Toward a Nonperturbative Topological String

A thesis presented
by

Andrew Neitzke

to
The Department of Physics
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Physics

Harvard University
Cambridge, Massachusetts
May 2005

Thesis advisor | Author

Cumrun Vafa | **Andrew Neitzke**

Toward a Nonperturbative Topological String

Abstract

We discuss three examples of nonperturbative phenomena in the topological string. First, we consider the computation of amplitudes in $\mathcal{N} = 4$ super Yang-Mills theory using the B model topological string as proposed by Witten. We give an argument suggesting that the computations using connected or disconnected D-instantons of the B model are in fact equivalent. Second, we formulate a conjecture that the squared modulus of the open topological string partition function can be defined nonperturbatively as the partition function of a mixed ensemble of BPS states in $d = 4$. This conjecture is an extension of a recent proposal for the closed topological string. In a particular example involving a non-compact Calabi-Yau threefold, we show that the conjecture passes some basic checks, and that the square of the open topological string amplitude has a natural interpretation in terms of 2-dimensional Yang-Mills theory, again generalizing known results for the closed string case. Third, we discuss an action for an abelian two-form gauge theory introduced by Hitchin which describes variations of G_2 structures in seven dimensions. Upon reducing to six dimensions this action splits into two pieces, one related to the complex structure and one related to the symplectic structure; we argue that these two pieces are related to the A and B model topological string theories. In this sense Hitchin's gauge theory is a candidate for a "topological M-theory" in seven dimensions. We also note that upon reduction to lower dimensions this two-form gauge theory naturally reduces to gauge theory descriptions of lower-dimensional gravity theories.

Contents

Acknowledgments

I am grateful to many people without whom this thesis could never have been written.

First, it goes without saying that I owe thanks to all the people with whom I have had the good fortune to collaborate during my time at Harvard. Their insights and their patience have been invaluable, and I have gained immeasurably from our repeated discussions and struggles to understand what we were talking about. Here I mean to include in particular Mina Aganagic, Robbert Dijkgraaf, Sergei Gukov, Luboš Motl and Cumrun Vafa, with whom I wrote the papers which are essentially reproduced in Chapters 2-4 of this thesis, as well as other papers not included here; but I also want to include Amer Iqbal and Mohammed Sheikh-Jabbari, with whom I collaborated on papers not included here.

I also owe an enormous debt to my advisor, Cumrun Vafa. His rare capacity to see to the heart of virtually every problem we encountered, and his unique understanding of the physics of the topological string, were instrumental in getting us past numerous roadblocks. In addition to physical insight, he also has a very good idea of how to treat a graduate student — he was willing to let me stray on my own when I wanted to, but when I needed advice he was always available. I am also grateful to him for his unhesitating willingness to take me on while I was still enrolled as a student in the Department of Mathematics.

Discussions with the rest of the high energy theory group were also crucial on many occasions, not to mention enormously enjoyable. Here I particularly want to thank Nima Arkani-Hamed and Shiraz Minwalla, two incredibly busy people who nevertheless always found the time to share their insights when I needed them, both on physics and on life as a physicist. Andy Strominger was also extremely helpful, and I also benefited from discussions with my fellow graduate students, particularly Matt Headrick, Daniel Jafferis, Alex Maloney, Marcus Spradlin, Anastasia Volovich, Martijn Wijnholt, and Xi Yin. I also want to thank all of the speakers at the Wednesday night postdoc seminar; we all learned a lot from their suffering.

Of course, all these wonderful conversations did not take place in a vacuum; none of them would have been possible without the continuous support of the Department of Physics and the theory group. In particular I want to thank Nancy Partridge and Rob Meyer, who were always ready to help, even when I began making utterly unreasonable requests that they send out a near-endless stream of letters, each with its own peculiar handling instructions. I also want to thank Joseph Harris in the Department of Mathematics for being utterly reasonable in handling my slightly unusual situation as a student in one department with an advisor in another, and Shing-Tung Yau for advising me, both while I was still in his department and after I moved across the road.

I owe special thanks to Marlys Fassett, Jody Kelman, Supinda Bunyavanich and Bob Griffin, for their continuous support and encouragement during the times when prospects looked bleak. Finally, I thank my parents, for their constant love and support, for always emphasizing that they would be proud of me even if topological M-theory did not quite work out, and for helping me keep things in perspective.

Chapter 1

Introduction and summary

The perturbative topological string can be viewed from several different perspectives. On the one hand it is a string theory in its own right, defined on target spaces with a rather special structure, which happens to have a very restricted set of observables and is (at the moment) mathematically more tractable than the physical superstring. On the other hand, the perturbative topological string can be viewed as an embedded subsector of the superstring, in the sense that the integral over moduli of Riemann surfaces which one does to compute any particular correlator in the perturbative string corresponds directly to a computation of some correlator in the superstring. In this identification it turns out that the topological string on M is related to the Type II superstring on $M \times \mathbb{R}^{3,1}$, with the topological string coupling related to the self-dual part of the graviphoton field strength in the physical string theory, $g_{\mathrm{top}} = F_+$; schematically, if we write the topological string free energy as $\mathcal{F}$, and the vector multiplet moduli of M as X, there is a term in the $\mathcal{N} = 2$ effective action of the form

$$\int \mathrm{d}^4 x \int \mathrm{d}^4\theta \, \mathcal{F}(g_{\mathrm{top}} = F_+, X). \tag{1.1}$$

Through this term, the perturbative topological string turns out to contain a lot of information about quantities in the superstring which are in some sense "BPS saturated." A recent review of this embedding and some of its consequences has appeared in [95].

On the other hand, there is by now fairly strong evidence that the perturbative superstring is the $g_s \to 0$ approximation of a theory which makes sense even at finite values of g_s. Given the embedding we just discussed, it is natural to ask whether the same might be true of the topological string. This is related (perhaps equivalent) to asking whether the physical string theory makes sense at finite values of F_+. No comprehensive answer to this question has yet emerged, but there are indications that the topological string indeed exhibits some nonperturbative phenomena. This thesis consists of an exploration of some of these phenomena.

Chapter 2: A computation in the twistor-string

A first indication that the nonperturbative topological string ought to make sense comes from Witten's twistor-string theory, proposed in [133]. That paper conjectures a correspondence between the perturbative $\mathcal{N} = 4$ super Yang-Mills theory in $d = 4$ and a string theory in twistor space $\mathbb{CP}^{3|4}$. More precisely, Witten conjectured that the string theory in question is the topological B model, and more precisely still, it is the *nonperturbative* physics of the B model which plays a crucial role: the perturbative B model corresponds to the *self-dual* sector of the $\mathcal{N} = 4$ super Yang-Mills theory, and contributions from D1-string instantons enhance it to the full theory including both helicities.

In Chapter 2 we consider a puzzle which arose in this context: which sorts of D1-string instantons should be included in the relation to $\mathcal{N} = 4$ super Yang-Mills? Do we consider only connected instantons, only disconnected instantons, or all of them? In fact, we give an argument which suggests that — at least for tree level amplitudes — considering disconnected instantons could give the *same* result as considering connected ones. The argument runs basically as follows: to evaluate a scattering amplitude one has to do a kind of contour integral over the moduli of the instantons under consideration. By Cauchy's theorem this integral can be localized on the divisor where the integrand has poles. Irrespective of whether we study connected or disconnected instantons, we see that there is a pole corresponding to degenerate configurations where the D-instantons form a "tree" of lines in $\mathbb{CP}^{3|4}$, so contour integration can be reduced to integrating the residue over the moduli of such trees. This residue turns out to be independent of whether we started with the connected or disconnected instantons.

This argument can be viewed as an ingredient in understanding how to compute in the nonperturbative topological B model string theory on $\mathbb{CP}^{3|4}$. However, it has not yet been converted into a rigorous proof of equivalence between the two computations of the tree level Yang-Mills amplitudes, because the words "a kind of contour integral" in the previous paragraph need to be given a definite meaning. At tree level, there are now rigorous formulas for the Yang-Mills amplitudes which morally represent the result of this contour integration (given in [107] for the connected instantons and [31] for the disconnected ones), but nobody has yet obtained these formulas rigorously from an integral over moduli space. The problem of defining this integral precisely can be viewed as the problem of defining the nonperturbative B model in this context.

The arguments given in this chapter were obtained in collaboration with Sergei Gukov and Luboš Motl. The text has appeared previously in

> S. Gukov, L. Motl and A. Neitzke. "Equivalence of twistor prescriptions for super Yang-Mills," `http://arxiv.org/pdf/hep-th/0404085`.

I also want to acknowledge Michal Fabinger, Peter Svrček, Cumrun Vafa, Anastasia Volovich and Edward Witten for discussions related to the material in this chapter.

Chapter 3: BPS microstates and the open topological string wave function

A second indication that the topological string might make sense beyond perturbation theory comes from a rather different direction, namely the conjecture of [98]. This paper argues that the squared modulus of the partition function of the topological string partition function on a Calabi-Yau threefold M, $|Z_{\mathrm{top}}|^2$, can be interpreted as computing the partition function of a particular ensemble of BPS black hole states in the Type II superstring on $M \times \mathbb{R}^{3,1}$. Because this relation involves the value of the partition function at a particular finite value of the moduli and couplings (essentially fixed by the attractor mechanism of $d = 4$, $\mathcal{N} = 2$ supergravity) it is essentially nonperturbative in character, and one could argue that it really *defines* what one should mean by the nonperturbative topological string.

This gives a candidate definition for $|Z_{\mathrm{top}}|^2$, but *not* for Z_{top} itself! This is presumably related to the fact that if one wants to turn on a finite graviphoton field strength

F in Minkowski signature one has to turn on both F_+ and F_-, with $F_- = \overline{F}_+$, so that in the $\mathcal{N} = 2$ effective action one gets not just (1.1) but also its complex conjugate. Indeed, in one example studied in [124, 10], one sees that Z_{top} seems to make sense only in perturbation theory around $g_{\mathrm{top}} = 0$, while $|Z_{\mathrm{top}}|^2$ has an extension to finite values of g_{top}; namely, it is the partition function of a 2-dimensional Yang-Mills theory,

$$|Z_{\mathrm{top}}|^2 = Z_{\mathrm{YM}}. \tag{1.2}$$

In Chapter 3 we explore the possibility of extending this conjecture to give a nonperturbative definition of the open topological string. By reconsidering the same example studied in [124, 10], we find that such an extension seems to exist; namely, we find that if we introduce branes in the topological string theory, the resulting $Z_{\mathrm{top}}^{\mathrm{open}}$ still satisfies

$$|Z_{\mathrm{top}}^{\mathrm{open}}|^2 = Z_{\mathrm{YM}}^{\mathrm{open}}, \tag{1.3}$$

where $Z_{\mathrm{YM}}^{\mathrm{open}}$ represents a 2-dimensional Yang-Mills partition function with a particular observable inserted. This then gives a nonperturbative completion of the square of the open topological string in this particular case. (More precisely, we give this completion only for a fixed value of the real part of the open string moduli — in other words, we consider branes wrapped on a particular Lagrangian cycle L which is not allowed to move, and give a completion of the dependence of the partition function on g_s and on a Wilson line around L.)

It is natural to expect that by an extension of the logic of [124, 10] this $Z_{\mathrm{YM}}^{\mathrm{open}}$ will also have an interpretation as counting BPS states, so at least in this case we would find that $|Z_{\mathrm{top}}^{\mathrm{open}}|^2$ counts BPS states; using the wave function property of the open topological string we in fact formulate a more general conjecture, in the spirit of [98]. The BPS states in question are states in a $1+1$ dimensional gauge theory obtained by wrapping D4-branes over a Lagrangian subspace of M. This conjecture is necessarily more tentative than that of [98], because we are lacking the analog of the attractor mechanism for these BPS states, so we have no spacetime argument for what their entropy should be.

The results in this chapter were obtained in collaboration with Mina Aganagic and Cumrun Vafa. The text has appeared previously in

> M. Aganagic, A. Neitzke, and C. Vafa. "BPS microstates and the open topological string wave function," `http://arxiv.org/pdf/hep-th/0504054`.

I also want to acknowledge Jacques Distler, Noam Elkies, Sergei Gukov, Marcos Mariño, Shiraz Minwalla, Luboš Motl, Hirosi Ooguri and Natalia Saulina for discussions related to the material in this chapter.

Chapter 4: Topological M-theory?

The last chapter of this thesis is somewhat more speculative than the first two. It concerns the possibility that the nonperturbative formulation of the topological string might be similar to that of the physical string, in the sense that the target space develops an extra dimension whose size is related to the coupling constant. We focus on the critical case, in which the target space is a Calabi-Yau threefold; in that case, it is natural to guess that after the extra dimension grows we should get a theory describing variations of G_2 structures. In Chapter 4 we collect some ideas and observations about the form such a theory might take, and how it fits into the broader context of theories of gravity where the basic fields are p-forms.

As with physical M-theory, we do not expect to understand topological M-theory directly from a microscopic perspective; instead we set our sights somewhat lower and look for the 7-dimensional analogue of the 11-dimensional supergravity action. This should be a theory for which the classical solutions are metrics of G_2 holonomy. Luckily, a natural candidate action has been provided by Hitchin in [74], which describes a kind of 2-form gauge theory in which the action is a complicated nonlinear function of the 3-form field-strength Φ. At the critical points in a fixed flux sector one finds that Φ is the associative 3-form of a G_2 holonomy metric. Furthermore, if we take the G_2 manifold to be of the form $X = M \times S^1$, the action on M roughly reduces to a sum of two pieces, one having to do with the holomorphic 3-form on M and one having to do with the symplectic structure. We argue that, at least classically, these two pieces are related to the B model and the A model topological strings on M respectively; they seem to correspond to reformulations of those theories which are naturally adapted to the problem of counting black hole states. In the B model case the reformulation that appears is essentially the one described in [98] which gives the squared modulus of the partition function. In the A model case the situation is less clear, but the action which one gets seems to be related to the "dual" description of that theory in terms of a $U(1)$ gauge theory in six dimensions, described in [80].

One surprising feature, essentially already noted in [74], is that in the Hamiltonian quantization of the seven-dimensional theory the A model and B model degrees of freedom show up as canonically conjugate variables. It is natural to suppose that this could be related to the S-duality of topological strings discussed in [94, 96, 83, 17]. At the moment, however, this is only a speculation; indeed, we have not even written a precise statement about the relation between the partition functions in seven dimensions and in six.

The ideas in this chapter were developed in collaboration with Robbert Dijkgraaf, Sergei Gukov and Cumrun Vafa. The text has appeared previously in

> R. Dijkgraaf, S. Gukov, A. Neitzke, and C. Vafa. "Topological M-theory as Unification of Form Theories of Gravity," `http://arxiv.org/pdf/hep-th/0411073`.

I also want to acknowledge Michael Atiyah, Jan de Boer, Robert Bryant, Claude LeBrun, Jan Louis, Hirosi Ooguri, Martni Roček, Lee Smolin, Cliff Taubes, Erik Verlinde, and Shing-Tung Yau for discussions related to the material in this chapter.

After the text of this chapter appeared on the Web archive, Vasily Pestun and Edward Witten [103] performed a one-loop test of the proposed relation discussed in Section 4.5.2 between the B model partition function and the holomorphic volume functional in six dimensions. They found that in order to obtain agreement one has to replace the holomorphic volume functional by an "extended" version which includes some additional fields. These additional fields can be understood as describing variations of *generalized* complex structures in the sense of [75, 65]. In retrospect this modified version of the conjecture is more natural than the one we originally proposed; indeed, the B model contains observables describing variations of these generalized complex structures [84]. It is likely that the conjecture in seven dimensions should be similarly modified, so that one considers generalized G_2 manifolds in the sense of [128] rather than ordinary ones.

Chapter 2

Equivalence of twistor prescriptions for super Yang-Mills

2.1 Introduction

Recently in [133] Witten proposed a new approach to perturbative gauge theories in four dimensions which, among other things, implies remarkable regularities in the perturbative scattering amplitudes of $\mathcal{N} = 4$ super Yang-Mills and leads to new ways of computing them. The scattering amplitudes in question depend on the momentum and polarization vectors of the external gluons, and are devilishly difficult to compute using the standard Feynman diagram techniques. For example, even computing a tree level amplitude with 4 external gluons of positive helicity and 3 gluons of negative helicity (such an amplitude will be denoted $\mathcal{A}_{[++++---]}$) requires summing over hundreds of different diagrams!

According to the conjecture of [133], perturbative $\mathcal{N} = 4$ super Yang-Mills theory can be described as a string theory in twistor space $\mathbb{CP}^{3|4}$. In this reformulation, the Yang-Mills scattering amplitudes are given by certain integrals over moduli spaces of holomorphic curves in $\mathbb{CP}^{3|4}$, which can be interpreted as D1-brane instantons. More precisely, for a tree level process involving q negative helicity gluons, the amplitude is given by an integral over moduli of curves of total degree d, where

$$d = q - 1. \tag{2.1}$$

For example, the simplest non-vanishing amplitude with $q = 2$ gluons of negative helicity[1] — the so-called maximally helicity violating (MHV) amplitude [101, 90] — can be computed by integrating over the moduli space of degree 1 curves in $\mathbb{CP}^{3|4}$ [133].

However, when one considers the next simplest case, $q = 3$, there is a puzzle. In the prescription of [133] this amplitude seems to involve a sum over two distinct contributions: one from an integral over connected degree 2 curves, and another from an integral over disconnected pairs of degree 1 curves; see Figure 2.1. Surprisingly, in the case of $\mathcal{A}_{[++---]}$, it was found that the contribution from connected degree 2 curves alone gives the full Yang-Mills amplitude, at least up to a multiplicative constant [106]. This computation was extended to all googly [108] and some non-MHV [107] amplitudes, again with the surprising result that connected degree d curves already account for the full Yang-Mills amplitude, without adding any disconnected curves.

On the other hand, there is some evidence that these tree level amplitudes can also be computed by considering only the contribution of curves which are "maximally

[1] We follow the conventions of [133] where a n-gluon scattering amplitude is called MHV if $n-2$ external gluons have positive helicity, and $\overline{\text{MHV}}$ (or "googly") if $n-2$ gluons have negative helicity.

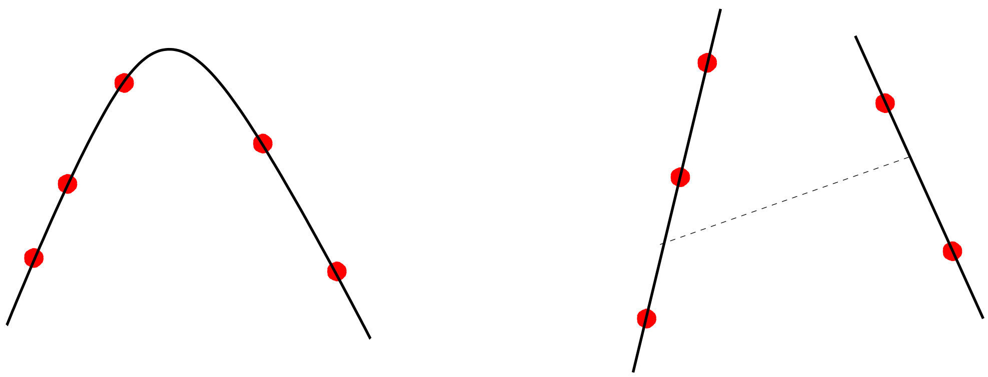

Figure 2.1: An instanton contribution: **(a)** from a connected curve of degree 2; **(b)** from a pair of degree 1 curves. The dotted line represents a propagator in holomorphic Chern-Simons theory.

disconnected," namely, they consist of d distinct degree 1 lines. Since degree 1 curves are associated with MHV amplitudes, this result suggests an alternative method of computing generic tree amplitudes from graphs with MHV vertices [31]. The number v of vertices is determined by the number of gluons with negative helicity; it is actually equal to the degree (2.1),

$$v = q - 1. \tag{2.2}$$

This approach leads to a spectacular simplification of the computations. For example, the 7-gluon amplitude $\mathcal{A}_{[++++---]}$ mentioned earlier can be computed using only 8 diagrams with MHV vertices. However, it also leads to a puzzle.

As we just discussed, the evidence so far in the literature suggests that rather than one prescription for Yang-Mills amplitudes there are at least two: one involving connected curves only, another involving maximally disconnected ones. We will refer to these as the "connected prescription" and the "disconnected prescription" respectively. These different prescriptions have so far not been related directly. In a sense, they seem to have complementary virtues: the connected prescription expresses the whole amplitude as a single integral, and from this form it is easier to prove some properties of the amplitude, such as the parity symmetry; on the other hand, the disconnected prescription leads to concrete and immediately useful formulas for the tree level amplitudes.

The purpose of this note is to argue that the connected and disconnected prescriptions are equivalent, at least for an appropriate choice of the integration contours, and to give an *a priori* explanation for this agreement. The explanation is that, in both prescriptions, the integral over the moduli space is localized to poles on a particular submoduli space. This submoduli space parameterizes configurations of intersecting degree 1 curves.

Let us illustrate this explanation in the simplest case of degree 2 curves. We have two different moduli spaces, $\overline{\mathcal{M}_{0,n,2}}$ and $\mathcal{M}_{\text{lines}}$, parameterizing respectively connected degree 2 curves in $\mathbb{CP}^{3|4}$ and disconnected pairs of lines in $\mathbb{CP}^{3|4}$, and integrands ω_{conn} and ω_{disc} on the two spaces (we will review the construction of these integrands in Section 2.2). Our job is to explain the equality

$$\int_{\overline{\mathcal{M}_{0,n,2}}} \omega_{\text{conn}} = \int_{\mathcal{M}_{\text{lines}}} \omega_{\text{disc}}. \tag{2.3}$$

The explanation begins by noting that both $\mathcal{M}_{\text{lines}}$ and $\overline{\mathcal{M}_{0,n,2}}$ contain a codimension-one "degeneration locus" $\mathcal{M}_{\text{int}}$ parameterizing the moduli of pairs of intersecting lines in $\mathbb{CP}^{3|4}$. In the case of $\mathcal{M}_{\text{lines}}$ we get such a degenerate configuration just by taking two lines in $\mathbb{CP}^{3|4}$ which happen to intersect. For $\overline{\mathcal{M}_{0,n,2}}$ we get such a degeneration by considering a hyperbola $xy = C$ in the limit $C \to 0$, appropriately embedded in $\mathbb{CP}^{3|4}$. The crucial point is that both ω_{conn} and ω_{disc} turn out to have a simple pole along $\mathcal{M}_{\text{int}}$, and furthermore the residue is the same in both cases.[2] Therefore, provided that the integration contours on $\mathcal{M}_{\text{lines}}$ and $\overline{\mathcal{M}_{0,n,2}}$ are chosen compatibly (so that they both encircle $\mathcal{M}_{\text{int}}$ and reduce to the same contour along it), the desired agreement follows.

The argument for general degree d proceeds along similar lines. In the moduli space $\overline{\mathcal{M}_{0,n,d}}$ we find a pole where a degree d curve degenerates into two intersecting curves of degrees d_1 and d_2; the integral over $\overline{\mathcal{M}_{0,n,d}}$ localizes to this sublocus; then inside this sublocus there is a pole where one of the two curves degenerates further, and so on until we reduce finally to the moduli space $\mathcal{M}_{\text{int}}$ of connected trees built from degree 1 curves. On the other hand, the integral over $\mathcal{M}_{\text{lines}}$ also reduces to the same $\mathcal{M}_{\text{int}}$, because the propagators connecting the different lines have poles when the lines intersect. Furthermore it turns out that the integrands on $\mathcal{M}_{\text{int}}$ coming from the two prescriptions are proportional. This establishes the agreement between these two prescriptions, again provided that the contours are chosen appropriately, and up to an overall constant which we do not fix.

This iterative argument pays a surprising dividend: for any $K = 0, \ldots, d-1$, we can define an "intermediate prescription," in which we integrate over configurations of $K + 1$ curves with total degree d. We will show that all of these intermediate prescriptions agree with the connected and disconnected prescriptions. They can also be understood diagrammatically: one sums over tree diagrams with $K + 1$ vertices, where each vertex is decorated with a degree. In these notations, vertices of degree 1 are the MHV vertices of [31], whereas vertices with $d > 1$ could be called "non-MHV

[2]We learned of the possibility of such an explanation from Edward Witten.

vertices". These intermediate prescriptions deserve further study.

For other recent work on the twistor string approach to Yang-Mills, see [106, 108, 107, 20, 21, 134] for the connected prescription, [31, 135, 55] for the disconnected prescription, and [13, 96, 94] for related topics.

2.1.1 Notation and moduli spaces

We always consider scattering amplitudes of n external gluons associated with the particular trace factor $\mathrm{Tr}\,(T_1 T_2 \ldots T_n)$.

We use a coordinate representation for the super twistor space $\mathbb{C}^{4|4}$. We unify the bosonic and fermionic indices into a superspace index $\mathbb{A}$ taking values in

$$\mathbb{A} \in \{1,\, 2,\, 3,\, 4\,|\,1', 2', 3', 4'\}. \tag{2.4}$$

The components of all objects with bosonic values of the superspace index are commuting, while components with fermionic (primed) values of the superspace index are anticommuting. The coordinates on the super twistor space will be denoted by $Z^{\mathbb{A}}$, which are related to the coordinates in the literature by

$$(Z^1, Z^2, Z^3, Z^4 | Z^{1'}, Z^{2'}, Z^{3'}, Z^{4'}) = (\lambda^1, \lambda^2, \mu^1, \mu^2 | \psi^1, \psi^2, \psi^3, \psi^4) \in \mathbb{C}^{4|4}. \tag{2.5}$$

We will also be considering various moduli spaces of curves in $\mathbb{CP}^{3|4}$ with marked points. We use the standard notation

$$\mathcal{M}_{0,n,d}(\mathbb{CP}^{3|4}) \tag{2.6}$$

for the moduli space of "genus 0, n-pointed curves of degree d in $\mathbb{CP}^{3|4}$." This moduli space has dimension $(4d+n)|(4d+4)$. As in [133] we realize it as the space of

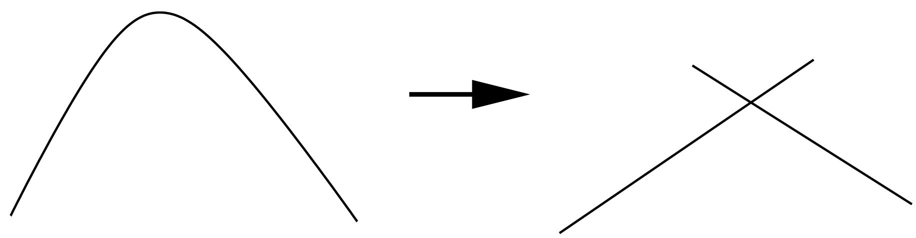

Figure 2.2: A curve of degree 2 can degenerate into a pair of intersecting lines.

automorphism classes of maps $\mathbb{CP}^1 \to \mathbb{CP}^{3|4}$, of degree d, with n marked points on $\mathbb{CP}^1$. Since the target space is always $\mathbb{CP}^{3|4}$ in this chapter, sometimes we abuse notation and write simply $\mathcal{M}_{0,n,d}$.

We will be interested in integrating over $\mathcal{M}_{0,n,d}(\mathbb{CP}^{3|4})$, so we need to understand the properties of this moduli space. First, $\mathcal{M}_{0,n,d}(\mathbb{CP}^{3|4})$ is non-compact, due to certain degenerations that a degree d curve with n marked points can have which are not simply described by a map $\mathbb{CP}^1 \to \mathbb{CP}^{3|4}$. One type of degeneration that will be important below is when a curve develops a node, *i.e.* splits into two components. There is a standard way of incorporating these degenerate curves into our moduli space of maps; one then obtains a larger compact space $\overline{\mathcal{M}_{0,n,d}}(\mathbb{CP}^{3|4})$, called the "moduli space of stable maps." This moduli space is a smooth algebraic variety, except for certain orbifold points which will not play an important role in this chapter.[3]

In particular, the "boundary" of this moduli space,

$$\overline{\mathcal{M}_{0,n,d}}(\mathbb{CP}^{3|4}) \setminus \mathcal{M}_{0,n,d}(\mathbb{CP}^{3|4}), \tag{2.7}$$

contains a codimension 1 divisor which parameterizes curves which have split into

[3]Strictly speaking this theorem has been proven when the target space is $\mathbb{CP}^3$ [53], not for the supermanifold $\mathbb{CP}^{3|4}$, but we do not expect any important differences.

two components. Similarly, for any K there is a subspace $\mathcal{M}_{\rm int}^K$ of codimension K that parameterizes reducible curves with K nodes, *i.e.* curves which have split up into $K+1$ intersecting components which intersect in a tree. This $\mathcal{M}_{\rm int}^K$ can be further decomposed into irreducible pieces,

$$\mathcal{M}_{\rm int} = \bigcup_{\Gamma} \mathcal{M}_{\rm int}^{\Gamma}, \tag{2.8}$$

where the different Γ label different shapes of the tree, together with different decompositions of d into individual degrees $\{d_i\}$, $\quad i = 1, 2, \ldots K+1, \quad d_i \geq d_{i+1}$, and different ways in which the n marked points can be distributed over the $K+1$ components. Some of these $\mathcal{M}_{\rm int}^{\Gamma}$ will play an important role in our discussion below.

2.2 Review of connected and disconnected prescriptions

Suppose we want to use the twistor prescription of [133] to evaluate a Yang-Mills amplitude with $q = d+1$ negative helicity gluons. All contributions to this amplitude are expected to involve holomorphic curves of total degree d, but *a priori* these can be either connected or disconnected. In this section we review the contributions which would be expected from the two most extreme cases: connected degree d curves and completely disconnected families of d degree 1 curves.

In both cases we will consider the Yang-Mills amplitude with arbitrary external scattering states. Via the Penrose transform these scattering states are described by twistor space wavefunctions,[4] which are $\overline{\partial}$-closed $(0,1)$ forms ϕ_i $(i = 1, \ldots, n)$ on

[4]Actually, the wavefunctions are not defined on all of $\mathbb{CP}^{3|4}$, but this distinction will not be

$\mathbb{CP}^{3|4}$. We always treat these ϕ_i as generic. In our computation, we will be focusing on poles which arise in integrals over moduli spaces of curves; we emphasize that the poles in question never come from the ϕ_i.

The prescriptions as we write them below are not gauge invariant. To make the amplitudes gauge invariant we would probably have to include additional diagrams in both prescriptions, involving cubic Chern-Simons interaction vertices. Nevertheless, both prescriptions make sense provided we choose a specific gauge for the gauge field, such as an axial gauge. In this gauge one expects that the cubic vertices do not contribute [133].[5]

2.2.1 Connected prescription

We first review the connected prescription for computation of n-point Yang-Mills amplitudes. The amplitude is obtained as an integral over degree d maps

$$P : \mathbb{CP}^1 \to \mathbb{CP}^{3|4}. \tag{2.9}$$

Such a map P can be written explicitly, in terms of the inhomogeneous coordinate σ on $\mathbb{CP}^1$, as

$$P^{\mathbb{A}}(\sigma) = Z^{\mathbb{A}} = \sum_{k=0}^{d} \beta_k^{\mathbb{A}} \sigma^k \tag{2.10}$$

The supermoduli of the degree d map P are $\beta_k^{\mathbb{A}}$; these span a space $\mathbb{C}^{4d+4|4d+4}$, which comes equipped with the natural measure

$$\mu_d = \prod_{k,\mathbb{A}} \mathrm{d}\beta_k^{\mathbb{A}}. \tag{2.11}$$

important for us.

[5]We thank Peter Svrček for reminding us of this point.

We also have a holomorphic n-form on $(\mathbb{CP}^1)^n$ given by the free-fermion correlator,

$$\omega(\sigma_1,\ldots,\sigma_n)=\prod_{i=1}^{n}\frac{\mathrm{d}\sigma_i}{\sigma_i-\sigma_{i+1}},\qquad \sigma_{n+1}\equiv\sigma_1. \tag{2.12}$$

Note that both μ and ω are invariant under the group $GL(2,\mathbb{C})$ that acts linearly on the homogeneous coordinates on $\mathbb{CP}^1$. Its action on σ is given by the usual expression

$$\sigma\mapsto\sigma'=\frac{a\sigma+b}{c\sigma+d},\qquad ad-bc\neq 0 \tag{2.13}$$

while its action on $\beta_k^{\mathbb{A}}$ is dictated by the invariance of $Z^{\mathbb{A}}$ in (2.10): the coefficients $\beta_k^{\mathbb{A}}$ may be reorganized (up to some combinatorial factors suppressed for simplicity) into a rank d tensor under $GL(2,\mathbb{C})$,

$$\{\beta_k^{\mathbb{A}}\}=\{\beta^{\mathbb{A}}_{I_1I_2\ldots I_d}\},\qquad I_l=1,2, \tag{2.14}$$

where the number of indices $I_l=2$ equals k, so that the action of $GL(2,\mathbb{C})$ on $\beta_k^{\mathbb{A}}$ becomes

$$\beta^{\mathbb{A}}_{I_1I_2\ldots I_d}\mapsto{\beta'}^{\mathbb{A}}_{I_1I_2\ldots I_d}=M_{I_1}^{I'_1}M_{I_2}^{I'_2}\ldots M_{I_d}^{I'_d}\beta^{\mathbb{A}}_{I'_1I'_2\ldots I'_d},\quad M_I^{I'}=\begin{pmatrix} d & -b\\ -c & a\end{pmatrix}. \tag{2.15}$$

Along with μ and ω we also have to include the external wavefunctions,[6]

$$\Phi=\prod_{i=1}^{n}\phi_i(P(\sigma_i)). \tag{2.16}$$

Putting everything together, the Yang-Mills amplitude is formally[7]

$$\int_{\overline{\mathcal{M}_{0,n,d}}}\frac{\mu_d\wedge\omega(\sigma_1,\ldots,\sigma_n)}{\mathrm{vol}(GL(2,\mathbb{C}))}\wedge\Phi. \tag{2.17}$$

[6]We write $\phi(P(\sigma_i))$ for the pullback of ϕ to moduli space via the evaluation map sending P to $P(\sigma_i)$.

[7]Here and below, by $\mathrm{vol}(GL(2,\mathbb{C}))$ we really mean the volume form on that group; this is just the standard quotient, when written in terms of an integral over the quotient space.

The expression (2.17) is formal for several reasons. The first and most serious reason is that we have to choose a contour for the integral over the coordinates $\beta_k^{\mathbb{A}}$ in $\overline{\mathcal{M}_{0,n,d}}$, and the proper choice of contour is not yet well understood. (We do not have to choose a contour for the integrals over σ, because the integrand includes both $\mathrm{d}\sigma$ from ω and $\mathrm{d}\bar{\sigma}$ from the external wavefunctions.) We will have more to say about the contour below; to match the disconnected prescription we will essentially use a contour around infinity (suitably defined) so that *all* residues are counted.

Second, we have to divide out by the action of $GL(2,\mathbb{C})$. A convenient gauge-fixing will be chosen below, but of course the amplitude is independent of the choice of gauge. We should perhaps mention that we consider $GL(2,\mathbb{C})$ over $\mathbb{C}$, *i.e.* we divide by the "holomorphic" volume form. This means that

- this symmetry will always be fixed by a set of holomorphic conditions;
- we will sum over all inequivalent solutions;
- only the holomorphic Jacobian will be included in the integrals.

These rules are compatible with the computations of [106, 108, 107].

2.2.2 Disconnected prescription

Now we describe the disconnected prescription for the same amplitudes, formulated in twistor space along the lines of the derivation given in [31]. In this prescription a tree level amplitude involving $d+1$ negative helicity gluons, with a particular cyclic ordering, is obtained as a sum over various tree diagrams with d vertices. In Figure 2.3 we show a representative example of a diagram Γ which contributes to

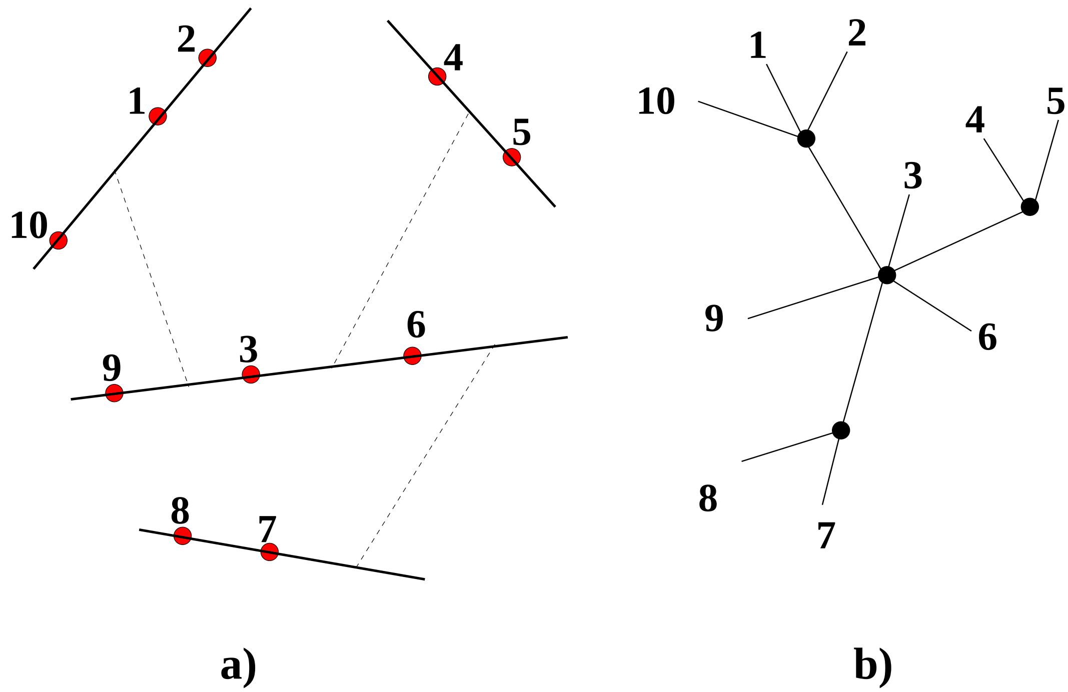

Figure 2.3: A contribution to Yang-Mills amplitudes with 5 positive and 5 negative helicity gluons, represented **(a)** as four disconnected lines in twistor space, **(b)** as a graph Γ with four MHV vertices.

amplitudes with 5 positive and 5 negative helicity gluons. The 10 external gluons are arranged cyclically around the index loop, and since there are 5 negative helicity gluons there are $5 - 1 = 4$ vertices. The vertices have arbitrary valence.[8] We have not specified which gluons have which helicities; the twistor space computation yields superspace expressions which generate the answers for all possible choices when suitably expanded in the fermionic coordinates.

Each vertex of Γ corresponds to a $\mathbb{CP}^1$ in $\mathbb{CP}^{3|4}$, equipped with marked points corresponding to internal or external lines attached to the vertex. To compute the

[8]Ultimately, it turns out that any diagram containing a vertex of valence ≤ 2 does not contribute to the amplitude [31].

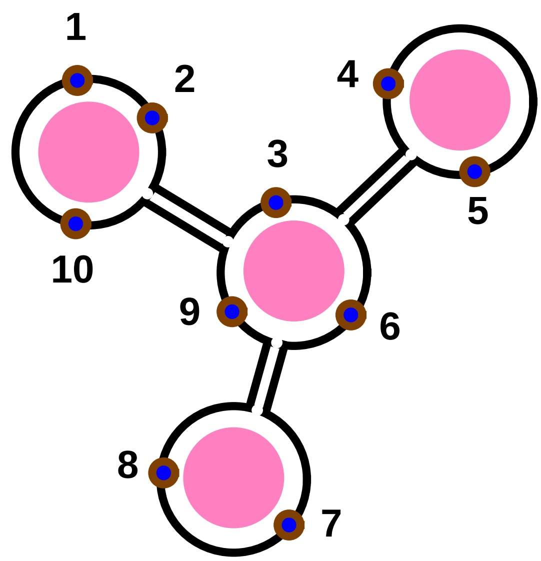

Figure 2.4: A different version of Figure 2.3, representing the same single-trace amplitude with the index line made manifest. The circles represent degree 1 curves in twistor space.

contribution of Γ to the amplitude we have to integrate over the moduli of these curves, given by d degree 1 maps

$$Q_i : \mathbb{CP}^1 \to \mathbb{CP}^{3|4}. \tag{2.18}$$

Each such map can be written

$$Q_i^{\mathbb{A}}(\sigma) = \sum_{k=0}^{1} \beta_{k,i}^{\mathbb{A}} \sigma^k \tag{2.19}$$

so there are a total of $8d|8d$ supermoduli $\beta_{k,i}^{\mathbb{A}}$ for these d maps, reduced to $4d|8d$ by the $GL(2,\mathbb{C})^d$ symmetry. We also have to integrate over the moduli for the marked points; if in the diagram Γ there are n_i marked points on the i-th $\mathbb{CP}^1$, then the full moduli space is

$$\mathcal{M}_{\text{lines}}^{\Gamma} = \prod_{i=1}^{d} \mathcal{M}_{0,n_i,1}(\mathbb{CP}^{3|4}). \tag{2.20}$$

As in the connected case there is a natural measure for the moduli of the curves,

$$\mu_{\text{lines}} = \prod_{k,\mathbb{A},i} \mathrm{d}\beta^{\mathbb{A}}_{k,i}. \tag{2.21}$$

There are several factors in the integrand which depend on the marked points. First, there is a free-fermion correlator for each curve; the points on the i-th $\mathbb{CP}^1$ come with a cyclic ordering as indicated in Figure 2.3, and if we label them $\sigma_1, \ldots, \sigma_{n_i}$, they contribute

$$\omega_i = \omega(\sigma_1, \ldots, \sigma_{n_i}) \tag{2.22}$$

with ω defined in (2.12). These free-fermion correlators contain $\mathrm{d}\sigma$ for each marked point.

Next we have to include the external wavefunctions: each external wavefunction ϕ_j is connected to a marked point σ on the i-th $\mathbb{CP}^1$, for some i, and the integrand includes the factor

$$\phi_j(Q_i(\sigma)) \tag{2.23}$$

just as in the connected prescription. But unlike the connected prescription, here we also have some marked points which are connected to internal propagators. Let us write $D(\cdot,\cdot)$ for the twistor space propagator, which is a $(0,2)$-form on $\mathbb{CP}^{3|4} \times \mathbb{CP}^{3|4}$. Each internal propagator is connected to two marked points σ, σ' on the i-th and i'-th $\mathbb{CP}^1$'s respectively, for some i, i', and contributes to the integrand a factor

$$D(Q_i(\sigma), Q_{i'}(\sigma')). \tag{2.24}$$

Let us write $\Phi \wedge D$ for the product of all the wavefunctions and propagators from (2.23), (2.24). Since every marked point is attached either to a propagator or to an external wavefunction, this $\Phi \wedge D$ includes one factor $\mathrm{d}\bar{\sigma}$ for each marked point.

Then the amplitude in the disconnected prescription is given by the sum over tree diagrams,

$$\sum_{\Gamma} \int_{\mathcal{M}^{\Gamma}_{\text{lines}}} \frac{\mu_{\text{lines}} \wedge \left(\prod_{i=1}^{d} \omega_i\right) \wedge \Phi \wedge D}{\text{vol}(GL(2,\mathbb{C}))^d}. \tag{2.25}$$

As with the connected prescription, to make this integral concrete we have to do two more things. First, we must gauge-fix the symmetry $GL(2,\mathbb{C})^d$ which acts separately on each $\mathbb{CP}^1$. Second, we must choose a contour for the integrals over the moduli $\beta^{\mathbb{A}}_{k,i}$.

In [31] it was argued that if one makes a particular choice of contour, and chooses external wavefunctions corresponding to gluons of fixed helicity and momentum, then the integral over $\mathcal{M}^{\Gamma}_{\text{lines}}$ in (2.25) can be evaluated by a simple rule. Namely, one first assigns $(+)$ and $(-)$ helicities to the endpoints of each propagator, consistent with the rule that each vertex should have exactly two $(-)$ helicities on it; for given Γ, there is at most one way to do this. (If there is no way to do it, then the diagram Γ just contributes zero.) Then each vertex gives a copy of the MHV amplitude — continued off-shell in a specific way to accommodate the internal lines — while each propagator carrying momentum q gives $1/q^2$.

For future use in section 2.4.2 we also mention a natural generalization of the disconnected prescription: instead of using d degree 1 curves we could use $K+1$ curves for some K, with total degree d, connected into a tree by K propagators. The integrand is then defined in a way precisely analogous to (2.25), except that the sum over Γ includes all choices for the degrees of the curves in addition to distributions of the marked points.

2.3 Matching the prescriptions in degree 2 case

2.3.1 The argument in degree 2 case

How can the disconnected and connected prescriptions give the same result? Let us consider next-to-maximally helicity violating amplitudes, $q = 3$, which come from degree 2 curves. We postpone the discussion of curves of higher degree to section 2.4.

The contribution of disconnected instantons comes from pairs of degree 1 curves connected by a single propagator, with n marked points distributed over the pair of curves. This moduli space has dimension $(8+n)|16$ (which includes $4|8$ for each degree 1 curve plus n for the marked points.) Different distributions of the marked points correspond to different MHV diagrams Γ.[9]

It was shown in [31] that for each Γ the integrand in (2.25) has a simple pole on the submoduli space $\mathcal{M}^{\Gamma}_{\text{int}}$, parameterizing degenerate configurations of intersecting lines of degree 1. This submoduli space has dimension $(7+n)|12$, because the condition that there exists an intersection in the bosonic space removes one bosonic modulus, and the condition that all four fermionic coordinates of the two lines coincide at this point removes four fermionic moduli.[10]

After contour-integrating to localize to $\mathcal{M}^{\Gamma}_{\text{int}}$, the sum (2.25) can be written as

$$\sum_{\Gamma} \int_{\mathcal{M}^{\Gamma}_{\text{int}}} \frac{1}{\text{vol}(GL(2,\mathbb{C}))^2} \left(\mu_{\text{int}} \wedge \left(\prod_{i=1}^{n_1} \frac{\mathrm{d}\sigma_i}{\sigma_i - \sigma_{i+1}} \right) \wedge \left(\prod_{j=1}^{n_2} \frac{\mathrm{d}\sigma'_j}{\sigma'_j - \sigma'_{j+1}} \right) \right) \wedge \Phi. \quad (2.26)$$

Here i and j run over the marked points on each $\mathbb{CP}^1$, including the point of inter-

[9] There are $n(n+1)/2$ such diagrams, although once we fix the external wavefunctions not every diagram gives a nonzero contribution to the sum (2.25); if the helicities are $---+++\cdots++$, then there are $2(n-3)$ diagrams which contribute.

[10] This fermionic delta-function guarantees the opposite helicity of the two endpoints of the propagators when one expands in fermions to evaluate a particular amplitude.

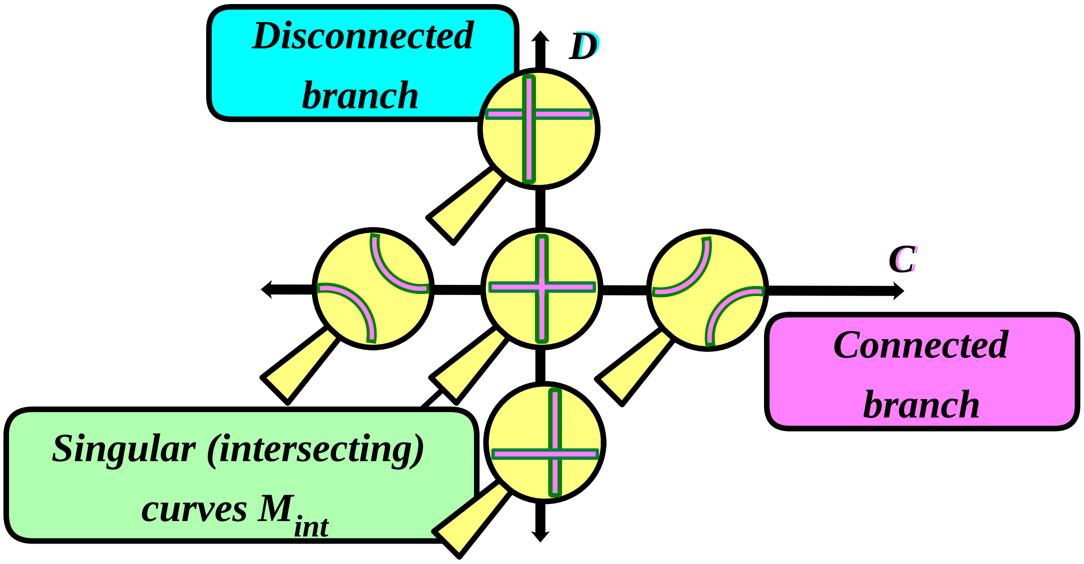

Figure 2.5: A degenerate configuration of two intersecting lines in $\mathbb{CP}^{3|4}$ can be deformed into a smooth connected curve of degree 2 or into two disconnected lines. The transition between the two branches of moduli space is reminiscent of a conifold transition.

section; so for a diagram with m external wavefunctions attached to the first line, $n_1 = m+1$ and $n_2 = n-m+1$. Also, $\sigma_{n_1+1} \equiv \sigma_1$ and $\sigma'_{n_2+1} \equiv \sigma'_1$. The measure $\mu_{\rm int}$ is completely determined by the symmetries of $\mathbb{CP}^{3|4}$.

On the other hand, from the connected prescription (2.17) we find

$$\int_{\overline{\mathcal{M}_{0,n,2}}} \frac{1}{\mathrm{vol}(GL(2,\mathbb{C}))} \left(\mu_2 \wedge \left(\prod_{i=1}^{n} \frac{\mathrm{d}\sigma_i}{\sigma_i - \sigma_{i+1}} \right) \right) \wedge \Phi. \tag{2.27}$$

We will reorganize the integral (2.27) over the $(8+n)|12$-dimensional space $\overline{\mathcal{M}_{0,n,2}}$ of conics in the following way: Locally, to any conic we will associate a pair of intersecting lines which are its "asymptotes." The moduli space of pairs of intersecting lines with n marked points is the $\mathcal{M}_{\rm int}$ which occurred in the disconnected prescription. This $\mathcal{M}_{\rm int}$ has dimension $(7+n)|12$, so in $\overline{\mathcal{M}_{0,n,2}}$ there is one more coordinate, which we call C; $C = 0$ corresponds to the singular conics, which coincide with their asymptotes.

This C can be thought of as a "deformation parameter" which resolves the singularity. We will find that the integrand has a pole at $C = 0$, *i.e.* along $\mathcal{M}_{\text{int}}$.

More precisely, $\mathcal{M}_{\text{int}}$ includes only those degenerations in which the marked points are distributed in a way corresponding to some MHV tree graph Γ. This just means the points are broken into two groups which are cyclically ordered — so e.g. if $n = 6$, there is a component of $\mathcal{M}_{\text{int}}$ with points $1, 2, 3$ on one line and $4, 5, 6$ on the other, but we do not include the degeneration which has $1, 2, 4$ on one line and $3, 5, 6$ on the other. Indeed, we will see that the latter degeneration does *not* give a pole. We will find poles only along $n(n+1)/2$ distinct components $\mathcal{M}^{\Gamma}_{\text{int}}$, which are in one-to-one correspondence with the diagrams Γ contributing to (2.26).

Moreover, we will show that the residue along $\mathcal{M}^{\Gamma}_{\text{int}}$ is precisely such that the integral (2.27) agrees with (2.26) after localizing. This will complete the argument for the equivalence in the degree 2 case.

2.3.2 Computing the residue in degree 2 case

In this section we show that the integral (2.27) over the moduli space $\overline{\mathcal{M}_{0,n,2}}$ of genus zero, degree 2 curves in $\mathbb{CP}^{3|4}$ with n marked points has a pole at the subspace $\mathcal{M}_{\text{int}}$ describing pairs of intersecting lines, and that it has the desired residue as discussed in the last section.

Let us start by fixing part of the $GL(2, \mathbb{C})$ symmetry reviewed in section 2.2.1. We use three generators of $GL(2, \mathbb{C})$ to impose the constraints

$$P^4(\sigma) = \sigma \qquad i.e. \qquad (\beta^4_0, \beta^4_1, \beta^4_2) = (0, 1, 0). \tag{2.28}$$

In other words, we are imposing the conditions that the two intersections of the hyperplane $Z^4=0$ with the curve have coordinates[11] $\sigma=0$ and $\sigma=\infty$, and normalizing the coefficients $\beta^4_{0,1,2}$. There is one more generator of $GL(2,\mathbb{C})$ to be fixed, the matrix

$$M = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}, \tag{2.29}$$

which acts as

$$\beta_k^{\mathbb{A}} \to \lambda^{2-2k}\beta_k^{\mathbb{A}}, \qquad \sigma \to \lambda^2\sigma. \tag{2.30}$$

This transformation preserves the gauge choice (2.28).

Factors from the measure on the moduli space

Using the freedom to divide all twistor coordinates $Z^{\mathbb{A}}$ by σ, we can write (2.10) as

$$P^{\mathbb{A}}(\sigma) = Z^{\mathbb{A}} = \sum_{k=0}^{2} \beta_k^{\mathbb{A}}\sigma^{k-1} = \beta_0^{\mathbb{A}}\sigma^{-1} + \beta_1^{\mathbb{A}} + \beta_2^{\mathbb{A}}\sigma, \tag{2.31}$$

which using (2.28) implies $P^4(\sigma) = 1$. As $\sigma \to \infty$ or $\sigma \to 0$, we can neglect the first or the last term in (2.31), respectively. So (2.31) describes a hyperbola that approaches two asymptotic lines in the superspace $\mathbb{C}^{3|4}$:

$$Z^{\mathbb{A}} = \beta_0^{\mathbb{A}}\sigma^{-1} + \beta_1^{\mathbb{A}}, \qquad Z^{\mathbb{A}} = \beta_1^{\mathbb{A}} + \beta_2^{\mathbb{A}}\sigma. \tag{2.32}$$

These two lines intersect at the point $Z^{\mathbb{A}} = \beta_1^{\mathbb{A}}$, while $\beta_0^{\mathbb{A}}$ and $\beta_2^{\mathbb{A}}$ with $\mathbb{A} \neq 4$ are the tangent vectors along these lines. It is important that for every conic $\Sigma := P_*(\mathbb{CP}^1) \subset \mathbb{CP}^{3|4}$ we can find a singular conic Σ' (a pair of intersecting lines) in $\mathcal{M}_{\text{int}}$ defining the asymptotes of Σ. This rule is not canonical; it depended on our choice to single out the points at infinity, *i.e.* the hyperplane $Z^4=0$.

[11]The point $\sigma = \infty$ can be written as $(1 : 0)$ in homogeneous coordinates, and therefore is completely nonsingular.

We want to express $\overline{\mathcal{M}_{0,n,2}}$ locally as a product of $\mathcal{M}_{\text{int}}$ and $\mathbb{C}$, with the extra $\mathbb{C}$ parameterized by the deformation parameter C. What are the appropriate coordinates? The $3|4$ parameters

$$\beta_1^{\mathbb{A}}, \qquad \mathbb{A} \neq 4, \tag{2.33}$$

describing the position of the intersection of the asymptotes, give coordinates on $\mathcal{M}_{\text{int}}$. The remaining $4|8$ coordinates on $\mathcal{M}_{\text{int}}$ are the directions of the two asymptotes; each asymptote gives us $2|4$ moduli. We want to describe these directions by "unit vectors" in a suitable sense. As we approach a generic point of $\mathcal{M}_{\text{int}}$, β_0^3 and β_2^3 are nonzero, and we may use them to normalize the direction vectors. In other words, the remaining $2|4$ plus $2|4$ coordinates on $\mathcal{M}_{\text{int}}$ may be chosen as

$$\frac{\beta_0^{\mathbb{A}}}{\beta_0^3} \quad \text{and} \quad \frac{\beta_2^{\mathbb{A}}}{\beta_2^3}, \qquad \mathbb{A} \in \{1,\, 2\,|1', 2', 3', 4'\}. \tag{2.34}$$

(Choosing different coordinates on $\mathcal{M}_{\text{int}}$ instead of (2.33) and (2.34) would not change the result below; the only change would be a C-independent Jacobian.)

Looking at our original coordinates on $\overline{\mathcal{M}_{0,n,2}}$, we still have two more bosonic components of β which are independent of our coordinates on $\mathcal{M}_{\text{int}}$, namely β_0^3 and β_2^3 themselves. We also have one unfixed generator of $GL(2,\mathbb{C})$ given in (2.30). This generator simply multiplies the ratio β_0^3/β_2^3 by λ^4, so we can use it to fix that ratio to a constant, such as

$$\frac{\beta_0^3}{\beta_2^3} = 1. \tag{2.35}$$

Now having fixed the full $GL(2,\mathbb{C})$ symmetry we can write the measure μ_2 from (2.11) as

$$(J/4) \prod_{k,\mathbb{A}} \mathrm{d}\beta_k^{\mathbb{A}}\ \delta(\beta_0^3/\beta_2^3 - 1)\ \delta(\beta_0^4)\ \delta(\beta_1^4 - 1)\ \delta(\beta_2^4). \tag{2.36}$$

Here J is the determinant of the Jacobian matrix of variations of the constraints with respect to the $GL(2,\mathbb{C})$ generators. If we parameterize the generators of $GL(2,\mathbb{C})$ by

$$M = \begin{pmatrix} 1+a & b \\ c & 1+d \end{pmatrix} \tag{2.37}$$

then this matrix is

$$\delta \begin{pmatrix} \beta_0^4 \\ \beta_1^4 \\ \beta_2^4 \\ \beta_0^3/\beta_2^3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ * & * & 2 & -2 \end{pmatrix} \begin{pmatrix} b \\ c \\ a \\ d \end{pmatrix} \tag{2.38}$$

and hence we get simply

$$J = -4. \tag{2.39}$$

The factor $J/4$ in (2.36) represents $1/\mathrm{vol}(GL(2,\mathbb{C}))$; we had to divide by 4 because the $\mathbb{Z}_4 \subset GL(2,\mathbb{C})$ generated by

$$M = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \tag{2.40}$$

is left unfixed by our gauge condition.

The three delta functions in (2.36) involving β_k^4 just eliminate the integrals over those variables, imposing (2.28). Let us also use $\delta(\beta_0^3/\beta_2^3-1)$ to eliminate β_0^3, imposing (2.35). Integrating over β_0^3 gives a factor β_2^3, so the measure becomes

$$-\beta_2^3\,\mathrm{d}\beta_2^3 \prod_{\mathbb{A}\neq 4} \mathrm{d}\beta_1^{\mathbb{A}} \prod_{k\in\{0,2\}} \prod_{\mathbb{A}\neq 3,4} \mathrm{d}\beta_k^{\mathbb{A}}. \tag{2.41}$$

We rewrite this as a measure for the single transverse coordinate β_2^3, times a measure on $\mathcal{M}_{\mathrm{int}}$, for which a full set of $7|12$ coordinates were given in (2.33), (2.34):

$$\left(-(\beta_2^3)^{1-4}\mathrm{d}\beta_2^3\right) \times \left(\prod_{\mathbb{A}\neq 4} \mathrm{d}\beta_1^{\mathbb{A}} \prod_{k\in\{0,2\}} \prod_{\mathbb{A}\neq 3,4} \mathrm{d}\left(\frac{\beta_k^{\mathbb{A}}}{\beta_k^3}\right)\right). \tag{2.42}$$

The extra power (-4) in $(\beta_2^3)^{-4}$ was calculated as $2_{k=0,2} \times (2_B - 4_F)$; the terms 2_B and -4_F arise from the redefined bosonic *and* fermionic measures involving $\beta_k^{\mathbb{A}}$, respectively.

The coordinate β_2^3 is related to the deformation parameter C — we will see that the natural definition of C is $(\beta_2^3)^2$. The measure $(\beta_2^3)^{-3}\mathrm{d}\beta_2^3$ occurring in (2.42) will be corrected to $\mathrm{d}\beta_2^3/\beta_2^3$ — the desired pole — once we include an extra factor $(\beta_2^3)^2$ which comes from the free-fermion correlator ω. We now turn to the analysis of this factor.

Factors from the fermion correlator

The integrand (2.27) contains the factor

$$\omega(\sigma_1, \ldots, \sigma_n) = \prod_{i=1}^{n} \frac{\mathrm{d}\sigma_i}{\sigma_i - \sigma_{i+1}}, \qquad \sigma_{n+1} \equiv \sigma_1. \tag{2.43}$$

We would like to investigate how this form behaves on conics that are degenerating into a pair of lines (*i.e.* near $\mathcal{M}_{\text{int}}$.) The result will be that along $\mathcal{M}_{\text{int}}$, ω factorizes into a product of two copies of ω defined on the two lines separately (with an extra marked σ on each line at the point of intersection), while transverse to $\mathcal{M}_{\text{int}}$, ω vanishes like $(\beta_2^3)^2$.

As the curve degenerates to a pair of lines, some of the n insertions approach one line and some approach the other. We consider the case where

$$\sigma_1, \ldots, \sigma_m \tag{2.44}$$

approach one asymptote while the remaining $(n-m)$ insertions

$$\sigma_{m+1}, \ldots, \sigma_n \tag{2.45}$$

approach the other. This is not the most general choice, since the σ_i come with a fixed cyclic ordering which is built into (2.43); our choice is characterized by the fact that as we run through the cyclic ordering we jump from the first line to the second and back only once. We will comment on other possibilities at the end.

With the $GL(2,\mathbb{C})$ gauge-fixing we chose above, as we approach some point of $\mathcal{M}_{\rm int}$, the coordinates σ_i do not remain finite; one of the lines is $\sigma \to 0$ while the other line is $\sigma \to \infty$. So we need to rescale the σ_i to get new coordinates $\hat{\sigma}_i$ on $\mathcal{M}_{\rm int}$ which label the positions of the marked points; we define $\hat{\sigma}_i$ so that $Z^{\mathbb{A}}$ defined in (2.32) remains constant as $\hat{\sigma}_i$ is kept fixed and $\beta_0^3, \beta_2^3 \to 0$. The correct redefinition is

$$\sigma_i = \left\{ \begin{array}{lll} (\beta_2^3)^{-1}\hat{\sigma}_i & \text{for} & i \in \{1, 2, \ldots m\} \\ \\ \beta_0^3 (\hat{\sigma}_i')^{-1} & \text{for} & i \in \{m+1, m+2, \ldots n\} \end{array} \right\}. \tag{2.46}$$

(We use two different symbols $\hat{\sigma}_i$ and $\hat{\sigma}_i'$ to distinguish the coordinates on the two different lines.) Rewriting ω from (2.43) in terms of $\hat{\sigma}_i$ and $\hat{\sigma}_i'$, we obtain

$$\omega(\hat{\sigma}_1, \ldots \hat{\sigma}_n') = \beta_0^3\beta_2^3 \left(\prod_{i=1}^{m-1} \frac{\mathrm{d}\hat{\sigma}_i}{\hat{\sigma}_i - \hat{\sigma}_{i+1}} \right) \frac{\mathrm{d}\hat{\sigma}_m}{\hat{\sigma}_1 \hat{\sigma}_m} \left(\prod_{i=m+1}^{n-1} \frac{\mathrm{d}\hat{\sigma}_i'}{\hat{\sigma}_i' - \hat{\sigma}_{i+1}'} \right) \frac{\mathrm{d}\hat{\sigma}_n'}{\hat{\sigma}_{m+1}' \hat{\sigma}_n'} + \ldots \tag{2.47}$$

where the intersection was defined to be at $\hat{\sigma} = \hat{\sigma}' = 0$. The dots in (2.47) indicate terms suppressed by powers of $\beta_0^3\beta_2^3$.

Most of the powers of β_0^3 and β_2^3 have canceled, but there is an extra factor of $\beta_0^3\beta_2^3$, which equals $(\beta_2^3)^2$ because of our gauge choice (2.35). Also, we obtained the expected free fermion contractions, including the $2+2$ contractions involving the intersection of the two lines at $\hat{\sigma} = \hat{\sigma}' = 0$.

Note that β_0^3 and β_2^3 always appeared in the combination

$$C = \beta_0^3\beta_2^3 \tag{2.48}$$

that is invariant under (2.30). This is the same C that we used in Figure 2.5; in fact, one can rewrite our curve in the form

$$xy = C \tag{2.49}$$

where x,y are coordinates on a plane in $\mathbb{CP}^{3|4}$. The limit $C \to 0$ describes the singular conics. Note that it is C rather than β_2^3 that is a good coordinate — this is because a simultaneous sign flip on β_0^3 and β_2^3 is the gauge transformation (2.30) with $\lambda = i$, which preserves our gauge choices (2.35).

Finally, it is easy to check that if we choose a different distribution of the marked points, the result comes out suppressed by additional powers of C. We are only interested in the leading terms, which are linear in C and will give the coefficient of $\mathrm{d}C/C$.

2.3.3 Finishing the proof in degree 2 case

Now we can collect the results from the previous two subsections. The powers of β_2^3 from (2.42) and (2.47) combine to give $\int \mathrm{d}\beta_2^3/\beta_2^3$, which is proportional to $\int \mathrm{d}C/C$. So as advertised, the integral (2.27) localizes after contour integration to an integral over $\mathcal{M}_{\mathrm{int}}$. The symmetries of $\mathbb{CP}^{3|4}$ determine the measure for the moduli of the two lines in $\mathcal{M}_{\mathrm{int}}$, which therefore agrees with the measure μ_{int} in (2.26) up to an overall constant; as for the integral over the marked points, comparing (2.47) and (2.26) we see that these measures are also identical. This completes the argument for equivalence in the $d = 2$ case.

Incidentally, one can also compare the measures on $\mathcal{M}_{\mathrm{int}}$ directly, without recourse to a symmetry argument. We have already computed the measure which arises from the connected prescription, in (2.42), so the job is to compute the measure μ_{int} which arises from the disconnected prescription. This computation is given (in greater generality) in section 2.4.4.

2.4 Higher degree

Now let us consider the connected prescription for general degree d. We will see that the fully disconnected description and the fully connected prescription are not only equivalent, they are just two extreme cases of a more general class of rules to calculate the amplitude. We will find d *a priori* different expressions for the scattering amplitude with $d+1$ negative-helicity gluons,

$$\mathcal{A}_{[K]}, \qquad K = 0, 1, 2, \ldots, d-1, \tag{2.50}$$

where $K+1$ denotes the total number of curves involved in the prescription.[12]

The organization of this section is as follows:

- subsection 2.4.1 outlines the argument that the completely connected and completely disconnected prescriptions agree;
- subsection 2.4.2 discusses the intermediate prescriptions with arbitrary K and their diagrammatic interpretation;
- subsection 2.4.3 generalizes the residue calculation of subsection 2.3.2 to the case of a degree d curve splitting into two curves of degrees d_1 and d_2;

[12] Later we will see that K also represents the codimension in moduli space on which the prescription is localized, or equivalently the number of internal propagators which appear in the prescription.

- subsection 2.4.4 shows that the residues occurring for any degeneration are actually independent of the chosen prescription, completing the argument.

2.4.1 The proof in higher degree case

Rather than showing directly that the connected prescription arising from a single connected degree d curve is equivalent to the disconnected prescription involving d lines, we will first show that it is equivalent to a computation involving two disconnected components of degrees d_1, d_2, such that

$$d_1 + d_2 = d. \tag{2.51}$$

The proof is a generalization of the computation we did in section 2.3.2: namely, in subsection 2.4.3 we will find a pole on each boundary divisor $\mathcal{M}^{\Gamma}_{\mathrm{int}}$, corresponding to a degeneration into intersecting curves,

$$\Sigma_d \longrightarrow \Sigma_{d_1} \cup \Sigma_{d_2}, \qquad d_1 + d_2 = d, \tag{2.52}$$

with a particular distribution of the marked points.

Next we want to show iteratively that this integral over curves with 2 irreducible components is equivalent to one over curves with 3 components, and so on until eventually we reach d components (all of which must have degree 1.) The idea which makes this iteration possible is the following: consider some locus $\mathcal{M}^{\Gamma}_{\mathrm{int}}$, corresponding to a particular degeneration of Σ into $K+1$ components, with a particular distribution of the marked points. This locus can be obtained as an intersection of K boundary divisors, $\mathcal{M}^{\Lambda_j}_{\mathrm{int}}$, each of which is associated with a degeneration of Σ_d into two irreducible

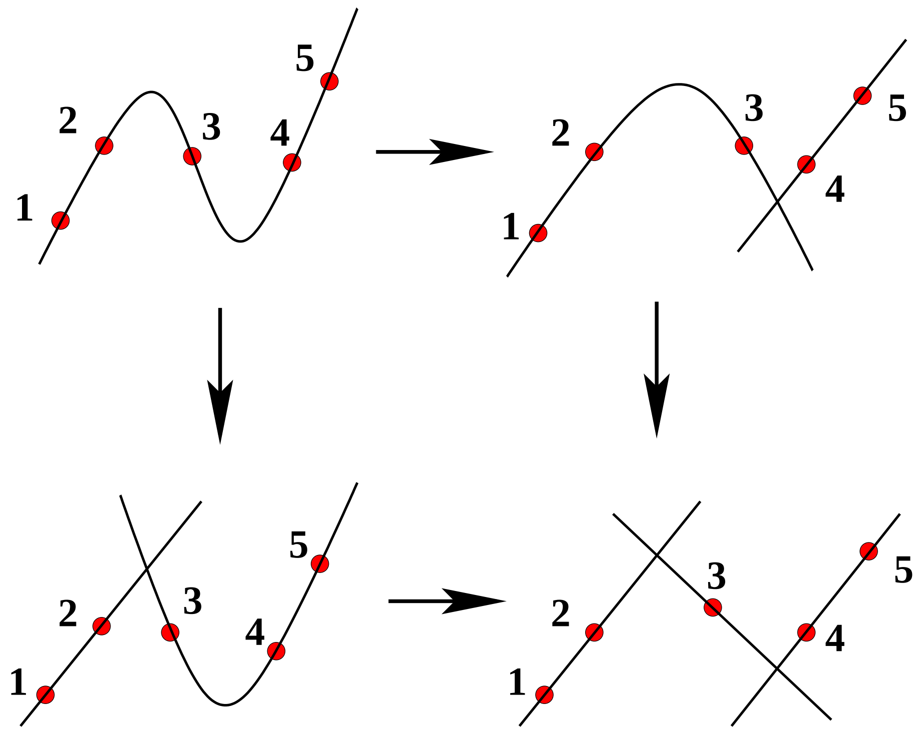

Figure 2.6: A degeneration of a degree 3 curve into three intersecting lines can be viewed as a two-step process. The moduli space of degree 3 maps with 5 marked points, $\overline{\mathcal{M}}_{0,5,3}$, contains divisors, $\mathcal{M}_{\text{int}}^{\Lambda_1}$ and $\mathcal{M}_{\text{int}}^{\Lambda_2}$, associated with degenerations into a degree 2 curve and a line, shown at the intermediate stages. The moduli space $\mathcal{M}_{\text{int}}^{\Gamma}$ of three intersecting lines (shown in the lower right corner) can be identified with the intersection $\mathcal{M}_{\text{int}}^{\Lambda_1} \cap \mathcal{M}_{\text{int}}^{\Lambda_2}$.

components,[13]

$$\mathcal{M}_{\text{int}}^{\Gamma} = \mathcal{M}_{\text{int}}^{\Lambda_1} \cap \cdots \cap \mathcal{M}_{\text{int}}^{\Lambda_K}. \tag{2.53}$$

An example is shown in Figure 2.6. In this sense, the problem of studying a general degeneration boils down to understanding the basic process (2.52).

So let's start with the integral over K-component curves and try to prove it agrees with an integral over $(K+1)$-component curves. In the K-component case we have

[13]We use Γ to denote a general degeneration into K+1 components, and Λ to denote a degeneration into just two components.

to integrate over various loci $\mathcal{M}_{\text{int}}^{\Gamma}$ as in (2.53). Since the various divisors $\mathcal{M}_{\text{int}}^{\Lambda}$ meet transversally [53], in integrating over each such $\mathcal{M}_{\text{int}}^{\Gamma}$ we will encounter poles wherever $\mathcal{M}_{\text{int}}^{\Gamma}$ intersects another divisor $\mathcal{M}_{\text{int}}^{\Lambda}$.[14] We choose our contour so that it picks up the residues at these poles. In this way we reduce the integral over $\mathcal{M}_{\text{int}}^{\Gamma}$ to the sum of integrals over all intersections $\mathcal{M}_{\text{int}}^{\Gamma} \cap \mathcal{M}_{\text{int}}^{\Lambda}$. Then we have to sum over all Γ describing K-component degenerations. What is the result of all this summation? From the perspective of the $(K+1)$-component degenerations — which we label by Γ' — the answer is clear: given some

$$\mathcal{M}_{\text{int}}^{\Gamma'} = \mathcal{M}_{\text{int}}^{\Lambda_1} \cap \ldots \cap \mathcal{M}_{\text{int}}^{\Lambda_K}, \tag{2.55}$$

there are K ways to make it by intersecting some $\mathcal{M}_{\text{int}}^{\Gamma}$ with some $\mathcal{M}_{\text{int}}^{\Lambda_i}$. Therefore we get a sum over all $(K+1)$-component degenerations, with an *overall* multiplicative factor K.

Finally, after repeating this process $d-1$ times, we arrive at an integral over the moduli space of connected trees consisting of d lines, with all possible shapes for the tree and all allowed distributions of marked points. But the arguments of [31] show that the disconnected prescription also reduces to such an integral, by a similar process of localization to poles. Furthermore, in section 2.4.3 we will see that the residues in these two computations agree; this will complete the proof.

[14] One way to understand this is to note that if we start with the full $\overline{\mathcal{M}_{0,n,d}}$ and look near such an intersection of K divisors, the integrand looks like

$$\frac{\mathrm{d}C_1}{C_1} \wedge \cdots \wedge \frac{\mathrm{d}C_K}{C_K} \wedge (\text{regular}). \tag{2.54}$$

We have already contour-integrated over $C_1, \ldots, C_{K-1}$ and thus restricted to $C_1 = \cdots = C_{K-1} = 0$, *i.e.* to $\mathcal{M}_{\text{int}}^{\Gamma}$; after doing this we get simply $\mathrm{d}C_K/C_K$, with a pole at $C_K = 0$, *i.e.* at $\mathcal{M}_{\text{int}}^{\Gamma} \cap \mathcal{M}_{\text{int}}^{\Lambda}$.

2.4.2 Intermediate prescriptions

In subsection 2.4.1 we encountered $d-1$ different moduli spaces $\mathcal{M}^K_{\text{int}}$ of singular curves, characterized by the number $K+1$ of components, which interpolated between the nonsingular degree d curve ($K=0$) and the tree of degree 1 curves ($K=d-1$). Furthermore we obtained integrals over each $\mathcal{M}^K_{\text{int}}$ by starting with the connected prescription ($K=0$) and successively localizing to poles. As a result of this localization all these integrals are equal; now we want to argue that the intermediate cases $K=1,\ldots,d-2$ can be naturally interpreted as coming from "intermediate prescriptions," involving integrals over the moduli of $K+1$ disconnected curves with K propagators connecting them. We defined these prescriptions at the end of section 2.2.2.

The argument is a generalization of the "heuristic" derivation of the computational rules for the disconnected prescription, given in [31]. Namely, starting from the intermediate prescription, note that the propagator $D(\cdot,\cdot)$ by definition satisfies

$$\bar{\partial} D = \Delta. \tag{2.56}$$

Here Δ is a $(0,3)$-form on $(\mathbb{CP}^{3|4})^2$ which is concentrated on the diagonal $\mathbb{CP}^{3|4}$: in inhomogeneous coordinates with $Z^4 = Z'^4 = 1$ it may be written

$$\Delta = \overline{\delta}(Z^1 - Z'^1)\,\overline{\delta}(Z^2 - Z'^2)\,\overline{\delta}(Z^3 - Z'^3)\,\delta^4(\psi - \psi'), \qquad \overline{\delta}(f) := \delta^2(f)\mathrm{d}\bar{f}. \tag{2.57}$$

The equation (2.56) means that $D(\cdot,\cdot)$ is meromorphic with a pole along the diagonal. The integral over $\mathcal{M}^K_{\text{int}}$ in the disconnected prescription contains K propagators (2.24); these factors therefore have poles when $Q_i(\sigma) = Q_{i'}(\sigma')$.

As in [31], we assume that K of the integrals over moduli of the disconnected curves are evaluated on contours which encircle these poles, in a suitable sense. Using (2.56), performing these contour integrals is equivalent to filling in the contour and replacing D by Δ. This localizes the integral to the sublocus of moduli space where all propagators have shrunk to zero length, which is exactly $\mathcal{M}_{\text{int}}^{K}$.

So finally we have d different prescriptions, involving summing over configurations with 1 curve (connected case), 2, 3, $\ldots$, d curves (maximally disconnected case); and we have argued that each of these prescriptions is equivalent, up to an overall rescaling. In this sense any of them can be used to calculate the Yang-Mills amplitudes.

Of course, another possibility is that the correct amplitudes are obtained by summing different contributions from various sorts of diagrams with various numbers of curves. We have argued that all such contributions are proportional to one another, so such a modified rule would only change the overall prefactor. Although we will not try to make the final verdict in this chapter, we believe that a more detailed analysis of the prescriptions (including the coefficients) should be able to resolve this uncertainty.

Diagrammatic interpretation and an example

Now let us discuss the diagrammatic interpretation of the intermediate prescriptions. We have seen that the K-th intermediate prescription is naturally localized on $\mathcal{M}_{\text{int}}^{K}$, which is a union of various $\mathcal{M}_{\text{int}}^{\Gamma}$. Here Γ describes the decomposition of the curve Σ_d into $K+1$ components and the distribution of marked points along these components. Equivalently, we could say that Γ describes a slight generalization of an MHV tree diagram: namely, it is a tree diagram with $K+1$ vertices, where each

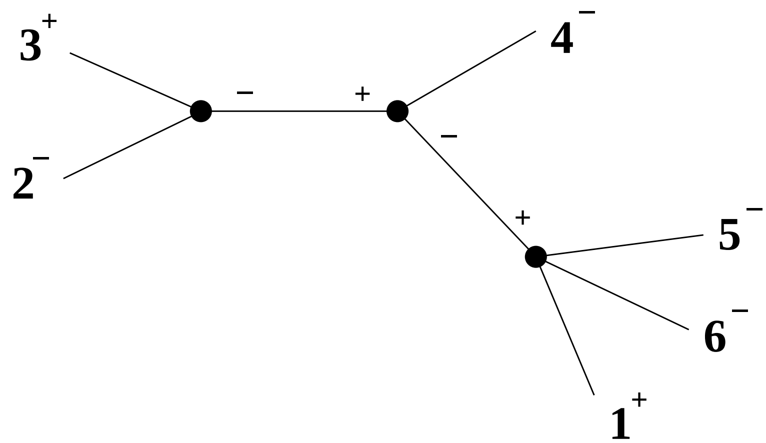

Figure 2.7: An MHV tree diagram contributing to $\mathcal{A}_{[+-+---]}$.

vertex now carries an internal index d_i, subject to the rule that $\sum d_i = d$. The MHV diagrams are the case where all $d_i = 1$.

It would be very useful if we could give a compact formula for the contribution of a general vertex with arbitrary d_i, analogous to the off-shell continuation of the MHV amplitude given in [31]. At the moment we do not possess such a formula, so we can only define the diagram Γ to be the integral over $\mathcal{M}^{\Gamma}_{\mathrm{int}}$ which we considered above. In this language, our localization argument relating different prescriptions becomes the statement that the contribution from a diagram Γ agrees with the sum over all Γ' obtained by "splitting a vertex" in Γ. In other words, Γ' should be obtained by replacing a vertex with index d by a pair of vertices with indices d_1, d_2, such that $d_1 + d_2 = d$, with a propagator connecting them. This is the diagrammatic analog of a degree d curve which degenerates into two curves with degrees d_1, d_2.

We can also repeat the combinatorics from subsection 2.4.1 in this language. Start with a diagram with K+1 vertices. This diagram contains K propagators. Therefore

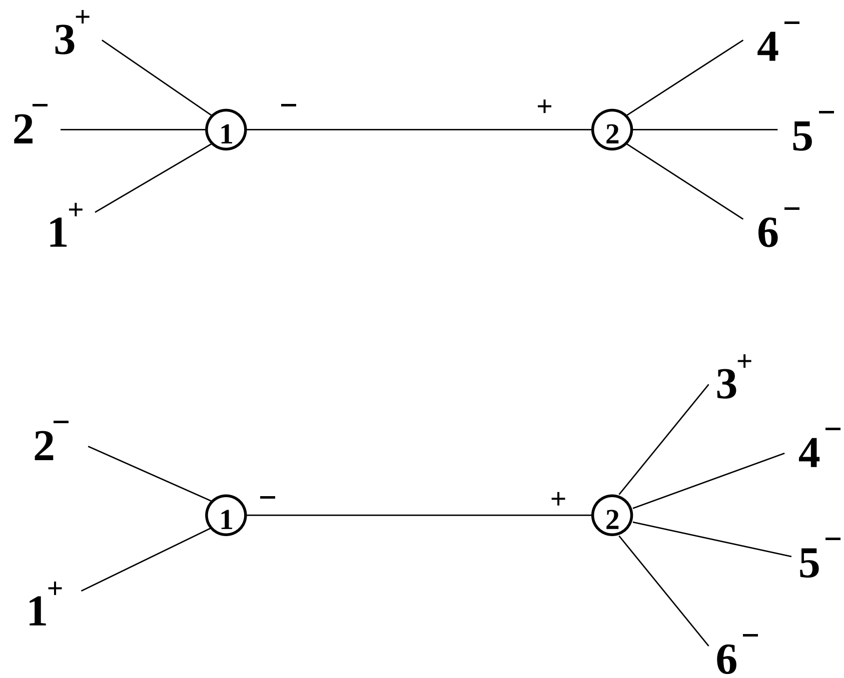

Figure 2.8: Two types of tree diagram with one MHV and one non-MHV (degree 2) vertex that contribute to the $\mathcal{A}_{[+-+---]}$ amplitude. In total, there are six diagrams of each kind. The number attached to each vertex represents the degree of the corresponding curve in twistor space.

there are K ways to shrink a single propagator and obtain a "parent" diagram with K vertices. Because a diagram with $K+1$ vertices has K parents, the sum over the daughters with $K+1$ vertices equals K times the sum over the parents with K vertices.

To illustrate how all this works when external wavefunctions of fixed helicity are included, let us consider a 6-gluon amplitude $\mathcal{A}_{[+-+---]}$. If we were to use the connected prescription, we would have to integrate over the moduli space $\overline{\mathcal{M}_{0,6,3}}$ of degree 3 curves. On the other hand, in the disconnected prescription one has to consider three degree 1 curves, which can be interpreted as MHV vertices in Yang-Mills theory [31]. Therefore, in this case one has to sum over all tree graphs with three MHV vertices connected by Yang-Mills propagators — see Figure 2.7. In total, there are 19 such graphs contributing to $\mathcal{A}_{[+-+---]}$.

Now let us consider the intermediate prescription with $K = 1$. This prescription leads to a sum over tree graphs with two vertices, one MHV and one non-MHV (the non-MHV vertex involves three insertions of negative helicity). Examples of such graphs with non-MHV vertices are shown in Figure 2.8. There are 12 such diagrams which contribute to $\mathcal{A}_{[+-+---]}$. Since each non-MHV vertex itself can be represented as a sum over tree diagrams with two MHV vertices, we should be able to reproduce the disconnected prescription if we split all non-MHV vertices into MHV ones. More precisely, in this decomposition we should encounter each MHV diagram twice (since in the disconnected prescription $K = 2$). Indeed, it is straightforward to check that the 12 non-MHV diagrams lead to 38 MHV graphs, in agreement with the general rule.

2.4.3 Computing the residue in higher degree case

Returning from our digression to discuss the intermediate prescriptions, in this section we show that the integral (2.17) over the moduli space $\overline{\mathcal{M}_{0,n,d}}$ which arises in the connected prescription has a pole along the codimension 1 divisor $\mathcal{M}^1_{\rm int}$ describing curves that are degenerated into 2 components. We further verify that the residue is the same as that which arises after localization of the $K = 1$ prescription on $\mathcal{M}^1_{\rm int}$, thus establishing the equivalence between connected and $K = 1$ prescriptions.

We want to study a degeneration in which the curve Σ_d degenerates into a pair of intersecting curves, Σ_{d_1} and Σ_{d_2}, of degree d_1 and d_2, as in (2.52). Using the projective

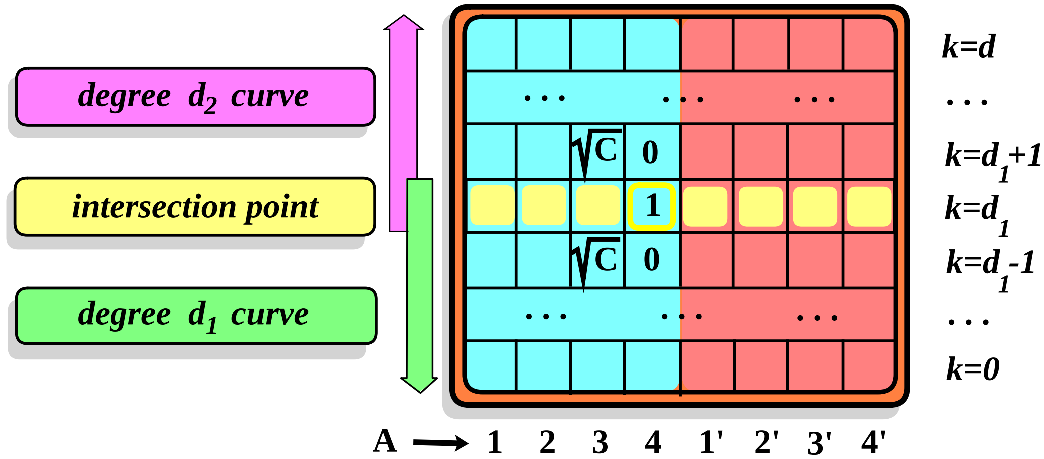

Figure 2.9: The organization of the coefficients $\beta_k^{\mathbb{A}}$ for a degree d curve degenerating into curves of degrees d_1 and d_2. The symmetry $GL(2,\mathbb{C})$ is fixed by setting three bosonic coefficients to the values $(0,1,0)$ and two others to $\sqrt{C}$; this C is the deformation parameter, which approaches zero in the degeneration limit.

symmetry to divide by σ^{d_1}, we can write the degree d map (2.10) as

$$Z^{\mathbb{A}}(\sigma) = \sum_{k=-d_1}^{d_2} \beta_{d_1+k}^{\mathbb{A}} \sigma^k. \tag{2.58}$$

We fix the $GL(2,\mathbb{C})$ symmetry similarly to the degree 2 case, namely by conditions based on (2.28) and (2.35):

$$(\beta_{d_1-1}^4, \beta_{d_1}^4, \beta_{d_1+1}^4) = (0,1,0), \qquad \frac{\beta_{d_1-1}^3}{\beta_{d_1+1}^3} = 1, \tag{2.59}$$

and define the deformation parameter $C := \beta_{d_1-1}^3 \beta_{d_1+1}^3$. As in degree 2, the singular limit will be $C \to 0$, or equivalently $\beta_{d_1+1}^3 \to 0$, and the question is how the other coefficients should scale in this limit.

The correct scaling is as follows: we take $\beta_{d_1-1}^3 = \beta_{d_1+1}^3 \to 0$ while holding finite the quantities

$$\alpha_k^{\mathbb{A}} := \frac{\beta_{d_1-k}^{\mathbb{A}}}{(\beta_{d_1-1}^3)^k}, \quad 0 \le k \le d_1; \qquad \alpha_k'^{\mathbb{A}} := \frac{\beta_{d_1+k}^{\mathbb{A}}}{(\beta_{d_1+1}^3)^k}, \quad 0 \le k \le d_2. \tag{2.60}$$

In that limit we obtain two curves,

$$\begin{aligned}\Sigma_{d_1} \quad &: \qquad Z^{\mathbb{A}}(\hat{\sigma}) = \sum_{k=0}^{d_1} \alpha_k^{\mathbb{A}} \hat{\sigma}^k, \\ \Sigma_{d_2} \quad &: \qquad Z^{\mathbb{A}}(\hat{\sigma}') = \sum_{k=0}^{d_2} \alpha_k'^{\mathbb{A}} \hat{\sigma}'^k.\end{aligned} \tag{2.61}$$

Namely, we obtain the points $Z^{\mathbb{A}}(\hat{\sigma})$ on Σ_{d_1} by holding fixed $\hat{\sigma} = \sigma/\beta^3_{d_1-1}$ in the limit, and we obtain the points $Z^{\mathbb{A}}(\hat{\sigma}')$ on Σ_{d_2} by holding fixed $\hat{\sigma}' = \sigma\beta^3_{d_1+1}$ in the same limit. See Figure 2.9.

Therefore the parameters $\alpha_k^{\mathbb{A}}, \alpha_k'^{\mathbb{A}}$ give coordinates on $\mathcal{M}^1_{\text{int}}$, specifying the moduli of the degenerated curve. (Note that $\alpha_0^{\mathbb{A}} = \alpha_0'^{\mathbb{A}}$; these shared coordinates specify the intersection point of Σ_{d_1} and Σ_{d_2}.)

Now we want to study how our integral (2.17) behaves near $\mathcal{M}^1_{\text{int}}$. As in section 2.3.2, we have to compute the Jacobian J from the gauge-fixing of $GL(2,\mathbb{C})$. The matrix of variations generalizing (2.38) is

$$\delta \begin{pmatrix} \beta^4_{d_1-1} \\ \beta^4_{d_1} \\ \beta^4_{d_1+1} \\ \beta^3_{d_1-1}/\beta^3_{d_1+1} \end{pmatrix} = \begin{pmatrix} d_2 & (d_1+2)\beta^4_{d_1+2} & 0 & 0 \\ (d_2+2)\beta^4_{d_1-2} & d_1 & 0 & 0 \\ 0 & 0 & d_1 & d_2 \\ * & * & 2 & -2 \end{pmatrix} \begin{pmatrix} b \\ c \\ a \\ d \end{pmatrix}. \tag{2.62}$$

In the singular limit, the $\beta^4_{d_1\pm2}$ terms in (2.62) vanish, and we get

$$J \to -2d_1d_2d. \tag{2.63}$$

The gauge-fixed integral includes the factor $J/2d$; the $2d$ comes from an unfixed subgroup of $GL(2,\mathbb{C})$, analogous to (2.40), which is $\mathbb{Z}_2 \times \mathbb{Z}_d$ if both d_1 and d_2 are even and $\mathbb{Z}_{2d}$ otherwise. Next we have to rewrite the integrand in terms of the new coordinates (2.60). One might be worried that switching to these coordinates will generate extra powers of C beyond what we had in the degree 2 case, spoiling the conclusion that there is a pole along $\mathcal{M}^1_{\text{int}}$. But this does not occur; if we increase d_1 by 1, for example, the integrand just acquires an extra integral over 4|4 variables:

$$\mu \to \mu \wedge \prod_{\mathbb{A}} \mathrm{d}\beta_0^{\mathbb{A}} = \mu \wedge \prod_{\mathbb{A}} \mathrm{d}\alpha_{d_1}^{\mathbb{A}} \tag{2.64}$$

The powers of $\beta^3_{d_1+1}$ simply cancel between the 4 bosons and 4 fermions! Unlike the coefficients $\beta^{\mathbb{A}}_{d_1}$ and $\beta^{\mathbb{A}}_{d_1\pm1}$, among which 5 special bosonic components have been used to gauge-fix the $GL(2,\mathbb{C})$ symmetry or to describe the parameter C, the additional moduli $\beta^{\mathbb{A}}_{d_1\pm k}$ with $k > 1$ come in full "supermultiplets" containing 4 bosons and 4 fermions. Therefore no new powers of C are generated in rescaling β's to α's, so the measure for the moduli of the degenerating curve still behaves as $\mathrm{d}C/C^2$ near $C = 0$. Similarly, the free fermion correlator ω factorizes,

$$\omega(\sigma) \to C\, \omega_1(\hat{\sigma}) \wedge \omega_2(\hat{\sigma}'), \tag{2.65}$$

just as in degree 2.

So we have a pole along $\mathcal{M}^1_{\text{int}}$, as in the degree 2 case, and after integrating around this pole the fully gauge-fixed measure for the moduli of the degenerate curve is

$$-d_1 d_2 \prod_{\mathbb{A}} \left({\prod_{k=0}^{d_1}}' \mathrm{d}\alpha_k^{\mathbb{A}} {\prod_{k=1}^{d_2}}' \mathrm{d}\alpha_k'^{\mathbb{A}} \right). \tag{2.66}$$

Here the symbol Π' indicates that we omit the 5 factors

$$\mathrm{d}\alpha_1^4,\ \mathrm{d}\alpha_1'^4,\ \mathrm{d}\alpha_0^4,\ \mathrm{d}\alpha_1^3,\ \mathrm{d}\alpha_1'^3; \tag{2.67}$$

there are no such α's among the coordinates on $\mathcal{M}^1_{\text{int}}$, because their corresponding

β's were used up in the gauge-fixing and in the transverse coordinate C, as shown in Figure 2.9.

2.4.4 Finishing the proof in higher degree case

Finally we have to check that the measure (2.66) agrees with the one coming from localization of the $K = 1$ prescription. From section 2.4.2 we know that the latter measure is obtained as follows: start with two curves of degree d_1, d_2,

$$Q^{\mathbb{A}}(\sigma) = \sum_{k=0}^{d_1} \alpha_k^{\mathbb{A}} \sigma^k,$$
$$Q'^{\mathbb{A}}(\sigma') = \sum_{k=0}^{d_2} \alpha_k'^{\mathbb{A}} \sigma'^k. \tag{2.68}$$

(The notation $\alpha_k^{\mathbb{A}}$, $\alpha_k'^{\mathbb{A}}$ we use here agrees with the notation we used above for the moduli of the curves obtained by a degeneration; compare (2.68) with (2.61), (2.60). The only difference is that here we do not necessarily have $\alpha_0^{\mathbb{A}} = \alpha_0'^{\mathbb{A}}$.) Then we have the standard measure (2.11) on the two curves separately, which before gauge-fixing is

$$\mu_{d_1} \wedge \mu_{d_2} = \prod_{\mathbb{A}} \left(\prod_{k=0}^{d_1} \mathrm{d}\alpha_k^{\mathbb{A}} \prod_{k=0}^{d_2} \mathrm{d}\alpha_k'^{\mathbb{A}} \right). \tag{2.69}$$

As explained in section 2.4.2, the requirement that the two curves actually intersect is enforced by a delta function which is coupled to one marked point on each curve,

$$\Delta(Q(\sigma), Q'(\sigma')). \tag{2.70}$$

To compare the measures (including this delta function) we have to gauge-fix the $GL(2, \mathbb{C})^2$ symmetry acting on the coefficients $\alpha_k^{\mathbb{A}}, \alpha_k'^{\mathbb{A}}$. There are many ways to do this; we choose a way that is as similar as possible to the gauge-fixing we used for the degenerating degree d curve, so that the unfixed moduli will match directly. Namely, we take

$$\alpha_0^i = \alpha_0'^i \quad \text{for } i \in \{2,3\}, \tag{2.71}$$

$$\alpha_0^4 = \alpha_0'^4 = 1, \tag{2.72}$$

$$\alpha_1^4 = \alpha_1'^4 = 0, \tag{2.73}$$

$$\alpha_1^3 = \alpha_1'^3 = 1. \tag{2.74}$$

The matrix of variations from this gauge-fixing is similar to (2.62), but since it is an 8×8 matrix we just write the answer here:

$$J = (d_1 d_2)^2 (\alpha_1^2 - \alpha_1'^2). \tag{2.75}$$

The gauge-fixing factor is $J/d_1 d_2$, because of the subgroup $\mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2} \subset GL(2,\mathbb{C}) \times GL(2,\mathbb{C})$, roots of unity acting on each curve separately; since this subgroup acts trivially it is unfixed by our gauge condition. Next we must include the integral over the delta function (2.70), which we write as

$$\int \mathrm{d}\sigma\ \mathrm{d}\sigma'\ \delta^{(3|4)} \left(\frac{Q^{\mathbb{A}}(\sigma)}{Q^4(\sigma)} - \frac{Q'^{\mathbb{A}}(\sigma')}{Q'^4(\sigma')} \right). \tag{2.76}$$

With our gauge choice, it is easy to study the behavior of this delta function in the vicinity of $\sigma = \sigma' = 0$.[15] One uses the Z^2 and Z^3 components of the delta function to set $\sigma = \sigma' = 0$, obtaining

$$\frac{1}{(\alpha_1^2 - \alpha_1'^2)}\ \delta(\alpha_0^1 - \alpha_0'^1) \prod_{\mathbb{A}=1'}^{4'} \delta(\alpha_0^{\mathbb{A}} - \alpha_0'^{\mathbb{A}}). \tag{2.77}$$

[15] Although our gauge choice was rigged so that studying $\sigma = \sigma' = 0$ would recover the desired moduli space of intersecting curves, it is not clear *a priori* from our arguments why one should consider only this region; this is related to the issue of the exact contour choice in the intermediate prescription, which we will not settle here. We are also integrating over the delta function as if it were real instead of holomorphic; similar manipulations were used in [134].

Note that the $1|4$ delta functions in (2.77), combined with the gauge conditions (2.71), (2.72), are enough to set all $\alpha_0'^{\mathbb{A}} = \alpha_0^{\mathbb{A}}$. This was the main motivation for this gauge-fixing; the point $\alpha_0^{\mathbb{A}}$ represents the intersection of the two curves, and the remaining moduli are precisely the ones we had for the degenerating degree d curve in (2.66). Therefore we easily see that the measures agree, including the prefactor $d_1 d_2$. (Although we have not been careful about overall constant factors, the absence of a relative factor here is important — it corresponds to the absence of prefactors weighing different diagrams in the intermediate prescriptions.)

This completes the argument for the equivalence between the connected and $K = 1$ prescriptions. It also sets up the iteration we described in section 2.4.1 to prove the equivalence of all prescriptions, by successive localization to poles in higher and higher codimension, corresponding to more and more degenerate curves.

One detail remains: we have to check that the residues we obtain are always independent of which prescription we started with. In other words, we have to prove that the measure for the integral over $K+1$-component trees obtained by some degeneration process always agrees with the measure coming from the disconnected prescription. As we know from section 2.4.2, the latter measure can be written as a product of measures for the individual curves, with delta-functions that guarantee the curves intersect. We just proved the agreement for $K = 1$. For general K we can work inductively; given a $K+1$-component tree on which some curve is further degenerating, just focus on the measure for that curve, and note that the delta-functions from the other curves are well behaved on moduli space near the degeneration we are studying. In this sense the degenerating curve can be isolated from the rest of the tree. The computation done above in the $K = 1$ case then shows that the measure after this degeneration agrees with that from the disconnected prescription. This then completes the argument for the equivalence of all prescriptions.

2.5 Conclusions and open questions

We have argued for the equivalence of the connected and disconnected twistorial formulae for the tree level scattering amplitudes of $\mathcal{N} = 4$ super Yang-Mills, provided that the contours are appropriately chosen. Using this equivalence we can now exploit the complementary virtues of the two prescriptions simultaneously. As we remarked in the introduction, the connected prescription minimizes the number of diagrams one has to sum, namely, there is only one; the amplitude is expressed as a single integral, which was the starting point for several theoretical developments [20, 21, 134]. The disconnected prescription involves more diagrams, but still a manageable number for some interesting amplitudes, and the contribution from each diagram can be immediately written down.

To conclude, we summarize some of the many remaining open problems in this area:

- **Contours I.** Is there a rigorous justification of the choice of contours in all these calculations? In our argument for the equivalence between connected and disconnected prescriptions we identified specific poles in the integral over moduli, and we roughly wanted a contour which encircles all of these poles. We believe it should be possible to show by a deformation argument that our choice of contour is equivalent to the one used in [107], thus completing the proof of equivalence, but this seems to be nontrivial; the computations in [107] depend on a particular method of evaluating the integral in the connected prescription by saturating delta-functions, and it is difficult to see which contour it corresponds to.

- **Contours II.** Once the residues are isolated in both prescriptions, we must still integrate over the degeneration locus $\mathcal{M}_{\text{int}}$, which requires yet another choice of contour; for example, the integration over t from 0 to ∞ in section 6 of [31] should have some *a priori* justification. This chapter has not addressed this question. Our argument for the equivalence requires that the contours on $\mathcal{M}_{\text{int}}$ are chosen to be equivalent in all prescriptions.

- **Explicit external wavefunctions.** Our derivation was rather formal. It did not depend on the particular form of the wavefunctions. Of course, it would be interesting to verify the picture by calculating the amplitudes involving particles with well-defined momenta *i.e.* $(\lambda, \tilde{\lambda}, \psi)$ using our generalized prescriptions. Unlike the MHV vertices in [31], one might expect that the $d > 1$ vertices will be ratios of polynomials involving both λ and $\tilde{\lambda}$. (Of course, it is also possible that one will not obtain any compact formula for the $d > 1$ vertices in this way.)

- **Derivation from the B-model.** Both connected and disconnected contributions seem to arise in the topological B-model of [133] as long as we use not only the degree d D1-instantons but also the propagators (and vertices) of the holomorphic Chern-Simons theory. Does our equivalence suggest that the D1-instantons are not independent of the Chern-Simons degrees of freedom?

- **Real versions.** The framework first proposed by Berkovits [20] and the topological A-model of [94] seem to prefer the real version of the twistor space, $\mathbb{RP}^{3|4}$, and correspondingly real values of the moduli. Is there a real variation of our procedures? One can imagine that the disconnected rules for the amplitudes might be derived from the cubic twistorial string field theory of [21] if K stringy propagators are expanded in component fields, so that the different parts of the worldsheet become effectively disconnected.

- **Choice of prescriptions.** According to our analysis, there is significant freedom to choose a twistor prescription for tree diagrams; we gave d different rules, all of which agree up to overall prefactors. Is this all one can say, or would a more sensitive study give more information about which is the "correct" prescription? Does this proliferation of prescriptions persist beyond tree level?

- **Loops and higher genus.** We only studied tree diagrams, corresponding to genus zero curves. What are the exact rules and equivalences between various formulae for loop and nonplanar amplitudes? Our analysis suggests that an investigation of possible degenerations of genus g curves should be relevant for the understanding of loop diagrams in the twistor string.

Chapter 3

BPS microstates and the open topological string partition function

3.1 Introduction

The connection between topological strings and 4-dimensional BPS black holes has been studied in recent years [42, 43, 44], leading to a conjecture [98] that identifies the mixed grand canonical partition function of BPS black hole states with the squared norm of the topological string wave function: $Z_{\mathrm{BH}} = |\psi_{\mathrm{top}}|^2$. This conjecture has been checked for certain Calabi-Yau threefolds [124, 10, 40]; see also the recent related work [112, 41, 111]. It is natural to ask how the conjecture generalizes to the case of open topological strings. Our primary aim in this chapter is to advance a conjecture about what the open topological string counts, and to check it in the case of certain non-compact Calabi-Yau spaces.

We will mainly concentrate on the Type IIA superstring (and correspondingly the topological A model) on a non-compact Calabi-Yau threefold. In the closed string context, one defines the mixed black hole ensemble by fixing the number of D4 and D6-branes (magnetic charges) while summing over all possible numbers of D2 and D0-branes bound to them (electric charges), weighed by chemical potentials; this was the setup investigated in [124, 10]. In our case the Type IIA background will additionally include a finite number of "background" D4-branes, which wrap Lagrangian 3-cycles of the Calabi-Yau and fill a 1+1 dimensional subspace of Minkowski spacetime. In the presence of these background D4-branes one gets a gauge theory in 1+1 dimensions, containing new BPS states. The role of the electric charges is played by open D2-branes, wrapped on holomorphic discs ending on the Lagrangian 3-cycles, while the magnetic charges are domain walls in the 1+1 dimensional theory. We conjecture that the full topological string amplitude, including contributions from open strings, is counting degeneracies of these BPS states, bound to D6, D4, D2 and D0-branes:

$$Z_{\mathrm{BPS}}^{\mathrm{open}} = |\psi_{\mathrm{top}}^{\mathrm{open}}|^2. \tag{3.1}$$

Here $Z_{\mathrm{BPS}}^{\mathrm{open}}$ is the partition function of a mixed grand canonical ensemble; in this ensemble the D6 and D4-brane charges, as well as the domain wall charge, are fixed (and related to the real part of the topological string moduli), while chemical potentials are turned on for the D2 and D0-branes (giving the imaginary parts of the moduli), including the open D2-branes.

Our proposal is necessarily more tentative than the one given in [98], because one of the major planks supporting the conjecture there is missing here: the large-charge macroscopic/gravitational description of the BPS states we are counting has not been studied, nor has the analogue of the attractor mechanism for these states, so we do not even have a classical derivation of the entropy. Further investigations in this direction would be extremely useful to check our conjecture.

Although we do not understand the macroscopic description of these BPS states, we can still compare $|\psi_{\text{top}}^{\text{open}}|^2$ to a partition function computed from their microscopic description, in cases where such a description is available. In this chapter we use such a description to check our proposal on a particular non-compact Calabi-Yau space supporting a compact Riemann surface. This case was previously discussed in [124, 10] where the closed string conjecture was verified. We find that our conjecture also holds in this case.

The organization of this chapter is as follows. In Section 3.2 we review the conjecture in the closed string case and review its confirmation in the context of local Riemann surfaces inside a Calabi-Yau. In Section 3.3 we explain the unexpected appearance of open topological string amplitudes in [10], reinterpreting them in terms of purely closed topological strings along the lines of the original conjecture [98]. In Section 3.4 we discuss the wave function nature of the open topological string. In Section 3.5 we introduce additional branes in our physical string background and state our main conjecture. In Section 3.6 we check the conjecture in the context of a local Calabi-Yau geometry near a Riemann surface with Lagrangian D-branes included. Most of the computations are relegated to the appendices: In Appendix 3.7 we fix some group theory conventions and review some basic group theory facts. In Appendix 3.8 we review the q-deformed Yang-Mills theory in 2 dimensions and the computation of its amplitudes by gluing, including insertion of some eigenvalue freezing operators important for this chapter. In Appendix 3.9 we express the wave function of q-deformed 2d Yang-Mills on the disc in terms of theta functions. Finally, Appendix 3.10 discusses many issues related to the large N limit of our computations, and the factorization of the BPS partition function at large N in terms of topological and anti-topological contributions. In particular, we give a physical explanation of the factorization of the q-deformed Yang-Mills amplitudes in the large N limit.

3.2 The closed string case

In [98] a duality was conjectured which relates counting of microstates of supersymmetric black holes which arise in compactification of type II string theory on a Calabi-Yau threefold X and closed topological string theory on X. In this section we review this conjecture and one case in which it has been explicitly checked.

Consider Type IIA on $X \times \mathbb{R}^{3,1}$. One can obtain charged BPS black holes in $\mathbb{R}^{3,1}$ by wrapping D6, D4, D2 and D0-branes over holomorphic cycles in X. The charges of the black hole are determined by the choice of holomorphic cycles; the intersection pairing in X gives rise to the electric-magnetic pairing in $\mathbb{R}^{3,1}$, and we refer to D6 and D4-brane charges as "magnetic" while D2 and D0-brane charges are "electric." Then one can define a mixed ensemble of BPS black hole states by fixing the D6 and D4-brane charges Q_6, Q_4, and summing over D2 and D0-brane charges with fixed chemical potentials φ_2, φ_0. One can write a partition function for this ensemble,

$$Z_{\mathrm{BH}}(Q_6, Q_4, \varphi_2, \varphi_0) = \sum_{Q_2, Q_0} \Omega_{Q_6,Q_4,Q_2,Q_0} e^{-Q_2\varphi_2 - Q_0\varphi_0}. \tag{3.2}$$

Here Ω_{Q_6,Q_4,Q_2,Q_0} is the contribution from BPS bound states with fixed D-brane charge.

The conjecture of [98] is that

$$Z_{\mathrm{BH}}(Q_6, Q_4, \varphi_2, \varphi_0) = |\psi_{\mathrm{top}}(g_{\mathrm{top}}, t)|^2, \tag{3.3}$$

where $\psi_{\mathrm{top}}(g_{\mathrm{top}}, t)$ denotes the A model topological string partition function, evaluated at the topological string coupling[1]

$$g_{\mathrm{top}} = \frac{4\pi i}{i\frac{\varphi_0}{\pi} + Q_6}, \tag{3.4}$$

and Kähler parameter

$$t = \frac{1}{2} g_{\mathrm{top}} \left(i\frac{\varphi_2}{\pi} + Q_4 \right). \tag{3.5}$$

The real parts of the parameters (3.4) and (3.5) are dictated by the “attractor mechanism” of $\mathcal{N} = 2, d = 4$ supergravity [52, 117], which relates the moduli of X near a black hole horizon to the black hole charges.

One can (at least formally) invert the relation (3.3) to recover the microcanonical degeneracies Ω from $|\psi_{\mathrm{top}}|^2$, via the integral formula

$$\Omega_{Q_6,Q_4,Q_2,Q_0} = \int d\varphi_2 \, d\varphi_0 \, e^{Q_0\varphi_0 + Q_2\varphi_2} |\psi_{\mathrm{top}}|^2. \tag{3.6}$$

This formula has a natural interpretation from the point of view of the wave function interpretation of ψ_{top} developed in [131, 45] as an interpretation of the holomorphic anomaly [23, 24]. Namely, (3.6) expresses Ω as the “Wigner function” (phase-space density) associated to ψ_{top}. The background-independent generalization of this transform and its relation to the counting of black hole states has been further elucidated in [126].

[1] Q_4 is naturally a class in $H_4(X, \mathbb{Z})$, which we are relating to $t \in H^2(X, \mathbb{C})$, and Q_6 is naturally a class in $H_6(X, \mathbb{Z})$, which we are relating to $H^0(X, \mathbb{C}) = \mathbb{C}$.

The formula (3.6) also illustrates a crucial point about the conjecture: in order to use it to compute Ω, one would need to know the full $|\psi_{\mathrm{top}}|^2$, not only its asymptotic expansion for $g_{\mathrm{top}} \ll 1$. Put another way, knowing the BPS degeneracies Ω is in some sense equivalent to having a nonperturbative completion of $|\psi_{\mathrm{top}}|^2$.

A solvable example

In this section we review the work of [124, 10] which argued that the conjecture (3.3) holds in the case where X is a particular non-compact Calabi-Yau threefold, namely the total space of a holomorphic vector bundle over a compact Riemann surface Σ of genus g,

$$X = \mathcal{L}_{-p} \oplus \mathcal{L}_{p+2g-2} \to \Sigma, \tag{3.7}$$

for some $p > 0$.[2]

The idea is that for this X one can use 2-dimensional Yang-Mills theory to compute Z_{BH}, as follows. Suppose we wrap N D4-branes on the holomorphic 4-cycle

$$\mathcal{D} = \mathcal{L}_{-p} \to \Sigma. \tag{3.8}$$

Then the theory on the D4-branes (in the Calabi-Yau directions) is the $\mathcal{N} = 4$ supersymmetric Yang-Mills theory, or more precisely a topologically twisted version of that theory, as explained in [26]. The path integral in this theory includes configurations in which D0-branes, and D2-branes wrapping Σ, are bound to the D4-branes. Hence the partition function of the 4-dimensional twisted supersymmetric gauge theory computes a sum over the mixed ensemble of BPS states which we considered

[2]By $\mathcal{L}_k$ we mean a holomorphic line bundle of degree k over Σ.

above. The D4 and D6-brane charges are

$$Q_4 = N[\mathcal{D}], \tag{3.9}$$

$$Q_6 = 0. \tag{3.10}$$

The chemical potentials for the brane charges are roughly given by the masses of the branes (for the D2-branes we turn on a Ramond-Ramond field θ):

$$\varphi_0 = 4\pi^2/g_s, \tag{3.11}$$

$$\varphi_2 = 2\pi p\theta/g_s. \tag{3.12}$$

Since the gauge theory sums over all brane charges we can now write[3]

$$Z_{\mathrm{YM}} = Z_{\mathrm{BH}}. \tag{3.13}$$

It was argued in [124] that, for the purpose of computing Z_{YM}, we can restrict to field configurations in the $\mathcal{N} = 4$ theory which are invariant under the $U(1)$ action on the fibers of $\mathcal{L}_{-p}$. One then obtains Z_{YM} as the partition function of a q-deformed Yang-Mills theory on Σ (see Appendix 3.8), where Σ has area p and the parameters are fixed by

$$\theta_{\mathrm{YM}} = \theta, \quad g_{\mathrm{YM}}^2 = g_s, \quad q = e^{-g_s}. \tag{3.14}$$

The q-deformed Yang-Mills theory is a relative of the ordinary Yang-Mills theory in two dimensions, and shares with that theory the property of being exactly solvable; the topological string on X is also exactly solvable to all orders in perturbation theory (using recent results of [28] in the case $g > 1$). Hence we can use X as a

[3] There are some subtleties because of the non-compactness of X, as noted in [10]: Z_{YM} turns out to give a sum over finitely many sectors, each with a g_s-dependent prefactor.

testing ground for (3.3). More precisely, since we do not have a good understanding of the nonperturbative topological string, what we can do is look at the asymptotic expansion of $|\psi_{\rm top}|^2$ in the limit $g_s \ll 1$, with t fixed. On the physical side this corresponds to taking φ_0, φ_2, and N to infinity with fixed ratios (this is a 't Hooft limit in the Yang-Mills theory.)

In this limit one finds that $Z_{\rm YM}$ factorizes into a sum of "conformal blocks," each given by the topological string on X, with some D-branes inserted as we will explain below:

$$Z_{\rm YM}(\varphi_0, \varphi_2, N) = \sum_{R'_1,\dots,R'_{|2g-2|}} \sum_{l\in\mathbb{Z}} \psi_{\rm top}^{R'_1,\dots,R'_{|2g-2|}}(g_{\rm top}, t + lpg_{\rm top}) \overline{\psi_{\rm top}^{R'_1,\dots,R'_{|2g-2|}}(g_{\rm top}, t - lpg_{\rm top})} + \mathcal{O}(e^{-N}). \tag{3.15}$$

Here t and $g_{\rm top}$ are as dictated by (3.4) and (3.5), namely,

$$g_{\rm top} = 4\pi^2/\varphi_0 = g_s, \tag{3.16}$$

$$t = \frac{1}{2} g_{\rm top} \left(\#(\Sigma \cap \mathcal{D})N + i\varphi_2/\pi\right) = \frac{1}{2} N(p + 2g - 2)g_s + ip\theta. \tag{3.17}$$

The index l was interpreted in [124] as measuring the Ramond-Ramond flux through Σ. The labels R'_i are subtler; they appear only when $g \neq 1$, in which case they were interpreted in [10] as running over boundary conditions on $|2g-2|$ infinite stacks of D-branes (which we christen "ghost" D-branes) in the topological string. Each stack lies on a Lagrangian submanifold of X, intersecting $\mathcal{D}$ in an S^1 in the fiber of $\mathcal{L}_{p+2g-2}$ over a point. The boundary conditions on each stack are specified by a choice of a Young diagram R'.[4]

[4]All primed quantities which appear in this chapter are associated to these ghost D-branes.

The form of (3.15) looks different from that of (3.3). Nevertheless, as we will explain in the next section, the sum over Young diagrams R'_i is indeed consistent with (3.3), when we take into account extra closed string moduli at infinity.

3.3 Revisiting the closed string theory

In this section we revisit the relation between 2-d Yang-Mills theory and the closed topological string, with the aim of giving a better interpretation to the sum over chiral blocks and the appearance of "ghost" D-branes.

As we reviewed in Section 3.2, the partition function of the twisted $U(N)$ Yang-Mills theory on $\mathcal{D} = \mathcal{L}_{-p} \to \Sigma$ factorizes at large N as a sum of blocks, each of which can be interpreted as the square of a topological string amplitude involving $2g-2$ infinite stacks of ghost branes. Introducing a $U(\infty)$-valued holonomy $U'_i = e^{u'_i}$ on each stack of ghost branes, we can rewrite (3.15) as

$$Z_{\text{YM}} = \sum_{l\in\mathbb{Z}} \int d_H u'_1 \cdots d_H u'_{2g-2} \; \psi^{\text{g}}_{\text{top}}(g_{\text{top}}, u', t + lpg_{\text{top}}) \overline{\psi^{\text{g}}_{\text{top}}(g_{\text{top}}, u', t - lpg_{\text{top}})}, \quad (3.18)$$

where

$$\psi^{\text{g}}_{\text{top}}(g_{\text{top}}, u', t) = \sum_{R'_1,\ldots R'_{2g-2}} \psi_{\text{top}}^{R'_1,\ldots,R'_{2g-2}}(g_{\text{top}}, t) e^{-\frac{1}{2}Ng_s \sum_{i=1}^{2g-2}|R'_i|} \prod_{i=1}^{2g-2} s_{R'_i}(e^{u'_i}). \quad (3.19)$$

For $g = 0$ the formula is similar, except that the role of ghost branes and ghost antibranes are reversed in the antitopological amplitude:

$$Z_{\text{YM}} = \sum_{l\in\mathbb{Z}} \int d_H u'_1 d_H u'_2 \; \psi^{\text{g}}_{\text{top}}(g_{\text{top}}, u', t + lpg_{\text{top}}) \overline{\psi^{\text{a}}_{\text{top}}(g_{\text{top}}, u', t - lpg_{\text{top}})}, \quad (3.20)$$

where

$$\psi^{\mathrm{g}}_{\mathrm{top}}(g_{\mathrm{top}}, u', t) = \sum_{R'_1, R'_2} \psi^{R'_1, R'_2}_{\mathrm{top}}(g_{\mathrm{top}}, t) e^{-\frac{1}{2} N g_s (|R'_1| + |R'_2|)} s_{R'_1}(e^{u'_1}) s_{R'_2}(e^{u'_2}), \tag{3.21}$$

$$\psi^{\mathrm{a}}_{\mathrm{top}}(g_{\mathrm{top}}, u', t) = \sum_{R'_1, R'_2} (-)^{|R'_1| + |R'_2|} \psi^{R'_1, R'_2}_{\mathrm{top}}(g_{\mathrm{top}}, t) e^{-\frac{1}{2} N g_s (|R'_1| + |R'_2|)} s_{R'^t_1}(e^{u'_1}) s_{R'^t_2}(e^{u'_2}). \tag{3.22}$$

The change from branes to antibranes is reflected in the signs $(-)^{|R'|}$ and the switch $R' \to R'^t$ between ψ^{g} and ψ^{a}, as in [7].

Now note that (3.18) and (3.20) look like the integral (3.6), that computes the microcanonical degeneracies by integrating over the imaginary part of each Kähler modulus while the real part is fixed by the corresponding magnetic charge. Indeed, the factor $e^{-\frac{1}{2} N g_s \sum_{i=1}^{2g-2} |R'_i|}$ could be absorbed in U', at the expense of making it non-unitary: this just amounts to giving u' a real part. This is reminiscent of the "attractor" formula (3.5), which says the real part of the Kähler modulus is related to the charge. So indeed, (3.18) could be consistent with the conjecture (3.3), if we somehow regard u' as an extra closed string modulus; then there would be electric and magnetic charges corresponding to it, and (3.18) says that Z_{YM} is the partition function of an ensemble in which we have fixed these charges. As we will now explain, this interpretation of u' is indeed plausible.

Open vs. closed

We explained above that the nonperturbative completion of the closed topological string appears to involve Lagrangian D-branes on the Calabi-Yau manifold. The appearance of open string amplitudes in this context is surprising, since in the physical string this would have half as much supersymmetry as we have available. As we will now argue, the correct interpretation involves not open but closed strings.

Namely, as was shown in [5], in the topological string, inserting non-compact D-branes is equivalent to turning on certain non-normalizable deformations of the Calabi-Yau. This is an open-closed duality of the topological string, generalizing the well-known duality for D-branes on compact cycles. This means that, at the level of the topological string, we can interpret the modulus U' in (3.18) as either corresponding to an open string configuration or to a boundary condition at infinity of the closed topological string. In the physical string theory, however, we do not have this freedom; since there are no Ramond-Ramond fluxes turned on, the only interpretation available is the closed string one.

The torus symmetries of the Calabi-Yau manifold can be used to constrain the types of deformation that we consider. Namely, the Lagrangian D-branes to which (3.19) corresponds respect the torus symmetries, and the gravitational backreaction they create does so as well. Such torus invariant deformations, normalizable and not, were studied in [5], so we can borrow the results of that paper. The topological string theory in [5] was described as the theory of a chiral boson on a Riemann surface, and the Lagrangian D-branes were coherent states of this chiral boson. (Note here that we are using the mirror B-model language. The global action of mirror symmetry on X is not relevant for us; this is merely a convenient language in which to describe the behavior near an asymptotic infinity.) The non-normalizable deformations of the Calabi-Yau near an asymptotic infinity[5] can be parameterized by the coherent states

[5]In the cases studied in [5] there is a clear notion of what "an asymptotic infinity" means: it means a toric 2-cycle which extends to infinity. In the cases we are considering here the situation is not as rigorously understood, but we will make some comments below.

of the chiral boson:

$$|\tau\rangle = \exp\left(\sum_{n>0} \tau_n \alpha_{-n}\right)|0\rangle, \tag{3.23}$$

where α_n are the chiral boson creation and annihilation operators.

The parameters τ are related to the D-brane holonomies by

$$\tau_n = g_s \mathrm{Tr} U'^n, \tag{3.24}$$

where Tr denotes the trace in the fundamental representation, The factor of g_s is needed to convert an open string amplitude in terms of U to a closed string amplitude in terms of t; it appears because a trace of U in the fundamental representation couples to a hole in the string worldsheet, and the hole is in turn weighted by g_s in the string perturbation expansion. In this sense the open string modulus U can be traded for the infinite collection of closed string moduli τ_n.

Actually, it is more convenient to reparameterize slightly by taking a logarithm, writing $\tau_n = e^{-t_n}$. The point is that the A model partition function turns out to be an expansion in e^{-t_n}, so the moduli t_n appear on the same footing as the Kähler volumes t of compact cycles. Indeed, we can think of them as representing Kähler volumes of classes in $H_2(X,\mathbb{Z})$ (with some appropriate notion of what $H_2(X,\mathbb{Z})$ means for this non-compact X.) What can we say about these classes? In the cases considered in [5], for each asymptotic infinity there is a holomorphic disc C which "ends" on it, and t_n represents a class which contains $n[C]$ as well as some extra contributions at infinity. In the open string language, the disc C can be thought of as ending on the Lagrangian branes which represent the deformations at this asymptotic infinity.

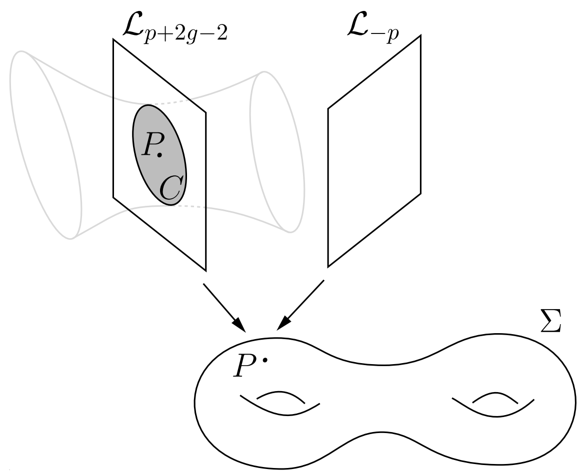

Figure 3.1: The disc C in the fiber of $\mathcal{L}_{p+2g-2}$ over a point P on the Riemann surface Σ; C meets $\mathcal{D}$ only at P, and the boundary of C lies on the Lagrangian submanifold representing this asymptotic infinity.

The attractor mechanism and ghost D-branes

Now we come to the interpretation of the shift $U' \to U' e^{-\frac{1}{2}Ng_s}$, or equivalently

$$\mathrm{Re}\ t_n = \frac{1}{2} nNg_s. \tag{3.25}$$

Such shifts have frequently appeared in the topological string in the presence of D-branes. Here we can understand the shift as a reflection of the attractor mechanism on the closed string moduli. Namely, in the case we are considering here, C is a disc in the fiber of $\mathcal{L}_{p+2g-2}$, which intersects $\mathcal{D}$ at one point, as shown in Figure 3.1. Then $\frac{1}{2}nNg_s$ is exactly the expected attractor value for the Kähler modulus t_n, as follows from (3.5), the fact that the D4-brane charge is $Q_4 = N[\mathcal{D}]$, and $\#(C \cap \mathcal{D}) = 1$. (Whatever the extra contributions at infinity to the class represented by t_n are, they have zero intersection number with $\mathcal{D}$, so they do not affect the attractor modulus.)

Why $2g-2$ asymptotic infinities

The discussion of the last few sections raises a natural question: why are there precisely $|2g-2|$ asymptotic infinities on X where we can have deformations?

In general we should have expected that in a non-compact Calabi-Yau we should include some closed string moduli coming from infinity. However, in problems with symmetries, it is natural to conjecture that the only relevant extra moduli from infinity are invariant under the corresponding symmetries. We will assume this here, and look for symmetries in our problem which simplify the task of specifying the closed string moduli coming from infinity.

A priori, one might have expected boundary moduli associated to the $\mathbb{C}^2$ fiber over each point of the Riemann surface. Here we have in addition D4-branes wrapping a line bundle over the Riemann surface. We claim that this implies that effectively we should view that direction as "compact," or more precisely, we should view it as a degenerate limit of a compact 4-cycle. After this reduction, we would expect to find boundary moduli corresponding to a $\mathbb{C}$ fiber over each point on the Riemann surface.

However, there are symmetries of the problem coming from meromorphic vector fields on the Riemann surface. Hence the variation of the data at infinity can be localized at poles or zeroes of such a vector field (deleting these points would give a well defined free action). A generic holomorphic vector field on a Riemann surface of genus $g>1$ is nonvanishing and well defined away from $2g-2$ poles, which we identify with places where the asymptotic boundary condition at infinity can be localized. The local picture is as shown in Figure 3.2.

So the closed string moduli at these $2g-2$ asymptotic infinities may be identified

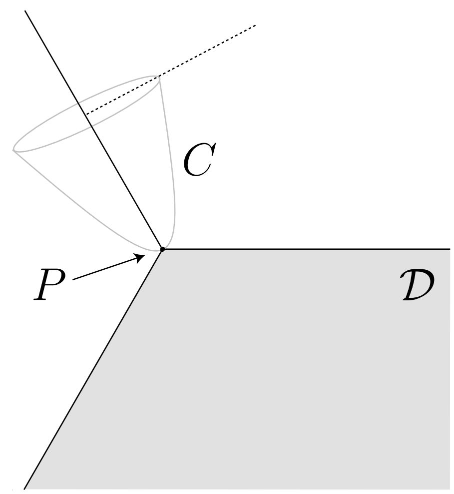

Figure 3.2: A rough toric representation of the behavior of X in a neighborhood of a singularity of the vector field v described in the text. Two of the three $U(1)$ actions making up the toric fiber are the rotations of the line bundles $\mathcal{L}_{-p} \oplus \mathcal{L}_{p+2g-2}$ and the third is the action of v. The toric base of the divisor $\mathcal{D}$ on which the D4-branes are wrapped is indicated, as is the base of the Lagrangian submanifold representing the asymptotic infinity. The disc C ends on this Lagrangian submanifold, meeting $\mathcal{D}$ at the single point P.

with the "ghost D-brane" contributions, as discussed above. In the case of genus 1 there are no fixed points, which is consistent with the fact that no ghost D-branes were needed in this case. For genus 0 we have a holomorphic vector field with 2 zeroes, which again suggests that we can localize the contribution from infinity at 2 points.

This is a heuristic argument, but we feel that it captures the correct physics.

3.4 The quantum mechanics of open strings

In Section 3.2 we reviewed the conjecture of [98] and its relation to the wave function nature of the closed topological string. In this section we recall the parallel statement for the open topological string. The fact that the open topological string partition function including non-compact branes is a wave function was first noticed in [5], and was crucial in that paper for the solution of the B model. In this section we will give two ways of understanding this wave function property: a direct route via canonical quantization of Chern-Simons theory, and a more indirect one via the holomorphic anomaly (background dependence) for open strings.

Canonical quantization in Chern-Simons

Recall that the topological A model string theory on M D-branes wrapped on a Lagrangian cycle L is the $U(M)$ Chern-Simons theory deformed by worldsheet instanton corrections:

$$S = S_{\mathrm{CS}} + S_{\mathrm{inst}}, \tag{3.26}$$

where

$$S_{\mathrm{CS}} = \frac{4\pi i}{k} \int_L \mathrm{Tr} \left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A \right), \tag{3.27}$$

and S_{inst} is the contribution from worldsheet instantons with boundaries on L. If L is non-compact, then we should consider it as having a boundary ∂L at infinity; the path integral on L then gives a wave function in the Hilbert space of Chern-Simons on the boundary. The case of interest for the rest of this chapter is $L \simeq \mathbb{R}^2 \times S^1$, which has $\partial L = T^2$; from now on we specialize to that case, although the discussion could be made more general.

To find which state the topological open string theory picks, we need to recall some facts about canonical quantization of the $U(M)$ Chern-Simons theory on $T^2 \times \mathbb{R}$, viewing $\mathbb{R}$ as the "time" direction. We will be brief here; see e.g. [48] for more details.

Integrating over the time component of the gauge field localizes the path integral to flat connections on T^2:

$$\int \mathcal{D}A'\,\delta(F')\,\exp\left(\frac{2\pi i}{k}\int_{T^2\times R}\mathrm{Tr}A'\,\partial_t\,A'dt\right). \tag{3.28}$$

Above A' is a connection on T^2, which we can write (up to conjugation) as

$$A' = u\,d\theta_u + v\,d\theta_v, \tag{3.29}$$

where u and v are the components of A' along two linearly independent cycles of T^2, with intersection number 1. From the action (3.28) we see that u and v are conjugate variables: upon quantization we thus expect

$$[u,v] = ig_{\mathrm{top}}, \tag{3.30}$$

where $g_{\mathrm{top}} = \frac{2\pi}{k+M}$. The familiar shift of k by M can be seen by carefully integrating over massive modes [48].

Since u and v are conjugate variables, in computing the Chern-Simons path integral on a manifold with T^2 boundary, we should fix either u or v on the boundary, but not both, and the wave function will depend on the variable we have chosen to fix. More generally, we could consider a mixed boundary condition where we fix $v+\overline{\tau}u$ where $\overline{\tau}$ is some parameter (the motivation for this notation will become clear later).

Note that in the present context $L \simeq \mathbb{R}^2\times S^1$ is a solid torus, so there is a unique 1-cycle $\eta \in H_1(T^2,\mathbb{Z})$ which collapses in the interior of L. There is thus a canonical choice of polarization for the wave function; namely, one can express it in terms of the holonomy around η, which we call v. In the next subsection we will relate this choice to the background dependence ("holomorphic anomaly") of the open topological string. We could have tried to choose the "cycle that survives" in the interior of L (corresponding to the holonomy u), but this is ambiguous up to the shift $u \mapsto u + nv$. This ambiguity will be related to the framing ambiguity of the open topological string.

It can be shown [48, 6] that the Chern-Simons path integral on the solid torus, without any insertions and with u fixed on the boundary, is given simply by

$$\psi^{\mathrm{open}}_{\mathrm{top}}(u) = \langle L|u\rangle = 1. \tag{3.31}$$

In the present context, the Chern-Simons action is deformed by worldsheet instantons wrapping holomorphic curves with boundaries on L [132]. Their contribution to S is given by the free energy of the gas of topological open strings:

$$S_{\mathrm{inst}}(u) = iF^{\mathrm{open}}_{\mathrm{top}}(u). \tag{3.32}$$

We now want to compute the path integral on L with the operator insertion

$$\exp S_{\mathrm{inst}}(u). \tag{3.33}$$

Since we are we are working in the basis of eigenstates of u, the insertion just acts by multiplication:

$$\psi^{\mathrm{open}}_{\mathrm{top}}(u) = \langle L|e^{S_{\mathrm{inst}}(u)}|u\rangle = e^{iF^{\mathrm{open}}_{\mathrm{top}}(u)}\langle L|u\rangle = e^{iF^{\mathrm{open}}_{\mathrm{top}}(u)}. \tag{3.34}$$

So we have identified the topological string partition function $e^{iF^{\mathrm{open}}_{\mathrm{top}}(u)}$ with a wave function.

Although v is the canonical choice, we will sometimes find it natural to write the wave function in terms of one of the holonomies $u+nv$ instead. The relation between different choices of variable in which to write the wave function is given by a Fourier transform: for example, to transform from u to v, one has

$$\psi_{\rm top}^{\rm open}(v) = \int d_H u \, e^{\frac{i}{g_{\rm top}} {\rm Tr} uv} \, \psi_{\rm top}^{\rm open}(u), \tag{3.35}$$

where $d_H u$ is the measure induced from the Haar measure on $U(M)$.

The freedom to choose a variable is crucial because there are some cases in which the Lagrangian cycle L can make a "flop transition." From the perspective of the boundary $\partial L = T^2$ nothing special happens at the transition, but in the interior of L the topology changes and in particular the cycle that collapses in the interior is different after the transition. An example of this phenomenon can be seen when X is a toric Calabi-Yau manifold. Moreover, in that case one can use the mirror B model to see that worldsheet instanton corrections eliminate the sharp transition: the different phases are smoothly connected. Thus, in the B model language there is a continuous change of variables which takes us from one choice of holonomy to another. This is related to the background dependence of open topological string amplitudes, to which we now turn.

Background dependence for the open topological string

In this section we take a brief detour to explain the background dependence of the open string topological string. It was conjectured in [5] that the open topological string partition function depends on a choice of "background" moduli, or equivalently, depends on the antiholomorphic coordinates of the moduli as well as the holomorphic ones. This conjecture was advanced in order to explain the fact that the open topological string behaves like a wave function, by analogy to what is known for the closed string case [131]. In the case considered in [5], the geometry of the Calabi-Yau is given (in the mirror B model) by a hypersurface in $\mathbb{C}^4$,

$$F(u,v) - xy = 0, \tag{3.36}$$

and the mirror of the Lagrangian brane is a brane on a holomorphic curve, specified by the condition $x = 0$ together with fixed choices of u, v satisfying $F(u,v) = 0$. As noted in [6, 5] the geometry with this D-brane included can be viewed as a special (degenerate) limit of a closed string geometry, with the D-brane serving as a source for the holomorphic 3-form; this source changes the usual equation $d\Omega = 0$ to

$$d\Omega = g_{\mathrm{top}}\delta(D), \tag{3.37}$$

where $\delta(D)$ denotes a delta function at the locus of the D-brane. We have already used this correspondence in Section 3.3, where we discussed how the "ghost branes" can be viewed as closed string moduli. Similarly, we can use it to interpret the holomorphic anomaly of closed strings as inducing a holomorphic anomaly (or equivalently a background dependence) for the open string partition function. Here we view the modulus of the open string, given by the choice of (u,v) on the surface $F(u,v) = 0$, as a closed string modulus. In fact, borrowing the closed string technology for background dependence developed in [131, 45] we immediately deduce that for a given background (u_0, v_0) the natural variable for the open string wave function is

$$v + \overline{\tau} u, \tag{3.38}$$

where

$$\tau = -\frac{\partial v}{\partial u}\bigg|_{(u_0,v_0)}. \tag{3.39}$$

Here we are considering v as a function of u through the implicit relation $F(u, v) = 0$, so τ is the slope of the tangent plane to the Riemann surface at (u_0, v_0). Note that $\tau = \partial^2 F/\partial u^2$.

The form (3.38) of the natural variable can be connected to our earlier discussion of the wave function nature of the Chern-Simons theory embedded in the open string; there too we claimed that there is a natural variable for the wave function, namely the holonomy around the cycle of T^2 which shrinks in the interior of the solid torus. In that classical picture (which neglects the effect of worldsheet instantons) the holonomy around the vanishing cycle is simply v; and choosing the background point near an asymptotic infinity of the quantum moduli space, where the classical picture becomes exact, one indeed gets $\tau \to 0$, so $v + \overline{\tau} u \to v$. More invariantly, the value of τ near an asymptotic infinity of the B model Riemann surface approaches the slope of the corresponding line in the A model toric diagram, and this slope indeed determines the collapsing cycle of the toric fiber.

Note that in order to go off the real locus $\tau = \overline{\tau}$ we need to recall that the Chern-Simons holonomies are complexified in the context of topological strings (to include the moduli which move the brane); in the geometric motivation we gave before we had essentially turned those off. It would be interesting to understand this relation off the real locus.

3.5 The open string conjecture

As we reviewed in Section 3.2, the closed topological string wave function on a Calabi-Yau space X is believed to compute the large-charge asymptotics of an index which counts BPS states in four dimensions, and this index has an interpretation as the Wigner function of ψ_{top}. On the other hand, we just saw in Section 3.4 that the open topological string partition function $\psi_{\text{top}}^{\text{open}}$ with non-compact D-branes is also naturally considered as a wave function. So we could construct a Wigner function from this wave function, and then a natural question is whether this Wigner function also has an interpretation as counting BPS states. We will argue that it does.

We embed the open topological string in the superstring in a familiar way [100]. Namely, consider D4-branes wrapping a special Lagrangian cycle $L \subset X$. Then there are open D2-branes ending on these D4-branes. These give rise to BPS particles in the 2-dimensional supersymmetric gauge theory on the non-compact directions of the D4-branes; we will interpret the charge $\mathcal{Q}_e$ as counting these BPS particles. The gauge theory in question also supports BPS domain walls; we will interpret $\mathcal{Q}_m$ as measuring the domain wall charge.

Altogether then, we will conjecture below that the open topological string, on a Calabi-Yau space X with Lagrangian branes included, computes the large-charge asymptotics of an index which counts open D2-branes, and their domain wall counterparts, bound to any number of closed D0, D2, D4 and D6-branes. Furthermore, we will describe one context in which some aspects of this proposal can be checked.

Calabi-Yau spaces with branes and BPS particles

Consider a Calabi-Yau manifold X containing a special Lagrangian 3-cycle L. We consider the Type IIA superstring on $X \times \mathbb{R}^{3,1}$, with M D4-branes on $L \times \mathbb{R}^{1,1}$, which we will call the "background branes." For simplicity, we assume L has the topology

$$L \simeq \mathbb{R}^2 \times S^1. \tag{3.40}$$

The dimensionally reduced theory on the $\mathbb{R}^{1,1}$ part of the background branes is a $(2,2)$ supersymmetric gauge theory. Its field content can be understood as follows [100]. Since $b_1(L) = 1$ it follows [92] that L has one real modulus r; this modulus pairs up with the Wilson line of the worldvolume gauge field $\oint A$ to give a complex field

$$u = r + i \oint A. \tag{3.41}$$

One also gets a gauge field in $\mathbb{R}^{1,1}$ by integrating the world-volume two-form B (which is the magnetic dual to the gauge field A on the D4-brane, defined by $d{*A} = dB$) over the S^1 of L. Since there are M D4-branes, the theory has (at least) a magnetic $U(1)^M$ gauge symmetry. The field u should be viewed as the lowest component of a twisted chiral multiplet, whose top component is the field strength of the magnetic gauge field in two dimensions.

There is an obvious way of getting BPS particles in this theory. Suppose for a moment that $M = 1$ (a single Lagrangian brane.) Let $\gamma \in H_1(L, \mathbb{Z})$ denote the homology class of the S^1 in L. Since the Calabi-Yau has no non-contractible 1-cycles, this γ is a boundary in X; so there exists some D with

$$[\partial D] = \gamma. \tag{3.42}$$

Open D2-branes wrapped on D give rise to particles charged under the $U(1)$ gauge field of the 2-dimensional theory; if D is a holomorphic disc, then these particles are BPS.

The conjecture

Now, to motivate our conjecture, recall from Section 3.2 that in the closed string case (without the background branes) we have the relation

$$Z_{\mathrm{BH}}(Q_6, Q_4, \varphi_2, \varphi_0) = \sum_{Q_0, Q_2} \Omega_{Q_6,Q_4,Q_2,Q_0} e^{-Q_2\varphi_2 - Q_0\varphi_0} = |\psi_{\mathrm{top}}(g_{\mathrm{top}}, t)|^2, \tag{3.43}$$

where $\varphi_2 = \mathrm{Im}\ 2\pi t/g_{\mathrm{top}}$ and $\varphi_0 = \mathrm{Im}\ 4\pi^2/g_{\mathrm{top}}$ as given in (3.4), (3.5). We wish to generalize this conjecture to the open topological string. What is the appropriate ensemble to consider? Since the closed D2-branes are "light electric states" in the closed string ensemble, which we sum over with chemical potentials, it is natural to try treating the open D2-branes in the same way. Thus, in formulating our conjecture we consider these BPS states as "electric charges," and sum over them with a chemical potential $\varphi_e^{\mathrm{open}}$. We also expect to have a "magnetic charge," which we fix to the some value $\mathcal{Q}_m^{\mathrm{open}}$; we will discuss these charges further below.[6] The partition function of the ensemble thus obtained is a simple generalization of (3.43),

$$Z_{\mathrm{BPS}}^{\mathrm{open}}(Q_6, Q_4, \mathcal{Q}_m^{\mathrm{open}}, \varphi_2, \varphi_0, \varphi_e^{\mathrm{open}}) = \sum_{Q_0, Q_2, \mathcal{Q}_e^{\mathrm{open}}} \Omega_{Q_6,Q_4,Q_2,Q_0,\mathcal{Q}_e^{\mathrm{open}},\mathcal{Q}_m^{\mathrm{open}}} e^{-Q_2\varphi_2 - Q_0\varphi_0 - \mathcal{Q}_e^{\mathrm{open}}\varphi_e^{\mathrm{open}}}. \tag{3.44}$$

We conjecture that the relation of $Z_{\mathrm{BPS}}^{\mathrm{open}}$ to the topological string is a direct generalization of (3.43),

$$Z_{\mathrm{BPS}}^{\mathrm{open}}(Q_6, Q_4, \mathcal{Q}_m^{\mathrm{open}}, \varphi_2, \varphi_0, \varphi_e^{\mathrm{open}}) = |\psi_{\mathrm{top}}^{\mathrm{open}}(g_{\mathrm{top}}, t, u)|^2, \tag{3.45}$$

where $\psi_{\mathrm{top}}^{\mathrm{open}}$ is the topological A model partition function on X, including open strings ending on M D-branes on L as well as closed strings.

[6] The terminology "electric" and "magnetic" here is chosen by analogy to the closed string case. The charges we are discussing here are both associated to point particles, which are not electric-magnetic duals in the theory on $\mathbb{R}^{1,1}$.

In this conjecture the closed string moduli g_{top}, t are determined by the attractor mechanism as before. What about the open string modulus u? The formula $\varphi_2 = \mathrm{Im}\ 2\pi t / g_{\mathrm{top}}$ for the closed D2-brane chemical potential suggests that the open D2-brane chemical potential should be related to u by

$$\varphi_e^{\mathrm{open}} = \mathrm{Im}\ 2\pi u / g_{\mathrm{top}}. \tag{3.46}$$

We will verify this identification of Im u in an explicit example below. The real part of u should be fixed by the charge $\mathcal{Q}_m^{\mathrm{open}}$, as we now discuss.

Adding magnetic charges

What is the spacetime meaning of the "magnetic" charge $\mathcal{Q}_m^{\mathrm{open}}$, and its relation to the real part of the modulus u? We can make a plausible guess by exploiting the symmetry between u and its conjugate v. Namely, as noted in Section 3.4, it is possible for L to undergo a flop transition to a new phase parameterized by a different parameter v (representing the holonomy of the gauge field around a new S^1 which was contractible in the old phase). The two phases are smoothly connected in the quantum topological string theory and also in the physical one, but they correspond to different classical descriptions of the physics. The most economical assumption would then be that the excitations which we are calling "electric" in one description are the same as the ones which we are calling "magnetic" in the other. In this section we explore the consequences of this assumption (without being too careful about the factors of i which appear.) We discuss only the open string sector, suppressing the closed strings, and drop the label "open" from our notation for simplicity.

First, we can write down the precise form of u, using the fact that $\psi_{\mathrm{top}}(v)$ is related to $\psi_{\text{top}}(u)$ by the Fourier transform (3.35), or equivalently

$$[u, v] = ig_{\text{top}}. \tag{3.47}$$

The dictionary between our statistical ensemble and the quantum-mechanical picture requires the relations

$$[\mathcal{Q}_e, \varphi_e] = 1 = [\mathcal{Q}_m, \varphi_m], \tag{3.48}$$

since we cannot fix the charges and the chemical potentials at the same time. On the other hand, we can fix the charges simultaneously, so

$$[\mathcal{Q}_e, \mathcal{Q}_m] = 0 = [\varphi_e, \varphi_m]. \tag{3.49}$$

The consistency of (3.47), (3.48), (3.49) with Im $u = g_{\text{top}}\varphi_e/2\pi$ then requires

$$\text{Re } u = \pi \mathcal{Q}_m, \qquad \text{Re } v = \pi \mathcal{Q}_e. \tag{3.50}$$

The equation (3.50) completes our conjecture (3.45), except that we have not been precise about how to fix the zero of Re u or Re v. We do not have a general proposal for how this should be done, although we will see how it works in an example below.

Note that the expectation value of v in the state corresponding to the open string wave function $\psi_{\text{top}}(u) = \exp(iF_{\text{top}}(u))$ is given by

$$v = g_s \partial_u F_{\text{top}}(u) \tag{3.51}$$

(the semi-classical version of this equation was discovered in [8]). This is precisely analogous to the special geometry relations of the closed string. In this sense (3.50) seems to describe an open string analogue of the attractor mechanism that fixes the moduli to values determined by charges of BPS states. It would be interesting to study this attractor mechanism directly in the physical theory.

Our identification of the parameters leads to two formulas for the Wigner function, i.e. the degeneracies of BPS states,

$$\Omega_{\mathcal{Q}_e,\mathcal{Q}_m} = \int d\varphi_e \, e^{-\mathcal{Q}_e \varphi_e} \, \psi(u = \frac{ig_{\text{top}}}{2\pi}\varphi_e + \pi \mathcal{Q}_m)\overline{\psi(u = \frac{ig_{\text{top}}}{2\pi}\varphi_e + \pi \mathcal{Q}_m)}, \tag{3.52}$$

$$= \int d\varphi_m \, e^{-\mathcal{Q}_m \varphi_m} \, \psi(v = \frac{ig_{\text{top}}}{2\pi}\varphi_m + \pi \mathcal{Q}_e)\overline{\psi(v = \frac{ig_{\text{top}}}{2\pi}\varphi_m + \pi \mathcal{Q}_m)}. \tag{3.53}$$

(The arguments we gave above about commutation relations are equivalent to the statement that these two formulas are indeed related by Fourier transforming $\psi(u) \leftrightarrow \psi(v)$.) Put another way, $\psi(u)$ and $\psi(v)$ sum over conjugate ensembles,

$$|\psi(u = \frac{ig_{\text{top}}}{2\pi}\varphi_e + \pi \mathcal{Q}_m)|^2 = \sum_{\mathcal{Q}_e} \Omega_{\mathcal{Q}_e,\mathcal{Q}_m} e^{-\mathcal{Q}_e \varphi_e}, \tag{3.54}$$

$$|\psi(v = \frac{ig_{\text{top}}}{2\pi}\varphi_m + \pi \mathcal{Q}_e)|^2 = \sum_{\mathcal{Q}_m} \Omega_{\mathcal{Q}_e,\mathcal{Q}_m} e^{-\mathcal{Q}_m \varphi_m}. \tag{3.55}$$

In the above we implicitly chose some framing for the open string wave function $\psi(u)$, and one could ask what is the meaning of changing the framing. As discussed in [5], the effect of shifting the framing by k units is $\psi^{(k)}(u) = e^{-ikg_{\text{top}}\partial_u^2}\psi(u)$. From this and (3.54) it follows that $\psi^{(k)}$ sums over an ensemble in which we have a chemical potential for dyons of charge $(1, k)$:

$$|\psi^{(k)}(u = \frac{ig_{\text{top}}}{2\pi}\varphi_e + \pi \mathcal{Q}_m)|^2 = \sum_{\mathcal{Q}_e} \Omega_{\mathcal{Q}_e,\mathcal{Q}_m + k\mathcal{Q}_e} e^{-\mathcal{Q}_e \varphi_e}. \tag{3.56}$$

So far we have discussed the magnetic charge $\mathcal{Q}_m$ abstractly in terms of its relation to the real part of the topological string modulus, but our assumption also leads to a natural description of the meaning of the magnetic charges in the physical theory. To understand this, note first that turning on electric charge $\mathcal{Q}_e$, arising from open D2-branes ending on $\gamma \subset L$, can be equivalently described as turning on magnetic flux on the background D4-brane. This is because the D2-brane ending on L looks

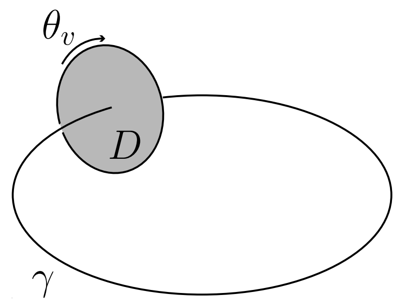

Figure 3.3: The 1-cycle γ and its dual cycle D inside L.

like a monopole string from the point of view of the gauge theory on the D4-brane. So, letting D denote any 2-cycle in L dual to γ (see Figure 3.3), we have

$$\int_{\mathbb{R}\times D} dF = 2\pi \mathcal{Q}_e, \tag{3.57}$$

where $\mathbb{R}$ denotes the spatial x-direction in $\mathbb{R}^{1,1}$. In particular, we could choose D to be the disc obtained by filling in the 1-cycle S^1 corresponding to v, which opens up after the flop transition. Then (3.57) is equivalent to

$$\int_{\mathbb{R}\times\partial D} F_{x\theta_v}\, dx\, d\theta_v = 2\pi \mathcal{Q}_e. \tag{3.58}$$

Alternatively, as $F_{x\theta_v} = \partial_x A_{\theta_v}$ and

$$\oint_{\partial D} A_{\theta_v}\, d\theta_v = \mathrm{Im}\ v, \tag{3.59}$$

we see that as we cross the D2-branes in the x direction v jumps by $2\pi i \mathcal{Q}_e$. Since exchanging electric and magnetic charges corresponds to exchanging u and v, it follows that turning on $\mathcal{Q}_m$ units of magnetic charge corresponds to having a domain wall where u jumps by $2\pi i \mathcal{Q}_m$ in going from $x = -\infty$ to $x = +\infty$. Hence these domain walls are the magnetic charges we were seeking.

Multiple Lagrangian branes

In the above discussion we have been assuming that we have a single background D4-brane. Let us now return to more general case $M \geq 1$. In this case $\mathcal{Q}^{\text{open}}_{e,m}$ label representations of $U(M)$.[7] By a straightforward generalization of the arguments given above, we see that the attractor values of the eigenvalues of u and v are (generalizing (3.50))

$$\text{Re } u_i = \pi(\hat{\mathcal{Q}}^{\text{open}}_m)_i, \qquad \text{Re } v_j = \pi(\hat{\mathcal{Q}}^{\text{open}}_e)_j. \tag{3.60}$$

Here $\hat{\mathcal{Q}}^{\text{open}}_{m,e}$ denote the highest weight vectors of the corresponding representations, shifted by the Weyl vector ρ (see Appendix 3.7). The rest of the discussion generalizes similarly.

3.6 A solvable example

After these general considerations we now return to the example we described in Section 3.2, where X is a rank 2 holomorphic vector bundle over Σ, and add background D4-branes on $L\times\mathbb{R}^{1,1}$ to the Type IIA theory. In this section we want to argue that one can use 2-dimensional Yang-Mills theory to compute $Z^{\text{open}}_{\text{BPS}}$, generalizing the discussion of Section 3.2. Our arguments will be heuristic, but they lead to a definite prescription which is natural and fits in well with our conjectures.

How is the discussion of Section 3.2 modified by the introduction of the background branes? The L we will consider meet $\mathcal{D}$ along a circle, which we call γ. Hence in the gauge theory on $\mathcal{D}$ there will be extra massless string states localized along γ, in the

[7] The gauge theory has at least a $U(1)^M$ symmetry, and since the degeneracies are symmetric under the symmetric group S_M, we can organize them into characters of representations $\mathcal{Q}^{\text{open}}_e$ of $U(M)$ (possibly with negative multiplicities.)

bifundamental of $U(M) \times U(N)$. By condensing these string states (going out along a Higgs branch), i.e. turning on a vacuum expectation value of the form

$$\begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \tag{3.61}$$

one can break the gauge symmetry along γ to $U(M) \times U(N-M)$, where the surviving $U(M)$ is the diagonal in $[U(M) \times U(M)] \times U(N-M)$.[8] We conjecture that from the point of view of the gauge theory on $\mathcal{D}$, the only effect of the interaction with the background branes comes from the fact that the $U(M)$ part of the gauge field along γ is identified with the $U(M)$ gauge field on the background branes, via this Higgsing to the diagonal. We can account for this by inserting a δ-function in the theory on $\mathcal{D}$, which freezes M of the eigenvalues of the holonomy $e^{i \oint_\gamma \mathcal{A}}$, identifying them with the holonomy on the background branes, which we call $e^{i\phi}$. The Weyl invariant way to write this delta function is

$$\delta_M \left(e^{i \oint_\gamma \mathcal{A}}, e^{i\phi} \right) = D(\oint_\gamma \mathcal{A})^{-1} \sum_{\sigma \in S_N} (-)^\sigma \prod_{j=1}^{M} \delta \left((e^{i \oint_\gamma \mathcal{A}})_{\sigma(j)}, e^{i\phi_j} \right), \tag{3.62}$$

where D denotes the Vandermonde determinant (3.109).

Let us write $Z_{\mathrm{YM}}^{\mathrm{open}}(\varphi_0, \varphi_2, \phi)$ for the partition function with this operator inserted (here "open" refers to the fact that it is related to the open topological string.) This

[8]We are considering only the case $M < N$; ultimately we will be interested in taking N large while M stays finite.

partition function sums over the open D2-branes which end on the Lagrangian branes, as well as over the D0 and D2-brane charges which one had without the Lagrangian branes; so altogether we should have

$$Z_{\mathrm{YM}}^{\mathrm{open}} = Z_{\mathrm{BPS}}^{\mathrm{open}}. \tag{3.63}$$

In this ensemble the chemical potential $\varphi_e^{\mathrm{open}}$ for the open D2-branes should roughly be their mass. This mass is given by the area of the disc on which they are wrapped, which is related by supersymmetry to the Wilson line on the background branes; with this as motivation we write

$$\varphi_e^{\mathrm{open}} = 2\pi\phi/g_s. \tag{3.64}$$

To compute $Z_{\mathrm{YM}}^{\mathrm{open}}$ it is convenient to reduce from the twisted $\mathcal{N} = 4$ theory on $\mathcal{D}$ to a q-deformed Yang-Mills theory on Σ, as was done in [124, 10]. How does the operator insertion $\delta_M\left(e^{i\oint_\gamma \mathcal{A}}, e^{i\phi}\right)$ translate to the reduced theory? There are two cases to consider:

1. γ lies in the fiber of $\mathcal{L}_{-p}$ over a point $P \in \Sigma$.

2. γ lies on the Riemann surface Σ.

In either case these Lagrangian branes can be locally modelled by the ones studied in [11, 8]. In case 1, where γ is in the fiber over P, the situation is basically straightforward: as explained in [124], the flux $\oint_\gamma \mathcal{A}$ shows up in the q-deformed Yang-Mills theory on Σ as a field Φ. The operator we have to insert in the q-deformed theory is therefore

$$\delta_M(e^{i\Phi(P)}, e^{i\phi}). \tag{3.65}$$

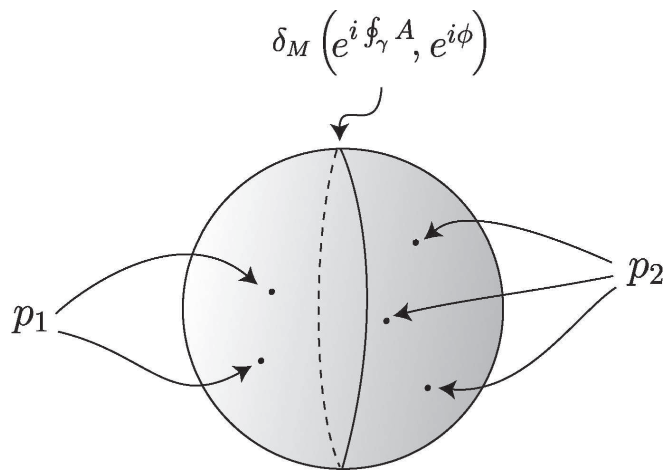

Figure 3.4: The operator $\delta_M\left(e^{i\oint_\gamma A}, e^{i\phi}\right)$ cuts Σ into two pieces.

The path integral gets localized on configurations where Φ is locally constant, so when there are no other operator insertions we can drop the P and write $\delta_M(e^{i\Phi}, e^{i\phi})$.

In case 2 the situation is a bit trickier, because of a subtlety which also appeared in [124]: namely, in performing the reduction one has to choose p points P_i on Σ, and at each such point one gets an operator corresponding to one unit of area in the Yang-Mills theory. The operator $\delta_M\left(e^{i\oint_\gamma \mathcal{A}}, e^{i\phi}\right)$ reduces to

$$\delta_M\left(e^{i\oint_\gamma A}, e^{i\phi}\right) \tag{3.66}$$

in two dimensions, but we have to specify how many of the p points go on each side of γ. Therefore there is a $\mathbb{Z}$-valued ambiguity in defining which operator we insert in the physical theory, parameterized by a choice of p_1 and p_2 with $p_1 + p_2 = p$. See Figure 3.4. This ambiguity should be understood as related to infrared regularization arising from the non-compactness of the situation; in the connection to the open topological string below, we will see that it is identified with the framing ambiguity.

Specializing to genus zero

Next we will investigate in detail the case when Σ has genus zero. So we specialize to Type IIA on $X \times \mathbb{R}^{3,1}$, where

$$X = \mathcal{O}(-p) \oplus \mathcal{O}(p-2) \to \mathbb{CP}^1, \tag{3.67}$$

with background D4-branes added on $L \times \mathbb{R}^{1,1}$. As we just explained, we can compute a mixed ensemble partition function $Z_{\mathrm{BPS}}^{\mathrm{open}}$ for this system by inserting an appropriate operator into the q-deformed Yang-Mills theory on an S^2 of area p. The parameters of the Yang-Mills theory are as in the closed case,

$$\theta_{\mathrm{YM}} = \theta, \quad g_{\mathrm{YM}}^2 = g_s, \quad q = e^{-g_s}. \tag{3.68}$$

We will show that for all p_1, p_2 we indeed have $Z_{\mathrm{YM}}^{\mathrm{open}} = |\psi_{\mathrm{top}}^{\mathrm{open}}|^2 + \mathcal{O}(e^{-N})$. We will also show that the identification of $Z_{\mathrm{YM}}^{\mathrm{open}}$ with $Z_{\mathrm{BPS}}^{\mathrm{open}}$ is consistent; namely, $Z_{\mathrm{BPS}}^{\mathrm{open}}$ should have an expansion where ϕ appears only in the form $e^{-2\pi\phi/g_s}$, and we will verify that $Z_{\mathrm{YM}}^{\mathrm{open}}$ indeed has such an expansion at least in the special case $p_1 = p_2 = 1$. These two results together give evidence for our conjecture (3.45).

Large N factorization on $\mathcal{O}(-p) \oplus \mathcal{O}(p-2) \to \mathbb{CP}^1$

We want to establish that

$$Z_{\mathrm{YM}}^{\mathrm{open}} = |\psi_{\mathrm{top}}^{\mathrm{open}}|^2 + \mathcal{O}(e^{-N}). \tag{3.69}$$

We compute $Z_{\mathrm{YM}}^{\mathrm{open}}$ using the gluing procedure described in Appendix 3.8: namely, we construct the sphere by gluing two discs together with the operator $\delta_M(e^{i\oint A}, e^{i\phi})$ in the middle. We use the fact that the Hilbert space of the 2-dimensional Yang-Mills theory is factorized at large N, $\mathcal{H} \simeq \mathcal{H}_+ \otimes \mathcal{H}_-$, and furthermore each component of the gluing procedure can be written in a factorized form. This factorization is described in detail in Appendix 3.10; the computation of $Z_{\rm YM}^{\rm open}$ we give below basically consists of fetching various results from that appendix and putting them together. We then compare this with the known form of the topological string amplitude and find the desired factorization; the final result is given in (3.81).

Branes in the base

Let us first discuss case 2, where to compute $Z_{\rm YM}^{\rm open}$ we have to insert a Wilson line freezing operator $\delta_M(e^{i\oint A}, e^{i\phi})$. This operator cuts the sphere into two pieces, with discrete areas p_1, p_2 such that $p_1 + p_2 = p$. The gluing computation of $Z_{\rm YM}^{\rm open}$ involves a zero area disc, an annulus of area p_1, the operator $\delta_M(e^{i\oint A}, e^{i\phi})$, an annulus of area p_2, and another zero area disc:

$$Z_{\rm YM}^{\rm open}(N, g_s, \theta, \phi) = \langle \Psi_0 | A_{p_1} \delta_M(e^{i\oint A}, e^{i\phi}) A_{p_2} | \Psi_0 \rangle. \tag{3.70}$$

Each of these pieces has been written in the factorized basis for $\mathcal{H}$ in Appendix 3.10: the disc is given in (3.172), the annulus in (3.170), and the Wilson line freezing operator in (3.171). Plugging in these factorizations and doing a little rearranging, we obtain the factorized form of $Z_{\rm YM}^{\rm open}$, schematically $Z_{\rm YM}^{\rm open} = Z_+ Z_-$, or more precisely (writing $q = e^{-g_s}$)

$$Z_{\rm YM}^{\rm open}(N, g_s, \theta, \phi) = Z_{\rm YM}^0(N, g_s, \theta, \phi) M(q)^2 \eta(q)^{2N} \times \\ \sum_{l \in \mathbb{Z}, R_1', R_2'} (-)^{|R_1'|+|R_2'|} q^{\frac{1}{2}(p_1+p_2)l^2} e^{iNlp\theta} Z_+^{R_1', R_2', l} Z_-^{R_1', R_2', l} + \mathcal{O}(e^{-N}), \tag{3.71}$$

where

$$Z_+^{R_1',R_2',l}(N,g_s,\theta,\phi) = q^{\frac{1}{2}N(|R_1'|+|R_2'|)}\times$$
$$\sum_{R_{1+},R_{2+},A_+} q^{\frac{1}{2}p_1\kappa_{R_{1+}}+\frac{1}{2}p_2\kappa_{R_{2+}}+(\frac{1}{2}N(p_1-1)+lp_1)|R_{1+}|+(\frac{1}{2}N(p_2-1)+lp_2)|R_{2+}|}\times$$
$$C_{0R_1'R_{1+}}C_{R_2'^tR_{2+}0}s_{R_{1+}/A_+}(e^{-i\phi})s_{R_{2+}/A_+}(e^{i\phi})e^{i\theta(p_1|R_{1+}|+p_2|R_{2+}|)}, \quad (3.72)$$

and similarly

$$Z_-^{R_1',R_2',l}(N,g_s,\theta,\phi) = q^{\frac{1}{2}N(|R_1'|+|R_2'|)}\times$$
$$\sum_{R_{1-},R_{2-},A_-} q^{\frac{1}{2}p_1\kappa_{R_{1-}}+\frac{1}{2}p_2\kappa_{R_{2-}}+(\frac{1}{2}N(p_1-1)-lp_1)|R_{1-}|+(\frac{1}{2}N(p_2-1)-lp_2)|R_{2-}|}\times$$
$$C_{0R_1'^tR_{1-}}C_{R_2'R_{2-}0}s_{R_{1-}/A_-}(e^{i\phi})s_{R_{2-}/A_-}(e^{-i\phi})e^{-i\theta(p_1|R_{1-}|+p_2|R_{2-}|)}. \quad (3.73)$$

The normalization factor $Z^0_{\mathrm{YM}}(N,g_s,\theta,\phi)$ will be fixed below.

Now we want to interpret the chiral blocks $Z_\pm^{R_1',R_2',l}(\phi)$ in terms of the topological string on X. This X can be represented torically by the picture in Figure 3.5, on which we also indicate the Lagrangian cycle L supporting M branes, one supporting a stack of infinitely many ghost branes, and one with a stack of infinitely many ghost antibranes. (See e.g. [11] for a review of the meaning of toric pictures such as this one.) The results of [7] give the topological string amplitude on this geometry as[9] (with $q = e^{-g_{\mathrm{top}}}$)

$$\psi^{\mathrm{g}}_{\mathrm{top}}(g_{\mathrm{top}},t,u,u') = \psi^0_{\mathrm{top}}(g_{\mathrm{top}},t,u)\sum_{R_1,R_2,A,R_1',R_2'}(-)^{|R_2'|}s_{R_1'}(e^{u_1'})s_{R_2'}(e^{u_2'})\times$$
$$q^{\frac{1}{2}p_1\kappa_{R_1}+\frac{1}{2}p_2\kappa_{R_2}}C_{0R_1'R_1}C_{R_2'^tR_20}s_{R_1/A}(e^{-u})s_{R_2/A}(e^{u})(-)^{p_1|R_1|+p_2|R_2|}e^{-t|R_2|}, \quad (3.74)$$

[9]Using the result as it appears in [7] one would actually get something slightly different from (3.74), namely, R_2/A would be replaced by R_2^t/A^t, and there would be an extra overall factor $(-)^{|R_2|+|A|}$. This difference is due to a typo in [7].

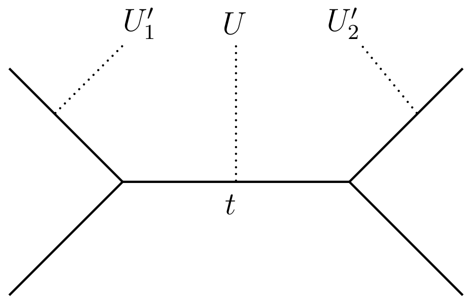

Figure 3.5: The vertex representation of $X = \mathcal{O}(-p) \oplus \mathcal{O}(p-2) \to \mathbb{CP}^1$, with a stack of M branes with complexified holonomy $U = e^u$, a stack of infinitely many ghost branes with complexified holonomy $U_1' = e^{u_1'}$, and a stack of infinitely many ghost antibranes with complexified holonomy $U_2' = e^{u_2'}$.

with the choice of p_1 and p_2 (subject to the constraint $p_1 + p_2 = p$) related to the choice of framing on the Lagrangian branes.[10] Similarly, if one swaps the ghost branes for ghost antibranes, one gets

$$\psi^{\mathrm{a}}_{\mathrm{top}}(g_{\mathrm{top}}, t, u, u') = \psi^0_{\mathrm{top}}(g_{\mathrm{top}}, t, u) \sum_{R_1, R_2, A, R_1', R_2'} (-)^{|R_1'|} s_{R_1'}(e^{u_1'}) s_{R_2'}(e^{u_2'}) \times$$

$$q^{\frac{1}{2}p_1\kappa_{R_1} + \frac{1}{2}p_2\kappa_{R_2}} C_{0R_1'^t R_1} C_{R_2' R_2 0} s_{R_1/A}(e^{-u}) s_{R_2/A}(e^u) (-)^{p_1|R_1| + p_2|R_2|} e^{-t|R_2|}. \quad (3.75)$$

Now to relate the chiral blocks $Z_\pm$ which make up Z_{YM} to the topological string amplitudes, we define

$$t = \frac{1}{2} N g_s (p-1) - ip\hat{\theta}, \quad (3.76)$$

$$u = \frac{1}{2} N g_s (p_1 - 1) - i(p_1\hat{\theta} - \phi), \quad (3.77)$$

$$u_1' = \frac{1}{2} N g_s + i\phi_1', \quad (3.78)$$

$$u_2' = \frac{1}{2} N g_s + i\phi_2', \quad (3.79)$$

$$g_{\mathrm{top}} = g_s. \quad (3.80)$$

[10] Strictly speaking, [7] considers the case $M = \infty$; but one can get finitely many branes by setting all but M components of the e^u and e^{-u} appearing in (3.74) to zero.

Here we introduced $\hat{\theta} = \theta + \pi$; this shift is meant to cancel the factor $(-)^{p|R|}$ in (3.74).[11]

The desired factorization is then basically straightforward to check. One begins with (3.71) which expresses Z_{YM} in terms of the chiral blocks, then relates the chiral blocks to $\psi_{\text{top}}^{\text{g}}$ and $\psi_{\text{top}}^{\text{a}}$ with the above choice of parameters, and converts the sums over R_1', R_2' into integrals over ϕ_1', ϕ_2' as discussed in Section 3.3. This essentially gives

$$\begin{aligned} Z_{\text{YM}}(N, g_s, \theta, \phi) = \sum_{l\in\mathbb{Z}} \int d_H\phi_1' d_H\phi_2' \times \\ \left(\psi_{\text{top}}^{\text{g}}\left(g_s, t + lpg_s, u + lp_1 g_s, u'\right)\right) \overline{\left(\psi_{\text{top}}^{\text{a}}\left(g_s, t - lpg_s, u - lp_1 g_s, u'\right)\right)}. \end{aligned} \tag{3.81}$$

In order to match the l-dependent terms in (3.71), though, one has to examine carefully the normalizations for the topological string and Yang-Mills amplitudes, as was done in [124, 10]. For the topological string we write

$$\psi_{\text{top}}^0(g_{\text{top}}, t) = M(q)\eta(q)^{2t/(p-2)g_{\text{top}}} \exp\left(-\frac{1}{6p(p-2)g_{\text{top}}^2}t^3 + \frac{p-2}{24p}t\right). \tag{3.82}$$

The meaning of this normalization factor was discussed in [10]. For the Yang-Mills theory, we write

$$Z_{\text{YM}}^0(N, g_s, \theta, \phi) = \exp\left(\frac{g_s(p-2)^2}{24p}(N - N^3) + N\frac{\hat{\theta}^2 p}{2g_s}\right). \tag{3.83}$$

Then the two chiral normalization factors multiply together to give the Yang-Mills

[11] The apparent asymmetry between p_1 and p_2 comes from the fact that we chose u to represent the complexified area of the disc which ends on the Lagrangian branes from the left; the disc which ends on them from the right has area $t - u = \frac{1}{2}Ng_s(p_2 - 1) - i(p_2\hat{\theta} + \phi)$.

normalization, up to some crucial l-dependent corrections:

$$\psi^0_{\mathrm{top}}(g_s, t + lpg_s, u + lp_1 g_s)\overline{\psi^0_{\mathrm{top}}(g_s, t - lpg_s, u - lp_1 g_s)} = \\ Z^0_{\mathrm{YM}}(N, g_s, \theta, \phi) M(q)^2 \eta(q)^{2N} q^{\frac{1}{2}pl^2} e^{iNlp\theta}. \quad (3.84)$$

These terms match the l-dependent terms in (3.71); they are exactly what is needed to make the factorization (3.81) work. So we have completed the factorization in case 2, corresponding to D-branes which intersect the base $\mathbb{CP}^1$ in X.

Branes in the fiber

We can also consider case 1, corresponding to D-branes which meet the fiber of $\mathcal{O}(-p) \to \mathbb{CP}^1$. In this case, in the Yang-Mills theory we insert the dual Wilson line freezing operator $\delta_M(e^{i\Phi}, e^{i\phi})$ at a point of $\mathbb{CP}^1$. Our discussion here will be more brief since the proof of the factorization runs along the same lines as case 2 above.

Again, we compute the Yang-Mills amplitude by gluing: we have to glue a disc containing the operator $\delta_M(e^{i\Phi}, e^{i\phi})$, an annulus of area p, and another disc, obtaining

$$Z^{\mathrm{open}}_{\mathrm{YM}}(N, g_s, \theta, \phi) = \langle \Psi_\phi | A_p | \Psi_0 \rangle. \quad (3.85)$$

Using the factorization results (3.170), (3.172) and (3.173) this becomes

$$Z^{\mathrm{open}}_{\mathrm{YM}}(N, g_s, \theta, \phi) = Z^0_{\mathrm{YM}}(N, g_s, \theta, \phi) M(q)^2 \eta(q)^{2N} \times \\ \sum_{l,m \in \mathbb{Z}, R'_1, R'_2} (-)^{|R_1|+|R_2|} q^{\frac{1}{2}pl^2} e^{iNlp\theta} \det(e^{im\phi}) Z_+^{R'_1, R'_2, l, m} Z_-^{R'_1, R'_2, l, m} + \mathcal{O}(e^{-N}), \quad (3.86)$$

with

$$Z_+^{R'_1, R'_2, l, m}(N, g_s, \theta, \phi) = q^{\frac{1}{2}N(|R'_1|+|R'_2|)} \sum_{R_+, T_+} q^{\frac{1}{2}p\kappa_{R_+} + \frac{1}{2}\kappa_{T_+} + (\frac{1}{2}N(p-2)+lp-m)|R_+| + (-\frac{1}{2}N - l)|T_+|} \times \\ C_{T_+ R'_1 R_+} C_{R'^t_2 R_+ 0} s_{T^t_+}(e^{-i\phi}) e^{i\theta p |R_+|}, \quad (3.87)$$

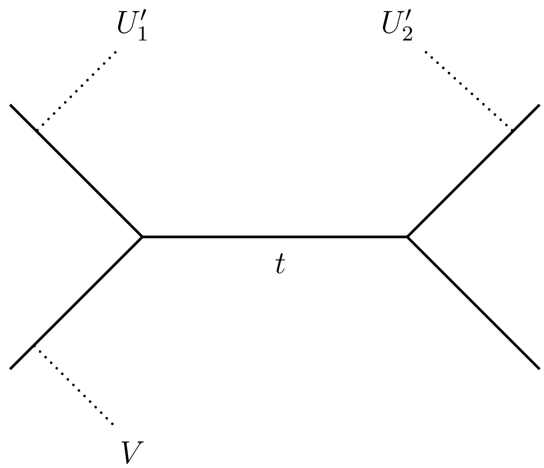

Figure 3.6: The vertex representation of X, with a stack of M Lagrangian branes with complexified holonomy $V = e^v$, a stack of infinitely many ghost branes with complexified holonomy $U_1' = e^{u_1'}$, and a stack of infinitely many ghost antibranes with complexified holonomy $U_2' = e^{u_2'}$.

and similarly

$$Z_{-}^{R_1',R_2',l,m}(N,g_s,\theta,\phi) = q^{\frac{1}{2}N(|R_1'|+|R_2'|)} \sum_{R_-,T_-} q^{\frac{1}{2}p\kappa_{R_-}+\frac{1}{2}\kappa_{T_-}+(\frac{1}{2}N(p-2)-lp+m)|R_-|+(-\frac{1}{2}N+l)|T_-|} \times \\ C_{T_- R_1'^t R_-} C_{R_2' R_- 0} s_{T_-^t}(e^{i\phi}) e^{-i\theta p |R_-|}. \quad (3.88)$$

As in case 2, we can now interpret these chiral blocks in terms of the topological string on X with M Lagrangian branes, now inserted on the external leg as indicated in Figure 3.6. Again from [7], the topological partition function in this geometry (with a particular choice of framing on the M external branes) is

$$\psi_{\mathrm{top}}^{\mathrm{g}}(g_{\mathrm{top}},t,v,u') = \psi_{\mathrm{top}}^{0}(g_{\mathrm{top}},t,v) \sum_{R,T} C_{TR_1'R} C_{R_2'R0} s_{R_1'}(e^{u_1'}) s_{R_2'}(e^{u_2'}) \times \\ (-)^{p|R|} q^{\frac{1}{2}p\kappa_R+\frac{1}{2}\kappa_T} e^{-t|R|} s_{T^t}(e^{-v}), \quad (3.89)$$

and similarly one can compute $\psi_{\mathrm{top}}^{\mathrm{a}}$ with the ghost branes exchanged for antibranes.

Now define

$$t = \frac{1}{2}Ng_s(p-2) - i\theta p, \tag{3.90}$$

$$u_1' = \frac{1}{2}Ng_s + i\phi_1', \tag{3.91}$$

$$u_2' = \frac{1}{2}Ng_s + i\phi_2', \tag{3.92}$$

$$v = -\frac{1}{2}Ng_s + i\phi, \tag{3.93}$$

$$g_{\text{top}} = g_s. \tag{3.94}$$

With this substitution and a suitable normalization one can relate (3.89) to the chiral blocks appearing in the factorization (3.86), similarly to what was done above in case 2, obtaining

$$Z_{\text{YM}}(N, g_s, \theta, \phi) = \sum_{l,m\in\mathbb{Z}} \int d_H\phi_1' d_H\phi_2' \times$$
$$\left(\psi^{\text{g}}_{\text{top}}\left(g_s, t + lpg_s, v - lg_s, u_1' - mg_s, u_2'\right)\right) \overline{\left(\psi^{\text{a}}_{\text{top}}\left(g_s, t - lpg_s, v + lg_s, u_1' + mg_s, u_2'\right)\right)}. \tag{3.95}$$

So finally we have found that $Z^{\text{open}}_{\text{YM}} = |\psi^{\text{open}}_{\text{top}}|^2 + \mathcal{O}(e^{-N})$, both for branes in the fiber and in the base.

Summing open D2-branes on $\mathcal{O}(-1) \oplus \mathcal{O}(-1) \to \mathbb{CP}^1$

Next we want to verify that $Z^{\text{open}}_{\text{YM}}$ can indeed be interpreted as counting open D2-branes with the chemical potential $\varphi^{\text{open}}_e = 2\pi\phi/g_s$. We consider case 2, where we have a Wilson line freezing operator $\delta_M\left(e^{i\oint_\gamma A}, e^{i\phi}\right)$ cutting the sphere into two pieces, with discrete areas p_1, p_2 such that $p_1 + p_2 = p$, and further specialize to the case $p_1 = p_2 = 1$.

As we did in the previous section, we compute the partition function $Z_{\mathrm{YM}}^{\mathrm{open}}$ of this Yang-Mills theory using the gluing procedure described in Appendix 3.8: namely, we construct the sphere by gluing two discs together with the operator $\delta_M(e^{i\oint A}, e^{i\phi})$ in the middle. However, unlike above where we used the splitting $\mathcal{H} = \mathcal{H}_+ \otimes \mathcal{H}_-$ to see the large N factorization of $Z_{\mathrm{YM}}^{\mathrm{open}}$, in this section we will write the explicit formula for $Z_{\mathrm{YM}}^{\mathrm{open}}$ at finite N.

The wave function of the q-deformed 2-dimensional Yang-Mills theory on the disc is a function of the eigenvalues ξ of the Wilson line around the boundary, evaluated in Appendix 3.9:

$$\Psi(\xi) = e^{-Ng_s/24}\Theta_N\left(\frac{1}{2\pi}(\xi+\theta), \frac{ig_s}{2\pi}\right), \tag{3.96}$$

where Θ_N denotes the theta function of $\mathbb{Z}^N$,

$$\Theta_N(z,\tau) = \sum_{\gamma\in\mathbb{Z}^N} e^{\pi i\tau\|\gamma\|^2} e^{2\pi i\langle\gamma,z\rangle}, \qquad \text{for } z\in\mathbb{R}^N,\ \mathrm{Im}\ \tau>0. \tag{3.97}$$

In our case we want to glue two such disc wave functions $\Psi_1(\xi)$, $\Psi_2(\xi)$ to one another with $\delta_M(e^{i\oint A}, e^{i\phi})$ sandwiched in the middle. The result of this gluing is given in (3.128), but we need a little notation first: we divide the lattice $\mathbb{Z}^N$ into $\mathbb{Z}^{N-M}\oplus\mathbb{Z}^M$, and correspondingly divide ξ into ζ and ϕ, with $N-M$ and M components respectively. Then the result of the gluing is

$$Z_{\mathrm{YM}}^{\mathrm{open}}(N,g_s,\theta,\phi) = \int_{[0,2\pi]^{N-M}} \frac{d\zeta}{2\pi}|D(\zeta)|^2\Psi_1(-\zeta,-\phi)\Psi_2(\zeta,\phi). \tag{3.98}$$

Because of the simple form of the wave function, the ζ dependence and ϕ dependence decouple, namely

$$\Psi(\zeta,\phi) = e^{-Ng_s/24}\Theta_{N-M}\left(\frac{1}{2\pi}(\zeta+\theta), \frac{ig_s}{2\pi}\right)\Theta_M\left(\frac{1}{2\pi}(\phi+\theta), \frac{ig_s}{2\pi}\right). \tag{3.99}$$

So write

$$f_{N-M}(\theta, g_s) = \int \frac{d\zeta}{2\pi} |D(\zeta)|^2 \Theta_{N-M}\left(\frac{1}{2\pi}(\zeta+\theta), \frac{ig_s}{2\pi}\right) \Theta_{N-M}\left(\frac{1}{2\pi}(-\zeta+\theta), \frac{ig_s}{2\pi}\right). \tag{3.100}$$

Then (3.98) becomes

$$Z^{\text{open}}_{\text{YM}}(N, g_s, \theta, \phi) = e^{-Ng_s/12} f_{N-M}(\theta, g_s) \Theta_M\left(\frac{1}{2\pi}(\phi+\theta), \frac{ig_s}{2\pi}\right) \Theta_M\left(\frac{1}{2\pi}(-\phi+\theta), \frac{ig_s}{2\pi}\right). \tag{3.101}$$

Now we can use the Poisson resummation property of the theta function,

$$\Theta_M(z, \tau) = \left(\frac{i}{\tau}\right)^{M/2} e^{-\pi i \|z\|^2/\tau} \Theta_M(z/\tau, -1/\tau), \tag{3.102}$$

to obtain

$$\begin{aligned} Z^{\text{open}}_{\text{YM}}(N, g_s, \theta, \phi) = e^{-Ng_s/12} f_{N-M}(\theta, g_s) \left(\frac{2\pi}{g_s}\right)^M e^{-\frac{1}{2g_s}(\|\phi+\theta\|^2 + \|\phi-\theta\|^2)} \times \\ \Theta_M\left(-\frac{i}{g_s}(\phi+\theta), \frac{2\pi i}{g_s}\right) \Theta_M\left(-\frac{i}{g_s}(-\phi+\theta), \frac{2\pi i}{g_s}\right). \end{aligned} \tag{3.103}$$

Expanding out these theta functions then gives $Z^{\text{open}}_{\text{YM}}(N, g_s, \theta, \phi)$ as an expansion in $e^{-2\pi\phi/g_s}$, up to a prefactor $e^{-\frac{1}{g_s}\|\phi\|^2}$. So up to this prefactor, we have verified that $Z^{\text{open}}_{\text{YM}}$ can indeed be interpreted as counting open D2-branes with the chemical potential $\varphi_2^{\text{open}} = 2\pi\phi/g_s$.

For completeness, let us briefly consider the leftover factor $f_{N-M}(\theta, g_s)$. Writing out using (3.109)

$$|D(\zeta)|^2 = \sum_{\sigma,\sigma' \in S_{N-M}} (-)^{\sigma\sigma'} e^{i\langle \zeta, \sigma(\rho) - \sigma'(\rho)\rangle} \tag{3.104}$$

(where $\rho = \rho_{N-M}$) and evaluating the integral using the definitions of the theta functions gives

$$f_{N-M}(\theta, g_s) = \sum_{\sigma,\sigma' \in S_{N-M}} (-)^{\sigma\sigma'} e^{-\frac{1}{2}g_s\|\sigma(\rho)-\sigma'(\rho)\|^2} \Theta_{N-M}\left(\frac{1}{2\pi}(-2\theta + ig_s(\sigma(\rho) - \sigma'(\rho))), \frac{ig_s}{\pi}\right). \tag{3.105}$$

So this can also be resummed to give an expansion in $e^{-4\pi^2/g_s}$ and $e^{-2\pi\theta/g_s}$, as one expects from the closed string sector of the conjecture.

3.7 Group theory

In this appendix we summarize our group theory conventions and a few useful formulas.

We use script letters $\mathcal{R}, \mathcal{P}, \mathcal{Q}, \ldots$ to denote representations of unitary groups such as $U(N)$, and capital letters $R, P, Q, \ldots$ to denote Young diagrams. Often Young diagrams will appear as the chiral and anti-chiral parts $R_\pm$ of a representation $\mathcal{R} = R_+\overline{R_-}[l]$, as described in Appendix 3.10.

The weight lattice of $U(N)$ is $\mathbb{Z}^N$, with its standard inner product $\langle, \rangle$. A highest weight representation $\mathcal{R}$ is characterized by a weight $(r_1, \ldots, r_N) \in \mathbb{Z}^N$, in a particular Weyl chamber; we make the standard choice of Weyl chamber, given by the constraint $r_1 \geq \cdots \geq r_N$. With this choice, the entries r_i correspond to the lengths of the rows of the extended Young diagram for the representation $\mathcal{R}$. The Weyl group W of $U(N)$ is the symmetric group, $W \simeq S_N$, which permutes the entries of $\mathbb{Z}^N$ in the obvious way.

We will use the symbol $\mathcal{R}$ both for the representation and for its highest weight. It is also convenient to introduce the symbol $\hat{\mathcal{R}}$ for $\mathcal{R} + \rho$, where ρ is half the sum of the positive roots of $U(N)$, concretely

$$\rho = \frac{1}{2}(N-1, N-3, \ldots, 3-N, 1-N). \tag{3.106}$$

We also write $\mathbf{1}$ for the "unit" vector,

$$\mathbf{1} = (1, 1, \ldots, 1, 1). \tag{3.107}$$

With this notation we can write the Weyl character formula,[12]

$$\mathrm{Tr}_{\mathcal{R}}(e^{i\xi}) = D(\xi)^{-1} \sum_{\sigma \in S_N} (-)^{\sigma} e^{i\langle \hat{\mathcal{R}}, \sigma(\xi) \rangle}, \tag{3.108}$$

where the denominator $D(\xi)$ is

$$D(\xi) = \sum_{\sigma \in S_N} (-)^{\sigma} e^{i\langle \xi, \sigma(\rho) \rangle} = \prod_{i<j} (e^{i(\xi_i - \xi_j)/2} - e^{-i(\xi_i - \xi_j)/2}). \tag{3.109}$$

In computing the q-deformed Yang-Mills amplitudes we will need to use the Hopf link invariant $S_{\mathcal{PQ}}$ of the level k Chern-Simons theory with gauge group $U(N)$. Define $g_s = \frac{2\pi}{N+k}$. There is a formula expressing $S_{\mathcal{PQ}}$ as a sum over the Weyl group $W \simeq S_N$:

$$S_{\mathcal{PQ}} = e^{-g_s(\|\rho\|^2 + N/24)} \sum_{\sigma \in S_N} (-)^{\sigma} e^{g_s \langle \hat{\mathcal{P}}, \sigma(\hat{\mathcal{Q}}) \rangle}. \tag{3.110}$$

(The standard formulas for $S_{\mathcal{PQ}}$ include a different normalization, but in the context where we will use $S_{\mathcal{PQ}}$ we will absorb this in other normalization factors.)

For any N_1, N_2 with $N_1 + N_2 = N$, let $\mathcal{Q}$ label a representation of $U(N_1)$ and $\mathcal{A}$ a representation of $U(N_2)$, while $\mathcal{R}$ is a representation of $U(N)$; then we define the branching coefficients $\mathcal{B}^{\mathcal{R}}_{\mathcal{QA}}$ by the rule that $\mathcal{R}$ decomposes under $U(N_1) \times U(N_2)$ as

$$\mathcal{R} \to \bigoplus_{\mathcal{Q}, \mathcal{A}} \mathcal{B}^{\mathcal{R}}_{\mathcal{QA}} [\mathcal{Q}, \mathcal{A}]. \tag{3.111}$$

[12] When N is odd, $D(\xi)^{-1}$ and $\sum_{\sigma \in S_N} (-)^{\sigma} e^{i\langle \hat{\mathcal{R}}, \sigma(\xi) \rangle}$ are not quite well defined as functions of the eigenvalues $e^{i\xi}$ — they change sign under $\xi_i \to \xi_i + 2\pi$. Nevertheless their product is still well defined.

We fix the normalization of the Casimir operators of $U(N)$ as follows: in a representation $\mathcal{R}$ with highest weight $(r_1, \ldots, r_N)$,

$$C_1(\mathcal{R}) = \langle \hat{\mathcal{R}}, \mathbf{1} \rangle = \sum_i r_i, \tag{3.112}$$

$$C_2(\mathcal{R}) = \|\hat{\mathcal{R}}\|^2 - \|\rho\|^2 = \sum_i r_i(r_i + N + 1 - 2i). \tag{3.113}$$

We write $N^R_{R_1 R_2}$ for the usual Littlewood-Richardson numbers, and also use a slight generalization which we write $N^R_{R_1 \cdots R_k}$. These numbers can be defined in various equivalent ways — for example, if we think of the Young diagrams R_i and R as representations of $GL(\infty)$, they are the tensor product coefficients, i.e.

$$R_1 \otimes \cdots \otimes R_k = \bigoplus_R N^R_{R_1 \cdots R_k} R. \tag{3.114}$$

In particular, $N^R_{R_1 \cdots R_k} = 0$ unless $\sum_{i=1}^k |R_i| = |R|$, where $|R|$ denotes the total number of boxes in the diagram R.

We write $s_R(x)$ for the "Schur function" associated with the Young diagram R: this is a symmetric polynomial in infinitely many variables, $x = (x_1, x_2, \ldots)$. It can be defined in various equivalent ways; one convenient way to think of it is as the character of the $\mathrm{Mat}(\infty, \mathbb{C})$ representation associated to R, evaluated on the diagonal matrix with entries $(x_1, x_2, \ldots)$. There is a bilinear inner product $\langle , \rangle$ on the ring of symmetric polynomials for which the Schur functions form an orthonormal basis, $\langle s_R, s_S \rangle = \delta_{RS}$; in terms of this inner product $N^R_{R_1 \cdots R_k} = \langle \prod_{i=1}^k s_{R_i}, s_R \rangle$. Viewing the x_i as eigenvalues, the inner product can be written as a formal integral of class functions over $U(\infty)$ (interpreted as an inverse limit of finite-dimensional groups with their normalized Haar measures),

$$\langle f, g \rangle = \int d_H \xi \, f(e^{-i\xi}) g(e^{i\xi}). \tag{3.115}$$

We also use the "skew Schur functions" $s_{R/A}(x)$, defined by

$$s_{R/A}(x) = \sum_Q N^R_{QA} s_Q(x). \tag{3.116}$$

See [89] for much more on Schur functions and skew Schur functions.

We also introduce an analog of the skew Schur function, a "skew trace" involving the branching $U(N) \to U(N_1) \times U(N_2)$ where $N_1 + N_2 = N$: this is a rule by which a representation of $U(N)$ and a representation of $U(N_2)$ induce a class function on $U(N_1)$, which we define by

$$\mathrm{Tr}_{\mathcal{R}/\mathcal{A}}(U) = \sum_{\mathcal{Q}} \mathcal{B}^{\mathcal{R}}_{\mathcal{Q}\mathcal{A}} \mathrm{Tr}_{\mathcal{Q}}(U). \tag{3.117}$$

Here $\mathcal{R}, \mathcal{Q}, \mathcal{A}$ denote representations of $U(N)$, $U(N_1)$, $U(N_2)$ respectively; $\mathcal{B}$ denotes the branching coefficients defined in (3.111); and $U \in U(N_1)$.

We will frequently encounter sums $\sum_{R'}$ over the set of all Young diagrams. A particularly useful identity for reducing such sums is

$$\sum_{R',S'} (-)^{|R'|} N^A_{R'S'A_1\cdots A_a} N^B_{R'^t S' B_1 \cdots B_b} = N^A_{A_1\cdots A_a} N^B_{B_1\cdots B_b}. \tag{3.118}$$

One can prove (3.118) using the "Cauchy identities" for Schur functions, given e.g. in [89],

$$\sum_{S'} s_{S'}(x) s_{S'}(y) = \prod_{i=1}^\infty \prod_{j=1}^\infty (1 - x_i y_j)^{-1}, \tag{3.119}$$

$$\sum_{R'} (-)^{|R'|} s_{R'}(x) s_{R'^t}(y) = \prod_{i=1}^\infty \prod_{j=1}^\infty (1 - x_i y_j). \tag{3.120}$$

3.8 The q-deformed 2-d Yang-Mills theory

In this section we review some facts about 2-dimensional Yang-Mills theory and its q-deformed cousin. We begin with the 2-dimensional Euclidean Yang-Mills action[13] for gauge group $G = U(N)$,

$$S_{\mathrm{YM}} = \frac{1}{2g_{\mathrm{YM}}^2}\left(\int_\Sigma d^2x\ \mathrm{Tr}F \wedge *F + \theta_{\mathrm{YM}} \int_\Sigma \mathrm{Tr}F\right). \tag{3.121}$$

It is often convenient to rewrite (3.121) by introducing an additional adjoint-valued scalar field Φ, which enters the action quadratically: namely, (3.121) is equivalent to

$$S_{\mathrm{YM}} = \frac{1}{2g_{\mathrm{YM}}^2}\left(2\int_\Sigma \mathrm{Tr}\Phi F - \int_\Sigma \mu\ \mathrm{Tr}\Phi^2 + \theta_{\mathrm{YM}} \int_\Sigma \mu\ \mathrm{Tr}\Phi\right), \tag{3.122}$$

where μ is the area element on Σ. Once we have introduced this Φ we can define the q-deformed theory: we use the same action S_{YM}, but we consider the fundamental variables to be the gauge connection and $e^{i\Phi}$, rather than the gauge connection and Φ. More precisely, since there is an ambiguity in recovering Φ from $e^{i\Phi}$, S_{YM} is not well defined as a function of Φ; to get a well defined expression inside the path integral one has to sum $e^{-S_{\mathrm{YM}}}$ over all "images" Φ. Equivalently, we integrate over all Φ, not just a fundamental domain, but we use the measure appropriate for an integral over $e^{i\Phi}$. This construction gives the q-deformed theory with $q = e^{-g_{\mathrm{YM}}^2}$, which is the one that naturally occurs in this chapter; to get a different value of q one would change the periodicity of Φ.

The partition function can be computed in various ways; here we will focus on the computation by cutting and pasting. In the case of the undeformed Yang-Mills theory, this procedure was reviewed in [38]; our treatment will be briefer, and is

[13] Note that our convention for θ_{YM} is not the usual one; $\theta_{\mathrm{YM}}^{\mathrm{usual}} = \frac{i\pi}{g_{\mathrm{YM}}^2}\theta_{\mathrm{YM}}$.

intended mostly to recall the new features that appear in the q-deformed case, as described in [10].

To get the cutting-and-pasting procedure started one first needs to know the Hilbert space $\mathcal{H}$ of the theory on S^1; as for the usual 2-d Yang-Mills theory, it is simply the space of class functions $\Psi(g)$, with $g \in G$ interpreted as the holonomy of the connection around S^1. The path integral over a surface with boundary S^1 thus gives a state $\Psi(g)$. Two such surfaces can be glued using the rule[14]

$$\langle \Psi_1 | \Psi_2 \rangle = \int_G d_H g \, \Psi_1(g^{-1}) \Psi_2(g), \tag{3.123}$$

with $d_H g$ the Haar measure. When $G = U(N)$ we can write these wave functions more concretely as functions of the eigenvalues $e^{i\xi_i}$, totally symmetric under the permutation group S_N, and the gluing rule becomes

$$\langle \Psi_1 | \Psi_2 \rangle = \int_{[0,2\pi]^N} \frac{d\xi}{2\pi} |D(\xi)|^2 \, \Psi_1(-\xi) \Psi_2(\xi), \tag{3.124}$$

where as in (3.109)

$$D(\xi) = \prod_{i<j} (e^{i(\xi_i - \xi_j)/2} - e^{-i(\xi_i - \xi_j)/2}). \tag{3.125}$$

A convenient basis of $\mathcal{H}$ (which in particular diagonalizes the Hamiltonian) is given by the characters $\mathrm{Tr}_{\mathcal{R}}(g)$ as $\mathcal{R}$ runs over all representations of G. In that basis the gluing rule becomes

$$\langle \mathcal{R}_1 | \mathcal{R}_2 \rangle = \delta_{\mathcal{R}_1 \mathcal{R}_2}. \tag{3.126}$$

Next we need the partition function on a few elementary surfaces, from which any Σ of interest to us can be pasted together:

[14] We will always use the notation $\langle | \rangle$ to stand for the (linear) gluing rule rather than the (sesquilinear) inner product on the Hilbert space. The two are the same when acting on real linear combinations of the characters $\mathrm{Tr}_{\mathcal{R}}(g)$ but differ for complex linear combinations.

The annulus. The annulus of area a has two boundaries, so it gives an operator $A_a : \mathcal{H} \to \mathcal{H}$. The gluing rule for an annulus can be obtained directly from the action by working out the Hamiltonian; it is [10]

$$\langle \mathcal{R}_1 | A_a | \mathcal{R}_2 \rangle = \delta_{\mathcal{R}_1 \mathcal{R}_2} e^{-a\left(\frac{1}{2} g_{\mathrm{YM}}^2 C_2(\mathcal{R}) - i\theta_{\mathrm{YM}} C_1(\mathcal{R})\right)}. \tag{3.127}$$

The Wilson line freezing operator. As discussed in Section 3.5, we will be particularly interested in computing amplitudes involving a particular operator, written $\delta_M(e^{i\oint A}, e^{i\phi})$, which has the effect of freezing M of the eigenvalues along a Wilson loop to the values $e^{i\phi_1}, \ldots, e^{i\phi_M}$. A natural guess for the gluing rule with $\delta_M(e^{i\oint A}, e^{i\phi})$ inserted can be obtained by splitting up the N eigenvalues ξ_i into $(\underbrace{\zeta}_{N-M}, \underbrace{\phi}_{M})$: namely, one freezes the ϕ eigenvalues in the gluing rule (3.124) and integrates only over the ζ eigenvalues, obtaining

$$\langle \Psi_1 | \delta_M(e^{i\oint A}, e^{i\phi}) | \Psi_2 \rangle = \int_{[0,2\pi]^{N-M}} \frac{d\zeta}{2\pi} |D(\zeta)|^2 \; \Psi_1(-\zeta, -\phi) \Psi_2(\zeta, \phi). \tag{3.128}$$

This integral has an interpretation in the representation basis. Namely, suppose $\Psi_j(\xi) = \mathrm{Tr}_{\mathcal{R}_j}(e^{i\xi})$. Then decomposing $\mathcal{R}_j$ under $U(N-M) \times U(M)$ gives

$$\Psi_j(\xi) = \sum_{\mathcal{A}_j, \mathcal{Q}_j} \mathcal{B}^{\mathcal{R}_j}_{\mathcal{A}_j \mathcal{Q}_j} \mathrm{Tr}_{\mathcal{A}_j}(e^{i\zeta}) \mathrm{Tr}_{\mathcal{Q}_j}(e^{i\phi}). \tag{3.129}$$

The integral over ζ then picks out the terms with $\mathcal{A}_1 = \mathcal{A}_2$, giving

$$\langle \mathcal{R}_1 | \delta_M(e^{i\oint A}, e^{i\phi}) | \mathcal{R}_2 \rangle = \sum_{\mathcal{Q}_1, \mathcal{Q}_2, \mathcal{A}} \mathcal{B}^{\mathcal{R}_1}_{\mathcal{A}\mathcal{Q}_1} \mathcal{B}^{\mathcal{R}_2}_{\mathcal{A}\mathcal{Q}_2} \mathrm{Tr}_{\mathcal{Q}_1}(e^{-i\phi}) \mathrm{Tr}_{\mathcal{Q}_2}(e^{i\phi}). \tag{3.130}$$

If we define the skew trace $\mathrm{Tr}_{\mathcal{R}/\mathcal{S}}$ as in (3.117), we can rewrite this as

$$\langle \mathcal{R}_1 | \delta_M(e^{i\oint A}, e^{i\phi}) | \mathcal{R}_2 \rangle = \sum_{\mathcal{A}} \mathrm{Tr}_{\mathcal{R}_1/\mathcal{A}}(e^{-i\phi}) \mathrm{Tr}_{\mathcal{R}_2/\mathcal{A}}(e^{i\phi}). \tag{3.131}$$

The disc. The disc of zero area gives a simple state $\Psi_0 \in \mathcal{H}$ on its boundary,

$$\Psi_0 = \sum_{\mathcal{R}} S_{\mathcal{R}0} |\mathcal{R}\rangle, \tag{3.132}$$

as was computed in [10]. (This should be compared to the analogous expression in the non-q-deformed Yang-Mills theory — there one would have replaced $S_{\mathcal{R}0}$ by $\dim \mathcal{R}$ up to overall normalization. Indeed, $S_{\mathcal{R}0}/S_{00}$ is the quantum dimension $\dim_q \mathcal{R}$.)

The dual Wilson line freezing operator. Also as discussed in Section 3.5, we need the operator $\delta_M(e^{i\Phi}, e^{i\phi})$ which freezes M of the eigenvalues of the dual Wilson line Φ at a point of Σ. The disc of zero area with this operator inserted gives a state $\Psi_\phi \in \mathcal{H}$ on its boundary, for which the natural formula is

$$\Psi_\phi = \sum_{\mathcal{R},\mathcal{S}} S_{\mathcal{R}\mathcal{S}} \mathrm{Tr}_{\mathcal{S}/0}(e^{i\phi}) |\mathcal{R}\rangle. \tag{3.133}$$

This is a straightforward generalization of the result of [10] in the case $M = N$, along the lines of what we did above for the Wilson line freezing operator. (In the special case $M = N$, the result of [10] just replaces $\mathrm{Tr}_{\mathcal{S}/0}$ by $\mathrm{Tr}_{\mathcal{S}}$ in the above.)

The trinion (pair of pants). The trinion has three boundaries, so it gives an element in $\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}$, namely

$$T = \sum_{\mathcal{R}} \frac{|\mathcal{R}\rangle \otimes |\mathcal{R}\rangle \otimes |\mathcal{R}\rangle}{S_{0\mathcal{R}}}. \tag{3.134}$$

It was computed in [10]; we include it here just for completeness since it would be relevant for Riemann surfaces of genus $g > 1$.

In addition to these local ingredients we will include an overall normalization factor Z^0_{YM} in the partition function of this q-deformed theory; we do not give a general rule for this normalization here, but in the example we consider in the text, it can be found in (3.83).

A q-deformation of 2-dimensional Yang-Mills theory has also been considered in [30] where it was formulated using a lattice regularization. That formulation is likely to be equivalent to the one discussed here and in [10]; at least the partition function appears to be the same on an arbitrary surface.

3.9 The disc wave function

Consider the q-deformed 2-d Yang-Mills theory on a disc of area p, with parameters fixed by

$$\theta_{\mathrm{YM}} = \theta, \quad g^2_{\mathrm{YM}} = g_s, \quad q = e^{-g_s}. \tag{3.135}$$

The path integral on this disc gives a state $\Psi(\xi)$ on the boundary, for which one can give a formula using the rules of Appendix 3.8; namely, it is a sum over irreducible representations $\mathcal{R}$ of $U(N)$,

$$\Psi(\xi) = \sum_{\mathcal{R}} S_{\mathcal{R}0} e^{-\frac{1}{2}pg_s C_2(\mathcal{R})} e^{i\theta p C_1(\mathcal{R})} \mathrm{Tr}_{\mathcal{R}} e^{i\xi}. \tag{3.136}$$

Our purpose in this section is to express this $\Psi(\xi)$ in terms of theta functions. As reviewed in Appendix 3.7, the irreducible representations of $\mathcal{R}$ can be labeled by their highest weights $\mathcal{R} = (r_1, \ldots, r_N) \in \mathbb{Z}^N$, subject to the constraint $r_1 \geq r_2 \geq \cdots \geq r_N$. We also write $\hat{\mathcal{R}} = \mathcal{R} + \rho$. Now we use the Weyl character formula (3.108), the modular S matrix formula (3.110), and the Casimirs (3.112), (3.113); altogether

(3.136) becomes

$$\Psi(\xi) = \sum_{\mathcal{R}} \left(e^{-g_s(\|\rho\|^2 + N/24)} \sum_{\sigma \in S_N} (-)^\sigma e^{g_s \langle \sigma(\hat{\mathcal{R}}), \rho \rangle} \right) e^{-\frac{1}{2} p g_s (\|\hat{\mathcal{R}}\|^2 - \|\rho\|^2) + i\theta p \langle \hat{\mathcal{R}}, \mathbf{1} \rangle} \times \left(D(\xi)^{-1} \sum_{\sigma' \in S_N} (-)^{\sigma'} e^{i \langle \sigma'(\hat{\mathcal{R}}), \xi \rangle} \right). \quad (3.137)$$

Writing $\tilde{\sigma} = \sigma \sigma'^{-1}$, and using the Weyl invariance of $\langle , \rangle$ and $\mathbf{1}$, we can rewrite this as

$$\Psi(\xi) = e^{-g_s(\|\rho\|^2 + N/24)} D(\xi)^{-1} e^{\frac{1}{2} p g_s \|\rho\|^2} \sum_{\mathcal{R}} \sum_{\sigma, \tilde{\sigma} \in S_N} (-)^{\tilde{\sigma}} e^{-\frac{1}{2} p g_s \|\hat{\mathcal{R}}\|^2 + i \langle \sigma(\hat{\mathcal{R}}), \tilde{\sigma}(\xi) + \theta p \mathbf{1} - i g_s \rho \rangle}. \quad (3.138)$$

Now we want to express this as a theta function. If $\mathcal{R}$ runs over all weight vectors in a given Weyl chamber, then it is easy to see that $\hat{\mathcal{R}}$ runs over all weight vectors in the *interior* of that chamber.[15] Since the Weyl group acts transitively to permute the Weyl chambers, the sum over σ and $\mathcal{R}$ can be combined into a single sum over $\gamma = \sigma(\hat{\mathcal{R}})$, where γ runs over the weight lattice $\mathbb{Z}^N$, or more precisely over those vectors in $\mathbb{Z}^N$ which are not in the boundary of any Weyl chamber. In terms of γ the sum becomes

$$\Psi(\xi) = e^{-g_s(\|\rho\|^2 + N/24)} D(\xi)^{-1} e^{\frac{1}{2} p g_s \|\rho\|^2} \sum_{\gamma} e^{-\frac{1}{2} p g_s \|\gamma\|^2 + i \langle \gamma, \theta p \mathbf{1} - i g_s \rho \rangle} \sum_{\tilde{\sigma} \in S_N} (-)^{\tilde{\sigma}} e^{i \langle \gamma, \tilde{\sigma}(\xi) \rangle}. \quad (3.139)$$

But now note that the sum over $\tilde{\sigma}$ vanishes if γ is fixed by some Weyl reflection $\tilde{\sigma}$, i.e. if it lies on the boundary of a Weyl chamber. Therefore we can extend the sum over γ to run over the whole weight lattice $\mathbb{Z}^N$. The sum can be written (now dropping

[15] If N is even, the weight lattice has to be shifted by $\frac{1}{2}\mathbf{1}$. This subtlety modifies some of our intermediate expressions but cancels out in the final result (3.143).

the $\tilde{}$ on σ for notational simplicity)

$$\Psi(\xi) = e^{-g_s(\|\rho\|^2+N/24)} D(\xi)^{-1} e^{\frac{1}{2}pg_s\|\rho\|^2} \sum_{\sigma\in S_N} (-)^\sigma \Theta_N \left(\frac{1}{2\pi}(\sigma(\xi) + \theta p \mathbf{1} - ig_s\rho), \frac{ipg_s}{2\pi} \right). \tag{3.140}$$

Here we have introduced the theta function of $\mathbb{Z}^N$,

$$\Theta_N(z,\tau) = \sum_{\gamma\in\mathbb{Z}^N} e^{\pi i\tau\|\gamma\|^2} e^{2\pi i\langle\gamma,z\rangle}, \qquad \text{for } z \in \mathbb{R}^N,\ \mathrm{Im}\ \tau > 0, \tag{3.141}$$

which obeys the functional equation

$$\Theta_N(z+\tau\lambda,\tau) = e^{-\pi i\tau\|\lambda\|^2} e^{-2\pi i\langle\lambda,z\rangle} \Theta_N(z,\tau) \qquad \text{for } \lambda \in \mathbb{Z}^N. \tag{3.142}$$

One can simplify (3.140) in the case $p = 1$; namely, in this case, one can apply (3.142) with $\lambda = -\rho$. After some straightforward algebra using (3.109) the sum over σ cancels the denominator $D(\xi)$, leaving

$$\Psi(\xi) = e^{-g_sN/24} \Theta_N \left(\frac{1}{2\pi}(\xi + \theta\mathbf{1}), \frac{ig_s}{2\pi} \right). \tag{3.143}$$

3.10 Factorization

In this appendix we give some mathematical results which are used in the text to establish the factorization of the 2-dimensional Yang-Mills amplitude with operator insertions into chiral and anti-chiral parts.

Coupled representations

An essential technical tool in studying the factorization of 2-d Yang-Mills into chiral and anti-chiral sectors, introduced in [64], is the notion of a *coupled representation* of $U(N)$. Here we review the notion of coupled representation.

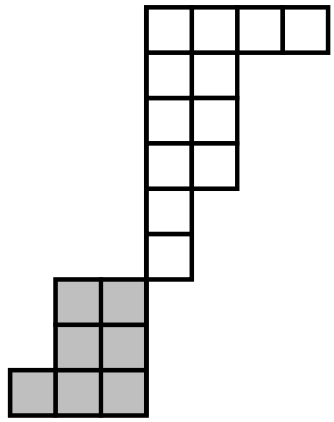

Figure 3.7: An extended Young diagram representing a representation of $U(N)$ (for $N = 9$) constructed by symmetrization and antisymmetrization over both fundamental representations (white boxes) and antifundamental representations (grey boxes).

Recall that the irreducible representations of $SU(N)$ correspond to Young diagrams with no more than N rows. Such a diagram can be specified by giving the lengths of the rows, $(\lambda_1, \ldots, \lambda_N)$, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$, and all $\lambda_i \geq 0$. Denote the fundamental representation by V. Then the representation of $SU(N)$ corresponding to λ is obtained as a particular subspace of $V^{\otimes|\lambda|}$, roughly by symmetrizing over the rows and antisymmetrizing over the columns. In the case of $SU(N)$ one can obtain all representations in this way, but for $U(N)$ one also needs to include copies of the antifundamental representation $\overline{V}$. This corresponds to considering "extended Young diagrams" which can include "antiboxes" as well as boxes, i.e. removing the constraint that all $\lambda_i \geq 0$, as shown in Figure 3.7.

Now we can describe the coupled representations of $U(N)$. The extended Young diagram for a coupled representation R is built from two Young diagrams R_+, R_-, just by putting boxes in the shape R_+ at the upper left, antiboxes in the shape of an upside-down version of R_- at the lower right, and zero-length rows in between

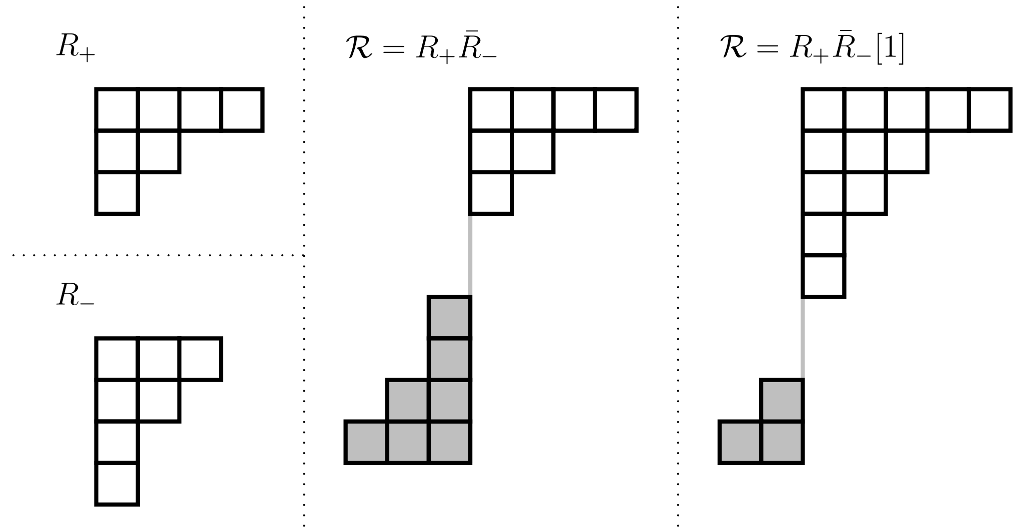

Figure 3.8: A coupled representation of $U(N)$ (for $N = 9$).

so that the total height of the diagram is N. An example is shown in Figure 3.8. (Note that this procedure only makes sense for sufficiently large N, namely, N has to be greater than the combined number of rows in R_+ and R_-. We consider coupled representations for which one of $R_\pm$ has more than $\frac{1}{2}N$ rows to be undefined.) We write the coupled representation $\mathcal{R} = R_+\overline{R_-}$. We will also need a slight generalization of this construction: denote by $R_+\overline{R_-}[l]$ the representation obtained by tensoring $R_+\overline{R_-}$ with l powers of the determinant representation of $U(N)$. This is equivalent to shifting the lengths of all rows by l.

The representations $R_+\overline{R_-}[l]$, where R_+ and R_- are Young diagrams with $\leq \frac{1}{2}N$ rows, are a basis for the representation ring of $U(N)$ (at least for N even.) Another such basis would be obtained by taking instead $R_+ \otimes \overline{R_-}$. The two are not the same, although $R_+\overline{R_-}$ is the principal component of $R_+ \otimes \overline{R_-}$; the relation between the two is given by

$$R_+ \otimes \overline{R_-} = \bigoplus_{S_\pm} \left[\sum_{S'} N^{R_+}_{S_+ S'} N^{R_-}_{S_- S'} \right] S_+ \overline{S_-} \tag{3.144}$$

(so long as $R_\pm$ each have $\leq \frac{1}{2}N$ rows; otherwise the right side would include $S_+\overline{S_-}$ where one of $S_\pm$ has more than $\frac{1}{2}N$ rows, which we have not defined.) Here S' and $S_\pm$ run over all (ordinary) Young diagrams. Note that the only S_+ that contribute are ones which are contained in R_+, and similarly for S_-, so the sum in (3.144) is finite. It gives the decomposition of $R_+ \otimes \overline{R_-}$ into irreducibles.

We will also need the inverse of (3.144). To get it, we use the fact that the sum over S' can be undone by summing over another auxiliary Young diagram R', using formula (3.118), here in the form

$$\sum_{R', S'} (-)^{|R'|} N^{A_+}_{B_+ R' S'} N^{A_-}_{B_- R'^t S'} = \delta^{A_+}_{B_+} \delta^{A_-}_{B_-}. \tag{3.145}$$

Applying this to (3.144) one obtains

$$S_+ \overline{S_-} = \bigoplus_{R_\pm} \left(\sum_{R'} (-)^{|R'|} N^{S_+}_{R_+ R'} N^{S_-}_{R_- R'^t} \right) R_+ \otimes \overline{R_-}. \tag{3.146}$$

Again here, R' and $R_\pm$ run over all ordinary Young diagrams, but only those $R_\pm$ which are contained in $S_\pm$ contribute, so the sum is finite.

One can also rewrite (3.144) in terms of characters as

$$\sum_{R_\pm} \mathrm{Tr}_{R_+ \otimes \overline{R_-}}(U) s_{R_+}(V_+) s_{R_-}(V_-) = \sum_{S_\pm} \left(\sum_{S'} s_{S_+ \otimes S'}(V_+) s_{S_- \otimes S'}(V_-) \right) \mathrm{Tr}_{S_+ \overline{S_-}}(U), \tag{3.147}$$

and (3.146) as

$$\sum_{S_\pm} \mathrm{Tr}_{S_+ \overline{S_-}}(U) s_{S_+}(V_+) s_{S_-}(V_-) = \sum_{R_\pm} \left(\sum_{R'} (-)^{|R'|} s_{R_+ \otimes R'}(V_+) s_{R_- \otimes R'^t}(V_-) \right) \mathrm{Tr}_{R_+ \otimes \overline{R_-}}(U). \tag{3.148}$$

It is useful to know how to decompose the Casimir operators for $\mathcal{R} = R_+\overline{R_-}[l]$,

$$C_1(R_+\overline{R_-}[l]) = |R_+| - |R_-| + Nl, \tag{3.149}$$

$$C_2(R_+\overline{R_-}[l]) = \kappa_{R_+} + \kappa_{R_-} + N(|R_+| + |R_-|) + 2l(|R_+| - |R_-|) + Nl^2. \tag{3.150}$$

Here we introduced

$$\kappa_R = \sum r_i^2 - \sum c_i^2, \tag{3.151}$$

where r_i are the lengths of the rows of the Young diagram R and c_i are the lengths of the columns. (So $\kappa_R = -\kappa_{R^t}$, where R^t denotes the transpose of the diagram.)

Branching rules

To understand the behavior of Yang-Mills theory when some eigenvalues are frozen, we need to understand the branching rules for coupled representations: how does a coupled representation of $U(N)$ decompose under $U(N) \to U(N_1) \times U(N_2)$, for $N_1 + N_2 = N$? In this section we will give formulas for these branching rules.

We begin with the case of a representation $\mathcal{R}$ of $U(N)$ which is given by an ordinary Young diagram, $\mathcal{R} = R$ (i.e. it can be constructed using only fundamental indices, without the need for antifundamentals.) In this case the branching rule is well known (it is given e.g. in [89] in the language of Schur functions),

$$R \to \bigoplus_{R_1,R_2} N^R_{R_1 R_2}[R_1, R_2]. \tag{3.152}$$

Here R_1 and R_2 run over all Young diagrams (but of course the coefficient $N^R_{R_1 R_2}$ is only nonzero if $|R_1| + |R_2| = |R|$.) The same rule holds for representations $\overline{R}$ constructed only from antifundamentals. Combining these two rules we can find the branching rule for tensor products,

$$R_+ \otimes \overline{R_-} \to \bigoplus_{R_{1\pm}, R_{2\pm}} N^{R_+}_{R_{1+} R_{2+}} N^{R_-}_{R_{1-} R_{2-}} [R_{1+} \otimes \overline{R_{1-}}, R_{2+} \otimes \overline{R_{2-}}]. \tag{3.153}$$

Now we can convert (3.153) into a branching rule for coupled representations. We start with a coupled representation $R_+\overline{R_-}$, apply (3.146) to write it in terms of tensor products, then apply (3.153) to decompose it under $U(N_1) \times U(N_2)$, then apply (3.144) to write the $U(N_2)$ part in terms of coupled representations again. This leads straightforwardly to

$$R_+\overline{R_-} \to \bigoplus_{A_\pm, Q_\pm} \left(\sum_{S', A'} (-)^{|S'|} N^{R_+}_{A_+ Q_+ S' A'} N^{R_-}_{A_- Q_- S'^t A'} \right) [Q_+ \otimes \overline{Q_-}, A_+\overline{A_-}]. \tag{3.154}$$

But using (3.118) the sums over S' and A' cancel one another, and we are left with

$$R_+\overline{R_-} \to \bigoplus_{A_\pm, Q_\pm} \left(N^{R_+}_{A_+ Q_+} N^{R_-}_{A_- Q_-} \right) [Q_+ \otimes \overline{Q_-}, A_+\overline{A_-}]. \tag{3.155}$$

Note that for this formula to make sense we need that the $A_+\overline{A_-}$ that appear are all well defined, which requires that $R_\pm$ are shorter than $\frac{1}{2}N_2$ rows.

Tensoring with powers of the determinant representation is also straightforward — writing this representation [1], it has the simple branching rule $[1] \to [[1], [1]]$. This leads to

$$R_+\overline{R_-}[l] \to \bigoplus_{A_\pm, Q_\pm} \left(N^{R_+}_{A_+ Q_+} N^{R_-}_{A_- Q_-} \right) \left[Q_+ \otimes \overline{Q_-} \otimes [l], A_+\overline{A_-}[l] \right]. \tag{3.156}$$

The form of (3.156) that we ultimately use will be, when $\mathcal{R} = R_+\overline{R_-}[l]$ and $\mathcal{A} = A_+\overline{A_-}[l]$,

$$\mathrm{Tr}_{\mathcal{R}/\mathcal{A}}(e^{i\phi}) = s_{R_+/A_+}(e^{i\phi}) s_{R_-/A_-}(e^{-i\phi}) \det(e^{il\phi}). \tag{3.157}$$

Here $e^{i\phi} \in U(N_1)$, and we emphasize again that (3.157) holds only when $R_\pm$ are shorter than $\frac{1}{2}N_2$ rows.

Factorization of $S_{\mathcal{P}\mathcal{Q}}$

In order to understand the large-N factorization of the q-deformed Yang-Mills with insertions, we need to study the properties of the modular matrix $S_{\mathcal{P}\mathcal{Q}}$ of the $U(N)$ Chern-Simons theory in the case where $\mathcal{P}$ and $\mathcal{Q}$ are coupled representations,

$$\mathcal{P} = P_+\overline{P_-}[l], \tag{3.158}$$

$$\mathcal{Q} = Q_+\overline{Q_-}[m]. \tag{3.159}$$

The most naive expectation would be that $S_{\mathcal{P}\mathcal{Q}}$ would be factorized into a piece depending on P_+, Q_+ and a piece depending on P_-, Q_-. As we will show below, the correct formula is a sum of such terms,

$$S_{\mathcal{P}\mathcal{Q}} = M(q)\eta(q)^N q^{-\frac{1}{2}(\kappa_{Q_+}+\kappa_{Q_-})+(-\frac{1}{2}N-m)|P_+|+(-\frac{1}{2}N+m)|P_-|+(-\frac{1}{2}N-l)|Q_+|+(-\frac{1}{2}N+l)|Q_-|-2lmN} \times \\ \sum_{R'} (-)^{|R'|} q^{N|R'|} C_{P_+Q_+^t R'} C_{P_-Q_-^t R'^t}, \tag{3.160}$$

where C is the topological vertex of [7] (in canonical framing.) This formula was already obtained in [10], in the special case $P_\pm = 0$, by direct computation using results from [79].

Here we will give a physical argument which explains the reason for the factorization in the more general case of arbitrary $P_\pm$ and $Q_\pm$.[16] We restrict to the case $l = m = 0$, i.e. $\mathcal{P} = P_+\overline{P_-}$ and $\mathcal{Q} = Q_+\overline{Q_-}$; the dependence on l and m is easily restored using (3.110) and (3.149). The idea is to realize the left side of (3.160) as the partition function of the A model topological string on the resolved conifold T^*S^3, with N branes wrapped on S^3 and four infinite stacks of non-compact branes. Via

[16] Our argument is not completely rigorous, but we hasten to add that the final result has been checked on a computer for a variety of representations $P_\pm$ and $Q_\pm$.

the geometric transition of [62] this is equal to the partition function of the A model on the deformed conifold $\mathcal{O}(-1) \oplus \mathcal{O}(-1) \to \mathbb{CP}^1$, including the four infinite stacks of non-compact branes [100]. The latter partition function can be computed using the topological vertex techniques of [7], which (as we will see) gives the right side of (3.160).[17]

So we begin with the geometry T^*S^3. As in [9], we represent it as a $T^2 \times \mathbb{R}$ fibration over $\mathbb{R}^3$. There are two lines l, l' in $\mathbb{R}^3$ over which an S^1 of the $T^2 \times \mathbb{R}$ fiber degenerates, which are shown in Figure 3.9. Also shown in that figure is the Lagrangian submanifold S^3, which is a T^2 fibration over a line interval connecting l and l'. Finally, we also indicate four Lagrangian submanifolds of T^*S^3, constructed as described in [11]. Each such submanifold has topology $S^1 \times \mathbb{R}^2$.

We will consider the topological A model on this geometry. On each Lagrangian submanifold we place an infinite stack of A model D-branes. There is then a $GL(\infty)$-valued complexified Wilson line on each stack, which couples to the open strings and thus enters the partition function. We write these four Wilson lines $U_\pm$ and $U'_\pm$, as indicated in the figure. We also include N D-branes on the Lagrangian submanifold S^3.

The A model partition function in this geometry can be computed following [100, 9]. The closed string partition function is essentially trivial — it just gives an overall function of q, which is potentially ambiguous due to the non-compactness of the Calabi-Yau. We set it to 1 here. The open string partition function receives contributions from annulus diagrams running along the lines l and l'. On each line

[17] Although the geometry $\mathcal{O}(-1) \oplus \mathcal{O}(-1) \to \mathbb{CP}^1$ is also considered in the main text, the role the topological string is playing here is quite different from the way it appears there. We are using it here only as an auxiliary tool to prove the factorization of $S_{\mathcal{PQ}}$.

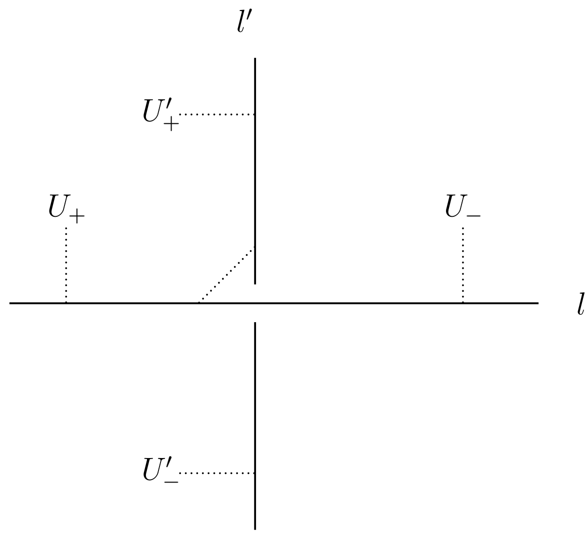

Figure 3.9: A toric representation of the resolved conifold geometry T^*S^3, with the Lagrangian submanifold S^3 indicated, as well as four noncompact Lagrangian submanifolds with topology $S^1 \times \mathbb{R}^2$. Each noncompact submanifold supports an infinite stack of branes with the $GL(\infty)$-valued complexified holonomy indicated.

there are three kinds of annuli which contribute: one kind with the two boundaries on the two infinite stacks of branes, and two kinds with one boundary on an infinite stack and one boundary on the N branes on S^3.

Integrating out the open string sector connecting the two infinite stacks, one gets a contribution to the partition function

$$\sum_R (-)^{|R|} s_R(U_+) s_{R^t}(U_-), \tag{3.161}$$

while the sectors connecting the infinite stacks to the N branes on S^3 contribute operators

$$\left(\sum_{P_+} s_{P_+}(U_+) \mathrm{Tr}_{P_+}(V)\right) \left(\sum_{P_-} s_{P_-}(U_-) \mathrm{Tr}_{P_-}(V^\dagger)\right), \tag{3.162}$$

with V representing the holonomy around the S^1 where the annuli over l meet S^3.

Combining (3.161) and (3.162), one obtains for the open string contribution from the branes on l

$$\sum_{R,P_\pm} (-)^{|R|} s_{R\otimes P_+}(U_+) s_{R^t \otimes P_-}(U_-) \mathrm{Tr}_{\mathcal{P}}(V). \tag{3.163}$$

Using the formula (3.148), (3.163) becomes

$$\sum_{P_\pm} s_{P_+}(U_+) s_{P_-}(U_-) \mathrm{Tr}_{\mathcal{P}}(V). \tag{3.164}$$

Similarly, from the branes on l' we obtain

$$\sum_{Q_\pm} s_{Q_+}(U'_+) s_{Q_-}(U'_-) \mathrm{Tr}_{\mathcal{Q}}(V'), \tag{3.165}$$

where V' is the holonomy on the S^1 where the annuli over l' meet S^3. Altogether, then, the contribution from open strings which involve the four infinite stacks of branes is

$$\sum_{P_\pm, Q_\pm} \left[s_{P_+}(U_+) s_{P_-}(U_-) s_{Q_+}(U'_+) s_{Q_-}(U'_-) \right] \mathrm{Tr}_{\mathcal{P}}(V) \mathrm{Tr}_{\mathcal{Q}}(V'). \tag{3.166}$$

We view this $s_P(V) s_Q(V')$ as a product of Wilson line operators deforming the open string theory on the N branes on S^3, namely the $U(N)$ Chern-Simons theory; these operators are arranged so as to give a Hopf link in S^3. The Chern-Simons amplitude with this link inserted is simply $S_{\mathcal{PQ}}$ [130], so the topological string partition function is finally

$$\psi_{\mathrm{top}} = \sum_{P_\pm, Q_\pm} \left[s_{P_+}(U_+) s_{P_-}(U_-) s_{Q_+}(U'_+) s_{Q_-}(U'_-) \right] S_{\mathcal{PQ}}. \tag{3.167}$$

To get the desired factorization of $S_{\mathcal{PQ}}$ we now compute this partition function in another way: namely, we consider the geometric transition [62] to the deformed conifold $\mathcal{O}(-1) \oplus \mathcal{O}(-1) \to \mathbb{CP}^1$, with the volume of $\mathbb{CP}^1$ given by $t = N g_s$, and with four infinite stacks of branes, as shown in Figure 3.10. Using the topological vertex of [7], the A model partition function in this geometry can be computed as[18]

$$\psi_{\mathrm{top}} = M(q)\eta(q)^N \sum_{P_\pm, Q_\pm} \left[s_{P_+}(U_+) s_{P_-}(U_-) s_{Q_+}(U'_+) s_{Q_-}(U'_-) \right] \times$$
$$q^{-\frac{1}{2}(\kappa(Q_+)+\kappa(Q_-)+N(|P_+|+|Q_+|+|P_-|+|Q_-|))} \sum_{R'} (-)^{|R'|} q^{N|R'|} C_{P_+ Q_+^t R'} C_{P_- Q_-^t R'^t}. \quad (3.168)$$

More precisely, the factors $M(q)\eta(q)^N$ in (3.168) do not appear in [7], so they deserve some extra comment. The factor $M(q)$ is associated with the closed string sector; namely, in the large volume limit, it was shown in [59, 49] that the closed A model partition function reduces to $M(q)^{\chi/2}$ on a Calabi-Yau threefold with Euler characteristic χ. In the non-compact case we are considering here χ is ambiguous, but the change in χ that occurs due to the geometric transition would naturally be $\Delta\chi = 2$ (a 3-cycle gets replaced by a 2-cycle). Thus, since we took $\chi = 0$ before the transition (we chose the closed string contribution in ψ_{top} to be 1), we should take $\chi = 2$ after the transition, which accounts for the factor $M(q)$. The factor $\eta(q)^N$ is not as easily understood, but is presumably associated with the fact that N D3-branes have disappeared in the transition; the same factor appeared in [110] associated to a single noncompact D3-brane. Comparing (3.167) and (3.168) one obtains the desired formula (3.160).

One can also compute a factorization formula for $1/S_{0\mathcal{P}}$, as was done in [10]:

$$\frac{1}{S_{0\mathcal{P}}} = M(q)^{-1}\eta(q)^{-N} q^{\frac{1}{2}N(|P_+|+|P_-|)} \frac{\sum_R C_{P_+ 0 R} q^{N|R|} C_{P_- 0 R}}{C^2_{00P_+} C^2_{00P_-}} \quad (3.169)$$

It would be interesting to know whether there is a physical argument for this factorization formula along the lines of the argument given above for (3.160).

[18] One could determine the framing factors in (3.168) from first principles; we determined them by requiring that the large N limit of our factorization formula be correct.

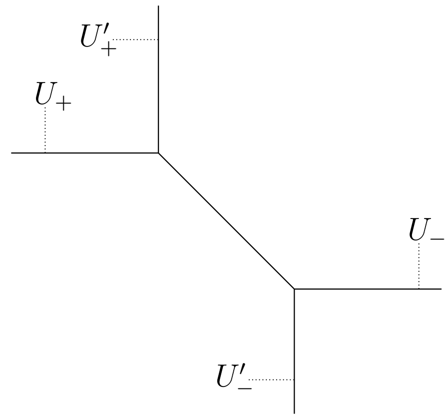

Figure 3.10: The geometry of Figure 3.9 after the geometric transition.

Factorization of q-deformed Yang-Mills amplitudes

Now we are ready to consider the large N factorization of the q-deformed Yang-Mills amplitudes with operator insertions. We will approach the problem by factorizing each of the elementary ingredients from Appendix 3.8 separately.

In the large N ('t Hooft) limit, a convenient basis for the Hilbert space $\mathcal{H}$ is given by the characters of the "coupled representations" which we introduced in Appendix 3.7; we write $\mathcal{R} = |R_+\overline{R_-}[l]\rangle$ for the states corresponding to the coupled representations. As was argued in [64], these representations are the only ones which contribute to the large N amplitudes, to all orders in $1/N$; the reason is that they are the only ones with $C_2(\mathcal{R}) \sim N$. All other representations are exponentially suppressed in the 't Hooft limit by the factors $e^{-\frac{1}{2}ag^2_{\mathrm{YM}}C_2(\mathcal{R})}$ which appear on a surface of area a as in Appendix 3.8 — in the large N limit they give contributions which are $\mathcal{O}(e^{-N})$.

In this factorized basis, the ingredients of the amplitudes can be written as follows:

The annulus. Using (3.127) together with (3.149) and (3.150), we obtain easily

$$\langle R_{1+}\overline{R_{1-}}[l_1]|A_a|R_{2+}\overline{R_{2-}}[l_2]\rangle = \delta_{R_{1+}R_{2+}}\delta_{R_{1-}R_{2-}}\delta_{l_1 l_2}e^{-a\frac{1}{2}g_{\mathrm{YM}}^2 Nl^2}e^{iNa\theta_{\mathrm{YM}}l}\times$$
$$e^{-a\left(\frac{1}{2}g_{\mathrm{YM}}^2(\kappa_{R_+}+N|R_+|+2l|R_+|)-i\theta_{\mathrm{YM}}|R_+|\right)}e^{-a\left(\frac{1}{2}g_{\mathrm{YM}}^2(\kappa_{R_-}+N|R_-|-2l|R_-|)+i\theta_{\mathrm{YM}}|R_-|\right)}. \quad (3.170)$$

The Wilson line freezing operator. From (3.131) and (3.157) we find

$$\langle R_{1+}\overline{R_{1-}}[l_1]|\delta_M(e^{i\oint A}, e^{i\phi})|R_{2+}\overline{R_{2-}}[l_2]\rangle =$$
$$\delta_{l_1,l_2}\left(\sum_{A_+} s_{R_{1+}/A_+}(e^{-i\phi})s_{R_{2+}/A_+}(e^{i\phi})\right)\left(\sum_{A_-} s_{R_{1-}/A_-}(e^{i\phi})s_{R_{2-}/A_-}(e^{-i\phi})\right). \quad (3.171)$$

The disc. From (3.132) and (3.160) this is

$$\Psi_0 = M(q)\eta(q)^N \sum_{l\in\mathbb{Z},R_\pm} q^{-\frac{1}{2}N(|R_+|+|R_-|)}\left[\sum_{R'}(-)^{|R'|}q^{N|R'|}C_{R_+0R'}C_{R_-0R'^t}\right]|R_+\overline{R_-}[l]\rangle. \quad (3.172)$$

The dual Wilson line freezing operator. From (3.133), (3.160) and (3.157) we get

$$\Psi_\phi = M(q)\eta(q)^N\times$$
$$\sum_{l,m\in\mathbb{Z},R_\pm,S_\pm} q^{-\frac{1}{2}(\kappa_{S_+}+\kappa_{S_-})+(-\frac{1}{2}N-m)|R_+|+(-\frac{1}{2}N+m)|R_-|+(-\frac{1}{2}N-l)|S_+|+(-\frac{1}{2}N+l)|S_-|-2lmN}\times$$
$$\left[\sum_{R'}(-)^{|R'|}q^{N|R'|}C_{R_+S_+^tR'}C_{R_-S_-^tR'^t}\right]s_{S_+}(e^{-i\phi})s_{S_-}(e^{i\phi})\det(e^{im\phi})|R_+\overline{R_-}[l]\rangle. \quad (3.173)$$

The trinion (pair of pants). From (3.134) and (3.169) this is

$$T = \sum_{l\in\mathbb{Z},R_\pm}\left[\sum_{R'} q^{\frac{1}{2}N(|R_+|+|R_-|)}\frac{C_{R_+0R'}q^{N|R'|}C_{R_-0R'}}{C_{00R_+}^2C_{00R_-}^2}\right]|R_+\overline{R_-}[l]\rangle\otimes|R_+\overline{R_-}[l]\rangle\otimes|R_+\overline{R_-}[l]\rangle. \quad (3.174)$$

Chapter 4

Topological M-theory as unification of form theories of gravity

4.1 Introduction

The search for a quantum theory of gravity has been a source of puzzles and inspirations for theoretical physics over the past few decades. The most successful approach to date is string theory; but, beautiful as it is, string theory has many extra aspects to it which were not asked for. These include the appearance of extra dimensions and the existence of an infinite tower of increasingly massive particles. These unexpected features have been, at least in some cases, a blessing in disguise; for example, the extra dimensions turned out to be a natural place to hide the microstates of black holes, and the infinite tower of particles was necessary in order for the AdS/CFT duality to make sense. Nevertheless, it is natural to ask whether there could be simpler theories of quantum gravity. If they exist, it might be possible to understand them more deeply, leading us to a better understanding of what it means to quantize gravity; furthermore, simple theories of gravity might end up being the backbone of the more complicated realistic theories of quantum gravity.

In the past decade, some realizations of this notion of a "simpler" theory of gravity have begun to emerge from a number of different directions. The common thread in all these descriptions is that, in the theories of gravity which appear, the metric is *not* one of the fundamental variables. Rather, these theories describe dynamics of gauge fields or higher p-forms, in terms of which the metric can be reconstructed. These theories generally have only a finite number of fields; we shall call them *form theories of gravity*.

Notable examples of form theories of gravity are[1] the description of 3-dimensional gravity in terms of Chern-Simons theory, the description of 4-dimensional gravity in terms of $SU(2)$ gauge theory coupled to other fields, the description of the target space theory of A model topological strings in terms of variations of the Kähler 2-form, and the description of the target space theory of the B model in terms of variations of the holomorphic 3-form.

Meanwhile, recent developments in the study of the topological A and B models suggest that we need a deeper understanding of these theories. On the one hand, they have been conjectured to be S-dual to one another [94, 96]. On the other hand, the A model has been related to a quantum gravitational foam [97, 80]. Moreover, their nonperturbative definition has begun to emerge through their deep connection

[1]One could also include in this list, as will be discussed later in this chapter, the case of 2-dimensional gravity in the target space of the non-critical $c = 1$ string; in that case one gets a theory involving a symplectic form on a 2-dimensional phase space, defining a Fermi surface, in term of which the metric and other fields can be reconstructed.

with the counting of BPS black hole states in 4 dimensions [98, 124]. There is also a somewhat older fact still in need of a satisfactory explanation: it has been known for a while that the holomorphic anomaly of topological strings [24] can be viewed as the statement that the partition function of topological string is a state in some 7-dimensional theory, with the Calabi-Yau 3-fold realized as the boundary of space [131] (see also [45]).

Parallel to the new discoveries about topological strings was the discovery of new actions for which the field space consists of "stable forms" [74]. The critical points of these actions can be used to construct special holonomy metrics. A particularly interesting example is a 3-form theory which constructs G_2 holonomy metrics in 7 dimensions. Interestingly enough, as we will explain, the Hamiltonian quantization of this theory looks a lot like a combination of the A and B model topological strings, which appear in terms of conjugate variables. All this hints at the existence of a "topological M-theory" in 7 dimensions, whose effective action leads to G_2 holonomy metrics and which can reduce to the topological A and B models.

The main aim of this chapter is to take the first steps in developing a unified picture of all these form theories of gravity. Our aim is rather modest; we limit ourselves to introducing some of the key ideas and how we think they may be related, without any claim to presenting a complete picture. The 7-dimensional theory will be the unifying principle; it generates the topological string theories as we just noted, and furthermore, the interesting gravitational form theories in 3 and 4 dimensions can be viewed as reductions of this 7-dimensional form theory near associative and coassociative cycles.

We will also find another common theme. The form theories of gravity naturally lead to calibrated geometries, which are the natural setting for the definition of supersymmetric cycles where branes can be wrapped. This link suggests an alternative way to view these form theories, which may indicate how to define them at the quantum level: they can be understood as counting the BPS states of wrapped branes of superstrings. Namely, recall that in the superstring there is an attractor mechanism relating the charges of the black hole (the homology class of the cycle they wrap on) to specific moduli of the internal theory (determining the metric of the internal manifold). We will see that the attractor mechanism can be viewed as a special case of the general idea of obtaining metrics from forms.

The organization of this chapter is as follows. In Section 4.2, we provide evidence for the existence of topological M-theory in 7 dimensions. In particular, we use the embedding of the topological string into the superstring to give a working definition of topological M-theory in terms of topological strings in 6 dimensions, with an extra circle bundle providing the "11-th" direction. We also give a more extensive discussion of how the very existence of topological M-theory could help resolve a number of puzzles for topological strings in 6 dimensions. In Section 4.3, we give a short review of some form gravity theories in dimensions 2, 3, 4 and 6. In Section 4.4, we discuss some new action principles constructed by Hitchin, which lead to effective theories of gravity in 6 and 7 dimensions. These gravity theories are related to special holonomy manifolds and depend on the mathematical notion of "stable form," so we begin by reviewing these topics; then we introduce Hitchin's actions in 6 and 7 dimensions, as well as a classical Hamiltonian formulation of the 7-dimensional theory. In Section

4.5, we argue that these new gravity theories in 6 dimensions are in fact reformulations of the target space dynamics of the A and B model topological string theories. In Section 4.6, we show how the 7-dimensional theory reduces classically to the 3, 4 and 6-dimensional gravity theories we reviewed in Section 3. In Section 4.7, we discuss canonical quantization of the 7-dimensional theory; we show that it is related to the A and B model topological strings, and we argue that this perspective could shed light on the topological S-duality conjecture. In Section 4.8, we reinterpret the gravitational form theories as computing the entropy of BPS black holes. In Section 4.9, we discuss a curious holographic connection between twistor theory and the topological G_2 gravity. In Section 4.10, we discuss an interesting connection between the phase space of topological M-theory and $\mathcal{N} = 1$ supersymmetric vacua in 4 dimensions. Finally, in Section 4.11 we discuss possible directions for further development of the ideas discussed in this chapter.

4.2 Evidence for topological M-theory

In order to define a notion of topological M-theory, we exploit the connection between the physical superstring and the physical M-theory. Recall that we know that topological strings make sense on Calabi-Yau 3-folds, and topological string computations can be embedded into the superstring. It is natural to expect that the dualities of the superstring, which found a natural geometric explanation in M-theory, descend to some dualities in topological theories, which might find a similar geometric explanation in topological M-theory. Thus a natural definition of topological M-theory is that it should be a theory with one extra dimension relative to the topological string, for a total of 7. Moreover, we should expect that *M-theory on* $M \times \mathbf{S}^1$ *is equivalent to topological strings on* M, where M is a Calabi-Yau manifold. More precisely, here we are referring to the topological A model on M. The worldsheets of A model strings are identified with M-theory membranes which wrap the $\mathbf{S}^1$. Later we will see that in some sense the M-theory formalism seems to automatically include the B model along with the A model, with the two topological string theories appearing as conjugate variables. The topological string should be a good guide to the meaning of topological M-theory, at least in the limit where the $\mathbf{S}^1$ has small radius. One would expect that the radius of the $\mathbf{S}^1$ gets mapped to the coupling constant of the topological string. Of course, topological M-theory should provide an, as yet not well-defined, nonperturbative definition of topological string theory.

So far we only discussed a constant size $\mathbf{S}^1$, but we could also consider the situation where the radius is varying, giving a more general 7-manifold. The only natural class of such manifolds which preserves supersymmetry and is purely geometric is the class of G_2 holonomy spaces; indeed, that there should be a topological theory on G_2 manifolds was already noted in [71], which studied Euclidean M2-brane instantons wrapping associative 3-cycles. So consider M-theory on a G_2 holonomy manifold X with a $U(1)$ action. This is equivalent to the Type IIA superstring, with D6 branes wrapping Lagrangian loci on the base where the circle fibration degenerates. *We define topological M-theory on* X *to be equivalent to A model topological strings on* $X/U(1)$*, with Lagrangian D-branes inserted where the circle fibration degenerates.* This way of defining a topological M-theory on G_2 was suggested in [8, 12].

In this setting, the worldsheets of the A model can end on the Lagrangian branes;

when lifted up to the full geometry of X these configurations correspond to honest closed 3-cycles which we identify as membrane worldvolumes. Moreover, string worldsheets which happen to wrap *holomorphic* cycles of the Calabi-Yau lift to membranes wrapping associative 3-cycles of the G_2 holonomy manifold. So, roughly speaking, we expect that topological M-theory should be classically a theory of G_2 holonomy metrics, which gets quantum corrected by membranes wrapping associative 3-cycles — in the same sense as the topological A model is classically a theory of Kähler metrics, which gets quantum corrected by strings wrapping holomorphic cycles. We can be a little more precise about the coupling between the membrane and the metric: recall that a G_2 manifold comes equipped with a 3-form Φ and a dual 4-form $G = *\Phi$, in terms of which the metric can be reconstructed. We will see that it is natural to consider this G as a field strength for a gauge potential, writing $G = G_0 + d\Gamma$; then Γ is a 3-form under which the membrane is charged.

So we have a workable definition of topological M-theory, which makes sense on 7-manifolds with G_2 holonomy, at least perturbatively to all orders in the radius of the circle. Thus the existence of the theory is established in the special cases where we have a $U(1)$ action on X; we conjecture that this can be extended to a theory which makes sense for arbitrary G_2 holonomy manifolds. This is analogous to what we do in the physical superstring; we do not have an *a priori* definition of M-theory on general backgrounds, but only in special situations.

Now that we have established the existence of a topological M-theory in 7 dimensions (more or less at the same level of rigor as for the usual superstring/M-theory relation), we can turn to the question of what new predictions this theory makes. Indeed, we now suggest that it may solve two puzzles which were previously encountered in the topological string.

There has been a longstanding prediction of the existence of a 7-dimensional topological theory from a very different perspective, namely the wavefunction property of the topological string partition function, which we now briefly recall in the context of the B model. The B model is a theory of variations $\delta\Omega$ of a holomorphic 3-form on a Calabi-Yau 3-fold X. Its partition function is written $Z_B(x; \Omega_0)$. Here x refers to the zero mode of $\delta\Omega$, $x \in H^{3,0}(X, \mathbb{C}) \oplus H^{2,1}(X, \mathbb{C})$, which is not integrated over in the B model. The other variable Ω_0 labels a point on the moduli space of complex structures on X; it specifies the background complex structure about which one perturbs. Studying the dependence of Z_B on Ω_0 one finds a "holomorphic anomaly equation" [24, 23], which is equivalent to the statement that Z_B is a *wavefunction* [131], defined on the phase space $H^3(X, \mathbb{R})$. Namely, different Ω_0 just correspond to different polarizations of this phase space, so $Z_B(x; \Omega_0)$ is related to $Z_B(x; \Omega_0')$ by a Fourier-type transform. This wavefunction behavior is mysterious from the point of view of the 6-dimensional theory on X. On the other hand, it would be natural in a 7-dimensional theory: namely, if X is realized as the boundary of a 7-manifold Y, then path integration over Y gives a wavefunction of the boundary conditions one fixes on X.

Another reason to expect a 7-dimensional description comes from the recent conjectures that the A model and B model are independent only perturbatively. Namely, each contains nonperturbative objects which could naturally couple to the fields of the other. The branes in the A model are wrapped on Lagrangian cycles, the volume of which are measured by some 3-form, and it is natural to identify this 3-form with the holomorphic 3-form Ω of the B model; conversely, the branes in the B model are wrapped on holomorphic cycles, whose volumes would be naturally measured by the Kähler form k of the A model. This observation has led to the conjecture [94, 96] that nonperturbatively both models should include both types of fields and branes, and in fact that the two could even be S-dual to one another. One is thus naturally led to search for a nonperturbative formulation of the topological string which would naturally unify the A and B model branes and fields. Such a unification is natural in the 7-dimensional context: near a boundary with unit normal direction dt, the 3- and 4-forms Φ, G defining the G_2 structure naturally combine the fields of the A and B model on the boundary,

$$\Phi = \mathrm{Re}\ \Omega + k \wedge dt, \tag{4.1}$$

$$G = \mathrm{Im}\ \Omega \wedge dt + \frac{1}{2} k \wedge k. \tag{4.2}$$

Later we will see that the A and B model fields are canonically conjugate in the Hamiltonian reduction of topological M-theory on $X \times \mathbb{R}$. In particular, the wavefunctions of the A model and the B model cannot be defined simultaneously.

4.3 Form theories of gravity in diverse dimensions

The long-wavelength action of the "topological M-theory" we are proposing will describe metrics equipped with a G_2 structure. In fact, as we will discuss in detail, the 7-dimensional metric in this theory is reconstructed from the 3-form Φ (or equivalently, from the 4-form $G = *\Phi$). This might at first seem exotic: the metric is not a fundamental field of this theory but rather can be reconstructed from Φ. However, similar constructions have appeared in lower dimensions, where it is believed at least in some cases that the reformulation in terms of forms ("form theory of gravity") is a better starting point for quantization: we know how to deal with gauge theories, and perhaps more general form theories, better than we know how to deal with gravity. Of course, *rewriting a classical theory of gravity in terms of classical forms is no guarantee that the corresponding quantum theory exists.* We are certainly not claiming that arbitrary form theories make sense at the quantum level!

Nevertheless, in low dimensions some special form gravity theories have been discussed in the literature, which we believe do exist in the quantum world — and moreover, as we will see, these theories are connected to topological M-theory, which we have already argued should exist.

In this section we review the form gravity theories in question. They describe various geometries in 2, 3, 4 and 6 dimensions. Here we will discuss mainly their classical description. It is more or less established that the theories discussed below in dimensions 2, 3, and 6 exist as quantum theories, at least perturbatively. In dimensions 2 and 6, this is guaranteed by topological string constructions. In dimension 3 also, the quantum theory should exist since it is known to lead to well defined invariants of 3-manifolds. The 4-dimensional theory, which gives self-dual gravity, is not known to exist in full generality, although for zero cosmological constant, it is related to the Euclidean $\mathcal{N} = 2$ string, which is known to exist perturbatively [99]. For the case of nonzero cosmological constant, we will give further evidence that the theory exists at the quantum level by relating it to topological M-theory later in this chapter.

4.3.1 $2D$ form gravity

By the $2D$ form theory of gravity we have in mind the theory which appears in the target space of non-critical bosonic strings, or more precisely, the description of the large N limit of matrix models in terms of the geometry of the eigenvalue distribution. The basic idea in this theory of gravity is to study fluctuations of a Fermi surface in a 2-dimensional phase space. The dynamical object is the area element ω, representing the phase space density, which is defined to be non-zero in some region R and vanishing outside. By a choice of coordinates, we can always write this area element as $\omega = dx \wedge dp$ inside R and zero elsewhere. Hence the data of the theory is specified by the boundary ∂R, which we can consider locally as the graph of a function $p(x)$. The study of the fluctuations of the boundary is equivalent to that of fluctuations of ω. Actually, in this gravity theory one does not allow arbitrary fluctuations; rather, one considers only those which preserve the integral

$$\oint p(x)\, dx = A. \tag{4.3}$$

Such fluctuations can be written $p(x) = p_0(x) + \partial\phi(x)$. In other words, the cohomology class of ω, or "zero mode," is fixed by A, and the "massive modes" captured by the field $\phi(x)$ are the dynamical degrees of freedom.

This gravity theory is related to the large N limit of matrix models, where x denotes the eigenvalue of the matrix and $p(x)\, dx$ denotes the eigenvalue density distribution. A gets interpreted as the rank N of the matrix. One can solve this theory using matrix models, or equivalently using W_∞ symmetries [5].

This theory can also be viewed [5, 7] as the effective theory of the B model topological string on the Calabi-Yau 3-fold $uv = F(x, p)$, where $F(x, p) = 0$ denotes the

Fermi surface. In this language, $\omega = dx \wedge dp$ is the reduction of the holomorphic 3-form to the (x, p) plane. In the B model one always fixes the cohomology class of the 3-form; here this reduces to fixing the area A as we described above.

4.3.2 $3D$ gravity theory as Chern-Simons gauge theory

Now we turn to the case of three dimensions. Pure gravity in three dimensions is topological, in a sense that it does not have propagating gravitons. In fact, we can write the usual Einstein-Hilbert action with cosmological constant Λ,

$$S_{\text{grav}} = \int_M \sqrt{g}\Big(R - 2\Lambda\Big), \tag{4.4}$$

in the first-order formalism,

$$S = \int_M \text{Tr}\Big(e \wedge F + \frac{\Lambda}{3} e \wedge e \wedge e\Big), \tag{4.5}$$

where $F = dA + A \wedge A$ is the field strength of an $SU(2)$ gauge connection A^i, and e^i is an $SU(2)$-valued 1-form on M. Notice that the gravity action (4.5) has the form of a BF theory, and does not involve a metric on the 3-manifold M. A metric (of Euclidean signature) can however be reconstructed from the fundamental fields — namely, given the $SU(2)$-valued 1-form e, one can write

$$g_{ab} = -\frac{1}{2}\text{Tr}(e_a e_b). \tag{4.6}$$

The equations of motion that follow from (4.5) are:

$$D_A e = 0, \tag{4.7}$$

$$F + \Lambda e \wedge e = 0. \tag{4.8}$$

The first equation says that A can be interpreted as the spin connection for the vielbein e^i, while the second equation is the Einstein equation with cosmological constant Λ.

Gravity in three dimensions has a well-known reformulation in terms of Chern-Simons gauge theory [4, 129],

$$S = \int_M \mathrm{Tr}\Big(\mathcal{A} \wedge d\mathcal{A} + \frac{2}{3}\mathcal{A} \wedge \mathcal{A} \wedge \mathcal{A}\Big), \tag{4.9}$$

where $\mathcal{A}$ is a gauge connection with values in the Lie algebra of the gauge group G. The gauge group G is determined by the cosmological constant, and can be viewed as the isometry group of the underlying geometric structure. Specifically, in the Euclidean theory, G is one of $SL(2,\mathbb{C})$, or $ISO(3)$, or $SU(2) \times SU(2)$, depending on the cosmological constant:

Cosmological Constant	$\Lambda < 0$	$\Lambda = 0$	$\Lambda > 0$
Gauge group G	$SL(2,\mathbb{C})$	$ISO(3)$	$SU(2) \times SU(2)$

The equations of motion that follow from the Chern-Simons action (4.9) imply that the gauge connection $\mathcal{A}$ is flat,

$$d\mathcal{A} + \mathcal{A} \wedge \mathcal{A} = 0. \tag{4.10}$$

Writing this equation in components one can reproduce the equations of motion (4.7). For example, if $\Lambda < 0$, one can write the complex gauge field $\mathcal{A}$ as $\mathcal{A}^k = w^k + ie^k$. Substituting this into (4.10) and combining the real and imaginary terms, we recognize the equations (4.7) with $\Lambda = -1$. Finally, we note that, in the Chern-Simons theory, the gauge transformation with a parameter ϵ has the form

$$\delta_\epsilon \mathcal{A} = d\epsilon - [\mathcal{A}, \epsilon]. \tag{4.11}$$

One can also describe the quantum version of $3D$ gravity directly via various discrete models. For example, given a triangulation Δ of M one can associate to each tetrahedron a quantum $6j$-symbol and, following Turaev and Viro [122], take the state sum

$$TV(\Delta) = \Big(- \frac{(q^{1/2} - q^{-1/2})^2}{2k} \Big)^V \sum_{j_e} \prod_{\text{edges}} [2j_e + 1]_q \prod_{\text{tetrahedra}} (6j)_q, \tag{4.12}$$

where V is the total number of vertices in the triangulation, and $[2j + 1]_q$ is the quantum dimension of the spin j representation of $SU(2)_q$,

$$[n]_q = \frac{q^{n/2} - q^{-n/2}}{q^{1/2} - q^{-1/2}} \tag{4.13}$$

One can prove [122] that the Turaev-Viro invariant is independent on the triangulation and, therefore, gives a topological invariant, $TV(M) = TV(\Delta)$. Furthermore, it has been shown by Turaev [120, 121] and Walker [127] that the Turaev-Viro invariant is equal to the square of the partition function in $SU(2)$ Chern-Simons theory (also known as the Reshetikhin-Turaev-Witten invariant [130, 105]):

$$TV(M) = |Z_{SU(2)}(M)|^2 \tag{4.14}$$

There is a similar relation between the $SL(2, \mathbb{C})$ Chern-Simons partition function and quantum invariants of hyperbolic 3-manifolds [66].

4.3.3 $4D$ 2-form gravity

In dimension four, there are several versions of "topological gravity". Here we review a theory known as 2-form gravity [104, 78, 118, 119, 2, 88, 93, 1], which also describes the self-dual sector of loop quantum gravity [115].

We begin by writing the action for Einstein's theory in a slightly unconventional way [81, 104, 32]:

$$S_{4D} = \int_{M^4} \Sigma^k \wedge F_k - \frac{\Lambda}{24}\Sigma^k \wedge \Sigma_k + \Psi_{ij}\Sigma^i \wedge \Sigma^j. \tag{4.15}$$

Here A^k is an $SU(2)$ gauge field, with curvature $F^k = dA^k + \epsilon^{ijk}A^j \wedge A^k$, and Σ^k is an $SU(2)$ triplet of 2-form fields, $k = 1, 2, 3$. The parameter Λ will be interpreted below as a cosmological constant. Finally, $\Psi_{ij} = \Psi_{(ij)}$ is a scalar field on M, transforming as a symmetric tensor of $SU(2)$.

To see the connection to ordinary general relativity, one constructs a metric out of the two-form field Σ^k as follows. The equation of motion from varying Ψ_{ij} implies that Σ^k obeys the constraint

$$\Sigma^{(i} \wedge \Sigma^{j)} - \frac{1}{3}\delta^{ij}\Sigma_k \wedge \Sigma^k = 0. \tag{4.16}$$

When (4.16) is satisfied the two-form Σ^k may be reexpressed in terms of a vierbein [32],

$$\Sigma^k = -\eta^k_{ab} e^a \wedge e^b. \tag{4.17}$$

Here e^a are vierbein 1-forms on M^4, $a = 1, \ldots, 4$, and η^k_{ab} is the 't Hooft symbol,

$$\eta^k_{ab} = \epsilon^k_{ab0} + \frac{1}{2}\epsilon^{ijk}\epsilon_{ijab}. \tag{4.18}$$

In other words,

$$\Sigma^1 = e^{12} - e^{34},$$

$$\Sigma^2 = e^{13} - e^{42},$$

$$\Sigma^3 = e^{14} - e^{23},$$

where $e^{ij} = e^i \wedge e^j$. The vierbein in turn determines a metric in the usual way:

$$g = \sum_{a=1}^{4} e^a \otimes e^a. \tag{4.19}$$

With respect to this metric g, the two-forms Σ^k ($k = 1, 2, 3$) are all self-dual in the sense that $\Sigma^k = *\Sigma^k$. (This just follows from their explicit expression (4.17) in terms of the vierbein.)

One can also write the metric directly in terms of Σ without first constructing the vierbein [32]:

$$\sqrt{g}\, g_{ab} = -\frac{1}{12} \Sigma^i_{aa_1} \Sigma^j_{ba_2} \Sigma^k_{a_3a_4} \epsilon^{ijk} \epsilon^{a_1a_2a_3a_4}. \tag{4.20}$$

Having constructed the metric g out of Σ, we now want to check that it obeys Einstein's equation on-shell. The equations of motion which follow from (4.15) are

$$D_A \Sigma = 0, \tag{4.21}$$

$$F_i = \Psi_{ij} \Sigma^j + \frac{\Lambda}{12} \Sigma_i. \tag{4.22}$$

The first equation $D_A \Sigma = 0$ says that A is the self-dual part of the spin connection defined by the metric g. The second equation then contains information about the Riemann curvature[2] acting on self-dual 2-forms Λ^2_+. Namely, since the Σ^k appearing on the right side are also self-dual two-forms, the Riemann curvature maps $\Lambda^2_+ \to \Lambda^2_+$. By the standard decomposition of the Riemann tensor, this implies that the trace-free part of the Ricci curvature vanishes. Then Ψ_{ij} is identified with the self-dual part of the Weyl curvature, and the last term gives the trace part of the Ricci tensor, consistent with the cosmological constant Λ.

[2]Here we are considering R with all indices down, R_{abcd}, as a symmetric map $\Lambda^2 \to \Lambda^2$.

So far we have seen that the action (4.15) reproduces Einstein's theory of gravity, in the sense that the classical solutions correspond exactly to Einstein metrics on M with cosmological constant Λ. Now let us consider the effect of dropping the field Ψ_{ij}, giving

$$S_{4D} = \int_{M^4} \Sigma^k \wedge F_k - \frac{\Lambda}{24} \Sigma^k \wedge \Sigma_k. \tag{4.23}$$

One can consider (4.23) as obtained by starting from (4.15), multiplying the last term by ϵ, and then taking the $\epsilon \to 0$ limit. Just when we reach $\epsilon = 0$ we seem to lose the constraint (4.16), which was the equation of motion for Ψ_{ij} and was crucial for the description of Σ^k in terms of the vierbein. However, at $\epsilon = 0$ something else happens: the action develops a large new symmetry,

$$\begin{aligned} \delta A_k &= \frac{\Lambda}{12} \theta_k, \\ \delta \Sigma_k &= D_A \theta_k. \end{aligned} \tag{4.24}$$

This new symmetry can be used to reimpose the constraint (4.16), so in this sense the $\epsilon \to 0$ limit is smooth and sensible to consider. The only change to the equations of motion is that the term $\Psi_{ij}\Sigma^j$ disappears from the right side of (4.21), leaving

$$\begin{aligned} D_A \Sigma &= 0, \\ F_i &= \frac{\Lambda}{12} \Sigma_i. \end{aligned} \tag{4.25}$$

As we mentioned above, the Ψ_{ij} term represents the self-dual part of the Weyl curvature; so (4.25) imply that the metric constructed from Σ is not only Einstein but also has vanishing self-dual Weyl curvature. In this sense the action (4.23) gives rise to anti-self-dual Einstein manifolds,

$$R_{ab} = \Lambda g_{ab}, \qquad W^{(+)}_{abcd} = 0. \tag{4.26}$$

Note that such manifolds are rather rare compared to ordinary Einstein manifolds; for example, with $\Lambda > 0$ there are just two smooth examples, namely $\mathbf{CP}^2$ and $\mathbf{S}^4$ [72]. With $\Lambda = 0$ the solutions are hyperkähler metrics in 4 dimensions; these are target spaces for the $\mathcal{N} = 2$ string (or equivalently the $\mathcal{N} = 4$ topological string), which provides a completion of the self-dual gravity theory in that case.

4.3.4 $6D$ form theories: Kähler and Kodaira-Spencer gravity

In dimension 6, two different form theories of gravity arise in ($\mathcal{N} = 2$) topological string theory. One, known as the Kähler gravity theory [25], describes the target space gravity (string field theory) of the topological A model. The second theory, called the Kodaira-Spencer theory of gravity [24], is the string field theory of the topological B model and describes variations of the complex structure. Below we review each of these theories in turn.

We begin with the B model. The basic field of the Kodaira-Spencer gravity theory is a vector-valued 1-form field A, for which the action is given by [24]

$$S_{\mathrm{KS}} = \frac{1}{2}\int_M A'\frac{1}{\partial}\overline{\partial}A' + \frac{1}{6}\int_M (A \wedge A)' \wedge A'. \tag{4.27}$$

Here, we use the standard notation $A' := (A \cdot \Omega_0)$ for the product with the background holomorphic $(3, 0)$ form. The field A then defines a variation of Ω, given by the formula

$$\Omega = \Omega_0 + A' + (A \wedge A)' + (A \wedge A \wedge A)'. \tag{4.28}$$

This expression for the variation of Ω follows from its local expression in complex coordinates, $\Omega = \Omega_{ijk} dz^i \wedge dz^j \wedge dz^k$, where A is interpreted as giving a variation of the 1-form dz^i:

$$dz^i \mapsto dz^i + A^i_{\overline{j}} d\overline{z}^{\overline{j}}. \tag{4.29}$$

In order that the "non-local term" $\frac{1}{\partial}\overline{\partial}A'$ in the action (4.27) make sense, A is not allowed to be an arbitrary vector-valued 1-form; rather, there is a constraint

$$\partial A' = 0. \tag{4.30}$$

Using this constraint we write

$$A' = x + \partial\phi, \qquad x \in H^{2,1}(M, \mathbb{C}). \tag{4.31}$$

Here the harmonic $x \in H^{2,1}(M, \mathbb{C})$ represents the massless modes (moduli) of Ω, which are frozen in the Kodaira-Spencer theory, while $\phi \in \Omega^{1,1}(M, \mathbb{C})$ represents the massive modes, which are the dynamical degrees of freedom. Substituting (4.31) into (4.27), we can write the Kodaira-Spencer action without non-local terms:

$$S_{\mathrm{KS}} = \frac{1}{2}\int_M \partial\phi \wedge \overline{\partial}\phi + \frac{1}{6}\int_M (A \wedge A)' \wedge A' \tag{4.32}$$

The equation of motion that follows from the action (4.32) is

$$\overline{\partial}A' + \partial(A \wedge A)' = 0. \tag{4.33}$$

Using (4.30) and (4.33) together one finds that the holomorphic 3-form (4.28) is closed on-shell,

$$d\Omega = 0. \tag{4.34}$$

Hence solutions to the Kodaira-Spencer field equations correspond to deformations Ω of the holomorphic 3-form Ω_0.

When we view ϕ as the dynamical degree of freedom we must note that it has a large shift symmetry,

$$\phi \mapsto \phi + \epsilon, \tag{4.35}$$

where $\partial\epsilon = 0$. This symmetry can be used to set the anti-holomorphic part of ϕ to zero, i.e. $\overline{\partial}\phi = 0$. In other words, ϕ should be viewed as the analog of the *chiral* boson in 2 dimensions; in this sense the Kodaira-Spencer theory is really a chiral theory. In fact, in the local geometry we discussed in Section 4.3.1, ϕ gets identified with a chiral boson on the Riemann surface $F(x,p) = 0$.

Although A' and the Kodaira-Spencer action depend on $\partial\phi$ rather than on ϕ itself, it turns out that D1-branes of the B model are charged under ϕ directly. To see this, consider a D1-brane wrapped on a 2-cycle E which moves to another 2-cycle E'. There is a 3-chain C which interpolates between E and E', and the variation of the action is given by (absorbing the string coupling constant into Ω)

$$\delta S = \int_C \Omega = \int_C \partial\phi = \int_C d\phi = \int_E \phi - \int_{E'} \phi. \tag{4.36}$$

This coupling also explains the fact that a D1-brane is a source for Ω (and hence shifts the integral of Ω on a 3-cycle linking it.) Namely, including such a source localized along E would modify the equations of motion to [5]

$$\overline{\partial} A' = \overline{\partial}\partial\phi = \delta^4_E, \tag{4.37}$$

so that the kinetic term $\phi\partial\overline{\partial}\phi$ from (4.32) becomes precisely $\int_E \phi$.

The fact that D1-branes couple to ϕ has an important consequence: there is an extra $H^{1,1}(M,\mathbb{C})$ worth of degrees of freedom in ϕ, corresponding to the freedom to shift ϕ by a harmonic form b, which does not affect A' but does figure into non-perturbative aspects of the B-model. Namely, the amplitudes involving D1-brane instantons, which should presumably be included in the nonperturbative definition of the B model, are sensitive to this shift. Thus the partition function of the B model is nonperturbatively a function both of $x \in H^{2,1}(M)$ and of $b \in H^{1,1}(M, \mathbb{C})$. The necessity of the extra field b was also recently noted in [22].

As we will discuss later in more detail, it is natural that in a nonperturbative definition the periods of Ω are quantized in units of g_s. There is an overall $1/g_s^2$ in front of the closed string action, so this will then give the appropriate $1/g_s$ coupling of the field ϕ to the D-branes. Because of this flux quantization, the corresponding "Wilson lines" b will be naturally periodic or $\mathbb{C}^*$ variables.

Having discussed the Kodaira-Spencer theory, let us now turn to another 6-dimensional form theory of gravity, namely the Kähler gravity theory, which describes variations of the Kähler structure on M. Its action is [25, 132]

$$S_{\text{Kahler}} = \int_M \left(\frac{1}{2} K \frac{1}{d^{c\dagger}} dK + \frac{1}{3} K \wedge K \wedge K \right), \tag{4.38}$$

where K is a variation of the (complexified) Kähler form on M, and $d^c = \partial - \overline{\partial}$. The Kähler gravity action (4.38) is invariant under gauge transformations of the form

$$\delta_\alpha K = d\alpha - d^{c\dagger}(K \wedge \alpha), \tag{4.39}$$

where α is a 1-form on M, such that $d^{c\dagger}\alpha = 0$. The equations of motion in the Kähler gravity theory are

$$dK + d^{c\dagger}(K \wedge K) = 0. \tag{4.40}$$

As in the Kodaira-Spencer theory, we can decompose K into massless and massive modes,

$$K = x + d^{c\dagger}\gamma, \qquad x \in H^{1,1}(M, \mathbb{C}) \tag{4.41}$$

where $x \in H^{1,1}(M, \mathbb{C})$ represents the Kähler moduli, which are not integrated over, and $\gamma \in \Omega^3(M)$ contains the massive modes of K. Using (4.41), we can write the Kähler gravity action (4.38) without non-local terms,

$$S_{\text{Kahler}} = \int_M \Big(\frac{1}{2}d\gamma \wedge d^{c\dagger}\gamma + \frac{1}{3}K \wedge K \wedge K\Big). \tag{4.42}$$

Just as in the B model, Lagrangian D-branes of the A model are charged under γ, implying that these branes are sources for K and hence modify the integral of K on 2-cycles which link them. This also implies that the partition function of the A model depends nonperturbatively on the choice of a cohomology class in $H^3(M)$ as well as on $x \in H^2(M)$. Here the same remarks about flux quantization hold as in the B model.

4.4 Hitchin's action functionals

In the previous section, we discussed various form theories of gravity which have appeared previously in the physics literature. Now we turn to some new ones. We will describe actions constructed by Hitchin [74, 73] for which the equations of motion yield special geometric structures: either holomorphic 3-forms Ω and symplectic structures k in 6 dimensions, or G_2 holonomy metrics in 7 dimensions. As for the form theories we considered above, the classical fields in these theories are real p-forms, from which the desired geometric structures can be reconstructed. In 6 dimensions one has a 3-form ρ and a 4-form σ; these forms will be interpreted as $\rho = \text{Re}\ \Omega$, $\sigma = \frac{1}{2}k \wedge k$.

In 7 dimensions, one just has a single 3-form Φ (or its dual 4-form G), interpreted as the associative 3-form (resp. coassociative 4-form) of the G_2 metric.

These action functionals have been used in the physics literature to construct new metrics with special holonomy [67, 36, 39]. In the present context, they should be regarded as effective actions for gravity theories. In 6 dimensions, we will see in Section 4.5 that these gravity theories are related to topological strings. The 7-dimensional action defines a new gravity theory which we identify as the low energy limit of topological M-theory.

4.4.1 Special holonomy manifolds and calibrations

In this subsection we briefly review the notion of special holonomy, which plays an important role in supersymmetric string compactifications, and which we expect to be important for topological string/membrane theories. In particular, we emphasize that the geometric structure on a special holonomy manifold X can be conveniently characterized by certain p-forms, invariant under the holonomy group, $\mathrm{Hol}(X)$.

Recall that for any n-dimensional Riemannian manifold X we have

$$\mathrm{Hol}(X) \subseteq SO(n). \tag{4.43}$$

The manifolds with special (reduced) holonomy are characterized by the condition $\mathrm{Hol}(X) \subset SO(n)$. Below we list some examples of special holonomy manifolds that will be relevant in what follows.

Metric	Holonomy	n	SUSY	Invariant p-forms
Calabi-Yau	$SU(3)$	6	1/4	$p=2$: K (Kähler) $p=3$: Ω
Exceptional	G_2	7	1/8	$p=3$: Φ (associative) $p=4$: $*\Phi$ (coassociative)
	$Spin(7)$	8	1/16	$p=4$: Ψ (Cayley)

Table 2: *Examples of special holonomy manifolds.*

All of these structures can be characterized by the existence of a covariantly constant spinor,

$$\nabla\xi = 0 \tag{4.44}$$

The existence of this ξ guarantees that superstring compactification on X preserves some fraction (also listed in the above table) of the original 32 supercharges, which is what makes such manifolds useful in string theory.

Another characteristic property of special holonomy manifolds is the existence of invariant forms, known as calibrations. Using the covariantly constant spinor ξ one can construct a p-form on X,

$$\omega^{(p)} = \xi^\dagger \gamma_{i_1 \ldots i_p} \xi. \tag{4.45}$$

By construction, such forms are covariantly constant and invariant under $\mathrm{Hol}(X)$. They are non-trivial only for special values of p: see Table 2 for a list of the invariant forms on manifolds of $SU(3)$, G_2, and $Spin(7)$ holonomy. These invariant forms play an important role in geometry and physics; in particular, they can be used to characterize minimal (supersymmetric) cycles in special holonomy manifolds. Indeed, if $S \subset X$ is a minimal submanifold of real dimension p, then its volume can be determined by integrating the invariant form $\omega^{(p)}$ over S,

$$\mathrm{Vol}(S) = \int_S \omega^{(p)}. \tag{4.46}$$

Such a manifold S is called calibrated, and the form $\omega^{(p)}$ is called a calibration. Notice the simplification that has occurred here. Ordinarily, in order to compute the volume, $\mathrm{Vol}(S) = \int d^p x \sqrt{g}$, we need to know the metric g; but the volume of a calibrated submanifold S is given by the simple formula (4.46) which does not involve the explicit form of the metric.

This phenomenon is a prototype of various situations in which the important geometric data can be characterized by differential forms rather than by a metric. This is essentially the same principle that was underlying the constructions of Section 4.3, where we discussed form theories of gravity in which the space-time geometry is encoded in tensor forms and/or gauge fields.

To illustrate further the idea that forms can capture the geometry, let us consider an example which will play an important role below. Let X be a manifold with G_2 holonomy. The existence of a G_2 holonomy metric is equivalent to the existence of an associative 3-form, Φ, which is closed and co-closed,

$$\begin{aligned} d\Phi &= 0 \\ d * \Phi &= 0, \end{aligned} \tag{4.47}$$

and which can be written in terms of an orthonormal vielbein e^i, $i = 1, \ldots, 7$, as

$$\Phi = \frac{1}{3!}\psi_{ijk} e^i \wedge e^j \wedge e^k. \tag{4.48}$$

Here ψ_{ijk} are the structure constants of the imaginary octonions: $\sigma_i \sigma_j = -\delta_{ij} + \psi_{ijk}\sigma_k$, $\sigma_i \in \mathrm{Im}\ (\mathbb{O})$. Conversely, writing Φ locally in the form (4.48) defines a metric g by the formula

$$g = \sum_{i=1}^{7} e^i \otimes e^i. \tag{4.49}$$

This g can be written more explicitly by first defining

$$B_{jk} = -\frac{1}{144}\Phi_{ji_1i_2}\Phi_{kj_3j_4}\Phi_{j_5j_6j_7}\epsilon^{j_1 \ldots j_7}, \tag{4.50}$$

in terms of which the metric has a simple form,

$$g_{ij} = \det(B)^{-1/9} B_{ij}. \tag{4.51}$$

Evaluating the determinant of B_{ij}, we get $\det(g) = \det(B)^{2/9}$, so (4.51) can be written in a more convenient form,

$$\sqrt{g}\, g_{jk} = -\frac{1}{144}\Phi_{ji_1i_2}\Phi_{ki_3i_4}\Phi_{i_5i_6i_7}\epsilon^{i_1 \ldots i_7}. \tag{4.52}$$

Notice that even if the 3-form Φ does not obey (4.47), we can still use (4.52) to construct a metric on X from Φ, as long as the 3-form Φ is non-degenerate in a suitable sense. (Of course, this metric will not have G_2 holonomy unless (4.47) is satisfied.) This construction naturally leads us to the notion of stable forms, which we now discuss.

4.4.2 Stable forms

Following [74], in this section we review the construction of action principles from p-forms. It is natural to define the action of a p-form ρ on a manifold X as a volume form $\phi(\rho)$ integrated over X. One might think that such a $\phi(\rho)$ is hard to construct, as one might have in mind the usual wedge product of ρ with itself, which gives a nonzero top-form only in rather special cases. In fact, this is not the only way to build a volume form out of p-forms; as we will see, all the actions of interest for us turn out to involve a volume element constructed in a rather non-trivial way from the p-form. For example, on a 7-manifold with G_2 holonomy, the volume form cannot be constructed as a wedge product of the associative 3-form Φ with itself; nevertheless, one can define $\phi(\Phi)$ as a volume form for the metric (4.51) constructed from Φ, as we will discuss in detail later.

The most general way to construct a volume form $\phi(\rho)$ from a p-form ρ is as follows: contract a number of $\rho_{i_1,\ldots,i_p}$'s with a number of epsilon tensors $\epsilon^{i_1,\ldots,i_n}$, to obtain some $W(\rho)$. Suppose we use k epsilon tensors in W; then W transforms as the k-th power of a volume, and we can define $\phi(\rho) = W(\rho)^{1/k}$ which is a volume form. It is easy to see that $\phi(\rho)$ scales as an n/p-th power of ρ (if W has q factors of ρ it will need $k = pq/n$ factors of ϵ).

Given such a volume form, one can define the action to be the total volume,

$$V(\rho) = \int_X \phi(\rho). \tag{4.53}$$

This $V(\rho)$ is a homogeneous function of ρ of degree $\frac{n}{p}$:

$$V(\lambda\rho) = \lambda^{\frac{n}{p}} V(\rho) \qquad \lambda \in \mathbb{R}. \tag{4.54}$$

In this chapter we will not be interested in arbitrary ways of putting together ρ and ϵ symbols to make $\phi(\rho)$. Rather, we will focus on cases in which there exists a notion of "generic" p-form. In such cases the generic p-form ρ defines an interesting geometric structure (*e.g.* an almost complex structure or a G_2 structure) even without imposing any additional constraints. Hence arbitrary variations of ρ can be thought of as variations of this structure, and as we will see, critical points of $V(\rho)$ imply integrability conditions on these geometric structures.

The notion of genericity we have in mind is known as *stability*, as described in [73] and reviewed below. The requirement of stability has drastically different consequences depending on the dimension n of the manifold and the degree p of the form. In most cases, as we will see, there are no stable forms at all; but for certain special values of n and p, stable forms can exist. Moreover, *all the calibrations in 6 and 7 dimensions that we discussed earlier turn out to be stable forms.* This deep "coincidence" makes the technology of stable forms a useful tool for the study of special holonomy. Nevertheless, possessing a stable form is far less restrictive than the requirement of special holonomy, needed for supersymmetry.

Let us now define the notion of stability precisely. Write V for the tangent space at a point x, so the space of p-forms at x is $\Lambda^p V^*$. Then a p-form ρ is said to be stable at x if $\rho(x)$ lies in an open orbit of the $GL(V)$ action on $\Lambda^p V^*$. We call ρ a stable form if ρ is stable at every point. In other words, *ρ is stable if all forms in a neighborhood of ρ are equivalent to ρ by a local $GL(n)$ action.*

Some special cases of stability are easy to understand. For example, there are no stable 0-forms, because under coordinate transformations the value of a function does not change. On the other hand, any nonzero n-form is stable, because by a linear transformation one can always map any volume form to any other one. Similarly, any nonzero 1-form or $(n-1)$-form is stable.

A less trivial case of stability is that of a 2-form, on a manifold of even dimension. In this case, viewing the 2-form as an antisymmetric matrix, stability just means that the determinant is nonzero; namely, this is the usual characterization of a presymplectic form, and any such form can be mapped to any other by a linear transformation, so they are indeed stable. Given such a stable form we can now construct its associated volume form: namely, we write $\phi(\omega) = \omega^{n/2}$. Note that the stability of ω is equivalent to $\phi(\omega) \neq 0$.

To understand the geometric structures defined by stable forms, it is useful to study the subgroup of $GL(n)$ which fixes them. In the case of a stable 2-form in even dimension this stabilizer is $Sp(n)$, corresponding to the fact that the 2-form defines a presymplectic structure. More generally, given a stable p-form, we can easily compute the dimension of the stabilizer: it is simply the dimension of $GL(n)$ minus the dimension of the space of p-forms. In the case $p = 2$, this counting gives $n^2 - \frac{n(n-1)}{2} = n(n+1)/2$, as expected.

Next we consider the case $p = 3$. The dimension of the space of 3-forms is

$$\dim \Lambda^3 V^* = n(n-1)(n-2)/6. \tag{4.55}$$

Already at this stage we see that there cannot be any stable 3-forms for large n, because $\dim GL(n) = n^2$ has a slower growth than $n^3/6$, so that $GL(n)$ cannot act locally transitively on the space of 3-forms. However, for $p = 3$ and small enough n the stability condition can be met. We have already discussed the cases $n = 3, 4$, and stable 3-forms also exist for $n = 6, 7, 8$. These special cases lead to interesting geometric structures; for example, consider the case $p = 3, n = 7$. Here the dimension counting gives

$$\begin{aligned} \dim GL(V) = n^2 &= 49, \\ \dim \Lambda^p V^* = \frac{n!}{p!(n-p)!} &= 35, \\ 14 &= \dim G_2. \end{aligned} \tag{4.56}$$

Indeed, the stabilizer of the 3-form in this case is G_2, so a stable 3-form in 7 dimensions defines a G_2 structure.

As we just discussed for $p = 3$, generically $\dim \Lambda^p V^*$ is much larger than $\dim GL(V) = n^2$. Hence stable forms exist only for special values of n and p [73]. The cases of interest for us in this chapter are $n = 7$ with $p = 3, 4$ and $n = 6$ with $p = 3, 4$. We now turn to the construction of volume functionals from stable forms in these cases.

4.4.3 3-form and 4-form actions in $6D$

We begin with the 6-dimensional case. In this case Hitchin constructed two action functionals $V_H(\rho)$ and $V_S(\sigma)$, depending respectively on a 3-form ρ and 4-form σ. When extremized, V_H and V_S yield respectively holomorphic and symplectic structures on M. In this section we introduce these action functionals and describe some of their properties.

Let us first construct $V_S(\sigma)$. Suppose σ is a stable 4-form; the stability condition in this case means there exists k such that $\sigma = \pm\frac{1}{2}k \wedge k$. We consider the $+$ case here. Interpreting this k as a candidate symplectic form, $V_S(\sigma)$ is defined to be the symplectic volume of M:

$$V_S(\sigma) = \frac{1}{6}\int_M k \wedge k \wedge k. \tag{4.57}$$

This action can also be written directly in terms of σ:

$$\begin{aligned} V_S(\sigma) &= \frac{1}{6}\int_M \sigma^{3/2} = \\ &= \int_M \sqrt{\tfrac{1}{384}\sigma_{a_1a_2b_1b_2}\sigma_{a_3a_4b_3b_4}\sigma_{a_5a_6b_5b_6}\epsilon^{a_1a_2a_3a_4a_5a_6}\epsilon^{b_1b_2b_3b_4b_5b_6}}, \end{aligned} \tag{4.58}$$

where $\epsilon^{a_1\cdots a_6}$ is the Levi-Civita tensor in six dimensions. As discussed before, the need to take a square root arises because to define a volume form we need to have exactly one net ϵ tensor.

We will be considering $V_S(\sigma)$ as the effective action of a gravity theory in six dimensions. We treat σ as a 4-form field strength for a 3-form gauge field γ: in other words, we hold the cohomology class of σ fixed,

$$\begin{aligned} &[\sigma] \in H^4(M,\mathbb{R}) \qquad \text{fixed, } \textit{i.e.} \\ &\sigma = \sigma_0 + d\gamma, \end{aligned} \tag{4.59}$$

where $d\sigma_0 = 0$. Now we want to find the classical solutions, i.e. critical points of $V_S(\sigma)$. Write

$$V_S(\sigma) = \frac{1}{3}\int_M \sigma \wedge k = \frac{1}{3}\int_M (\sigma_0 + d\gamma)\wedge k. \tag{4.60}$$

Varying γ then gives a term

$$\delta V_S = \frac{1}{3}\int_M d(\delta\gamma)\wedge k = -\frac{1}{3}\int_M \delta\gamma \wedge dk. \tag{4.61}$$

This is not the whole variation of V_S, because k also depends on σ; but it turns out that the extra term from the variation of k is just 1/2 of (4.61). This is a consequence of the fact (4.54) that $V_S(\sigma)$ is homogeneous as a function of σ. Altogether, the condition that $V_S(\sigma)$ is extremal under variations of γ is simply

$$dk = 0. \tag{4.62}$$

Hence the classical solutions of the gravity theory based on $V_S(\sigma)$ give symplectic structures on M.

Having discussed $V_S(\sigma)$, we now turn to $V_H(\rho)$. Suppose ρ is a stable 3-form. Provided that ρ is "positive" in a sense to be defined below, it is fixed by a subgroup of $GL(6,\mathbb{R})$ isomorphic to (two copies of) $SL(3,\mathbb{C})$; this ρ therefore determines a reduction of $GL(6)$ to $SL(3,\mathbb{C})$, which is the same as an almost complex structure on M. More concretely, we can find three complex 1-forms ζ_i for which

$$\rho = \frac{1}{2}(\zeta_1 \wedge \zeta_2 \wedge \zeta_3 + \overline{\zeta_1} \wedge \overline{\zeta_2} \wedge \overline{\zeta_3}), \tag{4.63}$$

and these ζ_i determine the almost complex structure. If locally there exist complex coordinates such that $dz_i = \zeta_i$, then the almost complex structure is integrable (it defines an actual complex structure.) Whether it is integrable or not, we can construct a $(3,0)$ form on M, namely

$$\Omega = \zeta_1 \wedge \zeta_2 \wedge \zeta_3. \tag{4.64}$$

This Ω can also be written

$$\Omega = \rho + i\widehat{\rho}(\rho), \tag{4.65}$$

where $\widehat{\rho}$ is defined as

$$\widehat{\rho} = -\frac{i}{2}(\zeta_1 \wedge \zeta_2 \wedge \zeta_3 - \overline{\zeta_1} \wedge \overline{\zeta_2} \wedge \overline{\zeta_3}). \tag{4.66}$$

Through (4.63), (4.64) and (4.66), we can regard Ω and $\widehat{\rho}$ as functions of ρ. The integrability condition is equivalent to the requirement that $d\Omega = 0$.

So far we have explained how a positive stable 3-form ρ determines an almost complex structure and a holomorphic 3-form Ω. Now $V_H(\rho)$ is defined to be the holomorphic volume:

$$V_H(\rho) = -\frac{i}{4}\int_M \Omega \wedge \overline{\Omega} = \frac{1}{2}\int_M \widehat{\rho}(\rho) \wedge \rho. \tag{4.67}$$

More concretely, using results from [74], this can be written

$$V_H(\rho) = \int_M \sqrt{-\tfrac{1}{6} K_a{}^b K_b{}^a}, \tag{4.68}$$

where[3]

$$K_a{}^b := \frac{1}{12}\rho_{a_1 a_2 a_3}\rho_{a_4 a_5 a}\epsilon^{a_1 a_2 a_3 a_4 a_5 b}. \tag{4.69}$$

As we did with V_S, we want to regard V_H as the effective action of some gravity theory in which ρ is treated as a field strength. So we start with a closed stable 3-form ρ_0 and allow it to vary in a fixed cohomology class,

$$\rho = \rho_0 + d\beta. \tag{4.70}$$

Then varying β, we obtain two terms, one from the variation of ρ and one from the variation of $\widehat{\rho}(\rho)$. As in the case of V_S, the homogeneity of $V_H(\rho)$ implies that these two terms are equal, and they combine to give

$$\delta V_H = \int_M d(\delta\gamma) \wedge \widehat{\rho} = -\int_M \delta\gamma \wedge d\widehat{\rho}. \tag{4.71}$$

Hence the equation of motion is

$$d\widehat{\rho} = 0.$$

[3] Having written this formula we can now explain the positivity condition on ρ to which we alluded earlier: the square root which appears in (4.68) should be real.

From (4.70) we also have $d\rho = 0$. So altogether on-shell we have $d\Omega = 0$, which is the condition for integrability of the almost complex structure, as explained above. In this sense, $V_H(\rho)$ is an action which generates complex structures together with holomorphic three-forms.

Finally, let us make one more observation about the functionals V_H and V_S. So far we have discussed them separately, but since they both exist on a 6-manifold M, it is natural to ask whether the structures they define are compatible with one another. Specifically, we would like to interpret k as the Kähler form on M, in the complex structure determined by Ω. For this interpretation to be consistent, there are two conditions which must be satisfied:

$$k \wedge \rho = 0, \tag{4.72}$$

and

$$2V_S(\sigma) = V_H(\rho). \tag{4.73}$$

The condition (4.72) expresses the requirement that k is of type $(1,1)$ in the complex structure determined by Ω, while (4.73) is the equality of the volume forms determined independently by the holomorphic and symplectic structures. Requiring (4.72)–(4.73), Ω and k together give an $SU(3)$ structure on M; if in addition $d\Omega = 0$, $dk = 0$, then M is Calabi-Yau, with Ω as holomorphic 3-form and k as Kähler form. When we discuss the Hamiltonian quantization of topological M-theory in Section 4.7, we will see one way in which these constraints could arise naturally.

4.4.4 3-form and 4-form actions in $7D$

Now let us discuss the 7-dimensional case. We will construct two functionals $V_7(\Phi)$, $V_7(G)$ depending on a 3-form or 4-form respectively, both of which generate G_2 holonomy metrics on a 7-manifold X.

The two cases are very similar to one another; we begin with the 3-form case. A stable 3-form $\Phi \in \Omega^3(X, \mathbb{R})$ determines a G_2 structure on X, because G_2 is the subgroup of $GL(7, \mathbb{R})$ fixing Φ at each point, as we explained in Section 4.4.2. There we gave the explicit expression for the metric g in terms of the 3-form Φ:

$$g_{jk} = B_{jk} \det(B)^{-1/9}, \tag{4.74}$$

where from (4.50),

$$B_{jk} = -\frac{1}{144} \Phi_{ji_1i_2} \Phi_{ki_3i_4} \Phi_{i_5i_6i_7} \epsilon^{i_1 \ldots i_7}. \tag{4.75}$$

We can thus introduce a volume functional, $V_7(\Phi)$, which is simply the volume of X as determined by g:

$$V_7(\Phi) = \int_X \sqrt{g_\Phi} = \int_X \Big(\det B \Big)^{1/9}, \tag{4.76}$$

where B is the symmetric tensor defined in (4.75).

In order to identify the critical points of the action functional (4.76), it is convenient to rewrite it slightly. For this we use the fact that since Φ determines the metric, it also determines any quantity which could be derived from the metric; in particular it determines a Hodge $*$-operator, which we write $*_\Phi$. Using this operator we can rewrite (4.76) as

$$V_7(\Phi) = \int_X \Phi \wedge *_\Phi \Phi. \tag{4.77}$$

As we did in the 6-dimensional cases, we regard Φ as a field strength for a 2-form gauge potential; in other words, we assume Φ is closed and vary it in a fixed cohomology class:

$$\begin{aligned} &[\Phi] \in H^3(X,\mathbb{R}) \qquad \text{fixed, } \textit{i.e.} \\ &\Phi = \Phi_0 + dB, \end{aligned} \tag{4.78}$$

with $d\Phi_0 = 0$, and B an arbitrary real 2-form on X. Using the homogeneity property (4.54) of the volume functional (4.77), we find

$$\frac{\delta V_7(\Phi)}{\delta \Phi} = \frac{7}{3} *_\Phi \Phi. \tag{4.79}$$

Hence critical points of $V_7(\Phi)$ in a fixed cohomology class give 3-forms which are closed and co-closed,

$$\begin{aligned} &d\Phi = 0, \\ &d *_\Phi \Phi = 0. \end{aligned} \tag{4.80}$$

These are precisely the conditions under which Φ is the associative 3-form for a G_2 holonomy metric on X.

So far we have discussed stable 3-forms, but the G_2 holonomy condition can also be obtained from a dual action based on a stable 4-form G,

$$V_7(G) = \int_X G \wedge *_G G. \tag{4.81}$$

It is this $V_7(G)$ which we propose to identify as the effective action of the 7-dimensional topological M-theory. As in the previous cases, we vary the 4-form G in a fixed cohomology class:

$$\begin{aligned} &[G] \in H^4(X,\mathbb{R}) \qquad \text{fixed, } i.e. \\ &G = G_0 + d\Gamma, \end{aligned} \tag{4.82}$$

where Γ is an arbitrary real 3-form on X, and G_0 is closed, $dG_0 = 0$. The condition that (4.81) is extremal under variations of Γ is then simply

$$\begin{aligned} &dG = 0, \\ &d *_G G = 0, \end{aligned} \tag{4.83}$$

which is again the condition (4.80) that X has G_2 holonomy, now written in terms of the coassociative 4-form $G = *_\Phi \Phi$. Just as with Φ, one can reconstruct the G_2 holonomy metric from G, using the expression of G in terms of an orthonormal vielbein,

$$G = e^{7346} - e^{7126} + e^{7135} - e^{7425} + e^{1342} + e^{5623} + e^{5641}. \tag{4.84}$$

The 4-form action (4.81) can also be written in a slightly different form. One introduces a fixed basis of the space $\wedge^2 V$ of bivectors in 7 dimensions: $e_a^{ij} = -e_a^{ji}$. Here $i, j = 1, \ldots, 7$ and $a = 1, \ldots, 21$, since the space of bivectors is 21-dimensional. Then define the matrix Q_{ab} by

$$Q_{ab} = \frac{1}{2} e_a^{ij} e_b^{kl} G_{ijkl}. \tag{4.85}$$

The action for G can then be written as

$$V_7(G) = \int_X (\det Q)^{\frac{1}{12}}. \tag{4.86}$$

Note that since Q is a matrix of rank 21, this action is indeed homogeneous of degree $\frac{21}{12} = \frac{7}{4}$ in G. It is a tempting thought that this action could be interpreted as a membrane version of Born-Infeld obtained by integrating out open (topological) membranes, since the exponent $\frac{1}{12}$ reminds one of a stringy one-loop determinant.

4.4.5 Hamiltonian flow

Now we shift gears to discuss a bridge between the $SU(3)$ structures and G_2 holonomy metrics considered in the last two subsections: we will describe a flow which constructs G_2 holonomy metrics from the $SU(3)$ structures which appeared there. This flow is essentially a Hamiltonian version of the Lagrangian field theories described in Section 4.4.4.

Suppose given a 6-manifold M, with stable forms $\rho \in \Omega^3(M,\mathbb{R})$ and $\sigma \in \Omega^4(M,\mathbb{R})$. As we discussed above, if ρ and σ satisfy the compatibility conditions (4.72) and (4.73), they define an $SU(3)$ structure on M and a corresponding metric. If ρ and σ are also both closed, one can extend the metric on M uniquely to a G_2 holonomy metric on $X = M \times (a,b)$ for some interval (a,b). Hitchin gave an elegant construction of this metric [74]: one takes the given ρ and σ as "initial data" on $M \times \{t_0\}$ and then lets ρ and σ flow according to

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= dk, \\ \frac{\partial \sigma}{\partial t} &= k \wedge \frac{\partial k}{\partial t} = -d\hat{\rho}. \end{aligned} \tag{4.87}$$

Here, as usual, $\sigma = \frac{1}{2} k \wedge k$, and t is the "time" direction normal to M.

The evolution equations (4.87) are equivalent to the G_2 holonomy conditions (4.80) for the 3-form

$$\Phi = \rho(t) + k(t) \wedge dt.$$

Moreover, (4.87) can be interpreted as Hamiltonian flow equations. Namely, one considers the variations of σ and ρ as spanning a phase space $\Omega^4_{exact}(M,\mathbb{R}) \times \Omega^3_{exact}(M,\mathbb{R})$; writing $\delta\sigma = d\beta$ and $\delta\rho = d\alpha$, the symplectic pairing on the phase space is

$$\langle \delta\sigma, \delta\rho \rangle = \int \alpha \wedge d\beta = -\int \beta \wedge d\alpha. \tag{4.88}$$

Then (4.87) are precisely the Hamiltonian flow equations with respect to

$$H = 2V_S(\sigma) - V_H(\rho), \tag{4.89}$$

where $V_H(\rho)$ and $V_S(\sigma)$ are the volume functionals (4.57) and (4.67) which we used to obtain $SU(3)$ structures in 6 dimensions.

4.5 Relating Hitchin's functionals in $6D$ to topological strings

In the last section we introduced two functionals $V_H(\rho)$, $V_S(\sigma)$ which, when extremized, generate respectively a symplectic form k and a closed holomorphic $(3,0)$ form Ω on a 6-manifold M. This is reminiscent of the topological A and B models, and one might wonder whether there is some relation. In this section we point out that such a relation does exist. Our arguments will be rigorous only at the classical level, but they suggest a natural extension to the quantum theories, which we will describe. One partcularly interesting feature will emerge: namely, $V_H(\rho)$ turns out to be equivalent not to the B model itself but to a combination of the B and $\overline{\text{B}}$ models.

4.5.1 Hitchin's V_S as the A model

We begin by discussing a relation between Hitchin's action functional (4.57),

$$V_S(\sigma) = \int_M \sigma^{3/2} \tag{4.90}$$

based on the closed 4-form σ, and the A model on M. As we discussed in Section 4.4.3, the solutions to the classical equations of motion coming from $V_S(\sigma)$ involve Kähler geometries, which are also the classical solutions of the A model Kähler gravity. In fact, more is true: the classical actions in both cases compute the volume of M. So at least at a superficial classical level, the two theories are equivalent. Moreover, we can argue that the small fluctuations in the two theories can be identified with one another. Namely, recall that in Hitchin's theory we write $\sigma = \sigma_0 + d\gamma$; then the action at quadratic order for the fluctuation γ includes $\int d\gamma \wedge d^{c\dagger}\gamma$, which nicely matches the action for the quadratic fluctuations in the Kähler gravity theory. So one would expect that the two should be identified.

Here we would like to take one more step in connecting the two theories: specifically, it has been recently argued [80] that the A model can be reformulated in terms of a topologically twisted $U(1)$ gauge theory on M, whose bosonic action contains the observables

$$S = \frac{g_s}{3}\int_M F \wedge F \wedge F + \int_M k_0 \wedge F \wedge F. \tag{4.91}$$

The partition function in this theory is a function of the fixed class k_0. The path integral can be defined as a sum over a gravitational "quantum foam" [80], *i.e.* over

Kähler geometries with quantized Kähler class,[4]

$$k = k_0 + g_s F, \tag{4.92}$$

or, equivalently, as a sum over ideal sheaves [91].

We claim that in the weak coupling ($g_s \to 0$) limit, the theory based on the action (4.91) is equivalent to the "gravity theory" based on Hitchin's action (4.90). Moreover, we show that fixing the BRST symmetries of the Hitchin action naturally leads to the description of the A model as a topologically twisted supersymmetric $U(1)$ gauge theory.

In order to show this, we begin with the action

$$S = \alpha \int_M \widetilde{F} \wedge \widetilde{F} \wedge \widetilde{F} - \beta \int_M \sigma \wedge \widetilde{F}, \tag{4.93}$$

where α and β are some coefficients (which will be related to g_s below), $\widetilde{F}$ is a 2-form on M, and σ is a 4-form that varies in a fixed cohomology class,

$$\begin{aligned} &[\sigma] \in H^4(M) \qquad \text{fixed, } \textit{i.e.} \\ &\sigma = \sigma_0 + d\gamma. \end{aligned} \tag{4.94}$$

At this point we do not make any assumptions about the 2-form $\widetilde{F}$; in particular, it need not be closed or co-closed.

First, let us integrate out the 2-form $\widetilde{F}$ in the action (4.93). The equation of motion for $\widetilde{F}$ has the form

$$3\alpha \widetilde{F} \wedge \widetilde{F} - \beta\sigma = 0. \tag{4.95}$$

[4]We choose our normalization so that F is integrally quantized: $\int_C F \in \mathbb{Z}$ for any closed 2-cycle $C \subset M$.

This equation implies that the 2-form $\widetilde{F}$ is a "square root" of σ, *i.e.* σ is a stable 4-form. Substituting $\widetilde{F}$ back into the action (4.93), we obtain precisely Hitchin's action (4.90), with the remaining path integral over a stable, closed 4-form σ. It is important to stress here that, since the action (4.93) is cubic in $\widetilde{F}$, the relation to Hitchin's action (4.90) holds only in the semi-classical limit. We return to this issue below, and show that this is precisely the limit $g_s \to 0$.

Similarly, starting with the action (4.93) and integrating out σ one can obtain the $U(1)$ gauge theory (4.91). In order to see this, one has to eliminate σ through its equations of motion, and then make a simple field redefinition. The equations of motion for σ are very simple. Since the dynamical variable γ appears as a Lagrange multiplier in (4.93), it leads to the constraint

$$d\widetilde{F} = 0, \tag{4.96}$$

which means that the 2-form $\widetilde{F}$ is closed and, therefore, can be interpreted as a curvature on a line bundle. The resulting action for $\widetilde{F}$ is

$$S = \alpha \int_M \widetilde{F} \wedge \widetilde{F} \wedge \widetilde{F} - \beta \int_M \sigma_0 \wedge \widetilde{F}. \tag{4.97}$$

In order to bring this action to the familiar form (4.91), it remains to do a simple change of variables. We introduce

$$F = \widetilde{F} - \xi k_0, \tag{4.98}$$

where ξ is a parameter and k_0 is the background Kähler form, such that $\sigma_0 = k_0 \wedge k_0$. Substituting (4.98) into (4.97), we get (up to a constant term) the action

$$S = \alpha \int_M F \wedge F \wedge F + 3\xi\alpha \int_M F \wedge F \wedge k_0 + \int_M \Big(3\xi^2 \alpha k_0 \wedge k_0 \wedge F - \beta \sigma_0 \wedge F \Big). \tag{4.99}$$

Comparing (4.99) with (4.91) determines the parameters α, β, and ξ:

$$\alpha = \frac{g_s}{3},$$
$$\xi = \frac{1}{g_s},$$
$$\beta = \frac{1}{g_s}. \tag{4.100}$$

With this choice of parameters, we find complete agreement between (4.99) and the $U(1)$ gauge theory action (4.91), including the numerical coefficients and the relation between the Kähler form k and the field F. Indeed, substituting $\xi = 1/g_s$ into (4.98), we get

$$\delta k = g_s F, \tag{4.101}$$

which is precisely the required quantization condition (4.92).

Summarizing, we find that (4.93) is equivalent to the gauge theory action (4.91) and, in the semi-classical limit, is also equal to Hitchin's action (4.90). In order to see when the semi-classical approximation is valid, it is convenient to write both terms in the action (4.93) with the same overall coefficient $1/\hbar$. To achieve this, we rescale

$$\widetilde{F} \to \gamma \widetilde{F}, \tag{4.102}$$

and set the coefficients in the two terms to be equal:

$$\frac{1}{\hbar} = \alpha\gamma^3 = \gamma. \tag{4.103}$$

In particular, the latter equality implies $\alpha = \frac{1}{\gamma^2}$. From the relations (4.100) and (4.103) it follows that the semiclassical limit, $\hbar \to 0$, corresponds to the limit $g_s \to 0$. Hence we conclude that the gauge theory action (4.91) is equivalent to Hitchin's action (4.90) precisely in the weak coupling limit.

BRST Symmetries and Gauge Fixing

As noted before, we really want to connect Hitchin's theory to the *topologically twisted* version of the supersymmetric $U(1)$ gauge theory. In order to do this let us describe the BRST symmetries of the theory (4.93), which, as we just established, is equivalent to the $U(1)$ gauge theory (4.91). First, notice that the partition sum over the quantum foam can be viewed as a vacuum expectation value,

$$\langle \exp\left(\frac{g_s}{3}\int \mathcal{O}_1 + \int \mathcal{O}_2\right)\rangle, \tag{4.104}$$

in the topological $U(1)$ gauge theory on M, where $\mathcal{O}_i$ are the topological observables:

$$\begin{aligned} \mathcal{O}_1 &= F\wedge F\wedge F, \\ \mathcal{O}_2 &= k_0\wedge F\wedge F, \\ &\vdots \end{aligned} \tag{4.105}$$

Following [16], one can reconstruct the action of this topological 6-dimensional theory by studying the BRST symmetries that preserve (4.104)–(4.105). Writing (locally) F as a curvature of a gauge connection A,

$$F = dA, \tag{4.106}$$

it is easy to see that (4.104)–(4.105) are invariant under the usual gauge transformations

$$\delta A = d\epsilon_0, \tag{4.107}$$

as well as under more general transformations

$$\delta A = \epsilon_1 \tag{4.108}$$

where the infinitesimal parameter ϵ_1 is a 1-form on M. The gauge fixing of the latter symmetry leads to a 1-form ghost field ψ. Since ϵ_1 itself has a gauge symmetry, $\epsilon_1 \sim \epsilon_1 + d\lambda$, one also has to introduce a commuting 0-form ϕ associated with this symmetry. Hence, already at this stage we see that the $6D$ topological theory in question should contain a gauge field and a scalar. The only such theory is a maximally supersymmetric topological gauge theory in six dimensions, *i.e.* a theory with $\mathcal{N}_T = 2$ topological supersymmetry. Equivalently, it is a theory with 16 real fermions, which can be identified with holomorphic $(p, 0)$-forms on M. Remember that on a Kähler manifold

$$S(M) \cong \Omega^{0,*}(M). \tag{4.109}$$

The complete BRST multiplet in this theory looks like:

$$\begin{array}{lrll}
\text{Bosons :} & 1-\text{form} & A & \\
 & \text{cplx. scalar} & \phi & \\
 & (3,0)-\text{form} & \varphi & \\
\text{Fermions :} & & \psi^{p,0}\ ,\ \psi^{0,p} & p=0,1,2,3
\end{array} \tag{4.110}$$

Under the action of the BRST operator s, these fields transform as [15, 76]:

$$\begin{array}{ll}
s\varphi^{0,3} = 0 & s\varphi^{3,0} = \psi^{3,0} \\
sA^{0,1} = \psi^{0,1} & sA^{1,0} = 0 \\
s\psi^{0,1} = 0 & s\psi^{1,0} = -\partial_A \phi \\
s\psi^{0,0} = (k \cdot F^{1,1}) & s\psi^{2,0} = F^{2,0}
\end{array} \tag{4.111}$$

This $\mathcal{N}_T = 2$ 6-dimensional topological $U(1)$ gauge theory has been extensively studied in the literature, see *e.g.* [15, 76, 46, 3, 27]. A reduction of this theory on a Kähler

4-manifold $M^4 \subset M$ leads to the $\mathcal{N}_T = 4$ topological gauge theory studied in [125].

Finally, we identify the symmetry of Hitchin's action (4.90) that corresponds to the BRST symmetry (4.108). In order to do this, we need to find how this symmetry acts on the 4-form field σ. Since in the $g_s \to 0$ limit the field F is identified with the variation of the Käher form (4.101) it follows that

$$\delta k = d\epsilon_1, \tag{4.112}$$

where $\sigma = k \wedge k$. It is easy to check that Hitchin's action (4.90) is indeed invariant under this symmetry,

$$\delta S_H = \frac{3}{2}\int_M k \wedge \delta\sigma = 3\int_M k \wedge k \wedge \delta k = 3\int_M \sigma \wedge d\epsilon_1 = 0. \tag{4.113}$$

We have thus recovered the topologically twisted $U(1)$ theory which was conjectured in [80] to be equivalent to the quantum foam description of the A model.

4.5.2 Hitchin's V_H as the B model

Now we want to discuss the relation between Hitchin's "holomorphic volume" functional $V_H(\rho)$ and the B model (see also the recent work [56], which proposes a relation similar to what we will propose below.) Classically, there is an obvious connection between the two, since solving the equations of motion of either one gives a closed holomorphic 3-form Ω. What about quantum mechanically? In order to understand this question we must first recall a subtle feature of the B model partition function.

Consider the B model on a Calabi-Yau 3-fold M. This model is obtained by topological twisting of the physical theory with a fixed "background" complex structure, determined by a holomorphic 3-form Ω. The topological observables in this model are the marginal operators ϕ_i representing infinitesimal deformations of the complex structure, where $i = 1, \ldots, h_{2,1}$. The partition function Z_B was defined in [24] to be the generating functional of correlations of marginal operators: namely, introducing $h_{2,1}$ variables x^i, $Z_B(x, g_s, \Omega_0)$ obeys

$$D_{i_1} \cdots D_{i_k} Z_B(x, g_s)|_{x=0} = \langle \phi_{i_1} \cdots \phi_{i_k} \rangle_{\Omega_0}. \tag{4.114}$$

More intrinsically, we can think of x as labeling an infinitesimal deformation, so that for fixed Ω_0, $Z_B(x, g_s, \Omega_0)$ is a function on the holomorphic tangent space $T_{\Omega_0}\mathcal{M}$ to the moduli space $\mathcal{M}$ of complex structures. By construction this Z_B is holomorphic in its dependence on x. But one gets one such function of x for every Ω_0, corresponding to all the different tangent spaces to $\mathcal{M}$, and one can ask how these different functions are related. This question was answered in [24], where the effect of an infinitesimal change in Ω_0 was found to be given by a "holomorphic anomaly equation."

This Ω_0 dependence of Z_B was later reinterpreted in [131] as the wavefunction property. To understand what this means, it is convenient to combine g_s and x into a "large phase space" of dimension $h_{2,1} + 1$. Changing g_s is equivalent to an overall rescaling of Ω_0 (which does not change the complex structure on X). So for fixed Ω_0, we can consider Z_B as a holomorphic function on $H^{3,0}(X, \mathbb{C}) \oplus H^{2,1}(X, \mathbb{C})$. Equivalently, Z_B is a function on the "phase space" $H^3(X, \mathbb{R})$, which depends only on the complex combination of periods

$$x_I = F_I - \tau_{IJ}(\Omega_0) X^J, \tag{4.115}$$

and not on the conjugate combination

$$\overline{x}_I = F_I - \overline{\tau}_{IJ}(\Omega_0)X^J. \tag{4.116}$$

This is similar to the idea of a wavefunction which depends only on q but not on its conjugate variable p; indeed, x_I and $\overline{x}_I$ are coordinates on $H^{3,0} \oplus H^{2,1}$ and on $H^{1,2} \oplus H^{0,3}$ respectively, and they are indeed conjugate with respect to the standard symplectic structure on $H^3(X, \mathbb{R})$. Note that since τ depends on Ω_0, changing Ω_0 changes the symplectic coordinate system.

Now, if one is given a wavefunction $\psi(q)$ as a function of q and one wants to express it as a function of p, there is a simple procedure for doing so: just take the Fourier transform. In fact, more generally, given $\psi(q)$ one can construct various different representations of the state, *e.g.* $\psi(p)$, $\psi(p+q)$, $\psi(p+iq)$ and so on. Each such representation corresponds to a different choice of symplectic coordinates inside the (p, q) phase space, and each can be obtained from $\psi(q)$ by an appropriate generalized Fourier transform. In [131] it was shown that the Ω_0 dependence of Z_B can be understood in exactly this way: starting from $Z_B(x, \Omega_0)$ one can obtain any other $Z_B(x, \Omega_0')$ by taking an appropriate Fourier transform! In this sense Z_B is a wavefunction obtained by quantization of the symplectic phase space $H^3(X, \mathbb{R})$, which has various different representations depending on one's choice of symplectic coordinates for $H^3(X, \mathbb{R})$.

Now what about Hitchin's gravity theory? Consider the partition function $Z_H([\rho])$ of Hitchin's 6-dimensional gravity theory, formally written

$$Z_H([\rho]) = \int D\beta \exp(V_H(\rho + d\beta)). \tag{4.117}$$

We do not expect that the formal expression (4.117) really captures the whole quantum theory, but the statement that Z_H depends on a class $[\rho] \in H^3(X, \mathbb{R})$ should be correct, as should the classical limit of (4.117). In comparing Z_H to Z_B we notice two points. First, unlike Z_B, Z_H does not depend on a choice of symplectic coordinates for $H^3(X, \mathbb{R})$. Second, Z_H depends on twice as many degrees of freedom as does Z_B (because Z_B depends on only half of the coordinates of $H^3(X, \mathbb{R})$ as explained above.) So Z_H cannot be equal to Z_B.

The situation changes drastically, however, if we combine the B model with the complex conjugate $\overline{\mathrm{B}}$ model. In that case we have two wavefunctions, Z_B and $\overline{Z_B}$, and we can consider the product state

$$\Psi = Z_B \otimes \overline{Z_B}. \tag{4.118}$$

(One could more generally consider a density matrix that is a sum of such pure product states.) This product state sits inside a doubled Hilbert space, obtained from quantization of a phase space which is also doubled, from $H^3(X, \mathbb{R})$ to $H^3(X, \mathbb{C})$. This doubled phase space has a polarization which does not depend on any arbitrary choice: namely, one can divide it into real and imaginary parts, and it is natural to ask for the representation of Ψ as a function of the real parts of all the periods, $\Psi(\mathrm{Re}\ X_I, \mathrm{Re}\ F^I)$. This gives a function on $H^3(X, \mathbb{R})$ which does not depend on any choice of symplectic coordinates. This is actually a standard construction in quantum mechanics: the function one obtains expresses the density in phase-space corresponding to the wavefunction Z_B, and is known as the "Wigner function" of Z_B. It is this Wigner function which we propose to identify with $Z_H([\rho])$.

We can give an explicit formula for the Wigner function if we start from a particular representation of Z_B, namely the one corresponding to a basis of A and B cycles $\{A_I, B^I\}$ in $H_3(X, \mathbb{Z})$. (This choice of polarization corresponds to a certain limit in the space of possible Ω_0; from now on we suppress Ω_0 in the notation.) Then Z_B can be written as $Z_B(X_I)$, a function of the A cycle periods X_I, and we denote the B cycle periods F^I. Writing $P_I = \text{Re } X_I$, $Q^I = \text{Re } F^I$, the Wigner function is given by

$$\Omega(P_I, Q^I) = \int d\Phi_I \; e^{-Q^I \Phi_I} \; |Z_B(P_I + i\Phi_I)|^2. \tag{4.119}$$

This can be identified with $Z_H([\rho])$ if we identify P_I and Q^I as the (real) periods of the class $[\rho] \in H^3(X, \mathbb{R})$.

At least at string tree level (which in this context means large ρ) we can verify that this identification is correct. Namely, in that limit, Z_B is dominated by the tree level free energy F_0, and writing $Z_B = e^{-\frac{i}{2}F_0}$, we can make a steepest descent approximation of the integral over Φ in (4.119). The argument of the exponential is

$$-\frac{i}{2}F_0(P_I + i\Phi_I) + \frac{i}{2}\overline{F_0(P_I + i\Phi_I)} - Q^I \Phi_I. \tag{4.120}$$

The value of Φ which extremizes (4.120) occurs when $Q^I = \text{Re } \partial F_0 / \partial X_I = \text{Re } F^I(P + i\Phi)$. At this Φ, (4.120) becomes

$$-\frac{i}{4}X_I F^I + \frac{i}{4}\overline{X_I}\overline{F^I} - (\text{Re } F^I)(\text{Im } X_I) = \frac{i}{4}X_I \overline{F^I} - \frac{i}{4}\overline{X_I} F^I. \tag{4.121}$$

But this is exactly the classical Hitchin action $V_H = -\frac{i}{4}\int \Omega \wedge \overline{\Omega}$, evaluated at the value of Ω for which $\text{Re } X_I = P_I$ and $\text{Re } F^I = Q^I$. This establishes the desired agreement between Z_H and the Wigner function of Z_B at tree level. In fact, the above relation between the topological string and and $\int \Omega \wedge \overline{\Omega}$ was already noted and used in [98], for the purpose of relating the topological string to $4D$ black hole entropy. We will discuss this connection in Section 4.8.

It seems likely that the agreement between Z_H and the Wigner function will also persist at one loop. The B model at one loop is known [24] to compute the Ray-Singer torsion of M, which is a ratio of determinants of $\overline{\partial}$ operators acting on forms of various degrees; these determinants should agree with those which appear in the quadratic expansion of V_H around a critical point. This basically follows from the fact that the kinetic term is given by $\int \partial\phi\overline{\partial}\phi$, where ϕ is a $(1,1)$ form and the complex structure is determined by the choice of critical point. The difference from the B model is just that here ϕ is not viewed as a chiral field, so we get both the B and $\overline{\mathrm{B}}$ contributions; the one-loop contribution in the B model alone is a chiral determinant, the holomorphic square root of the determinant of the Laplacian.

Finally, we note that, by introducing an extra 3-form field H, we can write the action functional $V_H(\rho)$ in a form that does not involve square roots, just as we did in (4.93) for the A-model,

$$S = \int_M \rho \wedge H + \int_M \alpha \cdot {K_a}^b {K_b}^a + \int_M (1 - \alpha \cdot (\rho \wedge H))\phi, \tag{4.122}$$

where ${K_a}^b$ is defined in (4.69). It is easy to see that integrating out the 3-form H and the Lagrange multiplier ϕ leads to the holomorphic volume action $V_H(\rho)$ of (4.68). The action (4.122) could be useful for a "quantum foam" description of the B model parallel to the one discussed above for the A model.

4.6 Reducing topological M-theory to form gravities

In this section, we want to argue that the various form gravity theories we reviewed earlier arise naturally on supersymmetric cycles in topological M-theory. We will discuss various examples, but the basic idea is always the same: we consider a "local model" of a complete 7-manifold X, obtained as the total space of an m-dimensional vector bundle over an n-dimensional supersymmetric (calibrated) cycle $M \subset X$, such that $m + n = 7$,

$$\begin{array}{ccc} \mathbb{R}^m & \to & X \\ & & \downarrow \\ & & M \end{array} . \tag{4.123}$$

This non-compact local model is intended to capture the dynamics of the 7-dimensional theory when the supersymmetric cycle M shrinks inside a global compact X. This idea is natural when one recalls that the geometry of X in the vicinity of a supersymmetric cycle M is completely dictated by the data on M. Thus the local gravity modes induce a lower-dimensional gravity theory on M. This is similar to what is familiar in the context of string theory: near singularities of Calabi-Yau manifolds one gets an effective lower-dimensional theory of gravity.

After making an appropriate ansatz, the three-form Φ on X induces a collection of p-form fields on M; the equations of motion of topological M-theory,

$$\begin{aligned} d\Phi &= 0, \\ d *_\Phi \Phi &= 0, \end{aligned} \tag{4.124}$$

lead to equations of motion for the p-form fields on M. These equations of motion can be interpreted as coming from a topological gravity theory on M.

Depending on the dimension n of M and the vector bundle we choose over it, we will have various ansaetze for Φ, leading to various gravity theories on M. For example, the cases $n = 3$ and $n = 4$ correspond respectively to associative and coassociative submanifolds, which are familiar examples of supersymmetric cycles in manifolds with G_2 holonomy. In these two cases, the corresponding vector bundle over M is either the spin bundle over M (when M is associative) or the bundle of self-dual 2-forms over M (when M is coassociative). In order to obtain the other two gravity theories, namely the cases $n = 2$ and $n = 6$, one has to assume that the bundle (4.123) splits into a trivial line bundle over M and a bundle of rank $m - 1$. In this case the holonomy group of X is reduced to $SU(3)$, so that locally X looks like a direct product,

$$X = \mathbb{R} \times Y, \tag{4.125}$$

where Y is a Calabi-Yau 3-fold of the form (4.123), with $n + m = 6$. Notice that apart from supersymmetric 2-cycles and 6-cycles, Calabi-Yau manifolds also contain supersymmetric 3-cycles and 4-cycles. A priori, the form gravity induced on the latter may be different from the gravity theory obtained on associative and coassociative cycles in a 7-manifold with the full holonomy G_2.

In the cases $n = 3, m = 4$ and $n = 4, m = 3$ we will be closely following two constructions of local G_2 manifolds given in [29, 57] and recently discussed in [39]. These two constructions have some common features which can be conveniently summarized in advance. We let y^i denote a local coordinate on the fiber $\mathbb{R}^m$, and write $r = y_i y^i$.

The ansatz is $SO(m)$ invariant, so that Φ depends only on r and the coordinates on M. We construct a basis of 1-forms in the fiber direction as

$$\alpha^i = D_A y^i = dy^i + (Ay)^i, \tag{4.126}$$

where A is the 1-form induced by a gauge connection on M which acts on the y^i in some representation.

The fact that Φ is a stable 3-form means that there exists a 7-dimensional vielbein e^i such that

$$\Phi = e^{567} + e^5 \wedge (e^{12} - e^{34}) + e^6 \wedge (e^{13} - e^{42}) + e^7 \wedge (e^{14} - e^{23}).$$

In the metric g determined by Φ, the e^i form an orthonormal basis. We define a triplet of 2-forms Σ^i by the formula (4.17) as in the 2-form gravity:

$$\begin{aligned} \Sigma^1 &= e^{12} - e^{34}, \\ \Sigma^2 &= e^{13} - e^{42}, \\ \Sigma^3 &= e^{14} - e^{23}. \end{aligned} \tag{4.127}$$

Then Φ is written

$$\Phi = e^{567} + e^i \wedge \Sigma^i. \tag{4.128}$$

To verify the equations of motion, we will also need the expression for $*_\Phi \Phi$, derived straightforwardly by expanding in the e^i:

$$*_\Phi \Phi = -\frac{1}{6}\Sigma_i \wedge \Sigma^i + \frac{1}{2}\epsilon^{ijk} e^i \wedge e^j \wedge \Sigma^k. \tag{4.129}$$

In fact, it is convenient to consider a slightly more general form of Φ: namely, rescaling the first three e^i by f and other four by g, we obtain

$$\Phi = f^3 e^{567} + f g^2 e^i \wedge \Sigma^i, \tag{4.130}$$

and

$$*_\Phi \Phi = -\frac{1}{6} g^4 \Sigma_i \wedge \Sigma^i + \frac{1}{2} f^2 g^2 \epsilon^{ijk} e^i \wedge e^j \wedge \Sigma^k. \tag{4.131}$$

4.6.1 $3D$ gravity on associative submanifolds

One local model for a G_2 space X is obtained by choosing X to be the total space of the spin bundle over a 3-manifold M. In this case, with our ansatz, the field content and equations of motion of topological M-theory on X reduce to those of Chern-Simons gravity on M; in particular, the condition that X has G_2 holonomy implies that M has constant sectional curvature.

First, let us show that the field content of topological M-theory on X can be naturally reduced to that of Chern-Simons gravity on M. This amounts to constructing an ansatz for the associative 3-form Φ in terms of forms on M. We write it in the general form (4.130) and then impose the condition that e^1, e^2, e^3, e^4 are constructed out of an $SU(2)$ connection on M acting on the spin bundle, as we explained earlier in (4.126): $e^i = \alpha^i$, $i = 1, \ldots, 4$. For convenience we also relabel e^5, e^6, e^7 as e^1, e^2, e^3, so finally the form of Φ is

$$\Phi = f^3 e^{123} + f g^2 e_i \wedge \Sigma^i, \tag{4.132}$$

where

$$\begin{aligned} \Sigma^1 &= \alpha^{12} - \alpha^{34}, \\ \Sigma^2 &= \alpha^{13} - \alpha^{42}, \\ \Sigma^3 &= \alpha^{14} - \alpha^{23}. \end{aligned} \tag{4.133}$$

Further assume that f, g depend only on the radial coordinate r, with $f(0) = g(0) = 1$. Then along M, the only fields (undetermined functions) in our ansatz are the dreibein e^i and the $SU(2)$ connection A^i. These are exactly the fields of 3-dimensional gravity in the first-order formalism, and they can be organized naturally into a complexified gauge field, as we discussed before.

Now we want to check that the equations of motion of topological M-theory reduce with our ansatz to those of 3-dimensional gravity. This amounts to evaluating $d\Phi$ and $d *_\Phi \Phi$ directly, using (4.131). One finds that if

$$\begin{aligned} f(r) &= \sqrt{3\Lambda}(1+r)^{1/3}, \\ g(r) &= 2(1+r)^{-1/6}, \end{aligned} \tag{4.134}$$

then $d\Phi = 0$ becomes equivalent to

$$\begin{aligned} de &= -A \wedge e - e \wedge A, \\ dA &= -A \wedge A - \Lambda e \wedge e. \end{aligned} \tag{4.135}$$

The conditions (4.135) are precisely the equations of motion in 3D Chern-Simons gravity,

$$d\mathcal{A} + \mathcal{A} \wedge \mathcal{A} = 0, \tag{4.136}$$

based on the gauge group G as indicated in Table 1. Furthermore one can check that $d *_\Phi \Phi = 0$ is automatic provided that (4.134)–(4.135) are satisfied. So with this particular ansatz, the equations of motion of topological M-theory do indeed reduce to those of 3-dimensional gravity.

4.6.2 $4D$ gravity on coassociative submanifolds

Another local model of a G_2 manifold is obtained by choosing X to be the bundle of self-dual 2-forms over a 4-manifold M. We will see that in this case the effective gravity theory on M is the 2-form gravity we considered in Section 4.3.3.

First, let us show that the field content of topological M-theory on X can be naturally reduced to that of 2-form gravity on M. This amounts to constructing an ansatz for the associative 3-form Φ in terms of forms on M. We write it in the general form (4.130) and then impose the condition that e^5, e^6, e^7 are constructed out of an $SU(2)$ connection on M acting on the bundle of self-dual 2-forms, as we explained earlier: $e^5 = \alpha_1$, $e^6 = \alpha_2$, $e^7 = \alpha_3$. Further assume that Σ^i are purely tangent to M, and that f, g depend only on the radial coordinate r, with $f(0) = g(0) = 1$.

Thus along M the associative 3-form Φ can be simply written:

$$\Phi = \alpha^{123} + \alpha^1 \wedge \Sigma^1 + \alpha^2 \wedge \Sigma^2 + \alpha^3 \wedge \Sigma^3. \tag{4.137}$$

Since we constructed both Φ and Σ from the vielbein, which determines the metric, the metrics on M which can be reconstructed from Φ and Σ must agree. It is gratifying that this can be seen explicitly, as we now do: recall the expression for the G_2 metric in terms of Φ from (4.52),

$$\sqrt{g}\, g_{jk} = -\frac{1}{144} \Phi_{j i_1 i_2} \Phi_{k i_3 i_4} \Phi_{i_5 i_6 i_7} \epsilon^{i_1 \ldots i_7}. \tag{4.138}$$

We wish to consider the components of the metric along M, g_{jk}, where $j, k = 1, \ldots, 4$. Normalizing the normal directions to have length scale 1, we can view g in (4.138) as the determinant of the 4-dimensional metric on M^4. Also, notice that if $j, k = 1, \ldots, 4$, then none of the Φ-components can be e^{567}. Hence, all the components of Φ in (4.138) should contain Σ^i, *cf.* (4.137).

Now, let us consider the combinatorial factors in (4.52). Since indices j and k are assumed to take values from 1 through 4, one of the indices i_1 or i_2 in $\Phi_{ji_1i_2}$ can take values 5, 6, or 7. Similarly, there are two choices to assign a normal direction to i_3 or i_4, and three choices in the last factor, $\Phi_{i_5i_6i_7}$. In total, we get a combinatorial factor $12 = 2 \cdot 2 \cdot 3$ and we can write the metric in the form

$$\sqrt{g}\, g_{ab} = -\frac{1}{12}\Sigma^i_{aa_1}\Sigma^j_{ba_2}\Sigma^k_{a_3a_4}\epsilon^{ijk}\epsilon^{a_1a_2a_3a_4}. \tag{4.139}$$

This is exactly the expression (4.20) for the metric on M in the 2-form gravity.

So we have written Φ in terms of an $SU(2)$ triplet of two-forms Σ^i, which by construction obey the constraint (4.16), and an $SU(2)$ gauge connection, which we used to define the α^i. These are precisely the fields of the two-form gravity theory we considered above, which has self-dual Einstein metrics on M as its classical solutions. Note that from the viewpoint of the 4-dimensional form theory of gravity, the $SU(2)$ gauge indices $i = 1, 2, 3$ and the 4-dimensional spacetime indices of Σ^i_{ab} are unrelated. However, we have seen that *in the context of topological M-theory the 3 $SU(2)$ gauge indices are unified with the 4 spacetime indices to give a 3-form in 7 dimensions.* This by itself is rather satisfactory, and suggestive of a deep role for topological M-theory in the context of 4-dimensional quantum gravity.

Next we want to argue that the field equations of topological M-theory reduce to those of the two-form gravity theory on M. First consider the equation $d\Phi = 0$. A direct computation shows that with the choice

$$\begin{aligned} f(r) &= (1+r)^{-1/4}, \\ g(r) &= \sqrt{2\Lambda}(1+r)^{1/4}, \end{aligned} \tag{4.140}$$

the condition $d\Phi = 0$ becomes equivalent to

$$\begin{aligned} D_A \Sigma &= 0, \\ F \wedge \Sigma &= 0. \end{aligned} \tag{4.141}$$

In fact, the latter equation follows from the former by applying D_A to both sides; so we just have to impose $D_A \Sigma = 0$, which means that A is the $SU(2)_+$ part of the spin connection. Note that this also implies that F is self-dual in the metric induced by Σ, $F = F_+$; this follows from the fact that F is $SU(2)_+$ valued, and the symmetry $R_{abcd} = R_{cdab}$ of the Riemann tensor, which is shared by F.

Similarly, one finds that $d *_\Phi \Phi = 0$ can be satisfied provided that

$$F_+ = \frac{\Lambda}{12}\Sigma. \tag{4.142}$$

So altogether, the equations of motion of topological M-theory imply

$$\begin{aligned} D_A \Sigma &= 0, \\ F &= \frac{\Lambda}{12}\Sigma. \end{aligned} \tag{4.143}$$

These agree precisely with (4.25). In sum, the field content and equations of motion of the self-dual version of two-form gravity agree with those of topological M-theory, when we make a special ansatz for Φ.

4.6.3 $6D$ topological strings

Finally let us consider the case $n = 6$. In this case X is a real line bundle over the 6-dimensional M, and we choose it to be trivial — either $X = M \times \mathbb{R}$ or its compactification $X = M \times \mathbf{S}^1$. Let $\mathbb{R}$ be parameterized by t. Then a natural ansatz for Φ is

$$\Phi = \rho + k \wedge dt, \tag{4.144}$$

where ρ and k are respectively a 3-form and 2-form on M. If Φ is a stable 3-form on X, then ρ and k are stable on X, so as we discussed earlier, they define respectively an almost complex structure and a presymplectic structure on X, and if we impose also the conditions (4.72)–(4.73) then these two structures are compatible. In that case they give an $SU(3)$ structure on X. The condition that this $SU(3)$ structure is integrable,

$$\begin{aligned} dk &= 0, \\ d(\rho + i\hat{\rho}) &= 0, \end{aligned} \tag{4.145}$$

is equivalent to the 7-dimensional equations of motion $d\Phi = 0$, $d *_\Phi \Phi = 0$. So with this ansatz, topological M-theory on X reduces to a theory on M, describing variations of k and ρ, for which the classical solutions are Calabi-Yau 3-folds.

What 6-dimensional theory is this? As they are usually conceived, neither the topological A model nor the topological B model alone fits the bill: at least perturbatively, the A model just describes variations of k, and the B model those of the holomorphic 3-form $\Omega = \rho + i\hat{\rho}$. The theory we are getting on M is a combination of the A and B models — with a slight coupling between them, expressed by the constraints (4.72)–(4.73). In support of this point of view, note that after imposing (4.72)–(4.73) the action $V_7(\Phi)$ can be simply expressed in terms of ρ and k: it becomes simply

$$V_7(\Phi) = 3V_S(k) + 2V_H(\rho), \tag{4.146}$$

where V_S and V_H are the 6-dimensional symplectic and holomorphic volume functionals introduced in Section 4.4.3. As we discussed in Section 4.5, these functionals correspond respectively to the A model and the B + $\overline{\mathrm{B}}$ model.

It is natural to conjecture that this 7-dimensional construction is in fact related to the nonperturbative completion of the topological string, which we expect to mix the A and B models, and to related phenomena such as the topological S-duality conjectured in [94, 96] (see also [83]). While this picture is far from complete, there is one encouraging sign, which we will describe further in the next section.

4.7 Canonical quantization of topological M-theory and S-duality

In the last section we discussed a possible relation between topological M-theory on $X = M \times \mathbf{S}^1$ and topological string theory on M. In particular, we found that the classical reduction, obtained just by considering fields which are independent of the coordinate along $\mathbf{S}^1$, leads to a combination of two systems, which are in the universality classes of the topological A and B models. Recently, there have been some hints that the A and B model could be coupled to one another. In this section we discuss how such a coupling could arise through canonical quantization, and how this relates to the notion that the topological string partition function is a wave function.

We begin by considering the 4-form version of topological M-theory. To perform the canonical quantization of Hitchin's action $V_7(G)$ is a nontrivial problem, because of the usual subtleties involved in quantizing a diffeomorphism invariant theory. Moreover, we should note that we are viewing Hitchin's action only as an effective action, which we are using just to extract some basic facts about the Hilbert space. For this purpose it is enough to work roughly, although a more precise treatment would certainly be desirable.

So let us consider the 7-dimensional gravity theory (4.81) on a manifold $X = M \times \mathbb{R}$, where M is a compact 6-manifold and $\mathbb{R}$ is the "time" direction. We decompose the 3-form gauge field Γ as

$$\Gamma = \gamma + \beta \wedge dt,$$

where γ and β are a 3-form and 2-form respectively, with components only along M. Similarly decompose G and $*_G G$ as

$$\begin{aligned} G &= \sigma + \widehat{\rho} \wedge dt, \\ *_G G &= \rho + k \wedge dt. \end{aligned} \tag{4.147}$$

Then write

$$\begin{aligned} G &= G_0 + d\Gamma \\ &= (\sigma_0 + d\gamma) + (\widehat{\rho}_0 + d\beta + \dot{\gamma}) \wedge dt, \end{aligned} \tag{4.148}$$

so that

$$\widehat{\rho} = \widehat{\rho}_0 + d\beta + \dot{\gamma},$$

$$\sigma = \sigma_0 + d\gamma. \tag{4.149}$$

The configuration space is spanned by the components (γ, β) of the gauge field Γ. Their conjugate momenta are

$$\pi_\gamma = \frac{7}{4}\rho, \qquad \pi_\beta = 0. \tag{4.150}$$

The longitudinal component β is an auxiliary field; it imposes the constraint

$$\frac{\delta V_7}{\delta \beta} = d\rho = 0, \tag{4.151}$$

which generates the spacelike, time-independent gauge transformations $\gamma \to \gamma + d\lambda$. Hence the reduced phase space which we obtain from canonical quantization of the 4-form theory is parameterized by (γ, ρ), where

$$\gamma \in \Omega^3(M)/\Omega^3_{exact}(M), \qquad \rho \in \Omega^3_{closed}(M). \tag{4.152}$$

Now let us compute the Hamiltonian. Suppose that we impose the conditions (4.72)–(4.73) (we will comment more on the role of these constraints later.) Then it is straightforward to verify that ρ and $\widehat{\rho}$ are related as in Section 4.4.3, and $\sigma = \frac{1}{2}k \wedge k$. The action V_7 from (4.81) becomes

$$V_7 = \int_X dt\,(2V_H(\rho) + 3V_S(\sigma(\gamma)))\,. \tag{4.153}$$

Using (4.150) we can construct the Hamiltonian,

$$H = 2V_H(\rho) + 3V_S(\sigma(\gamma)) - \dot{\gamma} \wedge \pi_\gamma = \frac{3}{2}(2V_S(\sigma(\gamma)) - V_H(\rho)). \tag{4.154}$$

Although our treatment has been rough, we can gain some confidence from the fact that the Hamiltonian we ultimately obtained at least gives classical equations of motion agreeing with the Lagrangian formulation; namely, it agrees with (4.89), which indeed defines a flow giving G_2 holonomy metrics. A more precise treatment (possibly starting from a different classically equivalent action) would require a better understanding of the constraints (4.72)–(4.73); we believe that they will turn out to be the diffeomorphism and Hamiltonian constraints, as usual in diffeomorphism invariant theories. Indeed, note that (4.73) is simply the constraint $H = 0$.

As usual in the Hamiltonian formalism, we treat ρ, γ as the canonical variables, where ρ is "momentum" and γ is "position." From (4.150) we see that they have canonical commutation relations

$$\{\delta\gamma, \delta\rho\} = \int_M \delta\gamma \wedge \delta\rho. \tag{4.155}$$

Recalling that V_H and V_S were identified respectively with the B and A models, we see that the Hamiltonian has split into a "kinetic term" involving the B model and a "potential term" involving the A model. Despite this splitting the A and B models are not independent; the fact that ρ and γ do not commute at the *same point* of M suggests that, for the quantum Calabi-Yau, the uncertainty principle would prevent measurements of the complex structure and Kähler structure at a given point from being done simultaneously. This is an interesting result which deserves more scrutiny.

One can also ask about the commutation relations between the zero modes of the A and B model fields, which might be of more direct interest, because these zero modes are observables on which the partition function depends. Some of these zero modes are already present in the heuristic construction of the phase space which we gave above.

For example, one can consider variations of γ which are closed, $d(\delta\gamma) = 0$; these induce no variation in the Kähler form, but nevertheless affect the nonperturbative A model partition function (via the coupling to Lagrangian branes) as we discussed in Section 4.3. These variations of γ up to gauge equivalence parameterize an $H^3(M, \mathbb{R})$ in the phase space, which via (4.155) is canonically conjugate to the $H^3(M, \mathbb{R})$ given by the cohomology class of ρ. This means that the A model variables and B model variables mix; the parameter playing the role of the Lagrangian D-brane tension in the A model is conjugate to the 3-form of the B model.

There is a dual version of the above discussion: if we had started from the 3-form version of topological M-theory instead of the 4-form version, we would have written

$$\rho = \rho_0 + dB, \tag{4.156}$$

where B is a 2-form on M. (This field is very closely related to the field that we denoted as ϕ that appeared in the B model topological string.) The phase space then turns out to be spanned by B and σ with the Poisson bracket given by

$$\{\delta B, \delta\sigma\} = \int_M \delta B \wedge \delta\sigma. \tag{4.157}$$

This pairing agrees with the one we obtained above, except that it includes different zero modes: instead of having two copies of $H^3(M, \mathbb{R})$ we now have the variations of B up to gauge equivalence which do not change ρ, parameterizing an $H^2(M, \mathbb{R})$, canonically conjugate to the $H^4(M, \mathbb{R})$ given by the cohomology class of σ. Hence the B-field which couples to the D1-brane of the B model is conjugate to the Kähler parameter of the A model.

In sum, we seem to be finding that even at the level of the zero modes, i.e. the observables, there is a sense in which the fields of the A and B models are conjugate to one another. It is natural to suspect that this is related to the conjectured S-duality between the A and B models, which would be interpreted as position/momentum exchange or electric/magnetic duality in topological M-theory. In particular, the fact that nonperturbative amplitudes of the B model involve the D1-brane and the B-field, and the fact that the nonperturbative amplitudes of the A model involve Lagrangian D-branes and the γ field, suggest that the full nonperturbative topological string is a single entity consisting of the A and B models together.

Clearly these ideas should be developed further, but we feel that there is a beautiful connection here, between the conjectured S-duality between the A and B models and the fact that topological M-theory treats their degrees of freedom as conjugate variables.

One might ask how this Hamiltonian quantization is related to the fact that the B model partition function is a wavefunction, reviewed in Section 4.5.2, which was one of our original motivations for introducing a 7-dimensional topological M-theory. In the zero mode sector we have found two conjugate copies of $H^3(X, \mathbb{R})$, which would be sufficient to account for both the phase spaces underlying the B model partition function and the $\overline{\mathrm{B}}$ model partition function. This is reminiscent of Section 4.5.2 where we saw that we could interpret the Wigner function as a wavefunction for the combined B and $\overline{\mathrm{B}}$ models, with the zero mode phase space $H^3(X, \mathbb{C})$, parameterized by the conjugate variables $\mathrm{Re}\ \Omega = \rho$ and $\mathrm{Im}\ \Omega = \widehat{\rho}$. On the other hand, as dicussed above, in the 7-dimensional theory γ is conjugate to ρ. We are thus naturally led to identify $\widehat{\rho} = \gamma$. This identification was in a sense predicted by the topological S-duality conjecture, since it says precisely that the Lagrangian D-branes of the A model are coupled to the imaginary part of Ω. It indeed follows semiclassically if we identify the wavefunction as being given by the Hitchin functional, $\Psi(\rho) \sim \exp V_H(\rho)$; then we get the necessary relation

$$\gamma|\Psi\rangle = \widehat{\rho}(\rho)|\Psi\rangle \tag{4.158}$$

using $\delta V_H/\delta\rho = \widehat{\rho}$. This relation between the potential and the wavefunction is not unexpected, since the function V_H is quadratic in ρ.

4.8 Form theories of gravity and the black hole attractor mechanism

In the previous sections we have discussed various theories of gravity in which one reconstructs geometric structures from p-forms on the spacetime M. Although this might seem like an unusual way to get these structures, a similar phenomenon occurs in superstring theory compactified on M: given a black hole charge, which can be represented as an integral cohomology class on M, the attractor mechanism fixes certain metric data $g_{\mu\nu}$ of M at the black hole horizon [117, 52, 50, 51]. In other words, it provides a map[5]

$$Q \mapsto g_{\mu\nu}. \tag{4.159}$$

In this section, we will discuss a relation between black holes and Hitchin's functionals. In particular, we argue that these functionals also lead to the map (4.159). In a sense, the metric flow of the internal manifold from spatial infinity to the black

[5]More precisely, it fixes some of the components of g; not all of the moduli are fixed by the attractor mechanism.

hole horizon can be viewed as a geodesic flow with respect to Hitchin's action. In fact, Hitchin's picture is more general: *it does not assume the metric to be of the Calabi-Yau form, but derives that from the same action principle which leads to the relation between the charge and metric. The usual attractor mechanism only discusses the zero mode sector of the metric, whereas Hitchin's action also deals with the massive modes.*

This link between form theories of gravity and BPS black holes can, in fact, lead to a fundamental nonperturbative definition of the gravitational form theory as counting black hole degeneracies with a fixed charge, as in the recent work [98, 124]. This interpretation of the gravitational form theories also "explains" why one fixes the cohomology class of the form and integrates only over massive modes; this corresponds simply to fixing the black hole charge. At least in the cases of $4D$ and $5D$ BPS black holes, we will show that this interpretation is correct at leading order in the black hole charge; this amounts to the statement that *the value of the extremized classical action agrees with the semiclassical black hole entropy.*

4.8.1 BPS black holes in 4 dimensions

We begin with the case of $4D$ BPS black holes in Type IIB string theory compactified on a Calabi-Yau 3-fold M. In Section 4.4.3 we defined Hitchin's "holomorphic volume" (4.67), a functional of a 3-form ρ in six dimensions:

$$V_H(\rho) = \frac{1}{2}\int_M \widehat{\rho} \wedge \rho = -\frac{i}{4}\int_M \Omega \wedge \overline{\Omega}. \tag{4.160}$$

Furthermore we noted that, if we hold the cohomology class $[\rho]$ fixed (writing $\rho = \rho_0 + d\beta$), the critical points of $V_H(\rho)$ yield holomorphic 3-forms on M with real part

ρ. So the process of minimizing V_H produces the imaginary part of Ω as a function of its real part. Remarkably, this is exactly what the attractor mechanism does: fixing the black hole charge C for the theory on $\mathbb{R}^4$, the attractor mechanism produces the value of Ω of the Calabi-Yau at the black hole horizon, and the real part of Ω is equal to C^*, the Poincare dual of C. Therefore it is natural to identify

$$[\rho] = C^*. \tag{4.161}$$

Note that the quantization of C matches the fact that ρ is naturally quantized, if we view it as the field strength of the 2-form potential β. So the holomorphic volume functional V_H is related to the attractor mechanism at least classically.

Furthermore, the classical value of the action also has a natural physical meaning: namely, after fixing C, the value of $\int \Omega \wedge \overline{\Omega}$ at the critical point gives the leading-order contribution to the black hole entropy at large C. Now consider the *quantum* theory with action V_H. The path integral formally defines a partition function $Z_H(C)$ depending on the charge,

$$Z_H(C) = \int_{[\rho]=C^*} D\rho \exp(V_H(\rho)). \tag{4.162}$$

We conjecture that this path integral computes the exact number of states of the black hole (or more precisely the index $Z_{\mathrm{BH}}(C)$ defined in [98], which counts the states with signs):

$$Z_{\mathrm{BH}}(C) = Z_H(C). \tag{4.163}$$

The main evidence for this conjecture is that if the path integral (4.162) exists, it would be a function of C whose leading asymptotics agree with the black hole entropy — it would be remarkable if there were two such functions with natural physical definitions and they were not equal. Conversely, one could *define* the nonperturbative quantum theory by the black hole entropy.

Additional evidence for the conjecture (4.163) comes by noticing that it is essentially the conjecture of [98], which identified $Z_{BH}(C)$ with a Wigner function constructed from the B model partition function Z_B. Namely, choose a splitting of $H^3(M, \mathbb{Z})$ into A and B cycles. Then splitting C into electric and magnetic charges, $C = (P, Q)$, one has [98]

$$Z_{\text{BH}}(C) = \int d\Phi \, e^{iQ^I \Phi_I} |Z_B(P + i\Phi)|^2. \tag{4.164}$$

On the other hand, as we already discussed in Section 4.5.2, there is indeed a relation (4.119) between the B model and Hitchin's theory,

$$Z_H(C) = \int d\Phi \, e^{iQ^I \Phi_I} |Z_B(P + i\Phi)|^2. \tag{4.165}$$

Recall that Hitchin's theory based on V_H is related not to the B model but to the B plus $\overline{\text{B}}$ model; this agrees well with the fact that this B plus $\overline{\text{B}}$ model also appears in the counting of black hole entropy. This makes one more confident that the connection between Hitchin's theory and the black hole is direct and deep.

4.8.2 BPS black holes in 5 dimensions

So far we have discussed a relation between V_H and counting of 4-dimensional BPS black hole states obtained from Type II string theory on M. But as described in Section 4.4.3, there is also the functional V_S which makes sense on the 6-manifold M; one could ask whether it is also related to black hole entropy. In this section we will argue that it is, and the black holes in question are the ones in the 5-dimensional theory obtained by compactifying M-theory on M. These BPS black holes can be constructed by wrapping M2-branes over 2-cycles of M, and are characterized by a charge $Q \in H_2(M, \mathbb{Z}) = H^4(M, \mathbb{Z})$ and a spin j. At first let us set $j = 0$. To connect the black hole counting to Hitchin's theory based on V_S, we identify

$$Q^* = [\sigma]. \tag{4.166}$$

The attractor value of the moduli in this case is given [51, 37] by a Kähler form k, such that $\frac{1}{2}k^2 = \sigma$; with this value of k, the volume of the Calabi-Yau is proportional to the entropy of the black hole,

$$S_{BH} \sim \int_M k^3 = \int_M \sigma^{3/2}. \tag{4.167}$$

In other words, the black hole entropy is given by the classical value of $V_S(\sigma)$. This is automatically consistent with the fact that the black hole entropy in five dimensions scales as $Q^{3/2}$. So, in parallel with what we did for V_H, we conjecture that the partition function $Z_S([\sigma])$ of the theory based on V_S counts BPS states of 5-dimensional black holes.

It is possible to extend the foregoing discussion to spinning black holes, by introducing an additional 6-form field J in the Hitchin action V_S. We denote the integral cohomology class of J by $j = [J] \in H^6(M, \mathbb{Z})$; this j can be naturally identified with the spin of the black hole. We consider the action

$$V_S(\sigma, J) = \int \sqrt{\sigma^3 - J^2}, \tag{4.168}$$

where

$$\sigma^3 - J^2 = (\sigma_{i_1 i_2 i_3 i_4}\sigma_{j_1 j_2 j_3 j_4}\sigma_{k_1 k_2 k_3 k_4} - J_{i_1 i_2 j_1 j_2 k_1 k_2} J_{i_3 i_4 j_3 j_4 k_3 k_4})\epsilon^{i_1 i_2 j_1 j_2 k_1 k_2}\epsilon^{i_3 i_4 j_3 j_4 k_3 k_4}.$$

It is easy to see that this modification does not change the attractor value of the Kähler form k, but changes the classical value of the action to $\sqrt{Q^3 - j^2}$, which agrees with the entropy of the spinning black hole.

We have just argued that the quantum theory based on the extended functional (4.168) should count the degeneracies of BPS black holes in five dimensions. On the other hand, since the perturbative A model counts exactly these degeneracies [60], one might expect a direct relation between the A model and (4.168). At least for $j = 0$, we have already encountered this relation in Section 4.5.1, where the quantum foam description of the A model was related to a Polyakov version of V_S.

4.8.3 Other cases

It is natural to conjecture that the relation between BPS objects and form gravity theories goes beyond the examples discussed above. In particular, it would be interesting to develop this story for the case of G_2 manifolds. For example, in M-theory compactified on a G_2 manifold, we can consider BPS domain walls formed from M5-branes wrapped on associative 3-cycles. It is natural to conjecture that the quantum version of Hitchin's 4-form theory is computing the degeneracies of these domain walls. In the Type IIB superstring compactified on a G_2 manifold, one can similarly ask about the degeneracy of BPS strings obtained by wrapping D5-branes on coassociative 4-cycles; one might expect a relation between this counting and the quantum version of Hitchin's 3-form theory.

4.9 Topological G_2, twistors, holography, and $4D$ gauge theories

In this section, we discuss possible dualities relating three different theories:

1. gauge theory on a Riemannian 4-manifold M;
2. topological A model on the twistor space, $T(M)$, of a 4-manifold M;
3. topological M-theory on a 7-manifold X,

$$\begin{array}{ccc} \mathbb{R}^3 & \rightarrow & X \\ & & \downarrow \\ & & M \end{array} \tag{4.169}$$

As we reviewed earlier, the 7-manifold X admits a natural metric with G_2 holonomy if M is a self-dual (i.e., with self-dual Weyl tensor) Einstein 4-manifold.[6] In that case, the $\mathbb{R}^3$ bundle (4.169) is the bundle of self-dual 2-forms on M. Let us compare this to the corresponding geometric structure on the twistor space $T(M)$.

First, let us recall the definition of the twistor space $T(M)$. Consider the space of self-dual 2-forms of norm 1. For each point on M this gives rise to a 2-sphere. The total space is the twistor space, $T(M)$, which has a canonical almost complex structure and also a canonical map to M, with fiber being the twistor sphere $\mathcal{P}^1$. There is a remarkable connection between self-dual metrics on M (not necessarily Einstein) and the integrability of the almost complex structure on $T(M)$: *$T(M)$ has an integrable complex structure if and only if M is self-dual* [14, 102]. Moreover, $T(M)$

[6] Such manifolds are also known as quaternionic Kähler manifolds of dimension 1.

admits a Kähler structure if and only if M is Einstein [72] (see also [87]). These are the necessary conditions for the existence of topological A and B models on $T(M)$.

In order to complete this to a string theory we also need conformal invariance, which is usually guaranteed by a Ricci flatness condition. This is not the case, however, for the twistor space $T(M)$, which is not Ricci-flat. One can complete $T(M)$ to a Ricci-flat supermanifold by including extra fermionic directions [133]. We want to explore another way of obtaining Ricci-flatness. As discussed above, the bundle X of self-dual 2-forms over M has a natural G_2 holonomy metric, so in particular it is Ricci-flat. On the other hand, the boundary of X is precisely the twistor space,

$$T(M) = \partial X. \tag{4.170}$$

In this sense we could view X as obtained by adjoining a radial direction to $T(M)$.

So we could define a topological string theory on the twistor space $T(M)$ as a holographic dual to topological M-theory on X. We note that the A model can be defined on 3-folds which are not necessarily Calabi-Yau; it has been studied in the mathematical literature on Gromov-Witten theory [58, 86, 18, 47]. Conversely, using the Gromov-Witten theory on $T(M)$ we can define topological M-theory on X, at least perturbatively.

This holographic duality is reminiscent of our original motivation to look for a 7-dimensional theory, which would naturally explain the observation that topological string partition function should be viewed as a wavefunction. We also note that, in the present case, the boundary 6-manifold is not stationary under Hitchin's Hamiltonian flow equations; this reflects the fact that $T(M)$ is not a Calabi-Yau.

Large N Holography and Gravitational Holography

We are familiar with examples of holography in the context of open-closed string dualities, where in the large N limit D-branes wrapping some cycles disappear and the theory is best described by a new geometry obtained by deleting the locus of the D-branes, replaced by suitable fluxes. Via this holographic duality, the open string gauge theory provides an answer to questions of gravity in a geometry obtained by a large N transition.

Another kind of duality — which is somewhat similar to holography — is a duality between M-theory on an interval and the heterotic $E_8 \times E_8$ theory living on the boundary [77]. In this duality, the coupling constant of the heterotic string controls the size of the interval. Even though the heterotic string "lives" on the boundary, it can be used, at least in principle, to study gravitational physics in the bulk.

The relation between the 7-dimensional topological M-theory on X and the topological string on a 6-manifold $T(M)$ is more similar to the heterotic/M-theory duality. In this sense, when we say that the partition function of a topological string theory on the twistor space can be regarded as a wave function in a 7-dimensional theory on X, what we mean is a "gravity/gravity holography."

Having said that, it is natural to ask: is there an open-closed string holographic duality in the present context? Given that we do not yet have a deep understanding of topological M-theory, we will limit ourselves to some string-motivated speculations below.

In order to have an open/closed duality, we need to be working in some context where D-branes exist. From the point of view of embedding of the G_2 theory in the physical M-theory, it is natural to compactify on one more circle and obtain a Type

IIA string theory compactification on a G_2 manifold. So let us consider Type IIA on a non-compact G_2 manifold X of the form (4.169), with N D-branes wrapped over the coassociative 4-manifold M (in the full superstring theory these could be viewed for example as spacetime-filling D6-branes). By analogy with geometric transitions in Calabi-Yau 3-folds [62, 61], in the large N limit we expect a transition to a new geometry which can be obtained by removing the locus of the D-branes,

$$X \setminus M.$$

This space is a real line bundle over $T(M)$. From the discussion earlier in this section, we expect topological M-theory on this 7-manifold to be related to topological string theory on $T(M)$. This leads to a natural conjecture that topological gauge theory on a 4-manifold M is a holographic dual to topological string theory on the twistor space $T(M)$. This dovetails in a natural way with the idea that the topological string partition function should be viewed as a wave function on the boundary of the 7-dimensional manifold with G_2 holonomy.

What kind of topological gauge theory in four dimensions should we expect? The most natural conjecture is that it is the self-dual Yang-Mills theory, which is related to the D-brane theory for $\mathcal{N} = 2$ strings. In other words, one might conjecture that the self-dual Yang-Mills on M is dual to topological strings on $T(M)$, so that the Kähler class of $\mathcal{P}^1 \in T(M)$ is identified with the 't Hooft parameter of the dual gauge theory, $t = Ng_s$. Below, we consider this duality in more detail for $M = \mathbf{S}^4$ and $T(M) = \mathbf{CP}^3$.

Topological string theory on $T(M) = \mathcal{P}^3$ is rather trivial due to the $U(1)$ charge conservation on the worldsheet. In particular, the free energy is simply given by the cubic classical triple intersection of $\mathbf{CP}^3$. This agrees with the fact that self-dual Yang-Mills is also trivial in perturbation theory. However, it is known that topological strings can be made more interesting by turning on higher charge (q, q) form operators with $q = 2, 3$. The most natural one is the volume form with $q = 3$, which preserves all the symmetries of $\mathcal{P}^3$. Once we add an operator $s\Phi_{3,3}$, the topological A model string on $\mathcal{P}^3$ becomes non-trivial and receives all order corrections. Thus, the partition function of the perturbed A model is a function of two independent variables,

$$Z^{top}(g_s, s^2 e^{-t}). \tag{4.171}$$

The fact that the combination $s^2 e^{-t}$ appears follows from charge conservation of the topological A model.

One possibility is to look for a deformation of the self-dual Yang-Mills corresponding to the deformation of the topological A model by the operator $s\Phi_{3,3}$. In a realization *á la* Siegel [113, 114, 34],

$$S = \int d^4x \,\mathrm{Tr}\, F \wedge G, \tag{4.172}$$

the self-dual Yang-Mills is written in terms of a $U(N)$ adjoint valued self-dual 2-form G and the curvature of a $U(N)$ connection, F. We can deform the action (4.172) by the term $\epsilon G \wedge G$, which (perturbatively) leads to the full Yang-Mills theory. It is natural to ask whether this deformation is dual to the deformation of the A model on $\mathcal{P}^3$ by $s\Phi_{3,3}$. Notice, that bosonic Yang-Mills on $\mathbf{S}^4$ has partition function which depends on the radius of the 4-sphere, R, the coupling constant of Yang-Mills theory, g^2_{YM}, and the rank N of the gauge group. Due to the running of the coupling constant only one combination of g^2_{YM} and R appears. It is not unreasonable to suppose that with a suitable choice of the parameter map (which should involve some kind of Fourier transform) we have

$$Z^{top}_{\mathcal{P}^3}(g_s, s^2 e^{-t}) \leftrightarrow Z^{YM}_{\mathbf{S}^4}(g^2_{YM}, N) \tag{4.173}$$

It would be very interesting to further develop and check this conjecture. If correct, it would allow one to place the appearance of the higher-dimensional twistor space $T(M)$ in the large N limit of gauge theory on M into the context of more familiar large N dualities, *e.g.* the duality between Chern-Simons theory on S^3 and topological strings on the 6-dimensional resolved conifold [62, 61].

4.10 Hitchin's Hamiltonian flow and geometry of $\mathcal{N} = 1$ string vacua

The geometric structures which appear in the 7-dimensional topological gravity are reminiscent of the geometries that arise in $\mathcal{N} = 1$ superstring compactifications. For example, 7-manifolds with G_2 holonomy are classical solutions in 7-dimensional topological gravity and, on the other hand, are $\mathcal{N} = 1$ vacua of M-theory. This relation can be extended to 6-manifolds with $SU(3)$ structure which play an important role in understanding the space of string vacua with minimal ($\mathcal{N} = 1$) supersymmetry, and which we briefly review in this appendix; see [68, 33, 82, 70, 85, 19, 54, 63, 69] for more details.

Let M be a 6-manifold with $SU(3)$ structure. Such M are characterized by the existence of a globally defined, $SU(3)$ invariant spinor ξ, which is the analog of the covariantly constant spinor one has on a Calabi-Yau manifold. In general, instead of

$\nabla\xi = 0$ we have

$$\nabla^{(T)}\xi = 0, \tag{4.174}$$

where $\nabla^{(T)}$ is a connection twisted by torsion T. Roughly speaking, the intrinsic torsion T represents the deviation from the Calabi-Yau condition. Its $SU(3)$ representation content involves five classes, usually denoted $\mathcal{W}_i$ [109, 35]:

$$T \in \mathcal{W}_1 \oplus \mathcal{W}_2 \oplus \mathcal{W}_3 \oplus \mathcal{W}_4 \oplus \mathcal{W}_5. \tag{4.175}$$

In order to describe the geometric meaning of each of these components, it is convenient to introduce a 2-form k and a 3-form Ω,

$$\begin{aligned} k &= -i\xi^\dagger \Gamma_{mn}\Gamma_7 \xi, \\ \Omega &= -i\xi^\dagger \Gamma_{mnp}(1+\Gamma_7)\xi, \end{aligned} \tag{4.176}$$

which satisfy

$$k \wedge \Omega = 0. \tag{4.177}$$

On a Calabi-Yau manifold, the 2-form k would be the usual Kähler form, while Ω would be the holomorphic volume form. In particular, M is a Calabi-Yau manifold if and only if $dk = 0$, $d\Omega = 0$. On a general manifold M with $SU(3)$ structure, these equations are modified by the components of torsion,

$$\begin{aligned} dk &= -\frac{3}{2}\mathrm{Im}\ (W_1\overline{\Omega}) + W_4 \wedge k + W_3, \\ d\Omega &= W_1 k^2 + W_2 \wedge k + \overline{W}_5 \wedge \Omega, \end{aligned} \tag{4.178}$$

where

$$
\begin{aligned}
& W_1 \in \Omega^0(M), \\
& W_2 \in \Omega^2(M), \\
& W_3 = \overline{W}_3 \in \Omega^{2,1}_{\text{prim}}(M) \oplus \Omega^{1,2}_{\text{prim}}(M), \\
& W_4 = \overline{W}_4 \in \Omega^1(M), \\
& W_5 \in \Omega^{1,0}(M).
\end{aligned}
\tag{4.179}
$$

A particularly interesting class of manifolds with $SU(3)$ structure are the so-called half-flat manifolds. In superstring theory, they play an important role in constructing realistic vacua with minimal ($\mathcal{N} = 1$) supersymmetry, and can be viewed as mirrors of Calabi-Yau manifolds with (a particular kind of) NS-NS fluxes [68]. Since under mirror symmetry 3-forms are mapped into forms of even degree, on half-flat manifolds one might expect "NS-NS fluxes" represented by forms of even degree [123]. In fact, as we explain in a moment, on a half-flat manifold M we have

$$
d(\text{Im } \Omega) \sim F_4^{NS}.
$$

Half-flat manifolds are defined by requiring certain torsion components to vanish,

$$
\text{Re } \mathcal{W}_1 = \text{Re } \mathcal{W}_2^- = \mathcal{W}_4 = \mathcal{W}_5 = 0. \tag{4.180}
$$

It is easy to see from (4.178) that this is equivalent to the conditions

$$
\begin{aligned}
& d(k \wedge k) = 0, \\
& d(\text{Re } \Omega) = 0.
\end{aligned}
\tag{4.181}
$$

If as usual we define $\sigma = \frac{1}{2} k \wedge k$ and $\Omega = \rho + i\widehat{\rho}$, we can write these equations in the familiar form

$$d\rho = 0,$$
$$d\sigma = 0, \tag{4.182}$$

with an additional constraint $\rho \wedge k = 0$. This is precisely the structure induced on a generic 6-dimensional hypersurface inside a G_2 manifold, where ρ is the pull-back of the associative 3-form Φ and σ is the pull-back of the coassociative 4-form $*\Phi$. In particular, using Hitchin's Hamiltonian flow which we reviewed in Section 4.4.5, a half-flat $SU(3)$ structure on M can always be thickened into a G_2 holonomy metric on $X = M \times (a, b)$.

So the phase space underlying Hitchin's Hamiltonian flow consists precisely of the half-flat manifolds which appear in $\mathcal{N} = 1$ string compactifications with fluxes and/or torsion,

$$\mathcal{P}_{\rm Hitchin} = \{M^6_{\rm half-flat}\}. \tag{4.183}$$

Moreover, the ground states are related to stationary solutions of Hitchin's flow equations, namely Calabi-Yau manifolds,

$$|{\rm vac}\rangle \qquad \Leftrightarrow \qquad M^6 = {\rm Calabi-Yau}. \tag{4.184}$$

It is tempting to speculate that all $\mathcal{N} = 1$ string vacua can be realized in topological M-theory.

4.11 Directions for future research

In this chapter we have discussed the fact that many theories of gravity fall into the general class of "form gravity theories," and that they seem to be unified into a 7-dimensional theory of gravity, topological M-theory, related to G_2 holonomy metrics. We have seen in particular that this 7-dimensional theory contains the A and B model topological strings, which appear as conjugate degrees of freedom. We have also seen connections with 3-dimensional Chern-Simons gravity and a 4-dimensional form theory of gravity — the topological sector of loop quantum gravity.

Intriguing as this list is, we view this as only a modest beginning: the connections we have outlined raise many new questions which need to be answered. In order to understand better the non-perturbative aspects of the A and B models, and particularly their implementation in the context of topological M-theory, we need to understand better the relation between these models and M-theory. In particular, it seems natural to try to explain the S-duality relating the A and B models using the S-duality of Type IIB superstrings. This could be embedded into the present discussion if we include one more dimension and consider 8-dimensional manifolds of special holonomy. The natural candidate in that dimension are manifolds with $Spin(7)$ holonomy. It seems that we also need to include this theory in our discussion of dualities to get a better handle on the S-duality of the A/B models.

Another natural question we have raised relates to the interpretation of the topological M-theory: does it indeed count domain walls? This is a very natural conjecture based on the links we found between form theories of gravity and the counting of black hole states. It would be important to develop this idea more thoroughly.

Another question raised by our work is whether one can reformulate the full M-theory in terms of form theories of gravity. This may not be as implausible as it may sound at first sight. For example, we do know that $\mathcal{N} = 2$ supergravity in 4 dimensions, which is a low energy limit of superstrings compactified on Calabi-Yau manifolds, has a simple low energy action: it is simply the covariantized volume form on $(4|4)$ chiral superspace [116]. In fact, more is true: we could include the Calabi-Yau internal space as and write the leading term in the effective action as the volume element in dimension $(10|4)$. The internal volume theory in this case would coincide with that of Hitchin. Indeed, this is related to the fact that topological string amplitudes compute F-terms in the corresponding supergravity theory. Given this link it is natural to speculate that the full M-theory does admit such a low energy formulation, which could be a basis of another way to quantize M-theory — rather in tune with the notion of quantum gravitational foam.

We have also discussed a speculation, motivated by topological M-theory, relating gauge theories on M^4 to topological strings on its twistor space. This connection, even though it needs to be stated more sharply, is rather gratifying, because it would give a holographic explanation of the fact that in the twistor correspondence a 4-dimensional theory gets related to a theory in higher dimensions. It would be very interesting to develop this conjectural relation; the potential rewards are clearly great, as a full understanding of the duality could lead to a large N solution of non-supersymmetric Yang-Mills.

Bibliography

[1] M. Abe, A. Nakamichi, and T. Ueno, "Gravitational instantons and moduli spaces in topological two form gravity," *Phys. Rev.* **D50** (1994) 7323–7334, hep-th/9408178.

[2] M. Abe, A. Nakamichi, and T. Ueno, "Moduli space of topological two form gravity," *Mod. Phys. Lett.* **A9** (1994) 895–901, hep-th/9306130.

[3] B. S. Acharya, M. O'Loughlin, and B. Spence, "Higher-dimensional analogues of Donaldson-Witten theory," *Nucl. Phys.* **B503** (1997) 657–674, hep-th/9705138.

[4] A. Achucarro and P. K. Townsend, "A Chern-Simons action for three-dimensional anti-de Sitter supergravity theories," *Phys. Lett.* **B180** (1986) 89.

[5] M. Aganagic, R. Dijkgraaf, A. Klemm, M. Marino, and C. Vafa, "Topological strings and integrable hierarchies," hep-th/0312085.

[6] M. Aganagic, A. Klemm, M. Marino, and C. Vafa, "Matrix model as a mirror of Chern-Simons theory," hep-th/0211098.

[7] M. Aganagic, A. Klemm, M. Marino, and C. Vafa, "The topological vertex," hep-th/0305132.

[8] M. Aganagic, A. Klemm, and C. Vafa, "Disk instantons, mirror symmetry and the duality web," *Z. Naturforsch.* **A57** (2002) 1–28, hep-th/0105045.

[9] M. Aganagic, M. Marino, and C. Vafa, "All loop topological string amplitudes from Chern-Simons theory," hep-th/0206164.

[10] M. Aganagic, H. Ooguri, N. Saulina, and C. Vafa, "Black holes, q-deformed 2d Yang-Mills, and non-perturbative topological strings," hep-th/0411280.

[11] M. Aganagic and C. Vafa, "Mirror symmetry, D-branes and counting holomorphic discs," hep-th/0012041.

[12] M. Aganagic and C. Vafa, "G_2 manifolds, mirror symmetry and geometric engineering," hep-th/0110171.

[13] M. Aganagic and C. Vafa, "Mirror symmetry and supermanifolds," hep-th/0403192.

[14] M. F. Atiyah, N. J. Hitchin, and I. M. Singer, "Self-duality in four-dimensional Riemannian geometry," *Proc. Roy. Soc. London Ser. A* **362** (1978), no. 1711, 425–461.

[15] L. Baulieu, H. Kanno, and I. M. Singer, "Special quantum field theories in eight and other dimensions," *Commun. Math. Phys.* **194** (1998) 149–175, hep-th/9704167.

[16] L. Baulieu and I. M. Singer, "Topological Yang-Mills symmetry," *Nucl. Phys. Proc. Suppl.* **5B** (1988) 12–19.

[17] L. Baulieu and A. Tanzini, "Topological symmetry of forms, $\mathcal{N} = 1$ supersymmetry and S-duality on special manifolds," hep-th/0412014.

[18] A. Bayer and Y. I. Manin, "(Semi)simple exercises in quantum cohomology," in *The Fano Conference*, pp. 143–173. Univ. Torino, Turin, 2004. math.AG/0103164.

[19] K. Becker, M. Becker, K. Dasgupta, and P. S. Green, "Compactifications of heterotic theory on non-Kähler complex manifolds. I," *JHEP* **04** (2003) 007, hep-th/0301161.

[20] N. Berkovits, "An alternative string theory in twistor space for $\mathcal{N} = 4$ super-Yang-Mills," hep-th/0402045.

[21] N. Berkovits and L. Motl, "Cubic twistorial string field theory," hep-th/0403187.

[22] N. Berkovits and E. Witten, "Conformal supergravity in twistor-string theory," hep-th/0406051.

[23] M. Bershadsky, S. Cecotti, H. Ooguri, and C. Vafa, "Holomorphic anomalies in topological field theories," *Nucl. Phys.* **B405** (1993) 279–304, hep-th/9302103.

[24] M. Bershadsky, S. Cecotti, H. Ooguri, and C. Vafa, "Kodaira-Spencer theory of gravity and exact results for quantum string amplitudes," *Commun. Math. Phys.* **165** (1994) 311–428, hep-th/9309140.

[25] M. Bershadsky and V. Sadov, "Theory of Kähler gravity," *Int. J. Mod. Phys.* **A11** (1996) 4689–4730, `hep-th/9410011`.

[26] M. Bershadsky, V. Sadov, and C. Vafa, "D-branes and topological field theories," *Nucl. Phys.* **B463** (1996) 420–434, `hep-th/9511222`.

[27] M. Blau and G. Thompson, "Euclidean SYM theories by time reduction and special holonomy manifolds," *Phys. Lett.* **B415** (1997) 242–252, `hep-th/9706225`.

[28] J. Bryan and R. Pandharipande, "The local Gromov-Witten theory of curves," `math.ag/0411037`.

[29] R. L. Bryant and S. M. Salamon, "On the construction of some complete metrics with exceptional holonomy," *Duke Math. J.* **58** (1989), no. 3, 829–850.

[30] E. Buffenoir and P. Roche, "Two-dimensional lattice gauge theory based on a quantum group," *Commun. Math. Phys.* **170** (1995) 669–698, `hep-th/9405126`.

[31] F. A. Cachazo, P. Svrcek, and E. Witten, "MHV vertices and tree amplitudes in gauge theory," `hep-th/0403047`.

[32] R. Capovilla, J. Dell, T. Jacobson, and L. Mason, "Selfdual two forms and gravity," *Class. Quant. Grav.* **8** (1991) 41–57.

[33] G. L. Cardoso, G. Curio, G. Dall'Agata, D. Lust, P. Manousselis, and G. Zoupanos, "Non-Kähler string backgrounds and their five torsion classes," *Nucl. Phys.* **B652** (2003) 5–34, `hep-th/0211118`.

[34] G. Chalmers and W. Siegel, "The self-dual sector of QCD amplitudes," *Phys. Rev.* **D54** (1996) 7628–7633, `hep-th/9606061`.

[35] S. Chiossi and S. M. Salamon, "The intrinsic torsion of $SU(3)$ and G_2 structures," `math.DG/0202282`.

[36] Z. W. Chong, M. Cvetic, G. W. Gibbons, H. Lu, C. N. Pope, and P. Wagner, "General metrics of G_2 holonomy and contraction limits," *Nucl. Phys.* **B638** (2002) 459–482, `hep-th/0204064`.

[37] A. Chou, R. Kallosh, J. Ramfeld, S. Rey, M. Shmakova, and W. K. Wong, "Critical points and phase transitions in 5d compactifications of m-theory," *Nucl. Phys.* **B508** (1997) 147–180, `hep-th/9704142`.

[38] S. Cordes, G. W. Moore, and S. Ramgoolam, "Lectures on 2-d Yang-Mills theory, equivariant cohomology and topological field theories," *Nucl. Phys. Proc. Suppl.* **41** (1995) 184–244, `hep-th/9411210`.

[39] M. Cvetic, G. W. Gibbons, H. Lu, and C. N. Pope, "Bianchi IX self-dual Einstein metrics and singular G_2 manifolds," *Class. Quant. Grav.* **20** (2003) 4239–4268, hep-th/0206151.

[40] A. Dabholkar, "Exact counting of black hole microstates," hep-th/0409148.

[41] A. Dabholkar, F. Denef, G. W. Moore, and B. Pioline, "Exact and asymptotic degeneracies of small black holes," hep-th/0502157.

[42] B. de Wit, G. Lopes Cardoso, and T. Mohaupt, "Corrections to macroscopic supersymmetric black-hole entropy," *Phys. Lett.* **B451** (1999) 309–316, hep-th/9812082.

[43] B. de Wit, G. Lopes Cardoso, and T. Mohaupt, "Deviations from the area law for supersymmetric black holes," *Fortsch. Phys.* **48** (2000) 49–64, hep-th/9904005.

[44] B. de Wit, G. Lopes Cardoso, and T. Mohaupt, "Macroscopic entropy formulae and non-holomorphic corrections for supersymmetric black holes," *Nucl. Phys.* **B567** (2000) 87–110, hep-th/9906094.

[45] R. Dijkgraaf, E. Verlinde, and M. Vonk, "On the partition sum of the NS five-brane," hep-th/0205281.

[46] S. K. Donaldson and R. P. Thomas, "Gauge theory in higher dimensions," in *The geometric universe (Oxford, 1996)*, pp. 31–47. Oxford Univ. Press, Oxford, 1998.

[47] T. Eguchi, K. Hori, and C.-S. Xiong, "Gravitational quantum cohomology," *Int. J. Mod. Phys.* **A12** (1997) 1743–1782, hep-th/9605225.

[48] S. Elitzur, G. W. Moore, A. Schwimmer, and N. Seiberg, "Remarks on the canonical quantization of the Chern-Simons-Witten theory," *Nucl. Phys.* **B326** (1989) 108.

[49] C. Faber and R. Pandharipande, "Logarithmic series and Hodge integrals in the tautological ring (with an appendix by D. Zagier)," math.ag/0002112.

[50] S. Ferrara and R. Kallosh, "Supersymmetry and attractors," *Phys. Rev.* **D54** (1996) 1514–1524, hep-th/9602136.

[51] S. Ferrara and R. Kallosh, "Universality of supersymmetric attractors," *Phys. Rev.* **D54** (1996) 1525–1534, hep-th/9603090.

[52] S. Ferrara, R. Kallosh, and A. Strominger, "$\mathcal{N} = 2$ extremal black holes," *Phys. Rev.* **D52** (1995) 5412–5416, hep-th/9508072.

[53] W. Fulton and R. Pandharipande, "Notes on stable maps and quantum cohomology," in *Algebraic geometry—Santa Cruz 1995*, vol. 62 of *Proc. Sympos. Pure Math.*, pp. 45–96. Amer. Math. Soc., Providence, RI, 1997. `alg-geom/9608011`.

[54] J. P. Gauntlett, D. Martelli, and D. Waldram, "Superstrings with intrinsic torsion," *Phys. Rev.* **D69** (2004) 086002, `hep-th/0302158`.

[55] G. Georgiou and V. V. Khoze, "Tree Amplitudes in Gauge Theory as Scalar MHV Diagrams," `hep-th/0404072`.

[56] A. A. Gerasimov and S. L. Shatashvili, "Towards integrability of topological strings. I: Three- forms on Calabi-Yau manifolds," `hep-th/0409238`.

[57] G. W. Gibbons, D. N. Page, and C. N. Pope, "Einstein metrics on S^3, $\mathbb{R}^3$ and $\mathbb{R}^4$ bundles," *Commun. Math. Phys.* **127** (1990) 529.

[58] A. Givental and B.-s. Kim, "Quantum cohomology of flag manifolds and Toda lattices," *Commun. Math. Phys.* **168** (1995) 609–642, `hep-th/9312096`.

[59] R. Gopakumar and C. Vafa, "M-theory and topological strings. I," `hep-th/9809187`.

[60] R. Gopakumar and C. Vafa, "M-theory and topological strings. II," `hep-th/9812127`.

[61] R. Gopakumar and C. Vafa, "Topological gravity as large N topological gauge theory," *Adv. Theor. Math. Phys.* **2** (1998) 413–442, `hep-th/9802016`.

[62] R. Gopakumar and C. Vafa, "On the gauge theory/geometry correspondence," *Adv. Theor. Math. Phys.* **3** (1999) 1415–1443, `hep-th/9811131`.

[63] M. Grana, R. Minasian, M. Petrini, and A. Tomasiello, "Supersymmetric backgrounds from generalized calabi-yau manifolds," *JHEP* **08** (2004) 046, `hep-th/0406137`.

[64] D. J. Gross and W. Taylor, "Two-dimensional QCD is a string theory," *Nucl. Phys.* **B400** (1993) 181–210, `hep-th/9301068`.

[65] M. Gualtieri, "Generalized complex geometry," `math.DG/0401221`.

[66] S. Gukov, "Three-dimensional quantum gravity, Chern-Simons theory, and the A-polynomial," *Commun. Math. Phys.* **255** (2005) 577–627, `hep-th/0306165`.

[67] S. Gukov, S.-T. Yau, and E. Zaslow, "Duality and fibrations on G_2 manifolds," `hep-th/0203217`.

[68] S. Gurrieri, J. Louis, A. Micu, and D. Waldram, "Mirror symmetry in generalized Calabi-Yau compactifications," *Nucl. Phys.* **B654** (2003) 61–113, `hep-th/0211102`.

[69] S. Gurrieri, A. Lukas, and A. Micu, "Heterotic on half-flat," *Phys. Rev.* **D70** (2004) 126009, `hep-th/0408121`.

[70] S. Gurrieri and A. Micu, "Type IIB theory on half-flat manifolds," *Class. Quant. Grav.* **20** (2003) 2181–2192, `hep-th/0212278`.

[71] J. A. Harvey and G. W. Moore, "Superpotentials and membrane instantons," `hep-th/9907026`.

[72] N. J. Hitchin, "Kählerian twistor spaces," *Proc. London Math. Soc. (3)* **43** (1981), no. 1, 133–150.

[73] N. J. Hitchin, "The geometry of three-forms in six and seven dimensions," `math.dg/0010054`.

[74] N. J. Hitchin, "Stable forms and special metrics," `math.dg/0107101`.

[75] N. J. Hitchin, "Generalized calabi-yau manifolds," *Quart. J. Math. Oxford Ser.* **54** (2003) 281–308, `math.dg/0209099`.

[76] C. Hofman and J.-S. Park, "Cohomological Yang-Mills theories on Kähler 3-folds," *Nucl. Phys.* **B600** (2001) 133–162, `hep-th/0010103`.

[77] P. Horava and E. Witten, "Heterotic and type I string dynamics from eleven dimensions," *Nucl. Phys.* **B460** (1996) 506–524, `hep-th/9510209`.

[78] G. T. Horowitz, "Exactly soluble diffeomorphism invariant theories," *Commun. Math. Phys.* **125** (1989) 417.

[79] A. Iqbal and A.-K. Kashani-Poor, "The vertex on a strip," `hep-th/0410174`.

[80] A. Iqbal, N. A. Nekrasov, A. Okounkov, and C. Vafa, "Quantum foam and topological strings," `hep-th/0312022`.

[81] W. Israel, "Differential forms in general relativity," *Comm. Dublin Inst. Adv. Stud. Ser. A* (1979), no. 26, 80.

[82] S. Kachru, M. B. Schulz, P. K. Tripathy, and S. P. Trivedi, "New supersymmetric string compactifications," *JHEP* **03** (2003) 061, `hep-th/0211182`.

[83] A. Kapustin, "Gauge theory, topological strings, and S-duality," `hep-th/0404041`.

[84] A. Kapustin, "Topological strings on noncommutative manifolds," *Int. J. Geom. Meth. Mod. Phys.* **1** (2004) 49–81, `hep-th/0310057`.

[85] P. Kaste, R. Minasian, M. Petrini, and A. Tomasiello, "Nontrivial RR two-form field strength and $SU(3)$-structure," *Fortsch. Phys.* **51** (2003) 764–768, `hep-th/0301063`.

[86] M. Kontsevich and Y. I. Manin, "Gromov-Witten classes, quantum cohomology, and enumerative geometry," *Commun. Math. Phys.* **164** (1994) 525–562, `hep-th/9402147`.

[87] C. LeBrun, "Twistors for tourists: a pocket guide for algebraic geometers," in *Algebraic geometry—Santa Cruz 1995*, vol. 62 of *Proc. Sympos. Pure Math.*, pp. 361–385. Amer. Math. Soc., Providence, RI, 1997.

[88] H. Y. Lee, A. Nakamichi, and T. Ueno, "Topological two form gravity in four-dimensions," *Phys. Rev.* **D47** (1993) 1563–1568, `hep-th/9205066`.

[89] I. G. Macdonald, *Symmetric functions and Hall polynomials.* Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, second ed., 1995. With contributions by A. Zelevinsky, Oxford Science Publications.

[90] M. L. Mangano, S. J. Parke, and Z. Xu, "Duality and multi-gluon scattering," *Nucl. Phys.* **B298** (1988) 653.

[91] D. Maulik, N. A. Nekrasov, A. Okounkov, and R. Pandharipande, "Gromov-Witten theory and Donaldson-Thomas theory," `math.AG/0312059`.

[92] R. C. McLean, "Deformations of calibrated submanifolds," *Comm. Anal. Geom.* **6** (1998), no. 4, 705–747.

[93] A. Nakamichi, "Wave function of the universe in topological and in einstein two form gravity," `hep-th/9303135`.

[94] A. Neitzke and C. Vafa, "$\mathcal{N} = 2$ strings and the twistorial Calabi-Yau," `hep-th/0402128`.

[95] A. Neitzke and C. Vafa, "Topological strings and their physical applications," `hep-th/0410178`.

[96] N. A. Nekrasov, H. Ooguri, and C. Vafa, "S-duality and topological strings," `hep-th/0403167`.

[97] A. Okounkov, N. Reshetikhin, and C. Vafa, "Quantum Calabi-Yau and classical crystals," `hep-th/0309208`.

[98] H. Ooguri, A. Strominger, and C. Vafa, "Black hole attractors and the topological string," `hep-th/0405146`.

[99] H. Ooguri and C. Vafa, "Geometry of $\mathcal{N} = 2$ strings," *Nucl. Phys.* **B361** (1991) 469–518.

[100] H. Ooguri and C. Vafa, "Knot invariants and topological strings," *Nucl. Phys.* **B577** (2000) 419–438, `hep-th/9912123`.

[101] S. J. Parke and T. R. Taylor, "An amplitude for n gluon scattering," *Phys. Rev. Lett.* **56** (1986) 2459.

[102] R. Penrose, "Nonlinear gravitons and curved twistor theory," *Gen. Rel. Grav.* **7** (1976) 31–52.

[103] V. Pestun and E. Witten, "The Hitchin functionals and the topological B-model at one loop," `hep-th/0503083`.

[104] M. J. Plebanski, "On the separation of Einsteinian substructures," *J. Math. Phys.* **18** (1977) 2511.

[105] N. Reshetikhin and V. G. Turaev, "Invariants of 3-manifolds via link polynomials and quantum groups," *Invent. Math.* **103** (1991) 547–597.

[106] R. Roiban, M. Spradlin, and A. Volovich, "A googly amplitude from the B-model in twistor space," `hep-th/0402016`.

[107] R. Roiban, M. Spradlin, and A. Volovich, "On the tree-level S-matrix of Yang-Mills theory," `hep-th/0403190`.

[108] R. Roiban and A. Volovich, "All googly amplitudes from the B-model in twistor space," `hep-th/0402121`.

[109] S. M. Salamon, "Almost parallel structures," `math.DG/0107146`.

[110] N. Saulina and C. Vafa, "D-branes as defects in the Calabi-Yau crystal," `hep-th/0404246`.

[111] A. Sen, "Black holes and the spectrum of half-BPS states in $\mathcal{N} = 4$ supersymmetric string theory," `hep-th/0504005`.

[112] A. Sen, "Black holes, elementary strings and holomorphic anomaly," `hep-th/0502126`.

[113] W. Siegel, "$\mathcal{N} = 2$, $\mathcal{N} = 4$ string theory is selfdual $\mathcal{N} = 4$ Yang-Mills theory," *Phys. Rev.* **D46** (1992) 3235–3238.

[114] W. Siegel, "Selfdual $\mathcal{N} = 8$ supergravity as closed $\mathcal{N} = 2$ ($\mathcal{N} = 4$) strings," *Phys. Rev.* **D47** (1993) 2504–2511, hep-th/9207043.

[115] L. Smolin, "An invitation to loop quantum gravity," hep-th/0408048.

[116] E. Sokatchev, "A superspace action for $\mathcal{N} = 2$ supergravity," *Phys. Lett.* **B100** (1981) 466.

[117] A. Strominger, "Macroscopic Entropy of $\mathcal{N} = 2$ Extremal Black Holes," *Phys. Lett.* **B383** (1996) 39–43, hep-th/9602111.

[118] C. G. Torre, "Perturbations of gravitational instantons," *Phys. Rev.* **D41** (1990) 3620.

[119] C. G. Torre, "A topological field theory of gravitational instantons," *Phys. Lett.* **B252** (1990) 242–246.

[120] V. G. Turaev, "Quantum invariants of 3-manifolds and a glimpse of shadow topology," *C. R. Acad. Sci. Paris Sér. I Math.* **313** (1991), no. 6, 395–398.

[121] V. G. Turaev, "Topology of Shadow," 1991. Preprint.

[122] V. G. Turaev and O. Y. Viro, "State sum invariants of 3-manifolds and quantum $6j$ symbols," *Topology* **31** (1992) 865–902.

[123] C. Vafa, "Superstrings and topological strings at large N," *J. Math. Phys.* **42** (2001) 2798–2817, hep-th/0008142.

[124] C. Vafa, "Two dimensional Yang-Mills, black holes and topological strings," hep-th/0406058.

[125] C. Vafa and E. Witten, "A strong coupling test of S duality," *Nucl. Phys.* **B431** (1994) 3–77, hep-th/9408074.

[126] E. Verlinde, "Attractors and the holomorphic anomaly," hep-th/0412139.

[127] K. Walker, "On Witten's 3-manifold invariants." Unpublished.

[128] F. Witt, "Generalised g_2 manifolds," math.dg/0411642.

[129] E. Witten, "(2+1)-dimensional gravity as an exactly soluble system," *Nucl. Phys.* **B311** (1988) 46.

[130] E. Witten, "Quantum field theory and the Jones polynomial," *Commun. Math. Phys.* **121** (1989) 351.

[131] E. Witten, "Quantum background independence in string theory," hep-th/9306122.

[132] E. Witten, "Chern-Simons gauge theory as a string theory," *Prog. Math.* **133** (1995) 637–678, `hep-th/9207094`.

[133] E. Witten, "Perturbative gauge theory as a string theory in twistor space," `hep-th/0312171`.

[134] E. Witten, "Parity invariance for strings in twistor space," `hep-th/0403199`.

[135] C.-J. Zhu, "The googly amplitudes in gauge theory," `hep-th/0403115`.