9112038.
- [15] M. Bucher, K.-M. Lee, and J. Preskill, *On Detecting Discrete Cheshire Charge*, *Nucl. Phys.* **B386** (1992) 27–42, hep-th/9112040.
- [16] A. Braverman, *Instanton Counting Via Affine Lie Algebras I: Equivariant J-Functions of (Affine) Flag Manifolds and Whittaker Vectors*, math.AG/0401409.
- [17] A. Braverman and P. Etinghof, *Instanton Counting Via Affine Lie Algebras II: from Whittaker Vectors to the Seiberg-Witten Prepotential*, math.AG/0409041.
- [18] N.R. Constable, J. Erdmenger, Z. Guralnik, and I. Kirsch, *Intersecting D3-Branes And Holography*, hep-th/0211222.
- [19] 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.
- [20] J.A. Harvey, G.W. Moore, and A. Strominger, *Reducing S Duality To T Duality*, *Phys. Rev.* **D52** (1995) 7161–7167.
- [21] N. Hitchin, *The Self-Duality Equations On A Riemann Surface*, *Proc. London Math. Soc.* (3) **55** (1987) 59–126.
- [22] C. Simpson, *Harmonic Bundles On Noncompact Curves*, *J. Am. Math. Soc.* **3** (1990) 713–770.

- [23] H. Konno, *Construction Of The Moduli Space Of Stable Parabolic Higgs Bundles On A Riemann Surface*, J. Math. Soc. Japan **45** (1993) 253–276.
- [24] H. Nakajima, *Hyper-Kahler Structures On Moduli Spaces Of Parabolic Higgs Bundles On Riemann Surfaces*, in ‘Moduli Of Vector Bundles’, Lecture Notes in Pure and Appl. Math., **179** (Marcel Dekker, New York, 1996), ed. M. Maruyama.
- [25] O. Biquard and P. Boalch, *Wild Non-Abelian Hodge Theory On Curves*, Compos. Math. **140** (2004) 179–204.
- [26] T. Springer, *A Construction Of Representations Of Weyl Groups*, Invent. Math. **44** (1978) 279–293.
- [27] D. Kazhdan and G. Lusztig, *Proof Of The Deligne-Langlands Conjecture For Hecke Algebras*, Invent. Math. **87** (1987) 153–215.
- [28] D. Kazhdan and G. Lusztig, *Equivariant K-Theory And Representations Of Hecke Algebras, II*, Invent. Math. **80** (1985) 209–231.
- [29] R. Bezrukavnikov, *Noncommutative Counterparts Of The Springer Resolution*, math.RT/0604445.
- [30] N. Chriss and V. Ginzburg, *Representation Theory And Complex Geometry*, Birkhauser, Boston, 1997.
- [31] M.F. Atiyah and R. Bielawski, *Nahm’s Equations, Configuration Spaces, and Flag Manifolds*, Bull. Braz. Math. Soc. (NS) **33** (2002) 157–76, math.RT/0110112.
- [32] D. Ben-Zvi and D. Nadler, to appear.
- [33] E. Witten, to appear.
- [34] P. Kronheimer, *A Hyper-Kahlerian Structure On Coadjoint Orbits Of A Semi-simple Complex Group*, J. London Math. Soc. **42** (1990) 193–208.
- [35] P. Kronheimer, *Instantons And The Geometry Of The Nilpotent Variety*, J. Differential Geom. **32** (1990) 473–490.
- [36] O. Biquard, *Sur les Équations de Nahm et la Structure de Poisson des Algèbres de Lie Semi-Simples Complexes*, Math. Ann. **304** (1996) 253.
- [37] A.G. Kovalev, *Nahm’s Equations And Complex Adjoint Orbits*, Quart. J. Math. Oxford Ser. (2) **47** (1996) 41–58.
- [38] A. Kapustin, *Wilson-’t Hooft Operators In Four-Dimensional Gauge Theories And S-Duality*, hep-th/0501015.
- [39] P. Goddard, J. Nuyts, and D.I. Olive, *Gauge Theories And Magnetic Charge*, Nucl. Phys. **B125** (1977) 1–28.
- [40] R. Bott and L. Tu, *Differential Forms In Algebraic Topology*, Springer, New York, 1982.
- [41] N. Dorey, C. Fraser, T.J. Hollowood and M.A.C. Kneipp, *S-Duality in $N = 4$ Supersymmetric Gauge Theories With Arbitrary Gauge Group*, Phys. Lett. B **383**, 422–428 (1996), arXiv:hep-th/9605069.
- [42] P. Argyres, A. Kapustin, N. Seiberg, *On S-Duality For Non-Simply-Laced Gauge Groups*, arXiv:hep-th/0603048.
- [43] E. Witten, *On S-Duality In Abelian Gauge Theory*, Selecta Math. **1** (1995) 383, hep-th/9505186.
- [44] T.H. Buscher, *Path Integral Derivation Of Quantum Duality In Nonlinear Sigma Models*, Phys. Lett. **201B** (1988) 466.
- [45] M. Rocek and E. Verlinde, *Duality, Quotients, And Currents*, Nucl. Phys. **B373** (1992) 630, hep-th/91110053.
- [46] E. Witten, *Dyons Of Electric Charge $e\theta/2\pi$* , Phys. Lett. **B86** (1979) 283–7.
- [47] G. ’t Hooft, *Magnetic Monopoles In Unified Gauge Theories*, Nucl. Phys. **B79** (1974) 276–284.
- [48] C. Vafa and E. Witten, *A Strong Coupling Test Of S-Duality*, Nucl. Phys. **B431** (1994) 3–77, hep-th/9408074.

- [49] N.J. Hitchin, A. Karlhede, U. Lindstrom, and M. Rocek, *Hyperkahler Metrics and Supersymmetry*, Commun. Math. Phys. **108** (1987) 535–589.
- [50] V.B. Mehta and C.S. Seshadri, *Moduli Of Vector Bundles On Curves With Parabolic Structures*, Math. Ann. **248** (1980) 205–39.
- [51] M.S. Narasimhan and C.S. Seshadri, *Stable And Unitary Vector Bundles On A Compact Riemann Surface*, Ann. Math. (2) **82** (1965) 540–67.
- [52] M.F. Atiyah and R. Bott, *The Yang-Mills Equations Over Riemann Surfaces*, Phil. Trans. Roy. Soc. London **A308** (1982) 523–615.
- [53] S. Kumar and M.S. Narasimhan, *Picard Group Of The Moduli Space Of G-Bundles*, Math. Ann. **308** (1997) 155–173.
- [54] Y. Laszlo and C. Sorger, *The Line Bundles On The Moduli Of Parabolic G-Bundles Over Curves And Their Sections*, alg-geom/9507002.
- [55] C. Pauly, *Fibrés Parabolique de Rang 2 et Fonctions Thêta Généralisées*, Math. Z. **128** (1998) 31–50.
- [56] L. Jeffrey, *The Verlinde Formula For Parabolic Bundles*, math.AG/0003150.
- [57] S. Donaldson, *Gluing Techniques In The Cohomology Of Moduli Spaces*, in ‘Topological Methods In Modern Mathematics’ (Publish or Perish Press, Houston, 1993), 137–170.
- [58] L. Jeffrey, *Extended Moduli Spaces Of Flat Connections On Riemann Surfaces*, Math. Annalen **298** (1994) 667–692.
- [59] E. Witten, *Quantum Field Theory And The Jones Polynomial*, Commun. Math. Phys. **121** (1989) 351.
- [60] H.U. Boden and Y. Hu, *Variations Of Moduli Of Parabolic Bundles*, Math. Annalen **301** (1995) 539–559.
- [61] H.U. Boden and K. Yokogawa, *Moduli Spaces Of Parabolic Higgs Bundles And Parabolic $K(D)$ Pairs*, Internat. J. Math. **7** (1996) 573–598.
- [62] P. Kronheimer, *The Construction Of ALE Spaces As Hyper-Kahler Quotients*, J. Differential Geom. **29** (1989) 665–83.
- [63] R. Bielawski, *Lie Groups, Nahm’s Equations, And Hyper-Kahler Manifolds*, math.DG/0509515.
- [64] W. Nahm, *The Construction Of All Self-Dual Multimonoles By The ADHM Method*, in ‘Monopoles And Quantum Field Theory’, ed. N.S. Craigie et. al. (World Scientific, 1982) 87–95.
- [65] D. Kazhdan and G. Lusztig, *A Topological Approach To Springer’s Representations*, Adv. Math. **38** (1980) 222–228.
- [66] C. H. Taubes, *Self-Dual Connections On 4-Manifolds With Indefinite Intersection Matrix*, J. Differential Geom. **19** (1984) 517–560.
- [67] A. Beauville, *Jacobiennes des Courbes Spectral et Systèmes Hamiltoniens Complètement Intégrables*, Acta Math. **164** (1990) 211–35.
- [68] E. Markman and R. Donagi, *Spectral Curves, Algebraically Completely Integrable Hamiltonian Systems, And Moduli Of Bundles*, in ‘Lecture Notes in Math.’, **1620** (Springer, 1996), 1–119, alg-geom/9507017.
- [69] M. Thaddeus and T. Hausel, *Mirror Symmetry, Langlands Duality, And The Hitchin System*, math.AG/0205236.
- [70] A. Braverman and R. Bezrukavnikov, *Geometric Langlands Correspondence For \mathcal{D} -Modules In Prime Characteristic: The $GL(N)$ Case*, math.AG/0602255.
- [71] R. Donagi and T. Pantev, *Langlands Duality For Hitchin Systems*, math.AG/0604617.
- [72] 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.
- [73] M. Kontsevich, *Deformation Quantization Of Algebraic Varieties*, math.AG/0106006.

- [74] A. Beilinson and J. Bernstein, *A Proof of Jantzen Conjectures*, Advances in Soviet Mathematics, **16**, Part 1, pp. 150, AMS, 1993.
- [75] D. Gaiitsgory, *Construction Of Central Elements In The Affine Hecke Algebra Via Nearby Cycles*, math.AG/9912074.
- [76] S. Coleman, *More About The Massive Schwinger Model*, Annals Phys. **101** (1976) 239.
- [77] N. Nitsure, *Cohomology Of The Moduli Of Parabolic Vector Bundles*, Proc. Indian Acad. Sci. Math. Sci. **95** (1986) 61–77.
- [78] A. Kapustin and S. Sethi, *The Higgs Branch Of Impurity Theories*, Adv. Theor. Math. Phys. **2** (1998) 571–591, hep-th/9804027.
- [79] O. DeWolfe, D.Z. Freedman, and H. Ooguri, *Holography And Defect Conformal Field Theories*, Phys. Rev. **D66** (2002) 025009, hep-th/0111135.
- [80] M. Douglas and G. Moore, *D-Branes, Quivers, And ALE Instantons*, hep-th/9603167.
- [81] E. Witten, *Some Comments On String Dynamics*, in I. Bars et. al., eds. (1996), ‘Future Perspectives In String Theory’, hep-th/9507121.
- [82] P.Z. Kobzok and A.F. Swann, *Classical Nilpotent Orbits As Hyper-Kahler Quotients*, Int. J. Math. **7** (1996) 193–210.
- [83] L. Dixon, J. Harvey, C. Vafa, and E. Witten, *Strings On Orbifolds*, Nucl. Phys. **B261** (1985) 678–686.
- [84] P. Aspinwall, *Enhanced Gauge Symmetries And K3 Surfaces*, hep-th/9507012.
- [85] J. Maldacena, *The Large N Limit Of Superconformal Field Theories And Supergravity*, Adv. Theor. Math. Phys. **2** (1998) 231–252, hep-th/9711200.
- [86] J. Maldacena, *Wilson Loops In Large N Field Theories*, Phys. Rev. Lett. **80** (1998) 4859–4862, 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] D. Berenstein, R. Corrado, W. Fischler, and J.M. Maldacena, *The Operator Product Expansion For Wilson Loops And Surfaces In The Large N Limit*, Phys. Rev. **D59** (1999) 105023, hep-th/9809188.
- [89] C.R. Graham and E. Witten, *Conformal Invariance Of Submanifold Observables In AdS-CFT Correspondence*, Nucl. Phys. **B546** (1999) 52–64, hep-th/9901021.
- [90] J.E. Humphreys, *Introduction To Lie Algebras And Representation Theory* (Springer-Verlag, New York, 1972).
- [91] B. Simon, *Representations Of Finite And Compact Groups* (American Mathematical Society, 1996).

DEPARTMENT OF PHYSICS, UNIVERSITY OF CALIFORNIA,
 SANTA BARBARA, CA 93106
E-mail address: `gukov@physics.ucsb.edu`

SCHOOL OF NATURAL SCIENCES, INSTITUTE FOR ADVANCED STUDY,
 PRINCETON, NJ 08540
E-mail address: `witten@ias.edu`