7

ISBN 92-95003-10-1

trieste - italy

the **abdus salam** international centre for theoretical physics

united nations educational, scientific and cultural organization

international atomic energy agency

ictp *lecture notes*

2001 SPRING SCHOOL ON SUPERSTRINGS AND RELATED MATTERS

2002

editors
C. Bachas
J. Maldacena
K.S. Narain
S. Randjbar-Daemi

ICTP *Lecture Notes*

2001 SPRING SCHOOL ON SUPERSTRINGS AND RELATED MATTERS

2 - 10 April 2001

Editors

C. Bachas
ENS, Paris
J. Maldacena
Harvard University, Cambridge, USA
K.S. Narain
The Abdus Salam ICTP, Trieste, Italy
S. Randjbar-Daemi
The Abdus Salam ICTP, Trieste, Italy

2001 SPRING SCHOOL ON SUPERSTRINGS AND RELATED MATTERS
– First edition

ISBN 92-95003-10-1

Printed in Trieste by The Abdus Salam ICTP Publications & Printing Section

PREFACE

One of the main missions of the Abdus Salam International Centre for Theoretical Physics in Trieste, Italy, founded in 1964 by Abdus Salam, is to foster the growth of advanced studies and research in developing countries. To this aim, the Centre organizes a large number of schools and workshops in a great variety of physical and mathematical disciplines.

Since unpublished material presented at the meetings might prove of great interest also to scientists who did not take part in the schools the Centre has decided to make it available through a new publication titled ICTP Lecture Note Series. It is hoped that this formally structured pedagogical material in advanced topics will be helpful to young students and researchers, in particular to those working under less favourable conditions.

The Centre is grateful to all lecturers and editors who kindly authorize the ICTP to publish their notes as a contribution to the series.

Since the initiative is new, comments and suggestions are most welcome and greatly appreciated. Information can be obtained from the Publications Section or by e-mail to "`pub_off@ictp.trieste.it`". The series is published in house and also made available on-line via the ICTP web site: "`http://www.ictp.trieste.it`".

M.A. Virasoro
Director

CONTENTS

Introduction

This proceedings contains the lectures given at the 2001 Trieste Spring School on String Theory. Several important and active areas of research in string theory related topics were covered in this school. One of the main topics of the School was the recently conjectured duality between gauge theory living on D-branes and and gravity (or more precisely string theory) living in the near horizon geometry around the D-branes. J. Maldacena gave a set of lectures on the gauge theory/gravity duality in different examples. M. Strassler's lectures dealt with a very interesting generalization of the gauge theory/gravity duality for the case of a confining gauge theory. D. Kutasov's lectures dealt with Little String Theories (LST) that are supposed to describe the physics of the NS5-branes. Using the holographic principle, interesting features of LST were deduced by describing the string theory in the background of NS5-branes.

E. Verlinde gave a set of lectures on holographic principle in the context of radiation dominated FRW universe. Other topics included lectures by R. Gopakumar on the solitons in non-commutative gauge theories that are relevant in the context of D-branes in the background on anti-symmetric tensor field, and lectures by M. Douglas on D-branes on Calabi-Yau spaces.

K.S. Narain
May, 2002

Introduction

This proceedings contains the lectures given at the 2001 Trieste Spring School on String Theory. Several important and active areas of research in string theory related topics were covered in this school. One of the main topics of the School was the recently conjectured duality between gauge theory living on D-branes and and gravity (or more precisely string theory) living in the near horizon geometry around the D-branes. J. Maldacena gave a set of lectures on the gauge theory/gravity duality in different examples. M. Strassler's lectures dealt with a very interesting generalization of the gauge theory/gravity duality for the case of a confining gauge theory. D. Kutasov's lectures dealt with Little String Theories (LST) that are supposed to describe the physics of the NS5-branes. Using the holographic principle, interesting features of LST were deduced by describing the string theory in the background of NS5-branes.

E. Verlinde gave a set of lectures on holographic principle in the context of radiation dominated FRW universe. Other topics included lectures by R. Gopakumar on the solitons in non-commutative gauge theories that are relevant in the context of D-branes in the background on anti-symmetric tensor field, and lectures by M. Douglas on D-branes on Calabi-Yau spaces.

K.S. Narain
May, 2002

CONTENTS

Large N Field Theories, String Theory and Gravity

Juan Maldacena*

Lyman Laboratory of Physics, Harvard University, Cambridge, USA

Lectures given at the
Spring School on Superstrings and Related Matters
Trieste, 2 – 10 April 2001

LNS027001

*malda@ias.edu

Abstract

We describe the holographic correspondence between field theories and string/M theory, focusing on the relation between compactifications of string/ M theory on Anti-de Sitter spaces and conformal field theories. We review the background for this correspondence and discuss its motivations and the evidence for its correctness. We describe the main results that have been derived from the correspondence in the regime that the field theory is approximated by classical or semiclassical gravity. We focus on the case of the $\mathcal{N} = 4$ supersymmetric gauge theory in four dimensions. These lecture notes are based on the Review written by O. Aharony, S. Gubser, J. Maldacena, H. Ooguri and Y. Oz, [1].

Contents

1 General introduction

These lecture notes are taken out of the review [1]. A more complete set of references is given there.

Even though though string theory is normally used as a theory of quantum gravity, it is not how string theory was originally discovered. String theory was discovered in an attempt to describe the large number of mesons and hadrons that were experimentally discovered in the 1960's. The idea was to view all these particles as different oscillation modes of a string. The string idea described well some features of the hadron spectrum. For example, the mass of the lightest hadron with a given spin obeys a relation like $m^2 \sim TJ^2 + const$. This is explained simply by assuming that the mass and angular momentum come from a rotating, relativistic string of tension T. It was later discovered that hadrons and mesons are actually made of quarks and that they are described by QCD.

QCD is a gauge theory based on the group $SU(3)$. This is sometimes stated by saying that quarks have three colors. QCD is asymptotically free, meaning that the effective coupling constant decreases as the energy increases. At low energies QCD becomes strongly coupled and it is not easy to perform calculations. One possible approach is to use numerical simulations on the lattice. This is at present the best available tool to do calculations in QCD at low energies. It was suggested by 't Hooft that the theory might simplify when the number of colors N is large [7]. The hope was that one could solve exactly the theory with $N = \infty$, and then one could do an expansion in $1/N = 1/3$. Furthermore, as explained in the next section, the diagrammatic expansion of the field theory suggests that the large N theory is a free string theory and that the string coupling constant is $1/N$. If the case with $N = 3$ is similar to the case with $N = \infty$ then this explains why the string model gave the correct relation between the mass and the angular momentum. In this way the large N limit connects gauge theories with string theories. The 't Hooft argument, reviewed below, is very general, so it suggests that different kinds of gauge theories will correspond to different string theories. In this review we will study this correspondence between string theories and the large N limit of field theories. We will see that the strings arising in the large N limit of field theories are the same as the strings describing quantum gravity. Namely, string theory in some backgrounds, including quantum gravity, is equivalent (dual) to a field theory.

Strings are not consistent in four flat dimensions. Indeed, if one wants to quantize a four dimensional string theory an anomaly appears that forces the introduction of an extra field, sometimes called the "Liouville" field [8]. This field on the string worldsheet may be interpreted as an extra dimension, so that the strings effectively move in five dimensions. One might qualitatively think of this new field as the "thickness" of the string. If this is the case, why do we say that the string moves in five dimensions? The reason is that, like any string theory, this theory will contain gravity, and the gravitational theory will live in as many dimensions as the number of fields we have on the string. It is crucial then that the five dimensional geometry is curved, so that it can correspond to a four dimensional field theory, as described in detail below.

The argument that gauge theories are related to string theories in the large N limit is very general and is valid for basically any gauge theory. In particular we could consider a gauge theory where the coupling does not run (as a function of the energy scale). Then, the theory is conformally invariant. It is quite hard to find quantum field theories that are conformally invariant. In supersymmetric theories it is sometimes possible to prove exact conformal invariance. A simple example, which will be the main example in this review, is the supersymmetric $SU(N)$ (or $U(N)$) gauge theory in four dimensions with four spinor supercharges ($\mathcal{N} = 4$). Four is the maximal possible number of supercharges for a field theory in four dimensions. Besides the gauge fields (gluons) this theory contains also four fermions and six scalar fields in the adjoint representation of the gauge group. The Lagrangian of such theories is completely determined by supersymmetry. There is a global $SU(4)$ R-symmetry that rotates the six scalar fields and the four fermions. The conformal group in four dimensions is $SO(4,2)$, including the usual Poincaré transformations as well as scale transformations and special conformal transformations (which include the inversion symmetry $x^\mu \to x^\mu/x^2$). These symmetries of the field theory should be reflected in the dual string theory. The simplest way for this to happen is if the five dimensional geometry has these symmetries. Locally there is only one space with $SO(4,2)$ isometries: five dimensional Anti-de-Sitter space, or AdS_5. Anti-de Sitter space is the maximally symmetric solution of Einstein's equations with a negative cosmological constant. In this supersymmetric case we expect the strings to also be supersymmetric. We said that superstrings move in ten dimensions. Now that we have added one more dimension it is not surprising any more to add five more to get to a ten dimensional space. Since the gauge theory has an $SU(4) \simeq SO(6)$ global symmetry it is rather natural that the extra five dimensional space should be a five sphere, S^5. So, we conclude that $\mathcal{N} = 4$ $U(N)$ Yang-Mills theory could be the same as ten dimensional superstring theory on $AdS_5 \times S^5$ [9]. Here we have presented a very heuristic argument for this equivalence; later we will be more precise and give more evidence for this correspondence.

The relationship we described between gauge theories and string theory on Anti-de-Sitter spaces was motivated by studies of D-branes and black holes in strings theory. D-branes are solitons in string theory [10]. They come in various dimensionalities. If they have zero spatial dimensions they are like ordinary localized, particle-type soliton solutions, analogous to the 't Hooft-Polyakov [11, 12] monopole in gauge theories. These are called D-zero-branes. If they have one extended dimension they are called D-one-branes or D-strings. They are much heavier than ordinary fundamental strings when the string coupling is small. In fact, the tension of all D-branes is proportional to $1/g_s$, where g_s is the string coupling constant. D-branes are defined in string perturbation theory in a very simple way: they are surfaces where open strings can end. These open strings have some massless modes, which describe the oscillations of the branes, a gauge field living on the brane, and their fermionic partners. If we have N coincident branes the open strings can start and end on different branes, so they carry two indices that run from one to N. This in turn implies that the low energy dynamics is described by a $U(N)$ gauge theory. D-p-branes are charged under $p+1$-form gauge potentials, in the same way that a 0-brane (particle) can be charged under a one-form gauge potential (as in electromagnetism). These $p+1$-form gauge potentials have $p+2$-form field strengths, and they are part of the massless closed string modes, which belong to the supergravity (SUGRA) multiplet containing the massless fields in flat space string theory (before we put in any D-branes). If we now add D-branes they generate a flux of the corresponding field strength, and this flux in turn contributes to the stress energy tensor so the geometry becomes curved. Indeed it is possible to find solutions of the supergravity equations carrying these fluxes. Supergravity is the low-energy limit of string theory, and it is believed that these solutions may be extended to solutions of the full string theory. These solutions are very similar to extremal charged black hole solutions in general relativity, except that in this case they are black branes with p extended spatial dimensions. Like black holes they contain event horizons.

If we consider a set of N coincident D-3-branes the near horizon geometry turns out to be $AdS_5 \times S^5$. On the other hand, the low energy dynamics on their worldvolume is governed by a $U(N)$ gauge theory with $\mathcal{N}=4$ supersymmetry [13]. These two pictures of D-branes are perturbatively valid for different regimes in the space of possible coupling constants. Perturbative field theory is valid when g_sN is small, while the low-energy gravitational description is perturbatively valid when the radius of curvature is much larger than the string scale, which turns out to imply that g_sN should be very large. As an object is brought closer and closer to the black brane horizon its energy measured by an outside observer is redshifted, due to the large gravitational potential, and the energy seems to be very small. On the other hand low energy excitations on the branes are governed by the Yang-Mills theory. So, it becomes natural to conjecture that Yang-Mills theory at strong coupling is describing the near horizon region of the black brane, whose geometry is $AdS_5 \times S^5$. The first indications that this is the case came from calculations of low energy graviton absorption cross sections [14, 15, 16]. It was noticed there that the calculation done using gravity and the calculation done using super Yang-Mills theory agreed. These calculations, in turn, were inspired by similar calculations for coincident D1-D5 branes. In this case the near horizon geometry involves $AdS_3 \times S^3$ and the low energy field theory living on the D-branes is a 1+1 dimensional conformal field theory. In this D1-D5 case there were numerous calculations that agreed between the field theory and gravity. First black hole entropy for extremal black holes was calculated in terms of the field theory in [17], and then agreement was shown for near extremal black holes [18, 19] and for absorption cross sections [20, 21, 22]. More generally, we will see that correlation functions in the gauge theory can be calculated using the string theory (or gravity for large $g_s N$) description, by considering the propagation of particles between different points in the boundary of AdS, the points where operators are inserted [23, 24].

Supergravities on AdS spaces were studied very extensively, see [25, 26] for reviews. See also [2, 3] for earlier hints of the correspondence.

One of the main points of these lectures will be that the strings coming from gauge theories are very much like the ordinary superstrings that have been studied during the last 20 years. The only particular feature is that they are moving on a curved geometry (anti-de Sitter space) which has a boundary at spatial infinity. The boundary is at an infinite spatial distance, but a light ray can go to the boundary and come back in finite time. Massive particles can never get to the boundary. The radius of curvature of Anti-de Sitter space depends on N so that large N corresponds to a large radius of curvature. Thus, by taking N to be large we can make the curvature as small as we want. The theory in AdS includes gravity, since any string theory includes gravity. So in the end we claim that there is an equivalence between a gravitational theory and a field theory. However, the mapping between the gravitational and field theory degrees of freedom is quite non-trivial since the field theory lives in a lower dimension. In some sense the field theory (or at least the set of local observables in the field theory) lives on the boundary of spacetime. One could argue that in general any quantum gravity theory in AdS defines a conformal field theory (CFT) "on the boundary". In some sense the situation is similar to the correspondence between three dimensional Chern-Simons theory and a WZW model on the boundary [27]. This is a topological theory in three dimensions that induces a normal (non-topological) field theory on the boundary. A theory which includes gravity is in some sense topological since one is integrating over all metrics and therefore the theory does not depend on the metric. Similarly, in a quantum gravity theory we do not have any local observables. Notice that when we say that the theory includes "gravity on AdS" we are considering any finite energy excitation, even black holes in AdS. So this is really a sum over all spacetimes that are asymptotic to AdS at the boundary. This is analogous to the usual flat space discussion of quantum gravity, where asymptotic flatness is required, but the spacetime could have any topology as long as it is asymptotically flat. The asymptotically AdS case as well as the asymptotically flat cases are special in the sense that one can choose a natural time and an associated Hamiltonian to define the quantum theory. Since black holes might be present this time coordinate is not necessarily globally well-defined, but it is certainly well-defined at infinity. If we assume that the conjecture we made above is valid, then the $U(N)$ Yang-Mills theory gives a non-perturbative definition of string theory on AdS. And, by taking the limit $N \to \infty$, we can extract the (ten dimensional string theory) flat space physics, a procedure which is in principle (but not in detail) similar to the one used in matrix theory [28].

The fact that the field theory lives in a lower dimensional space blends in perfectly with some previous speculations about quantum gravity. It was suggested [29, 30] that quantum gravity theories should be holographic, in the sense that physics in some region can be described by a theory at the boundary with no more than one degree of freedom per Planck area. This "holographic" principle comes from thinking about the Bekenstein bound which states that the maximum amount of entropy in some region is given by the area of the region in Planck units [31]. The reason for this bound is that otherwise black hole formation could violate the second law of thermodynamics. We will see that the correspondence between field theories and string theory on AdS space (including gravity) is a concrete realization of this holographic principle.

Other reviews of this subject are [32, 33, 34, 35, 1].

2 The correspondence

In this section we will present an argument connecting type IIB string theory compactified on $AdS_5 \times S^5$ to $\mathcal{N} = 4$ super-Yang-Mills theory [9]. Let us start with type IIB string theory in flat, ten dimensional Minkowski space. Consider N parallel D3 branes that are sitting together or very close to each other (the precise meaning of "very close" will be defined below). The D3 branes are extended along a $(3+1)$ dimensional plane in $(9+1)$ dimensional spacetime. String theory on this background contains two kinds of perturbative excitations, closed strings and open strings. The closed strings are the excitations of empty space and the open strings end on the D-branes and describe excitations of the D-branes. If we consider the system at low energies, energies lower than the string scale $1/l_s$, then only the massless string states can be excited, and we can write an effective Lagrangian describing their interactions. The closed string massless states give a gravity supermultiplet in ten dimensions, and their low-energy effective Lagrangian is that of type IIB supergravity. The open string massless states give an $\mathcal{N}=4$ vector supermultiplet in $(3+1)$ dimensions, and their low-energy effective Lagrangian is that of $\mathcal{N}=4$ $U(N)$ super-Yang-Mills theory [13, 36].

The complete effective action of the massless modes will have the form

$$S = S_{\text{bulk}} + S_{\text{brane}} + S_{\text{int}}. \tag{1}$$

S_{bulk} is the action of ten dimensional supergravity, plus some higher derivative corrections. Note that the Lagrangian (1) involves only the massless fields but it takes into account the effects of integrating out the massive fields. It is not renormalizable (even for the fields on the brane), and it should only be understood as an effective description in the Wilsonian sense, i.e. we integrate out all massive degrees of freedom but we do not integrate out the massless ones. The brane action S_{brane} is defined on the $(3+1)$ dimensional brane worldvolume, and it contains the $\mathcal{N}=4$ super-Yang-Mills Lagrangian plus some higher derivative corrections, for example terms of the form $\alpha'^2\text{Tr}(F^4)$. Finally, S_{int} describes the interactions between the brane modes and the bulk modes. The leading terms in this interaction Lagrangian can be obtained by covariantizing the brane action, introducing the background metric for the brane [37].

We can expand the bulk action as a free quadratic part describing the propagation of free massless modes (including the graviton), plus some interactions which are proportional to positive powers of the square root of the Newton constant. Schematically we have

$$S_{bulk} \sim \frac{1}{2\kappa^2}\int \sqrt{g}\mathcal{R} \sim \int (\partial h)^2 + \kappa(\partial h)^2 h + \cdots, \tag{2}$$

where we have written the metric as $g = \eta + \kappa h$. We indicate explicitly the dependence on the graviton, but the other terms in the Lagrangian, involving other fields, can be expanded in a similar way. Similarly, the interaction Lagrangian S_{int} is proportional to positive powers of κ. If we take the low energy limit, all interaction terms proportional to κ drop out. This is the well known fact that gravity becomes free at long distances (low energies).

In order to see more clearly what happens in this low energy limit it is convenient to keep the energy fixed and send $l_s \to 0$ ($\alpha' \to 0$) keeping all the dimensionless parameters fixed, including the string coupling constant and N. In this limit the coupling $\kappa \sim g_s \alpha'^2 \to 0$, so that the interaction Lagrangian relating the bulk and the brane vanishes. In addition all the higher derivative terms in the brane action vanish, leaving just the pure $\mathcal{N} = 4$ $U(N)$ gauge theory in $3 + 1$ dimensions, which is known to be a conformal field theory. And, the supergravity theory in the bulk becomes free. So, in this low energy limit we have two decoupled systems. On the one hand we have free gravity in the bulk and on the other hand we have the four dimensional gauge theory.

Next, we consider the same system from a different point of view. D-branes are massive charged objects which act as a source for the various supergravity fields. We can find a D3 brane solution [38] of supergravity, of the form

$$\begin{aligned} ds^2 &= f^{-1/2}(-dt^2 + dx_1^2 + dx_2^2 + dx_3^2) + f^{1/2}(dr^2 + r^2 d\Omega_5^2) \ , \\ F_5 &= (1 + *) dt dx_1 dx_2 dx_3 df^{-1} \ , \\ f &= 1 + \frac{R^4}{r^4} \ , \qquad R^4 \equiv 4\pi g_s \alpha'^2 N \ . \end{aligned} \tag{3}$$

Note that since g_{tt} is non-constant, the energy E_p of an object as measured by an observer at a constant position r and the energy E measured by an observer at infinity are related by the redshift factor

$$E = f^{-1/4} E_p \ . \tag{4}$$

This means that the same object brought closer and closer to $r = 0$ would appear to have lower and lower energy for the observer at infinity. Now we take the low energy limit in the background described by equation (3). There are two kinds of low energy excitations (from the point of view of an observer at infinity). We can have massless particles propagating in the bulk region with wavelengths that becomes very large, or we can have any kind of excitation that we bring closer and closer to $r = 0$. In the low energy limit these two types of excitations decouple from each other. The bulk massless particles decouple from the near horizon region (around $r = 0$) because the low energy absorption cross section goes like $\sigma \sim \omega^3 R^8$ [14, 15], where ω is the energy. This can be understood from the fact that in this limit the wavelength of the particle becomes much bigger than the typical gravitational size of the brane (which is of order R). Similarly, the excitations that live very close to $r=0$ find it harder and harder to climb the gravitational potential and escape to the asymptotic region. In conclusion, the low energy theory consists of two decoupled pieces, one is free bulk supergravity and the second is the near horizon region of the geometry. In the near horizon region, $r \ll R$, we can approximate $f \sim R^4/r^4$, and the geometry becomes

$$ds^2 = \frac{r^2}{R^2}(-dt^2 + dx_1^2 + dx_2^2 + dx_3^2) + R^2\frac{dr^2}{r^2} + R^2 d\Omega_5^2, \tag{5}$$

which is the geometry of $AdS_5 \times S^5$.

We see that both from the point of view of a field theory of open strings living on the brane, and from the point of view of the supergravity description, we have two decoupled theories in the low-energy limit. In both cases one of the decoupled systems is supergravity in flat space. So, it is natural to identify the second system which appears in both descriptions. Thus, we are led to the conjecture that $\mathcal{N}=4$ $U(N)$ *super-Yang-Mills theory in* $3+1$ *dimensions is the same as (or dual to) type IIB superstring theory on* $AdS_5 \times S^5$ [9].

We could be a bit more precise about the near horizon limit and how it is being taken. Suppose that we take $\alpha' \to 0$, as we did when we discussed the field theory living on the brane. We want to keep fixed the energies of the objects in the throat (the near-horizon region) in string units, so that we can consider arbitrary excited string states there. This implies that $\sqrt{\alpha'}E_p \sim$ fixed. For small α' (4) reduces to $E \sim E_p r/\sqrt{\alpha'}$. Since we want to keep fixed the energy measured from infinity, which is the way energies are measured in the field theory, we need to take $r \to 0$ keeping r/α' fixed. It is then convenient to define a new variable $U \equiv r/\alpha'$, so that the metric becomes

$$ds^2 = \alpha'\left[\frac{U^2}{\sqrt{4\pi g_s N}}(-dt^2 + dx_1^2 + dx_2^2 + dx_3^2) + \sqrt{4\pi g_s N}\frac{dU^2}{U^2} + \sqrt{4\pi g_s N}d\Omega_5^2\right]. \tag{6}$$

This can also be seen by considering a D3 brane sitting at $\vec{r}$. This corresponds to giving a vacuum expectation value to one of the scalars in the Yang-Mills theory. When we take the $\alpha' \to 0$ limit we want to keep the mass of the "W-boson" fixed. This mass, which is the mass of the string stretching between the branes sitting at $\vec{r}=0$ and the one at $\vec{r}$, is proportional to $U = r/\alpha'$, so this quantity should remain fixed in the decoupling limit.

A $U(N)$ gauge theory is essentially equivalent to a free $U(1)$ vector multiplet times an $SU(N)$ gauge theory, up to some $\mathbb{Z}_N$ identifications (which affect only global issues). In the dual string theory all modes interact with gravity, so there are no decoupled modes. Therefore, the bulk AdS theory is describing the $SU(N)$ part of the gauge theory. In fact we were not precise when we said that there were two sets of excitations at low energies, the excitations in the asymptotic flat space and the excitations in the near horizon region. There are also some zero modes which live in the region connecting the "throat" (the near horizon region) with the bulk, which correspond to the $U(1)$ degrees of freedom mentioned above. The $U(1)$ vector supermultiplet includes six scalars which are related to the center of mass motion of all the branes [39]. From the AdS point of view these zero modes live at the boundary, and it looks like we might or might not decide to include them in the AdS theory. Depending on this choice we could have a correspondence to an $SU(N)$ or a $U(N)$ theory. The $U(1)$ center of mass degree of freedom is related to the topological theory of B-fields on AdS [40]; if one imposes local boundary conditions for these B-fields at the boundary of AdS one finds a $U(1)$ gauge field living at the boundary [41], as is familiar in Chern-Simons theories [27, 42]. These modes living at the boundary are sometimes called singletons (or doubletons) [43, 44, 45, 46, 47, 48, 49, 50, 51].

Anti-de-Sitter space has a large group of isometries, which is $SO(4,2)$ for the case at hand. This is the same group as the conformal group in $3+1$ dimensions. Thus, the fact that the low-energy field theory on the brane is conformal is reflected in the fact that the near horizon geometry is Anti-de-Sitter space. We also have some supersymmetries. The number of supersymmetries is twice that of the full solution (3) containing the asymptotic region [39]. This doubling of supersymmetries is viewed in the field theory as a consequence of superconformal invariance, since the superconformal algebra has twice as many fermionic generators as the corresponding Poincare superalgebra. We also have an $SO(6)$ symmetry which rotates the S^5. This can be identified with the $SU(4)_R$ R-symmetry group of the field theory. In fact, the whole supergroup is the same for the $\mathcal{N}=4$ field theory and the $AdS_5 \times S^5$ geometry, so both sides of the conjecture have the same spacetime symmetries. We will discuss in more detail the matching between the two sides of the correspondence in section 3.

In the above derivation the field theory is naturally defined on $\mathbb{R}^{3,1}$, but we could also think of the conformal field theory as defined on $S^3 \times \mathbb{R}$ by redefining the Hamiltonian. Since the isometries of AdS are in one to one correspondence with the generators of the conformal group of the field theory, we can conclude that this new Hamiltonian $\frac{1}{2}(P_0 + K_0)$ can be associated on AdS to the generator of translations in global time. This formulation of the conjecture is more useful since in the global coordinates there is no horizon. When we put the field theory on S^3 the Coulomb branch is lifted and there is a unique ground state. This is due to the fact that the scalars ϕ^I in the field theory are conformally coupled, so there is a term of the form $\int d^4x \mathrm{Tr}(\phi^2)\mathcal{R}$ in the Lagrangian, where $\mathcal{R}$ is the curvature of the four-dimensional space on which the theory is defined. Due to the positive curvature of S^3 this leads to a mass term for the scalars [24], lifting the moduli space.

The parameter N appears on the string theory side as the flux of the five-form Ramond-Ramond field strength on the S^5,

$$\int_{S^5} F_5 = N. \tag{7}$$

From the physics of D-branes we know that the Yang-Mills coupling is related to the string coupling through [10, 52]

$$\tau \equiv \frac{4\pi i}{g_{YM}^2} + \frac{\theta}{2\pi} = \frac{i}{g_s} + \frac{\chi}{2\pi} \ , \tag{8}$$

where we have also included the relationship of the θ angle to the expectation value of the RR scalar χ. We have written the couplings in this fashion because both the gauge theory and the string theory have an $SL(2,\mathbb{Z})$ self-duality symmetry under which $\tau \to (a\tau + b)/(c\tau + d)$ (where a, b, c, d are integers with $ad - bc = 1$). In fact, $SL(2,\mathbb{Z})$ is a conjectured strong-weak coupling duality symmetry of type IIB string theory in flat space [53], and it should also be a symmetry in the present context since all the fields that are being turned on in the $AdS_5 \times S^5$ background (the metric and the five form field strength) are invariant under this symmetry. The connection between the $SL(2,\mathbb{Z})$ duality symmetries of type IIB string theory and $\mathcal{N} = 4$ SYM was noted in [54, 55, 56]. The string theory seems to have a parameter that does not appear in the gauge theory, namely α', which sets the string tension and all other scales in the string theory. However, this is not really a parameter in the theory if we do not compare it to other scales in the theory, since only relative scales are meaningful. In fact, only the ratio of the radius of curvature to α' is a parameter, but not α' and the radius of curvature independently. Thus, α' will disappear from any final physical quantity we compute in this theory. It is sometimes convenient, especially when one is doing gravity calculations, to set the radius of curvature to one. This can be achieved by writing the metric as $ds^2 = R^2 d\tilde{s}^2$, and rewriting everything in terms of $\tilde{g}$. With these conventions $G_N \sim 1/N^2$ and $\alpha' \sim 1/\sqrt{g_s N}$. This implies that any quantity calculated purely in terms of the gravity solution, without including stringy effects, will be independent of $g_s N$ and will depend only on N. α' corrections to the gravity results give corrections which are proportional to powers of $1/\sqrt{g_s N}$.

Now, let us address the question of the validity of various approximations. The analysis of loop diagrams in the field theory shows that we can trust the perturbative analysis in the Yang-Mills theory when

$$g_{YM}^2 N \sim g_s N \sim \frac{R^4}{l_s^4} \ll 1. \tag{9}$$

Note that we need $g_{YM}^2 N$ to be small and not just g_{YM}^2. On the other hand, the classical gravity description becomes reliable when the radius of curvature R of AdS and of S^5 becomes large compared to the string length,

$$\frac{R^4}{l_s^4} \sim g_s N \sim g_{YM}^2 N \gg 1. \tag{10}$$

We see that the gravity regime (10) and the perturbative field theory regime (9) are perfectly incompatible. In this fashion we avoid any obvious contradiction due to the fact that the two theories look very different. This is the reason that this correspondence is called a "duality". The two theories are conjectured to be exactly the same, but when one side is weakly coupled the other is strongly coupled and vice versa. This makes the correspondence both hard to prove and useful, as we can solve a strongly coupled gauge theory via classical supergravity. Notice that in (9)(10) we implicitly assumed that $g_s < 1$. If $g_s > 1$ we can perform an $SL(2,\mathbb{Z})$ duality transformation and get conditions similar to (9)(10) but with $g_s \to 1/g_s$. So, we cannot get into the gravity regime (10) by taking N small ($N = 1, 2, ..$) and g_s very large, since in that case the D-string becomes light and renders the gravity approximation invalid. Another way to see this is to note that the radius of curvature in Planck units is $R^4/l_p^4 \sim N$. So, it is always necessary, but not sufficient, to have large N in order to have a weakly coupled supergravity description.

One might wonder why the above argument was not a proof rather than a conjecture. It is not a proof because we did not treat the string theory non-perturbatively (not even non-perturbatively in α'). We could also consider different forms of the conjecture. In its weakest form the gravity description would be valid for large $g_s N$, but the full string theory on AdS might not agree with the field theory. A not so weak form would say that the conjecture is valid even for finite $g_s N$, but only in the $N \to \infty$ limit (so that the α' corrections would agree with the field theory, but the g_s corrections may not). The strong form of the conjecture, which is the most interesting one and which we will assume here, is that the two theories are exactly the same for all values of g_s and N. In this conjecture the spacetime is only required to be asymptotic to $AdS_5 \times S^5$ as we approach the boundary. In the interior we can have all kinds of processes; gravitons, highly excited fundamental string states, D-branes, black holes, etc. Even the topology of spacetime can change in the interior. The Yang-Mills theory is supposed to effectively sum over all spacetimes which are asymptotic to $AdS_5 \times S^5$. This is completely analogous to the usual conditions of asymptotic flatness. We can have black holes and all kinds of topology changing processes, as long as spacetime is asymptotically flat. In this case asymptotic flatness is replaced by the asymptotic AdS behavior.

2.1 Brane probes and multicenter solutions

The moduli space of vacua of the $\mathcal{N} = 4$ $U(N)$ gauge theory is $(\mathbb{R}^6)^N/S_N$, parametrizing the positions of the N branes in the six dimensional transverse space. In the supergravity solution one can replace

$$f \propto \frac{N}{r^4} \to \sum_{i=1}^{N} \frac{1}{|\vec{r} - \vec{r}_i|^4}, \qquad (11)$$

and still have a solution to the supergravity equations. We see that if $|\vec{r}| \gg |\vec{r}_i|$ then the two solutions are basically the same, while when we go to $r \sim r_i$ the solution starts looking like the solution of a single brane. Of course, we cannot trust the supergravity solution for a single brane (since the curvature in Planck units is proportional to a negative power of N). What we can do is separate the N branes into groups of N_i branes with $g_s N_i \gg 1$ for all i. Then we can trust the gravity solution everywhere.

Another possibility is to separate just one brane (or a small number of branes) from a group of N branes. Then we can view this brane as a D3-brane in the AdS_5 background which is generated by the other branes (as described above). A string stretching between the brane probe and the N branes appears in the gravity description as a string stretching between the D3-brane and the horizon of AdS. This seems a bit surprising at first since the proper distance to the horizon is infinite. However, we get a finite result for the energy of this state once we remember to include the redshift factor. The D3-branes in AdS (like any D3-branes in string theory) are described at low energies by the Born-Infeld action, which is the Yang-Mills action plus some higher derivative corrections. This seems to contradict, at first sight, the fact that the dual field theory (coming from the original branes) is just the pure Yang-Mills theory. In order to understand this point more precisely let us write explicitly the bosonic part of the Born-Infeld action for a D-3 brane in AdS [37],

$$
\begin{aligned}
S = & -\frac{1}{(2\pi)^3 g_s \alpha'^2} \int d^4x f^{-1} \Big[\\
& \sqrt{-\det(\eta_{\alpha\beta} + f\partial_\alpha r \partial_\beta r + r^2 f g_{ij} \partial_\alpha \theta^i \partial_\beta \theta^j + 2\pi\alpha' \sqrt{f} F_{\alpha\beta})} - 1 \Big] , \\
f = & \frac{4\pi g_s \alpha'^2 N}{r^4} ,
\end{aligned} \tag{12}
$$

where θ^i are angular coordinates on the 5-sphere. We can easily check that if we define a new coordinate $U = r/\alpha'$, then all the α' dependence drops out of this action. Since U (which has dimensions of energy) corresponds to the mass of the W bosons in this configuration, it is the natural way to express the Higgs expectation value that breaks $U(N+1)$ to $U(N) \times U(1)$. In fact, the action (12) is precisely the low-energy effective action in the field theory for the massless $U(1)$ degrees of freedom, that we obtain after integrating out the massive degrees of freedom (W bosons). We can expand (12) in powers of ∂U and we see that the quadratic term does not have any correction, which is consistent with the non-renormalization theorem for $\mathcal{N} = 4$ super-Yang-Mills [57]. The $(\partial U)^4$ term has only a one-loop correction, and this is also consistent with another non-renormalization theorem [58]. This one-loop correction can be evaluated explicitly in the gauge theory and the result agrees with the supergravity result [59]. It is possible to argue, using broken conformal invariance, that all terms in (12) are determined by the $(\partial U)^4$ term [9]. Since the massive degrees of freedom that we are integrating out have a mass proportional to U, the action (12) makes sense as long as the energies involved are much smaller than U. In particular, we need $\partial U/U \ll U$. Since (12) has the form $\mathcal{L}(g_s N (\partial U)^2/U^4)$, the higher order terms in (12) could become important in the supergravity regime, when $g_s N \gg 1$. The Born Infeld action (12), as always, makes sense only when the curvature of the brane is small, but the deviations from a straight flat brane could be large. In this regime we can keep the non-linear terms in (12) while we still neglect the massive string modes and similar effects. Further gauge theory calculations for effective actions of D-brane probes include [60, 61, 62].

2.2 The field $\leftrightarrow$ operator correspondence

A conformal field theory does not have asymptotic states or an S-matrix, so the natural objects to consider are operators. For example, in $\mathcal{N} = 4$ super-Yang-Mills we have a deformation by a marginal operator which changes the value of the coupling constant. Changing the coupling constant in the field theory is related by (8) to changing the coupling constant in the string theory, which is then related to the expectation value of the dilaton. The expectation value of the dilaton is set by the boundary condition for the dilaton at infinity. So, changing the gauge theory coupling constant corresponds to changing the boundary value of the dilaton. More precisely, let us denote by $\mathcal{O}$ the corresponding operator. We can consider adding the term $\int d^4x \phi_0(\vec{x})\mathcal{O}(\vec{x})$ to the Lagrangian (for simplicity we assume that such a term was not present in the original Lagrangian, otherwise we consider $\phi_0(\vec{x})$ to be the total coefficient of $\mathcal{O}(\vec{x})$ in the Lagrangian). According to the discussion above, it is natural to assume that this will change the boundary condition of the dilaton at the boundary of AdS to $\phi(\vec{x}, z)|_{z=0} = \phi_0(\vec{x})$, in the coordinate system

$$ds^2 = R_{AdS}^2 \frac{-dt^2 + dx_1^2 + \cdots + dx_3^3 + dz^2}{z^2}.$$

More precisely, as argued in [23, 24], it is natural to propose that

$$\langle e^{\int d^4x \phi_0(\vec{x})\mathcal{O}(\vec{x})} \rangle_{CFT} = \mathcal{Z}_{string}\left[\phi(\vec{x}, z)\Big|_{z=0} = \phi_0(\vec{x}) \right], \tag{13}$$

where the left-hand side is the generating function of correlation functions in the field theory, i.e. ϕ_0 is an arbitrary function and we can calculate correlation functions of $\mathcal{O}$ by taking functional derivatives with respect to ϕ_0 and then setting $\phi_0 = 0$. The right-hand side is the full partition function of string theory with the boundary condition that the field ϕ has the value ϕ_0 on the boundary of AdS. Notice that ϕ_0 is a function of the four variables parametrizing the boundary of AdS_5.

A formula like (13) is valid in general, for any field ϕ. Therefore, each field propagating on AdS space is in a one to one correspondence with an operator in the field theory. There is a relation between the mass of the field ϕ and the scaling dimension of the operator in the conformal field theory. Let us describe this more generally in AdS_{d+1}. The wave equation in Euclidean space for a field of mass m has two independent solutions, which behave like $z^{d-\Delta}$ and z^{Δ} for small z (close to the boundary of AdS), where

$$\Delta = \frac{d}{2} + \sqrt{\frac{d^2}{4} + R^2 m^2}. \tag{14}$$

Therefore, in order to get consistent behavior for a massive field, the boundary condition on the field in the right-hand side of (13) should in general be changed to

$$\phi(\vec{x}, \epsilon) = \epsilon^{d-\Delta} \phi_0(\vec{x}), \tag{15}$$

and eventually we would take the limit where $\epsilon \to 0$. Since ϕ is dimensionless, we see that ϕ_0 has dimensions of $[\text{length}]^{\Delta-d}$ which implies, through the left-hand side of (13), that the associated operator $\mathcal{O}$ has dimension Δ (14). A more detailed derivation of this relation will be given in section 4, where we will verify that the two-point correlation function of the operator $\mathcal{O}$ behaves as that of an operator of dimension Δ [23, 24]. A similar relation between fields on AdS and operators in the field theory exists also for non-scalar fields, including fermions and tensors on AdS space.

Correlation functions in the gauge theory can be computed from (13) by differentiating with respect to ϕ_0. Each differentiation brings down an insertion $\mathcal{O}$, which sends a ϕ particle (a closed string state) into the bulk. Feynman diagrams can be used to compute the interactions of particles in the bulk. In the limit where classical supergravity is applicable, the only diagrams that contribute are the tree-level diagrams of the gravity theory (see for instance figure 1).

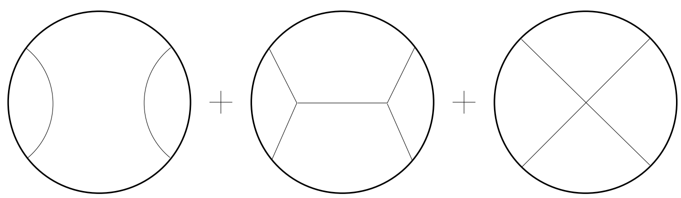


Figure 1: Correlation functions can be calculated (in the large g_sN limit) in terms of supergravity Feynman diagrams. Here we see the leading contribution coming from a disconnected diagram plus connected pieces involving interactions of the supergravity fields in the bulk of AdS. At tree level, these diagrams and those related to them by crossing are the only ones that contribute to the four-point function.

This method of defining the correlation functions of a field theory which is dual to a gravity theory in the bulk of AdS space is quite general, and it applies in principle to any theory of gravity [24]. Any local field theory contains the stress tensor as an operator. Since the correspondence described above matches the stress-energy tensor with the graviton, this implies that the AdS theory includes gravity. It should be a well defined quantum theory of gravity since we should be able to compute loop diagrams. String theory provides such a theory. But if a new way of defining quantum gravity theories comes along we could consider those gravity theories in AdS, and they should correspond to some conformal field theory "on the boundary". In particular, we could consider backgrounds of string theory of the form $AdS_5 \times M^5$ where M^5 is any Einstein manifold [63, 64, 65]. Depending on the choice of M^5 we get different dual conformal field theories. Similarly, this discussion can be extended to any AdS_{d+1} space, corresponding to a conformal field theory in d spacetime dimensions (for $d > 1$).

2.3 Holography

In this section we will describe how the AdS/CFT correspondence gives a holographic description of physics in AdS spaces.

Let us start by explaining the Bekenstein bound, which states that the maximum entropy in a region of space is $S_{max} = \text{Area}/4G_N$ [31], where the area is that of the boundary of the region. Suppose that we had a state with more entropy than S_{max}, then we show that we could violate the second law of thermodynamics. We can throw in some extra matter such that we form a black hole. The entropy should not decrease. But if a black hole forms inside the region its entropy is just the area of its horizon, which is smaller than the area of the boundary of the region (which by our assumption is smaller than the initial entropy). So, the second law has been violated.

Note that this bound implies that the number of degrees of freedom inside some region grows as the area of the boundary of a region and not like the volume of the region. In standard quantum field theories this is certainly not possible. Attempting to understand this behavior leads to the "holographic principle", which states that in a quantum gravity theory all physics within some volume can be described in terms of some theory on the boundary which has less than one degree of freedom per Planck area [29, 30] (so that its entropy satisfies the Bekenstein bound).

In the AdS/CFT correspondence we are describing physics in the bulk of AdS space by a field theory of one less dimension (which can be thought of as living on the boundary), so it looks like holography. However, it is hard to check what the number of degrees of freedom per Planck area is, since the theory, being conformal, has an infinite number of degrees of freedom, and the area of the boundary of AdS space is also infinite. Thus, in order to compare things properly we should introduce a cutoff on the number of degrees of freedom in the field theory and see what it corresponds to in the gravity theory. For this purpose let us write the metric of AdS as

$$ds^2 = R^2 \left[-\left(\frac{1+r^2}{1-r^2}\right)^2 dt^2 + \frac{4}{(1-r^2)^2}(dr^2 + r^2 d\Omega^2) \right] . \tag{16}$$

In these coordinates the boundary of AdS is at $r = 1$. We saw above that when we calculate correlation functions we have to specify boundary conditions at $r = 1 - \delta$ and then take the limit of $\delta \to 0$. It is clear by studying the action of the conformal group on Poincaré coordinates that the radial position plays the role of some energy scale, since we approach the boundary when we do a conformal transformation that localizes objects in the CFT. So, the limit $\delta \to 0$ corresponds to going to the UV of the field theory. When we are close to the boundary we could also use the Poincaré coordinates

$$ds^2 = R^2 \frac{-dt^2 + d\vec{x}^2 + dz^2}{z^2}, \tag{17}$$

in which the boundary is at $z = 0$. If we consider a particle or wave propagating in (17) or (16) we see that its motion is independent of R in the supergravity approximation. Furthermore, if we are in Euclidean space and we have a wave that has some spatial extent λ in the $\vec{x}$ directions, it will also have an extent λ in the z direction. This can be seen from (17) by eliminating λ through the change of variables $x \to \lambda x$, $z \to \lambda z$. This implies that a cutoff at

$$z \sim \delta \tag{18}$$

corresponds to a UV cutoff in the field theory at distances δ, with no factors of R (δ here is dimensionless, in the field theory it is measured in terms of the radius of the S^4 or S^3 that the theory lives on). Equation (18) is called the UV-IR relation [66].

Consider the case of $\mathcal{N} = 4$ SYM on a three-sphere of radius one. We can estimate the number of degrees of freedom in the field theory with a UV cutoff δ. We get

$$S \sim N^2 \delta^{-3}, \tag{19}$$

since the number of cells into which we divide the three-sphere is of order $1/\delta^3$. In the gravity solution (16) the area in Planck units of the surface at $r = 1 - \delta$, for $\delta \ll 1$, is

$$\frac{\text{Area}}{4G_N} = \frac{V_{S^5} R^3 \delta^{-3}}{4G_N} \sim N^2 \delta^{-3}. \tag{20}$$

Thus, we see that the AdS/CFT correspondence saturates the holographic bound [66].

One could be a little suspicious of the statement that gravity in AdS is holographic, since it does not seem to be saying much because in AdS space the volume and the boundary area of a given region scale in the same fashion as we increase the size of the region. In fact, *any* field theory in AdS would be holographic in the sense that the number of degrees of freedom within some (large enough) volume is proportional to the area (and also to the volume). What makes this case different is that we have the additional parameter R, and then we can take AdS spaces of different radii (corresponding to different values of N in the SYM theory), and then we can ask whether the number of degrees of freedom goes like the volume or the area, since these have a different dependence on R.

One might get confused by the fact that the surface $r = 1 - \delta$ is really nine dimensional as opposed to four dimensional. From the form of the full metric on $AdS_5 \times S^5$ we see that as we take $\delta \to 0$ the physical size of four of the dimensions of this nine dimensional space grow, while the other five, the S^5, remain constant. So, we see that the theory on this nine dimensional surface becomes effectively four dimensional, since we need to multiply the metric by a factor that goes to zero as we approach the boundary in order to define a finite metric for the four dimensional gauge theory.

3 Tests of the AdS/CFT correspondence

In this section we review the direct tests of the AdS/CFT correspondence. In section 2 we saw how string theory on AdS defines a partition function which can be used to define a field theory. Here we will review the evidence showing that this field theory is indeed the same as the conjectured dual field theory. We will focus here only on tests of the correspondence between the $\mathcal{N} = 4$ $SU(N)$ SYM theory and the type IIB string theory compactified on $AdS_5 \times S^5$; most of the tests described here can be generalized also to cases in other dimensions and/or with less supersymmetry, which will be described below.

As described in section 2, the AdS/CFT correspondence is a strong/weak coupling duality. In the 't Hooft large N limit, it relates the region of weak field theory coupling $\lambda = g_{YM}^2 N$ in the SYM theory to the region of high curvature (in string units) in the string theory, and vice versa. Thus, a direct comparison of correlation functions is generally not possible, since (with our current knowledge) we can only compute most of them perturbatively in λ on the field theory side and perturbatively in $1/\sqrt{\lambda}$ on the string theory side. For example, as described below, we can compute the equation of state of the SYM theory and also the quark-anti-quark potential both for small λ and for large λ, and we obtain different answers, which we do not know how to compare since we can only compute them perturbatively on both sides. A similar situation arises also in many field theory dualities that were analyzed in the last few years (such as the electric/magnetic $SL(2, \mathbb{Z})$ duality of the $\mathcal{N} = 4$ SYM theory itself), and it was realized that there are several properties of these theories which do not depend on the coupling, so they can be compared to test the duality. These are:

- The global symmetries of the theory, which cannot change as we change the coupling (except for extreme values of the coupling). As discussed in section 2, in the case of the AdS/CFT correspondence we have the same supergroup $SU(2,2|4)$ (whose bosonic subgroup is $SO(4,2) \times SU(4)$) as the global symmetry of both theories. Also, both theories are believed to have a non-perturbative $SL(2,\mathbb{Z})$ duality symmetry acting on their coupling constant τ. These are the only symmetries of the theory on $\mathbb{R}^4$. Additional $\mathbb{Z}_N$ symmetries arise when the theories are compactified on non-simply-connected manifolds, and these were also successfully matched in [67, 40][1].

- Some correlation functions, which are usually related to anomalies, are protected from any quantum corrections and do not depend on λ. The matching of these correlation functions will be described in section 3.2 below.

- The spectrum of chiral operators does not change as the coupling varies, and it will be compared in section 3.1 below.

- The moduli space of the theory also does not depend on the coupling. In the $SU(N)$ field theory the moduli space is $\mathbb{R}^{6(N-1)}/S_N$, parametrized by the eigenvalues of six commuting traceless $N \times N$ matrices. On the AdS side it is not clear exactly how to define the moduli space. As described in section 2.1, there is a background of string theory corresponding to any point in the field theory moduli space, but it is not clear how to see that this is the exact moduli space on the string theory side (especially since high curvatures arise for generic points in the moduli space).

- The qualitative behavior of the theory upon deformations by relevant or marginal operators also does not depend on the coupling (at least for chiral operators whose dimension does not depend on the coupling, and in the absence of phase transitions).

  There are many more qualitative tests of the correspondence, such as the existence of confinement for the finite temperature theory [68], which we will not discuss in this section. We will also not discuss here tests involving the behavior of the theory on its moduli space [60, 69, 61].

[1]Unlike most of the other tests described here, this test actually tests the finite N duality and not just the large N limit.

3.1 The spectrum of chiral primary operators

3.1.1 The field theory spectrum

The $\mathcal{N} = 4$ supersymmetry algebra in $d = 4$ has four generators Q^A_α (and their complex conjugates $\bar{Q}_{\dot{\alpha}A}$), where α is a Weyl-spinor index (in the **2** of the $SO(3,1)$ Lorentz group) and A is an index in the **4** of the $SU(4)_R$ R-symmetry group (lower indices A will be taken to transform in the $\bar{\mathbf{4}}$ representation). They obey the algebra

$$\begin{aligned} \{Q^A_\alpha, \bar{Q}_{\dot{\alpha}B}\} &= 2(\sigma^\mu)_{\alpha\dot{\alpha}} P_\mu \delta^A_B, \\ \{Q^A_\alpha, Q^B_\beta\} &= \{\bar{Q}_{\dot{\alpha}A}, \bar{Q}_{\dot{\beta}B}\} = 0, \end{aligned} \tag{21}$$

where σ^i $(i = 1, 2, 3)$ are the Pauli matrices and $(\sigma^0)_{\alpha\dot{\alpha}} = -\delta_{\alpha\dot{\alpha}}$ (we use the conventions of Wess and Bagger [70]).

$\mathcal{N} = 4$ supersymmetry in four dimensions has a unique multiplet which does not include spins greater than one, which is the vector multiplet. It includes a vector field A_μ (μ is a vector index of the $SO(3,1)$ Lorentz group), four complex Weyl fermions $\lambda_{\alpha A}$ (in the $\bar{\mathbf{4}}$ of $SU(4)_R$), and six real scalars ϕ^I (where I is an index in the **6** of $SU(4)_R$). The classical action of the supersymmetry generators on these fields is schematically given (for on-shell fields) by

$$\begin{aligned} [Q^A_\alpha, \phi^I] &\sim \lambda_{\alpha B}, \\ \{Q^A_\alpha, \lambda_{\beta B}\} &\sim (\sigma^{\mu\nu})_{\alpha\beta} F_{\mu\nu} + \epsilon_{\alpha\beta}[\phi^I, \phi^J], \\ \{Q^A_\alpha, \bar{\lambda}^B_{\dot{\beta}}\} &\sim (\sigma^\mu)_{\alpha\dot{\beta}} \mathcal{D}_\mu \phi^I, \\ [Q^A_\alpha, A_\mu] &\sim (\sigma_\mu)_{\alpha\dot{\alpha}} \bar{\lambda}^A_{\dot{\beta}} \epsilon^{\dot{\alpha}\dot{\beta}}, \end{aligned} \tag{22}$$

with similar expressions for the action of the $\bar{Q}$'s, where $\sigma^{\mu\nu}$ are the generators of the Lorentz group in the spinor representation, $\mathcal{D}_\mu$ is the covariant derivative, the field strength $F_{\mu\nu} \equiv [\mathcal{D}_\mu, \mathcal{D}_\nu]$, and we have suppressed the $SU(4)$ Clebsch-Gordan coefficients corresponding to the products $\mathbf{4} \times \mathbf{6} \to \bar{\mathbf{4}}$, $\mathbf{4} \times \bar{\mathbf{4}} \to \mathbf{1} + \mathbf{15}$ and $\mathbf{4} \times \mathbf{4} \to \mathbf{6}$ in the first three lines of (22).

An $\mathcal{N} = 4$ supersymmetric field theory is uniquely determined by specifying the gauge group, and its field content is a vector multiplet in the adjoint of the gauge group. Such a field theory is equivalent to an $\mathcal{N} = 2$ theory with one hypermultiplet in the adjoint representation, or to an $\mathcal{N} = 1$ theory with three chiral multiplets Φ^i in the adjoint representation (in the $\mathbf{3}_{2/3}$ of the $SU(3) \times U(1)_R \subset SU(4)_R$ which is left unbroken by the choice of a single $\mathcal{N} = 1$ SUSY generator) and a superpotential of the form $W \propto \epsilon_{ijk} \mathrm{Tr}(\Phi^i \Phi^j \Phi^k)$. The interactions of the theory include a scalar potential proportional to $\sum_{I,J} \mathrm{Tr}([\phi^I, \phi^J]^2)$, such that the moduli space of the theory is the space of commuting matrices ϕ^I $(I = 1, \cdots, 6)$.

The spectrum of operators in this theory includes all the gauge invariant quantities that can be formed from the fields described above. In this section we will focus on local operators which involve fields taken at the same point in space-time. For the $SU(N)$ theory described above, properties of the adjoint representation of $SU(N)$ determine that such operators necessarily involve a product of traces of products of fields (or the sum of such products). It is natural to divide the operators into single-trace operators and multiple-trace operators. In the 't Hooft large N limit correlation functions involving multiple-trace operators are suppressed by powers of N compared to those of single-trace operators involving the same fields. We will discuss here in detail only the single-trace operators; the multiple-trace operators appear in operator product expansions of products of single-trace operators.

It is natural to classify the operators in a conformal theory into primary operators and their descendants. In a superconformal theory it is also natural to distinguish between chiral primary operators, which are in short representations of the superconformal algebra and are annihilated by some of the supercharges, and non-chiral primary operators. Representations of the superconformal algebra are formed by starting with some state of lowest dimension, which is annihilated by the operators S and K_μ, and acting on it with the operators Q and P_μ. The $\mathcal{N} = 4$ supersymmetry algebra involves 16 real supercharges. A generic primary representation of the superconformal algebra will thus include 2^{16} primaries of the conformal algebra, generated by acting on the lowest state with products of different supercharges; acting with additional supercharges always leads to descendants of the conformal algebra (i.e. derivatives). Since the supercharges have helicities $\pm 1/2$, the primary fields in such representations will have a range of helicities between $\lambda - 4$ (if the lowest dimension operator ψ has helicity λ) and $\lambda + 4$ (acting with more than 8 supercharges of the same helicity either annihilates the state or leads to a conformal descendant). In non-generic representations of the superconformal algebra a product of less than 16 different Q's annihilates the lowest dimension operator, and the range of helicities appearing is smaller. In particular, in the small representations of the $\mathcal{N} = 4$ superconformal algebra only up to 4 Q's of the same helicity acting on the lowest dimension operator give a non-zero result, and the range of helicities is between $\lambda - 2$ and $\lambda + 2$. For the $\mathcal{N} = 4$ supersymmetry algebra (not including the conformal algebra) it is known that medium representations, whose range of helicities is 6, can also exist (they arise, for instance, on the moduli space of the $SU(N)$ $\mathcal{N} = 4$ SYM theory [71, 72, 73, 74, 75, 76, 77, 78]); it is not clear if such medium representations of the superconformal algebra [79] can appear in physical theories or not (there are no known examples). More details on the structure of representations of the $\mathcal{N}=4$ superconformal algebra may be found in [80, 81, 82, 83, 84, 85, 79] and references therein.

In the $U(1)$ $\mathcal{N}=4$ SYM theory (which is a free theory), the only gauge-invariant "single trace" operators are the fields of the vector multiplet itself (which are $\phi^I, \lambda_A, \bar{\lambda}^A$ and $F_{\mu\nu} = \partial_{[\mu}A_{\nu]}$). These operators form an ultra-short representation of the $\mathcal{N}=4$ algebra whose range of helicities is from (-1) to 1 (acting with more than two supercharges of the same helicity on any of these states gives either zero or derivatives, which are descendants of the conformal algebra). All other local gauge invariant operators in the theory involve derivatives or products of these operators. This representation is usually called the doubleton representation, and it does not appear in the $SU(N)$ SYM theory (though the representations which do appear can all be formed by tensor products of the doubleton representation). In the context of AdS space one can think of this multiplet as living purely on the boundary of the space [86, 87, 88, 89, 90, 46, 91, 92, 93, 94, 95], as expected for the $U(1)$ part of the original $U(N)$ gauge group of the D3-branes (see the discussion in section 2).

There is no known simple systematic way to compute the full spectrum of chiral primary operators of the $\mathcal{N}=4$ $SU(N)$ SYM theory, so we will settle for presenting the known chiral primary operators. The lowest component of a superconformal-primary multiplet is characterized by the fact that it cannot be written as a supercharge Q acting on any other operator. Looking at the action of the supersymmetry charges (22) suggests that generally operators built from the fermions and the gauge fields will be descendants (given by Q acting on some other fields), so one would expect the lowest components of the chiral primary representations to be built only from the scalar fields, and this turns out to be correct.

Let us analyze the behavior of operators of the form $\mathcal{O}^{I_1 I_2 \cdots I_n} \equiv \mathrm{Tr}(\phi^{I_1}\phi^{I_2}\cdots\phi^{I_n})$. First we can ask if this operator can be written as $\{Q, \psi\}$ for any field ψ. In the SUSY algebra (22) only commutators of ϕ^I's appear on the right-hand side, so we see that if some of the indices are antisymmetric the field will be a descendant. Thus, only symmetric combinations of the indices will be lowest components of primary multiplets. Next, we should ask if the multiplet built on such an operator is a (short) chiral primary multiplet or not. There are several different ways to answer this question. One possibility is to use the relation between the dimension of chiral primary operators and their R-symmetry representation [96, 97, 98, 99, 100], and to check if this relation is obeyed in the free field theory, where $[\mathcal{O}^{I_1 I_2 \cdots I_n}] = n$. In this way we find that the representation is chiral primary if and only if the indices form a symmetric traceless product of n **6**'s (traceless representations are defined as those who give zero when any two indices are contracted). This is a representation of weight $(0,n,0)$ of $SU(4)_R$; in this section we will refer to $SU(4)_R$ representations either by their dimensions in boldface or by their weights.

Another way to check this is to see if by acting with Q's on these operators we get the most general possible states or not, namely if the representation contains "null vectors" or not (it turns out that in all the relevant cases "null vectors" appear already at the first level by acting with a single Q, though in principle there could be representations where "null vectors" appear only at higher levels). Using the SUSY algebra (22) it is easy to see that for symmetric traceless representations we get "null vectors" while for other representations we do not. For instance, let us analyze in detail the case $n=2$. The symmetric product of two $\mathbf{6}$'s is given by $\mathbf{6}\times\mathbf{6}\rightarrow\mathbf{1}+\mathbf{20'}$. The field in the $\mathbf{1}$ representation is $\text{Tr}(\phi^I\phi^I)$, for which $[Q^A_\alpha,\text{Tr}(\phi^I\phi^I)]\sim C^{AJB}\text{Tr}(\lambda_{\alpha B}\phi^J)$ where C^{AIB} is a Clebsch-Gordan coefficient for $\bar{\mathbf{4}}\times\mathbf{6}\rightarrow\mathbf{4}$. The right-hand side is in the $\mathbf{4}$ representation, which is the most general representation that can appear in the product $\mathbf{4}\times\mathbf{1}$, so we find no null vectors at this level. On the other hand, if we look at the symmetric traceless product $\text{Tr}(\phi^{\{I}\phi^{J\}})\equiv\text{Tr}(\phi^I\phi^J)-\frac{1}{6}\delta^{IJ}\text{Tr}(\phi^K\phi^K)$ in the $\mathbf{20'}$ representation, we find that $\{Q^A_\alpha,\text{Tr}(\phi^{\{I}\phi^{J\}})\}\sim\text{Tr}(\lambda_{\alpha B}\phi^K)$ with the right-hand side being in the $\mathbf{20}$ representation (appearing in $\bar{\mathbf{4}}\times\mathbf{6}\rightarrow\mathbf{4}+\mathbf{20}$), while the left-hand side could in principle be in the $\mathbf{4}\times\mathbf{20'}\rightarrow\mathbf{20}+\mathbf{60}$. Since the $\mathbf{60}$ does not appear on the right-hand side (it is a "null vector") we identify that the representation built on the $\mathbf{20'}$ is a short representation of the SUSY algebra. By similar manipulations (see [24, 101, 81, 84] for more details) one can verify that chiral primary representations correspond exactly to symmetric traceless products of $\mathbf{6}$'s.

It is possible to analyze the chiral primary spectrum also by using $\mathcal{N}=1$ subalgebras of the $\mathcal{N}=4$ algebra. If we use an $\mathcal{N}=1$ subalgebra of the $\mathcal{N}=4$ algebra, as described above, the operators $\mathcal{O}_n$ include the chiral operators of the form $\text{Tr}(\Phi^{i_1}\Phi^{i_2}\cdots\Phi^{i_n})$ (in a representation of $SU(3)$ which is a symmetric product of $\mathbf{3}$'s), but for a particular choice of the $\mathcal{N}=1$ subalgebra not all the operators $\mathcal{O}_n$ appear to be chiral (a short multiplet of the $\mathcal{N}=4$ algebra includes both short and long multiplets of the $\mathcal{N}=1$ subalgebra).

The last issue we should discuss is what is the range of values of n. The product of more than N commuting[2] $N\times N$ matrices can always be written as a sum of products of traces of less than N of the matrices, so it does not

[2] We can limit the discussion to commuting matrices since, as discussed above, commutators always lead to descendants, and we can write any product of matrices as a product of commuting matrices plus terms with commutators.

form an independent operator. This means that for $n > N$ we can express the operator $\mathcal{O}^{I_1 I_2 \cdots I_n}$ in terms of other operators, up to operators including commutators which (as explained above) are descendants of the SUSY algebra. Thus, we find that the short chiral primary representations are built on the operators $\mathcal{O}_n = \mathcal{O}^{\{I_1 I_2 \cdots I_n\}}$ with $n = 2, 3, \cdots, N$, for which the indices are in the symmetric traceless product of n **6**'s (in a $U(N)$ theory we would find the same spectrum with the additional representation corresponding to $n = 1$). The superconformal algebra determines the dimension of these fields to be $[\mathcal{O}_n] = n$, which is the same as their value in the free field theory. We argued above that these are the only short chiral primary representations in the $SU(N)$ gauge theory, but we will not attempt to rigorously prove this here.

The full chiral primary representations are obtained by acting on the fields $\mathcal{O}_n$ by the generators Q and P of the supersymmetry algebra. The representation built on $\mathcal{O}_n$ contains a total of $256 \times \frac{1}{12} n^2 (n^2 - 1)$ primary states, of which half are bosonic and half are fermionic. Since these multiplets are built on a field of helicity zero, they will contain primary fields of helicities between (-2) and 2. The highest dimension primary field in the multiplet is (generically) of the form $Q^4 \bar{Q}^4 \mathcal{O}_n$, and its dimension is $n + 4$. There is an elegant way to write these multiplets as traces of products of "twisted chiral $\mathcal{N} = 4$ superfields" [101, 81]; see also [102] which checks some components of these superfields against the couplings to supergravity modes predicted on the basis of the DBI action for D3-branes in anti-de Sitter space [4].

It is easy to find the form of all the fields in such a multiplet by using the algebra (22). For example, let us analyze here in detail the bosonic primary fields of dimension $n + 1$ in the multiplet. To get a field of dimension $n + 1$ we need to act on $\mathcal{O}_n$ with two supercharges (recall that $[Q] = \frac{1}{2}$). If we act with two supercharges Q^A_α of the same chirality, their Lorentz indices can be either antisymmetrized or symmetrized. In the first case we get a Lorentz scalar field in the $(2, n-2, 0)$ representation of $SU(4)_R$, which is of the schematic form

$$\epsilon^{\alpha\beta}\{Q_\alpha, [Q_\beta, \mathcal{O}_n]\} \sim \epsilon^{\alpha\beta}\mathrm{Tr}(\lambda_{\alpha A}\lambda_{\beta B}\phi^{J_1}\cdots\phi^{J_{n-2}}) + \mathrm{Tr}([\phi^{K_1}, \phi^{K_2}]\phi^{L_1}\cdots\phi^{L_{n-1}}). \tag{23}$$

Using an $\mathcal{N} = 1$ subalgebra some of these operators may be written as the lowest components of the chiral superfields $\mathrm{Tr}(W_\alpha^2 \Phi^{j_1} \cdots \Phi^{j_{n-2}})$. In the second case we get an anti-symmetric 2-form of the Lorentz group, in the $(0, n-1, 0)$ representation of $SU(4)_R$, of the form

$$\{Q_{\{\alpha}, [Q_{\beta\}}, \mathcal{O}_n]\} \sim \mathrm{Tr}((\sigma^{\mu\nu})_{\alpha\beta}F_{\mu\nu}\phi^{J_1}\cdots\phi^{J_{n-1}}) + \mathrm{Tr}(\lambda_{\alpha A}\lambda_{\beta B}\phi^{K_1}\cdots\phi^{K_{n-2}}). \tag{24}$$

Both of these fields are complex, with the complex conjugate fields given by the action of two $\bar{Q}$'s. Acting with one Q and one $\bar{Q}$ on the state $\mathcal{O}_n$ gives a (real) Lorentz-vector field in the $(1, n-2, 1)$ representation of $SU(4)_R$, of the form

$$\{Q_\alpha, [\bar{Q}_{\dot{\alpha}}, \mathcal{O}_n]\} \sim \text{Tr}(\lambda_{\alpha A} \bar{\lambda}^B_{\dot{\alpha}} \phi^{J_1} \cdots \phi^{J_{n-2}}) + (\sigma^\mu)_{\alpha\dot{\alpha}} \text{Tr}((\mathcal{D}_\mu \phi^J) \phi^{K_1} \cdots \phi^{K_{n-1}}). \tag{25}$$

At dimension $n+2$ (acting with four supercharges) we find :

- A complex scalar field in the $(0, n-2, 0)$ representation, given by $Q^4 \mathcal{O}_n$, of the form $\text{Tr}(F^2_{\mu\nu} \phi^{I_1} \cdots \phi^{I_{n-2}}) + \cdots$.

- A real scalar field in the $(2, n-4, 2)$ representation, given by $Q^2 \bar{Q}^2 \mathcal{O}_n$, of the form $\epsilon^{\alpha\beta} \epsilon^{\dot{\alpha}\dot{\beta}} \text{Tr}(\lambda_{\alpha A_1} \lambda_{\beta A_2} \bar{\lambda}^{B_1}_{\dot{\alpha}} \bar{\lambda}^{B_2}_{\dot{\beta}} \phi^{I_1} \cdots \phi^{I_{n-4}}) + \cdots$.

- A complex vector field in the $(1, n-4, 1)$ representation, given by $Q^3 \bar{Q} \mathcal{O}_n$, of the form $\text{Tr}(F_{\mu\nu} \mathcal{D}^\nu \phi^J \phi^{I_1} \cdots \phi^{I_{n-2}}) + \cdots$.

- An complex anti-symmetric 2-form field in the $(2, n-3, 0)$ representation, given by $Q^2 \bar{Q}^2 \mathcal{O}_n$, of the form $\text{Tr}(F_{\mu\nu} [\phi^{J_1}, \phi^{J_2}] \phi^{I_1} \cdots \phi^{I_{n-2}}) + \cdots$.

- A symmetric tensor field in the $(0, n-2, 0)$ representation, given by $Q^2 \bar{Q}^2 \mathcal{O}_n$, of the form $\text{Tr}(\mathcal{D}_{\{\mu} \phi^J \mathcal{D}_{\nu\}} \phi^K \phi^{I_1} \cdots \phi^{I_{n-2}}) + \cdots$.

The spectrum of primary fields at dimension $n+3$ is similar to that of dimension $n+1$ (the same fields appear but in smaller $SU(4)_R$ representations), and at dimension $n+4$ there is a single primary field, which is a real scalar in the $(0, n-4, 0)$ representation, given by $Q^4 \bar{Q}^4 \mathcal{O}_n$, of the form $\text{Tr}(F^4_{\mu\nu} \phi^{I_1} \cdots \phi^{I_{n-4}}) + \cdots$. Note that fields with more than four $F_{\mu\nu}$'s or more than eight λ's are always descendants or non-chiral primaries.

For $n = 2, 3$ the short multiplets are even shorter since some of the representations appearing above vanish. In particular, for $n = 2$ the highest-dimension primaries in the chiral primary multiplet have dimension $n+2 = 4$. The $n = 2$ representation includes the currents of the superconformal algebra. It includes a vector of dimension 3 in the **15** representation which is the $SU(4)_R$ R-symmetry current, and a symmetric tensor field of dimension 4 which is the energy-momentum tensor (the other currents of the superconformal algebra are descendants of these). The $n = 2$ multiplet also includes a complex scalar field which is an $SU(4)_R$-singlet, whose real part is the Lagrangian density coupling to $\frac{1}{4g^2_{YM}}$ (of the form $\text{Tr}(F^2_{\mu\nu}) + \cdots$) and whose imaginary part is the Lagrangian density coupling to θ (of the form $\text{Tr}(F \wedge F)$). For later use we note that the chiral primary multiplets which contain scalars of dimension $\Delta \leq 4$ are the $n = 2$ multiplet (which has a scalar in the $\mathbf{20'}$ of dimension 2, a complex scalar in the **10** of dimension 3, and a complex scalar in the **1** of dimension 4), the $n = 3$ multiplet (which contains a scalar in the **50** of dimension 3 and a complex scalar in the **45** of dimension 4), and the $n = 4$ multiplet which contains a scalar in the **105** of dimension 4.

3.1.2 The string theory spectrum and the matching

As discussed in section 2.2, fields on AdS_5 are in a one-to-one correspondence with operators in the dual conformal field theory. Thus, the spectrum of operators described in section 3.1.1 should agree with the spectrum of fields of type IIB string theory on $AdS_5 \times S^5$. Fields on AdS naturally lie in the same multiplets of the conformal group as primary operators; the second Casimir of these representations is $C_2 = \Delta(\Delta - 4)$ for a primary scalar field of dimension Δ in the field theory, and $C_2 = m^2 R^2$ for a field of mass m on an AdS_5 space with a radius of curvature R. Single-trace operators in the field theory may be identified with single-particle states in AdS_5, while multiple-trace operators correspond to multi-particle states.

Unfortunately, it is not known how to compute the full spectrum of type IIB string theory on $AdS_5 \times S^5$. In fact, the only known states are the states which arise from the dimensional reduction of the ten-dimensional type IIB supergravity multiplet. These fields all have helicities between (-2) and 2, so it is clear that they all lie in small multiplets of the superconformal algebra, and we will describe below how they match with the small multiplets of the field theory described above. String theory on $AdS_5 \times S^5$ is expected to have many additional states, with masses of the order of the string scale $1/l_s$ or of the Planck scale $1/l_p$. Such states would correspond (using the mass/dimension relation described above) to operators in the field theory with dimensions of order $\Delta \sim (g_s N)^{1/4}$ or $\Delta \sim N^{1/4}$ for large $N, g_s N$. Presumably none of these states are in small multiplets of the superconformal algebra (at least, this would be the prediction of the AdS/CFT correspondence).

The spectrum of type IIB supergravity compactified on $AdS_5 \times S^5$ was computed in [103]. The computation involves expanding the ten dimensional fields in appropriate spherical harmonics on S^5, plugging them into the supergravity equations of motion, linearized around the $AdS_5 \times S^5$ background, and diagonalizing the equations to give equations of motion for free (massless or massive) fields[3]. For example, the ten dimensional dilaton

[3]The fields arising from different spherical harmonics are related by a "spectrum generating algebra", see [104].

field τ may be expanded as $\tau(x,y) = \sum_{k=0}^{\infty} \tau^k(x) Y^k(y)$ where x is a coordinate on AdS_5, y is a coordinate on S^5, and the Y^k are the scalar spherical harmonics on S^5. These spherical harmonics are in representations corresponding to symmetric traceless products of **6**'s of $SU(4)_R$; they may be written as $Y^k(y) \sim y^{I_1} y^{I_2} \cdots y^{I_k}$ where the y^I, for $I = 1, 2, \cdots, 6$ and with $\sum_{I=1}^{6} (y^I)^2 = 1$, are coordinates on S^5. Thus, we find a field $\tau^k(x)$ on AdS_5 in each such $(0,k,0)$ representation of $SU(4)_R$, and the equations of motion determine the mass of this field to be $m_k^2 = k(k+4)/R^2$. A similar expansion may be performed for all other fields.

If we organize the results of [103] into representations of the superconformal algebra [80], we find representations of the form described in the previous section, which are built on a lowest dimension field which is a scalar in the $(0,n,0)$ representation of $SU(4)_R$ for $n = 2, 3, \cdots, \infty$. The lowest dimension scalar field in each representation turns out to arise from a linear combination of spherical harmonic modes of the S^5 components of the graviton h_a^a (expanded around the $AdS_5 \times S^5$ vacuum) and the 4-form field D_{abcd}, where a, b, c, d are indices on S^5. The scalar fields of dimension $n+1$ correspond to 2-form fields B_{ab} with indices in the S^5. The symmetric tensor fields arise from the expansion of the AdS_5-components of the graviton. The dilaton fields described above are the complex scalar fields arising with dimension $n+2$ in the multiplet (as described in the previous subsection).

In particular, the $n = 2$ representation is called the supergraviton representation, and it includes the field content of $d = 5, \mathcal{N} = 8$ gauged supergravity. The field/operator correspondence matches this representation to the representation including the superconformal currents in the field theory. It includes a massless graviton field, which (as expected) corresponds to the energy-momentum tensor in the field theory, and massless $SU(4)_R$ gauge fields which correspond to (or couple to) the global $SU(4)_R$ currents in the field theory.

In the naive dimensional reduction of the type IIB supergravity fields, the $n = 1$ doubleton representation, corresponding to a free $U(1)$ vector multiplet in the dual theory, also appears. However, the modes of this multiplet are all pure gauge modes in the bulk of AdS_5, and they may be set to zero there. This is one of the reasons why it seems more natural to view the corresponding gauge theory as an $SU(N)$ gauge theory and not a $U(N)$ theory. It may be possible (and perhaps even natural) to add the doubleton representation to the theory (even though it does not include modes which propagate in the bulk of AdS_5, but instead it is equivalent to a topological theory in the bulk) to obtain a theory which is dual to the $U(N)$ gauge theory, but this will not affect most of our discussion in this review so we will ignore this possibility here.

Comparing the results described above with the results of section 3.1.1, we see that we find the same spectrum of chiral primary operators for $n = 2, 3, \cdots, N$. The supergravity results cannot be trusted for masses above the order of the string scale (which corresponds to $n \sim (g_s N)^{1/4}$) or the Planck scale (which corresponds to $n \sim N^{1/4}$), so the results agree within their range of validity. The field theory results suggest that the exact spectrum of chiral representations in type IIB string theory on $AdS_5 \times S^5$ actually matches the naive supergravity spectrum up to a mass scale $m^2 \sim N^2/R^2 \sim N^{3/2} M_p^2$ which is much higher than the string scale and the Planck scale, and that there are no chiral fields above this scale. It is not known how to check this prediction; tree-level string theory is certainly not enough for this since when $g_s = 0$ we must take $N = \infty$ to obtain a finite value of $g_s N$. Thus, with our current knowledge the matching of chiral primaries of the $\mathcal{N} = 4$ SYM theory with those of string theory on $AdS_5 \times S^5$ tests the duality only in the large N limit. In some generalizations of the AdS/CFT correspondence the string coupling goes to zero at the boundary even for finite N, and then classical string theory should lead to exactly the same spectrum of chiral operators as the field theory. This happens in particular for the near-horizon limit of NS5-branes, in which case the exact spectrum was successfully compared in [105]. In other instances of the AdS/CFT correspondence (such as the ones discussed in [106, 107, 108]) there exist also additional chiral primary multiplets with n of order N, and these have been successfully matched with wrapped branes on the string theory side.

The fact that there seem to be no non-chiral fields on AdS_5 with a mass below the string scale suggests that for large N and large $g_s N$, the dimension of all non-chiral operators in the field theory, such as $\mathrm{Tr}(\phi^I \phi^I)$, grows at least as $(g_s N)^{1/4} \sim (g_{YM}^2 N)^{1/4}$. The reason for this behavior on the field theory side is not clear; it is a prediction of the AdS/CFT correspondence.

3.2 Matching of correlation functions and anomalies

The classical $\mathcal{N} = 4$ theory has a scale invariance symmetry and an $SU(4)_R$ R-symmetry, and (unlike many other theories) these symmetries are exact also in the full quantum theory. However, when the theory is coupled to external gravitational or $SU(4)_R$ gauge fields, these symmetries are broken by quantum effects. In field theory this breaking comes from one-loop diagrams and does not receive any further corrections; thus it can be computed also in the strong coupling regime and compared with the results from string theory on AdS space.

We will begin by discussing the anomaly associated with the $SU(4)_R$ global currents. These currents are chiral since the fermions $\lambda_{\alpha A}$ are in the

$\bar{\mathbf{4}}$ representation while the fermions of the opposite chirality $\bar{\lambda}^A_{\dot{\alpha}}$ are in the $\mathbf{4}$ representation. Thus, if we gauge the $SU(4)_R$ global symmetry, we will find an Adler-Bell-Jackiw anomaly from the triangle diagram of three $SU(4)_R$ currents, which is proportional to the number of charged fermions. In the $SU(N)$ gauge theory this number is N^2-1. The anomaly can be expressed either in terms of the 3-point function of the $SU(4)_R$ global currents,

$$\left\langle J^a_\mu(x) J^b_\nu(y) J^c_\rho(z)\right\rangle_- = -\frac{N^2-1}{32\pi^6} i d^{abc} \frac{\mathrm{Tr}\left[\gamma_5\gamma_\mu(\not{x}-\not{y})\gamma_\nu(\not{y}-\not{z})\gamma_\rho(\not{z}-\not{x})\right]}{(x-y)^4(y-z)^4(z-x)^4}, \tag{26}$$

where $d^{abc}=2\mathrm{Tr}(T^a\{T^b,T^c\})$ and we take only the negative parity component of the correlator, or in terms of the non-conservation of the $SU(4)_R$ current when the theory is coupled to external $SU(4)_R$ gauge fields $F^a_{\mu\nu}$,

$$(\mathcal{D}^\mu J_\mu)^a = \frac{N^2-1}{384\pi^2} i d^{abc}\epsilon^{\mu\nu\rho\sigma}F^b_{\mu\nu}F^c_{\rho\sigma}. \tag{27}$$

How can we see this effect in string theory on $AdS_5\times S^5$? One way to see it is, of course, to use the general prescription of section 4 to compute the 3-point function (26), and indeed one finds [109, 110] the correct answer to leading order in the large N limit (namely, one recovers the term proportional to N^2). It is more illuminating, however, to consider directly the meaning of the anomaly (27) from the point of view of the AdS theory [24]. In the AdS theory we have gauge fields A^a_μ which couple, as explained above, to the $SU(4)_R$ global currents J^a_μ of the gauge theory, but the anomaly means that when we turn on non-zero field strengths for these fields the theory should no longer be gauge invariant. This effect is precisely reproduced by a Chern-Simons term which exists in the low-energy supergravity theory arising from the compactification of type IIB supergravity on $AdS_5\times S^5$, which is of the form

$$\frac{iN^2}{96\pi^2}\int_{AdS_5} d^5x(d^{abc}\epsilon^{\mu\nu\lambda\rho\sigma}A^a_\mu\partial_\nu A^b_\lambda\partial_\rho A^c_\sigma+\cdots). \tag{28}$$

This term is gauge invariant up to total derivatives, which means that if we take a gauge transformation $A^a_\mu\to A^a_\mu+(\mathcal{D}_\mu\Lambda)^a$ for which Λ does not vanish on the boundary of AdS_5, the action will change by a boundary term of the form

$$-\frac{iN^2}{384\pi^2}\int_{\partial AdS_5} d^4x\epsilon^{\mu\nu\rho\sigma}d^{abc}\Lambda^a F^b_{\mu\nu}F^c_{\rho\sigma}. \tag{29}$$

From this we can read off the anomaly in $(\mathcal{D}^\mu J_\mu)$ since, when we have a coupling of the form $\int d^4x A^\mu_a J^a_\mu$, the change in the action under a gauge transformation is given by $\int d^4x(\mathcal{D}^\mu\Lambda)_a J^a_\mu = -\int d^4x\Lambda_a(\mathcal{D}^\mu J^a_\mu)$, and we find exact agreement with (27) for large N.

The other anomaly in the $\mathcal{N}=4$ SYM theory is the conformal (or Weyl) anomaly (see [111, 112] and references therein), indicating the breakdown of conformal invariance when the theory is coupled to a curved external metric (there is a similar breakdown of conformal invariance when the theory is coupled to external $SU(4)_R$ gauge fields, which we will not discuss here). The conformal anomaly is related to the 2-point and 3-point functions of the energy-momentum tensor [113, 114, 115, 116]. In four dimensions, the general form of the conformal anomaly is

$$\langle g^{\mu\nu} T_{\mu\nu} \rangle = -aE_4 - cI_4, \tag{30}$$

where

$$\begin{aligned} E_4 &= \frac{1}{16\pi^2}(\mathcal{R}^2_{\mu\nu\rho\sigma} - 4\mathcal{R}^2_{\mu\nu} + \mathcal{R}^2), \\ I_4 &= -\frac{1}{16\pi^2}(\mathcal{R}^2_{\mu\nu\rho\sigma} - 2\mathcal{R}^2_{\mu\nu} + \frac{1}{3}\mathcal{R}^2), \end{aligned} \tag{31}$$

where $\mathcal{R}_{\mu\nu\rho\sigma}$ is the curvature tensor, $\mathcal{R}_{\mu\nu} \equiv \mathcal{R}^\rho_{\mu\rho\nu}$ is the Riemann tensor, and $\mathcal{R} \equiv \mathcal{R}^\mu_\mu$ is the scalar curvature. A free field computation in the $SU(N)$ $\mathcal{N}=4$ SYM theory leads to $a = c = (N^2-1)/4$. In supersymmetric theories the supersymmetry algebra relates $g^{\mu\nu}T_{\mu\nu}$ to derivatives of the R-symmetry current, so it is protected from any quantum corrections. Thus, the same result should be obtained in type IIB string theory on $AdS_5 \times S^5$, and to leading order in the large N limit it should be obtained from type IIB supergravity on $AdS_5 \times S^5$. This was indeed found to be true in [117, 118, 119, 120][4], where the conformal anomaly was shown to arise from subtleties in the regularization of the (divergent) supergravity action on AdS space. The result of [117, 118, 119, 120] implies that a computation using gravity on AdS_5 always gives rise to theories with $a = c$, so generalizations of the AdS/CFT correspondence which have (for large N) a supergravity approximation are limited to conformal theories which have $a = c$ in the large N limit. Of course, if we do not require the string theory to have a supergravity approximation then there is no such restriction.

For both of the anomalies we described the field theory and string theory computations agree for the leading terms, which are of order N^2. Thus, they are successful tests of the duality in the large N limit. For other instances of the AdS/CFT correspondence there are corrections to anomalies at order $1/N \sim g_s(\alpha'/R^2)^2$; such corrections were discussed in [122] and successfully compared in [123, 124, 125][5]. It would be interesting to compare other

[4] A generalization with more varying fields may be found in [121].

[5] Computing such corrections tests the conjecture that the correspondence holds order by order in $1/N$; however, this is weaker than the statement that the correspondence holds for finite N, since the $1/N$ expansion is not expected to converge.

corrections to the large N result.

4 Correlation functions

A useful statement of the AdS/CFT correspondence is that the partition function of string theory on $AdS_5 \times S^5$ should coincide with the partition function of $\mathcal{N} = 4$ super-Yang-Mills theory "on the boundary" of AdS_5 [23, 24]. The basic idea was explained in section 2.2, but before summarizing the actual calculations of Green's functions, it seems worthwhile to motivate the methodology from a somewhat different perspective.

Throughout this section, we approximate the string theory partition function by $e^{-I_{SUGRA}}$, where I_{SUGRA} is the supergravity action evaluated on $AdS_5 \times S^5$ (or on small deformations of this space). This approximation amounts to ignoring all the stringy α' corrections that cure the divergences of supergravity, and also all the loop corrections, which are controlled essentially by the gravitational coupling $\kappa \sim g_{st}\alpha'^2$. On the gauge theory side, as explained in section 2.2, this approximation amounts to taking both N and $g_{YM}^2 N$ large, and the basic relation becomes

$$e^{-I_{SUGRA}} \simeq Z_{\text{string}} = Z_{\text{gauge}} = e^{-W} \ , \tag{32}$$

where W is the generating functional for connected Green's functions in the gauge theory. At finite temperature, $W = \beta F$ where β is the inverse temperature and F is the free energy of the gauge theory. When we apply this relation to a Schwarzschild black hole in AdS_5, which is thought to be reflected in the gauge theory by a thermal state at the Hawking temperature of the black hole, we arrive at the relation $I_{SUGRA} \simeq \beta F$. Calculating the free energy of a black hole from the Euclidean supergravity action has a long tradition in the supergravity literature [126], so the main claim that is being made here is that the dual gauge theory provides a description of the state of the black hole which is physically equivalent to the one in string theory. We will discuss the finite temperature case further in section 6, and devote the rest of this section to the partition function of the field theory on $\mathbb{R}^4$.

The main technical idea behind the bulk-boundary correspondence is that the boundary values of string theory fields (in particular, supergravity fields) act as sources for gauge-invariant operators in the field theory. From a D-brane perspective, we think of closed string states in the bulk as sourcing gauge singlet operators on the brane which originate as composite operators built from open strings. We will write the bulk fields generically as $\phi(\vec{x}, z)$ (in the coordinate system (17)), with value $\phi_0(\vec{x})$ for $z = \epsilon$. The true boundary of anti-de Sitter space is $z = 0$, and $\epsilon \neq 0$ serves as a cutoff which will eventually be removed. In the supergravity approximation, we think of choosing the values ϕ_0 arbitrarily and then extremizing the action $I_{SUGRA}[\phi]$ in the region $z > \epsilon$ subject to these boundary conditions. In short, we solve the equations of motion in the bulk subject to Dirichlet boundary conditions on the boundary, and evaluate the action on the solution. If there is more than one solution, then we have more than one saddle point contributing to the string theory partition function, and we must determine which is most important. In this section, multiple saddle points will not be a problem. So, we can write

$$W_{\rm gauge}[\phi_0] = -\log \left\langle e^{\int d^4x\, \phi_0(x)\mathcal{O}(x)} \right\rangle_{CFT} \simeq \underset{\phi|_{z=\epsilon}=\phi_0}{\rm extremum}\, I_{SUGRA}[\phi]\ . \qquad (33)$$

That is, the generator of connected Green's functions in the gauge theory, in the large $N, g^2_{YM}N$ limit, is the on-shell supergravity action.

Note that in (33) we have not attempted to be prescient about inserting factors of ϵ. Instead our strategy will be to use (33) without modification to compute two-point functions of $\mathcal{O}$, and then perform a wave-function renormalization on either $\mathcal{O}$ or ϕ so that the final answer is independent of the cutoff. This approach should be workable even in a space (with boundary) which is not asymptotically anti-de Sitter, corresponding to a field theory which does not have a conformal fixed point in the ultraviolet.

A remark is in order regarding the relation of (33) to the old approach of extracting Green's functions from an absorption cross-section [16]. In absorption calculations one is keeping the whole D3-brane geometry, not just the near-horizon $AdS_5 \times S^5$ throat. The usual treatment is to split the space into a near region (the throat) and a far region. The incoming wave from asymptotically flat infinity can be regarded as fixing the value of a supergravity field at the outer boundary of the near region. As usual, the supergravity description is valid at large N and large 't Hooft coupling. At small 't Hooft coupling, there is a different description of the process: a cluster of D3-branes sits at some location in flat ten-dimensional space, and the incoming wave impinges upon it. In the low-energy limit, the value of the supergravity field which the D3-branes feel is the same as the value in the curved space description at the boundary of the near horizon region. Equation (33) is just a mathematical expression of the fact that the throat geometry should respond identically to the perturbed supergravity fields as the low-energy theory on the D3-branes.

Following [23, 24], a number of papers—notably [127, 128, 109, 129, 110, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141]—have undertaken the program of extracting explicit n-point correlation functions of gauge singlet operators by developing both sides of (33) in a power series in ϕ_0.

Because the right-hand side is the extremization of a classical action, the power series has a graphical representation in terms of tree-level Feynman graphs for fields in the supergravity. There is one difference: in ordinary Feynman graphs one assigns the wavefunctions of asymptotic states to the external legs of the graph, but in the present case the external leg factors reflect the boundary values ϕ_0. They are special limits of the usual gravity propagators in the bulk, and are called bulk-to-boundary propagators. We will encounter their explicit form in the next two sections.

4.1 Two-point functions

For two-point functions, only the part of the action which is quadratic in the relevant field perturbation is needed. For massive scalar fields in AdS_5, this has the generic form

$$S = \eta \int d^5x \sqrt{g} \left[\tfrac{1}{2}(\partial\phi)^2 + \tfrac{1}{2}m^2\phi^2 \right] , \tag{34}$$

where η is some normalization which in principle follows from the ten-dimensional origin of the action. The bulk-to-boundary propagator is a particular solution of the equation of motion, $(\Box - m^2)\phi = 0$, which has special asymptotic properties. We will start by considering the momentum space propagator, which is useful for computing the two-point function and also in situations where the bulk geometry loses conformal invariance; then, we will discuss the position space propagator, which has proven more convenient for the study of higher point correlators in the conformal case. We will always work in Euclidean space[6]. A coordinate system in the bulk of AdS_5 such that

$$ds^2 = \frac{R^2}{z^2}\left(d\vec{x}^2 + dz^2 \right) \tag{35}$$

provides manifest Euclidean symmetry on the directions parametrized by $\vec{x}$. To avoid divergences associated with the small z region of integration in (34), we will employ an explicit cutoff, $z \geq \epsilon$.

A complete set of solutions for the linearized equation of motion, $(\Box - m^2)\phi = 0$, is given by $\phi = e^{i\vec{p}\cdot\vec{x}} Z(pz)$, where the function $Z(u)$ satisfies the radial equation

$$\left[u^5 \partial_u \frac{1}{u^3} \partial_u - u^2 - m^2R^2 \right] Z(u) = 0 \ . \tag{36}$$

[6] The results may be analytically continued to give the correlation functions of the field theory on Minkowskian $\mathbb{R}^4$, which corresponds to the Poincaré coordinates of AdS space.

There are two independent solutions to (36), namely $Z(u) = u^2 I_{\Delta-2}(u)$ and $Z(u) = u^2 K_{\Delta-2}(u)$, where I_ν and K_ν are Bessel functions and

$$\Delta = 2 + \sqrt{4 + m^2 R^2} \ . \tag{37}$$

The second solution is selected by the requirement of regularity in the interior: $I_{\Delta-2}(u)$ increases exponentially as $u \to \infty$ and does not lead to a finite action configuration. Imposing the boundary condition $\phi(\vec{x}, z) = \phi_0(\vec{x}) = e^{i\vec{p}\cdot\vec{x}}$ at $z = \epsilon$, we find the bulk-to-boundary propagator

$$\phi(\vec{x}, z) = K_{\vec{p}}(\vec{x}, z) = \frac{(pz)^2 K_{\Delta-2}(pz)}{(p\epsilon)^2 K_{\Delta-2}(p\epsilon)} e^{i\vec{p}\cdot\vec{x}} \ . \tag{38}$$

To compute a two-point function of the operator $\mathcal{O}$ for which ϕ_0 is a source, we write

$$\begin{aligned} \langle \mathcal{O}(\vec{p}) \mathcal{O}(\vec{q}) \rangle &= \left. \frac{\partial^2 W \left[\phi_0 = \lambda_1 e^{i\vec{p}\cdot x} + \lambda_2 e^{i\vec{q}\cdot x} \right]}{\partial\lambda_1 \partial\lambda_2} \right|_{\lambda_1 = \lambda_2 = 0} \\ &= (\text{leading analytic terms in } (\epsilon p)^2) \\ &\quad - \eta \epsilon^{2\Delta-8} (2\Delta - 4) \frac{\Gamma(3-\Delta)}{\Gamma(\Delta-1)} \delta^4(\vec{p} + \vec{q}) \left(\frac{\vec{p}}{2} \right)^{2\Delta-4} \\ &\quad + (\text{higher order terms in } (\epsilon p)^2), \\ \langle \mathcal{O}(\vec{x}) \mathcal{O}(\vec{y}) \rangle &= \eta \epsilon^{2\Delta-8} \frac{2\Delta-4}{\Delta} \frac{\Gamma(\Delta+1)}{\pi^2 \Gamma(\Delta-2)} \frac{1}{|\vec{x} - \vec{y}|^{2\Delta}} \ . \end{aligned} \tag{39}$$

Several explanatory remarks are in order:

- To establish the second equality in (39) we have used (38), substituted in (34), performed the integral and expanded in ϵ. The leading analytic terms give rise to contact terms in position space, and the higher order terms are unimportant in the limit where we remove the cutoff. Only the leading nonanalytic term is essential. We have given the expression for generic real values of Δ. Expanding around integer $\Delta \geq 2$ one obtains finite expressions involving $\log \epsilon p$.

- The Fourier transforms used to obtain the last line are singular, but they can be defined for generic complex Δ by analytic continuation and for positive integer Δ by expanding around a pole and dropping divergent terms, in the spirit of differential regularization [142]. The result is a pure power law dependence on the separation $|\vec{x} - \vec{y}|$, as required by conformal invariance.

- We have assumed a coupling $\int d^4x\, \phi(\vec{x}, z = \epsilon)\mathcal{O}(\vec{x})$ to compute the Green's functions. The explicit powers of the cutoff in the final position space answer can be eliminated by absorbing a factor of $\epsilon^{\Delta-4}$ into the definition of $\mathcal{O}$. From here on we will take that convention, which amounts to inserting a factor of $\epsilon^{4-\Delta}$ on the right-hand side of (38). In fact, precise matchings between the normalizations in field theory and in string theory for all the chiral primary operators have not been worked out. In part this is due to the difficulty of determining the coupling of bulk fields to field theory operators (or in stringy terms, the coupling of closed string states to composite open string operators on the brane). See [15] for an early approach to this problem. For the dilaton, the graviton, and their superpartners (including gauge fields in AdS_5), the couplings can be worked out explicitly. In some of these cases all normalizations have been worked out unambiguously and checked against field theory predictions (see for example [23, 109, 134]).

- The mass-dimension relation (37) holds even for string states that are not included in the Kaluza-Klein supergravity reduction: the mass and the dimension are just different expressions of the second Casimir of $SO(4,2)$. For instance, excited string states, with $m \sim 1/\sqrt{\alpha'}$, are expected to correspond to operators with dimension $\Delta \sim (g_{YM}^2 N)^{1/4}$. The remarkable fact is that all the string theory modes with $m \sim 1/R$ (which is to say, all closed string states which arise from massless ten dimensional fields) fall in short multiplets of the supergroup $SU(2,2|4)$. All other states have a much larger mass. The operators in short multiplets have algebraically protected dimensions. The obvious conclusion is that all operators whose dimensions are not algebraically protected have large dimension in the strong 't Hooft coupling, large N limit to which supergravity applies. This is no longer true for theories of reduced supersymmetry: the supergroup gets smaller, but the Kaluza-Klein states are roughly as numerous as before, and some of them escape the short multiplets and live in long multiplets of the smaller supergroups. They still have a mass on the order of $1/R$, and typically correspond to dimensions which are finite (in the large $g_{YM}^2 N$ limit) but irrational.

Correlation functions of non-scalar operators have been widely studied following [24]; the literature includes [143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153]. For $\mathcal{N} = 4$ super-Yang-Mills theory, all correlation functions of fields in chiral multiplets should follow by application of supersymmetries once those of the chiral primary fields are known, so in this case it should be enough to study the scalars. It is worthwhile to note however that the mass-dimension formula changes for particles with spin. In fact the definition of mass has some convention-dependence. Conventions seem fairly uniform in the literature, and a table of mass-dimension relations in AdS_{d+1} with unit radius was made in [154] from the various sources cited above (see also [101]):

- scalars: $\Delta_{\pm} = \frac{1}{2}(d \pm \sqrt{d^2 + 4m^2})$,
- spinors: $\Delta = \frac{1}{2}(d + 2|m|)$,
- vectors: $\Delta_{\pm} = \frac{1}{2}(d \pm \sqrt{(d-2)^2 + 4m^2})$,
- p-forms: $\Delta = \frac{1}{2}(d \pm \sqrt{(d-2p)^2 + 4m^2})$,
- first-order $(d/2)$-forms (d even): $\Delta = \frac{1}{2}(d + 2|m|)$,
- spin-3/2: $\Delta = \frac{1}{2}(d + 2|m|)$,
- massless spin-2: $\Delta = d$.

In the case of fields with second order lagrangians, we have not attempted to pick which of $\Delta_{\pm}$ is the physical dimension. Usually the choice $\Delta = \Delta_+$ is clear from the unitarity bound, but in some cases (notably $m^2 = 15/4$ in AdS_5) there is a genuine ambiguity. In practice this ambiguity is usually resolved by appealing to some special algebraic property of the relevant fields, such as transformation under supersymmetry or a global bosonic symmetry.

For brevity we will omit a further discussion of higher spins, and instead refer the reader to the (extensive) literature.

4.2 Three-point functions

Working with bulk-to-boundary propagators in the momentum representation is convenient for two-point functions, but for higher point functions position space is preferred because the full conformal invariance is more obvious. (However, for non-conformal examples of the bulk-boundary correspondence, the momentum representation seems uniformly more convenient). The boundary behavior of position space bulk-to-boundary propagators is specified in a slightly more subtle way: following [109] we require

$$K_\Delta(\vec{x}, z; \vec{y}) \to z^{4-\Delta} \delta^4(\vec{x} - \vec{y}) \quad \text{as} \quad z \to 0 \; . \tag{40}$$

Here $\vec{y}$ is the point on the boundary where we insert the operator, and $(\vec{x}, z)$ is a point in the bulk. The unique regular K_Δ solving the equation of motion and satisfying (40) is

$$K_\Delta(\vec{x}, z; \vec{y}) = \frac{\Gamma(\Delta)}{\pi^2\Gamma(\Delta-2)} \left(\frac{z}{z^2 + (\vec{x}-\vec{y})^2} \right)^\Delta \ . \tag{41}$$

At a fixed cutoff, $z = \epsilon$, the bulk-to-boundary propagator $K_\Delta(\vec{x}, \epsilon; \vec{y})$ is a continuous function which approximates $\epsilon^{4-\Delta}\delta^4(\vec{x}-\vec{y})$ better and better as $\epsilon \to 0$. Thus at any finite ϵ, the Fourier transform of (41) only approximately coincides with (38) (modified by the factor of $\epsilon^{4-\Delta}$ as explained after (39)). This apparently innocuous subtlety turned out to be important for two-point functions, as discovered in [109]. A correct prescription is to specify boundary conditions at finite $z = \epsilon$, cut off all bulk integrals at that boundary, and only afterwards take $\epsilon \to 0$. That is what we have done in (39). Calculating two-point functions directly using the position-space propagators (40), but cutting the bulk integrals off again at ϵ, and finally taking the same $\epsilon \to 0$ answer, one arrives at a different answer. This is not surprising since the $z = \epsilon$ boundary conditions were not used consistently. The authors of [109] checked that using the cutoff consistently (i.e. with the momentum space propagators) gave two-point functions $\langle \mathcal{O}(\vec{x}_1)\mathcal{O}(\vec{x}_2)\rangle$ a normalization such that Ward identities involving the three-point function $\langle \mathcal{O}(\vec{x}_1)\mathcal{O}(\vec{x}_2)J_\mu(\vec{x}_3)\rangle$, where J_μ is a conserved current, were obeyed. Two-point functions are uniquely difficult because of the poor convergence properties of the integrals over z. The integrals involved in three-point functions are sufficiently benign that one can ignore the issue of how to impose the cutoff.

If one has a Euclidean bulk action for three scalar fields ϕ_1, ϕ_2, and ϕ_3, of the form

$$S = \int d^5x \sqrt{g} \left[\sum_i \tfrac{1}{2}(\partial\phi_i)^2 + \tfrac{1}{2}m_i^2\phi_i^2 + \lambda\phi_1\phi_2\phi_3 \right] \ , \tag{42}$$

and if the ϕ_i couple to operators in the field theory by interaction terms $\int d^4x \, \phi_i \mathcal{O}_i$, then the calculation of $\langle \mathcal{O}_1\mathcal{O}_2\mathcal{O}_3\rangle$ reduces, via (33), to the evaluation of the graph shown in figure 2. That is,

$$\begin{aligned}\langle \mathcal{O}_1(\vec{x}_1)\mathcal{O}_2(\vec{x}_2)\mathcal{O}_3(\vec{x}_3)\rangle &= -\lambda \int d^5x \sqrt{g} K_{\Delta_1}(x;\vec{x}_1)K_{\Delta_2}(x;\vec{x}_2)K_{\Delta_3}(x;\vec{x}_3) \\ &= \frac{\lambda a_1}{|\vec{x}_1-\vec{x}_2|^{\Delta_1+\Delta_2-\Delta_3}|\vec{x}_1-\vec{x}_3|^{\Delta_1+\Delta_3-\Delta_2}|\vec{x}_2-\vec{x}_3|^{\Delta_2+\Delta_3-\Delta_1}} \ ,\end{aligned} \tag{43}$$

for some constant a_1. The dependence on the $\vec{x}_i$ is dictated by the conformal invariance, but the only way to compute a_1 is by performing the integral over

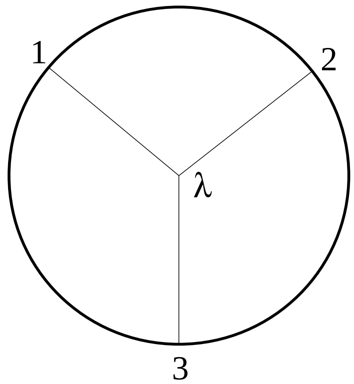


Figure 2: The Feynman graph for the three-point function as computed in supergravity. The legs correspond to factors of K_{Δ_i}, and the cubic vertex to a factor of λ. The position of the vertex is integrated over AdS_5.

x. The result [109] is

$$a_1 = -\frac{\Gamma\left[\frac{1}{2}(\Delta_1+\Delta_2-\Delta_3)\right]\Gamma\left[\frac{1}{2}(\Delta_1+\Delta_3-\Delta_2)\right]\Gamma\left[\frac{1}{2}(\Delta_2+\Delta_3-\Delta_1)\right]}{2\pi^4\Gamma(\Delta_1-2)\Gamma(\Delta_2-2)\Gamma(\Delta_3-2)} \cdot \Gamma\left[\frac{1}{2}(\Delta_1+\Delta_2+\Delta_3)-2\right] \ . \tag{44}$$

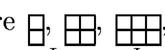
In principle one could also have couplings of the form $\phi_1\partial\phi_2\partial\phi_3$. This leads only to a modification of the constant a_1.

The main technical difficulty with three-point functions is that one must figure out the cubic couplings of supergravity fields. Because of the difficulties in writing down a covariant action for type IIB supergravity in ten dimensions (see however [155, 156, 157]), it is most straightforward to read off these "cubic couplings" from quadratic terms in the equations of motion. In flat ten-dimensional space these terms can be read off directly from the original type IIB supergravity papers [158, 159]. For $AdS_5 \times S^5$, one must instead expand in fluctuations around the background metric and five-form field strength. The old literature [103] only dealt with the linearized equations of motion; for 3-point functions it is necessary to go to one higher order of perturbation theory. This was done for a restricted set of fields in [132]. The fields considered were those dual to operators of the form $\mathrm{Tr}\phi^{(J_1}\phi^{J_2}\dots\phi^{J_\ell)}$ in field theory, where the parentheses indicate a symmetrized traceless product. These operators are the chiral primaries of the gauge theory: all other single trace operators of protected dimension descend from these by commuting with supersymmetry generators. Only the metric and the five-form are involved in the dual supergravity fields, and we are interested only in modes which are scalars in AdS_5. The result of [132] is that the equations of motion for the scalar modes $\tilde{s}_I$ dual to

$$\mathcal{O}^I = \mathcal{C}^I_{J_1 \ldots J_\ell} \mathrm{Tr} \phi^{(J_1} \ldots \phi^{J_\ell)} \tag{45}$$

follow from an action of the form

$$S = \frac{4N^2}{(2\pi)^5} \int d^5x \sqrt{g} \Bigg\{ \sum_I \frac{A_I (w^I)^2}{2} \left[-(\nabla \tilde{s}_I)^2 - l(l-4)\tilde{s}_I^2 \right] + \sum_{I_1, I_2, I_3} \frac{\mathcal{G}_{I_1 I_2 I_3} w^{I_1} w^{I_2} w^{I_3}}{3} \tilde{s}_{I_1} \tilde{s}_{I_2} \tilde{s}_{I_3} \Bigg\} . \tag{46}$$

Derivative couplings of the form $\tilde{s}\partial\tilde{s}\partial\tilde{s}$ are expected *a priori* to enter into (46), but an appropriate field redefinition eliminates them. The notation in (45) and (46) requires some explanation. I is an index which runs over the weight vectors of all possible representations constructed as symmetric traceless products of the **6** of $SU(4)_R$. These are the representations whose Young diagrams are , $\cdots$. $\mathcal{C}^I_{J_1 \ldots J_\ell}$ is a basis transformation matrix, chosen so that $\mathcal{C}^I_{J_1 \ldots J_\ell} \mathcal{C}^J_{J_1 \ldots J_\ell} = \delta^{IJ}$. As commented in the previous section, there is generally a normalization ambiguity on how supergravity fields couple to operators in the gauge theory. We have taken the coupling to be $\int d^4x \, \tilde{s}_I \mathcal{O}^I$, and the normalization ambiguity is represented by the "leg factors" w^I. It is the combination $s^I = w^I \tilde{s}^I$ rather than $\tilde{s}^I$ itself which has a definite relation to supergravity fields. We refer the reader to [132] for explicit expressions for A_I and the symmetric tensor $\mathcal{G}_{I_1 I_2 I_3}$. To get rid of factors of w^I, we introduce operators $\mathcal{O}^I = \tilde{w}^I \mathcal{O}^I$. One can choose $\tilde{w}^I$ so that a two-point function computation along the lines of section 4.1 leads to

$$\langle \mathcal{O}^{I_1}(\vec{x}) \mathcal{O}^{I_2}(0) \rangle = \frac{\delta^{I_1 I_2}}{x^{2\Delta_1}} . \tag{47}$$

With this choice, the three-point function, as calculated using (43), is

$$\langle \mathcal{O}^{I_1}(\vec{x_1}) \mathcal{O}^{I_2}(\vec{x_2}) \mathcal{O}^{I_3}(\vec{x_3}) \rangle = \frac{1}{N} \frac{\sqrt{\Delta_1 \Delta_2 \Delta_3} \langle \mathcal{C}^{I_1} \mathcal{C}^{I_2} \mathcal{C}^{I_3} \rangle}{|\vec{x}_1 - \vec{x}_2|^{\Delta_1 + \Delta_2 - \Delta_3} |\vec{x}_1 - \vec{x}_3|^{\Delta_1 + \Delta_3 - \Delta_2} |\vec{x}_2 - \vec{x}_3|^{\Delta_2 + \Delta_3 - \Delta_1}} , \tag{48}$$

where we have defined

$$\langle \mathcal{C}^{I_1} \mathcal{C}^{I_2} \mathcal{C}^{I_3} \rangle = \mathcal{C}^{I_1}_{J_1 \cdots J_i K_1 \cdots K_j} \mathcal{C}^{I_2}_{J_1 \cdots J_i L_1 \cdots L_k} \mathcal{C}^{I_3}_{K_1 \cdots K_j L_1 \cdots L_k} . \tag{49}$$

Remarkably, (48) is the same result one obtains from free field theory by Wick contracting all the ϕ^J fields in the three operators. This suggests that there is a non-renormalization theorem for this correlation function, but such a theorem has not yet been proven (see however comments at the end of section 3.2). It is worth emphasizing that the normalization ambiguity in the bulk-boundary coupling is circumvented essentially by considering invariant ratios of three-point functions and two-point functions, into which the "leg factors" w^I do not enter. This is the same strategy as was pursued in comparing matrix models of quantum gravity to Liouville theory.

4.3 Four-point functions

The calculation of four-point functions is difficult because there are several graphs which contribute, and some of them inevitably involve bulk-to-bulk propagators of fields with spin. The computation of four-point functions of the operators $\mathcal{O}_\phi$ and $\mathcal{O}_C$ dual to the dilaton and the axion was completed in [160]. See also [128, 133, 135, 136, 161, 162, 139, 137, 163, 5] for earlier contributions. One of the main technical results, further developed in [164], is that diagrams involving an internal propagator can be reduced by integration over one of the bulk vertices to a sum of quartic graphs expressible in terms of the functions

$$\begin{aligned} D_{\Delta_1\Delta_2\Delta_3\Delta_4}(\vec{x}_1,\vec{x}_2,\vec{x}_3,\vec{x}_4) &= \int d^5x\,\sqrt{g}\prod_{i=1}^{4}\tilde{K}_{\Delta_i}(\vec{x},z;\vec{x}_i), \\ \tilde{K}_\Delta(\vec{x},z;\vec{y}) &= \left(\frac{z}{z^2+(\vec{x}-\vec{y})^2}\right)^\Delta . \end{aligned} \tag{50}$$

The integration is over the bulk point $(\vec{x},z)$. There are two independent conformally invariant combinations of the $\vec{x}_i$:

$$s = \frac{1}{2}\frac{\vec{x}_{13}^2\vec{x}_{24}^2}{\vec{x}_{12}^2\vec{x}_{34}^2+\vec{x}_{14}^2\vec{x}_{23}^2} \qquad t = \frac{\vec{x}_{12}^2\vec{x}_{34}^2-\vec{x}_{14}^2\vec{x}_{23}^2}{\vec{x}_{12}^2\vec{x}_{34}^2+\vec{x}_{14}^2\vec{x}_{23}^2} . \tag{51}$$

One can write the connected four-point function as

$$\begin{aligned} \langle\mathcal{O}_\phi(\vec{x}_1)\mathcal{O}_C(\vec{x}_2)\mathcal{O}_\phi(\vec{x}_3)\mathcal{O}_C(\vec{x}_4)\rangle &= \left(\frac{6}{\pi^2}\right)^4\Bigg[16\vec{x}_{24}^2\left(\frac{1}{2s}-1\right)D_{4455}+\frac{64}{9}\frac{\vec{x}_{24}^2}{\vec{x}_{13}^2}\frac{1}{s}D_{3355} \\ &+\frac{16}{3}\frac{\vec{x}_{24}^2}{\vec{x}_{13}^2}\frac{1}{s}D_{2255}-14D_{4444}-\frac{46}{9\vec{x}_{13}^2}D_{3344}-\frac{40}{9\vec{x}_{13}^2}D_{2244}-\frac{8}{3\vec{x}_{13}^6}D_{1144}+64\vec{x}_{24}^2D_{4455}\Bigg] . \end{aligned} \tag{52}$$

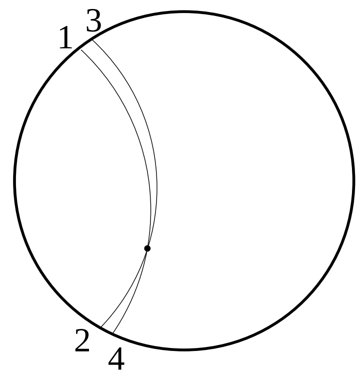


Figure 3: A nearly degenerate quartic graph contributing to the four-point function in the limit $|\vec{x}_{13}|, |\vec{x}_{24}| \ll |\vec{x}_{12}|$.

An interesting limit of (52) is to take two pairs of points close together. Following [160], let us take the pairs $(\vec{x}_1, \vec{x}_3)$ and $(\vec{x}_2, \vec{x}_4)$ close together while holding $\vec{x}_1$ and $\vec{x}_2$ a fixed distance apart. Then the existence of an OPE expansion implies that

$$\langle \mathcal{O}_{\Delta_1}(\vec{x}_1)\mathcal{O}_{\Delta_2}(\vec{x}_2)\mathcal{O}_{\Delta_3}(\vec{x}_3)\mathcal{O}_{\Delta_4}(\vec{x}_4)\rangle = \sum_{n,m} \frac{\alpha_n \langle \mathcal{O}_n(\vec{x}_1)\mathcal{O}_m(\vec{x}_2)\rangle \beta_m}{\vec{x}_{13}^{\Delta_1+\Delta_3-\Delta_m}\vec{x}_{24}^{\Delta_2+\Delta_4-\Delta_n}}, \quad (53)$$

at least as an asymptotic series, and hopefully even with a finite radius of convergence for $\vec{x}_{13}$ and $\vec{x}_{24}$. The operators $\mathcal{O}_n$ are the ones that appear in the OPE of $\mathcal{O}_1$ with $\mathcal{O}_3$, and the operators $\mathcal{O}_m$ are the ones that appear in the OPE of $\mathcal{O}_2$ with $\mathcal{O}_4$. $\mathcal{O}_\phi$ and $\mathcal{O}_C$ are descendants of chiral primaries, and so have protected dimensions. The product of descendants of chiral fields is not itself necessarily the descendent of a chiral field: an appropriately normal ordered product $: \mathcal{O}_\phi \mathcal{O}_\phi :$ is expected to have an unprotected dimension of the form $8 + O(1/N^2)$. This is the natural result from the field theory point of view because there are $O(N^2)$ degrees of freedom contributing to each factor, and the commutation relations between them are non-trivial only a fraction $1/N^2$ of the time. From the supergravity point of view, a composite operator like $: \mathcal{O}_\phi \mathcal{O}_\phi :$ corresponds to a two-particle bulk state, and the $O(1/N^2) = O(\kappa^2/R^8)$ correction to the mass is interpreted as the correction to the mass of the two-particle state from gravitational binding energy. Roughly one is thinking of graviton exchange between the legs of figure 3 that are nearly coincident.

If (53) is expanded in inverse powers of N, then the $O(1/N^2)$ correction to Δ_n and Δ_m shows up to leading order as a term proportional to a logarithm of some combination of the separations $\vec{x}_{ij}$. Logarithms also appear in the expansion of (52) in the $|\vec{x}_{13}|, |\vec{x}_{24}| \ll |\vec{x}_{12}|$ limit in which (53) applies: the leading log in this limit is $\frac{1}{(\vec{x}_{12})^{16}} \log\left(\frac{\vec{x}_{13}\vec{x}_{24}}{\vec{x}_{12}^2}\right)$. This is the correct form to be interpreted in terms of the propagation of a two-particle state dual to an operator whose dimension is slightly different from 8.

5 Wilson loops

In this section we consider Wilson loop operators in the gauge theory. The Wilson loop operator

$$W(\mathcal{C}) = \mathrm{Tr}\left[P \exp\left(i \oint_{\mathcal{C}} A\right)\right] \tag{54}$$

depends on a loop $\mathcal{C}$ embedded in four dimensional space, and it involves the path-ordered integral of the gauge connection along the contour. The trace is taken over some representation of the gauge group; we will discuss here only the case of the fundamental representation (see [165] for a discussion of other representations). From the expectation value of the Wilson loop operator $\langle W(\mathcal{C})\rangle$ we can calculate the quark-antiquark potential. For this purpose we consider a rectangular loop with sides of length T and L in Euclidean space. Then, viewing T as the time direction, it is clear that for large T the expectation value will behave as e^{-TE} where E is the lowest possible energy of the quark-anti-quark configuration. Thus, we have

$$\langle W\rangle \sim e^{-TV(L)} , \tag{55}$$

where $V(L)$ is the quark anti-quark potential. For large N and large $g_{YM}^2 N$, the AdS/CFT correspondence maps the computation of $\langle W\rangle$ in the CFT into a problem of finding a minimum surface in AdS [166, 167].

5.1 Wilson loops and minimum surfaces

In QCD, we expect the Wilson loop to be related to the string running from the quark to the antiquark. We expect this string to be analogous to the string in our configuration, which is a superstring which lives in ten dimensions, and which can stretch between two points on the boundary of AdS. In order to motivate this prescription let us consider the following situation. We start with the gauge group $U(N+1)$, and we break it to $U(N) \times U(1)$ by giving an expectation value to one of the scalars. This corresponds, as discussed in section 2, to having a D3 brane sitting at some radial position U in AdS, and at a point on S^5. The off-diagonal states, transforming in the $\mathbf{N}$ of $U(N)$, get a mass proportional to U, $m = U/2\pi$. So, from the point of view of the $U(N)$ gauge theory, we can view these states as massive quarks, which act as a source for the various $U(N)$ fields. Since they are charged they will act as a source for the vector fields. In order to get a non-dynamical source (an "external quark" with no fluctuations of its own, which will correspond precisely to the Wilson loop operator) we need to take $m \to \infty$, which means U should also go to infinity. Thus, the string should end on the boundary of AdS space.

These stretched strings will also act as a source for the scalar fields. The coupling to the scalar fields can be seen qualitatively by viewing the quarks as strings stretching between the N branes and the single separated brane. These strings will pull the N branes and will cause a deformation of the branes, which is described by the scalar fields. A more formal argument for this coupling is that these states are BPS, and the coupling to the scalar (Higgs) fields is determined by supersymmetry. Finally, one can see this coupling explicitly by writing the full $U(N+1)$ Lagrangian, putting in the Higgs expectation value and calculating the equation of motion for the massive fields [166]. The precise definition of the Wilson loop operator corresponding to the superstring will actually include also the field theory fermions, which will imply some particular boundary conditions for the worldsheet fermions at the boundary of AdS. However, this will not affect the leading order computations we describe here.

So, the final conclusion is that the stretched strings couple to the operator

$$W(\mathcal{C}) = \mathrm{Tr}\left[P\exp\left(\oint (iA_\mu \dot{x}^\mu + \theta^I \phi^I \sqrt{\dot{x}^2})d\tau\right)\right], \tag{56}$$

where $x^\mu(\tau)$ is any parametrization of the loop and θ^I $(I = 1, \cdots, 6)$ is a unit vector in $\mathbb{R}^6$ (the point on S^5 where the string is sitting). This is the expression when the signature of $\mathbb{R}^4$ is Euclidean. In the Minkowski signature case, the phase factor associated to the trajectory of the quark has an extra factor "i" in front of θ^I [7].

Generalizing the prescription of section 4 for computing correlation functions, the discussion above implies that in order to compute the expectation value of the operator (56) in $\mathcal{N} = 4$ SYM we should consider the string theory partition function on $AdS_5 \times S^5$, with the condition that we have a string worldsheet ending on the loop $\mathcal{C}$, as in figure 4 [167, 166]. In the supergravity regime, when $g_s N$ is large, the leading contribution to this partition function will come from the area of the string worldsheet. This area is

[7] The difference in the factor of i between the Euclidean and the Minkowski cases can be traced to the analytic continuation of $\sqrt{\dot{x}^2}$. A detailed derivation of (56) can be found in [168].

measured with the AdS metric, and it is generally not the same as the area enclosed by the loop $\mathcal{C}$ in four dimensions.

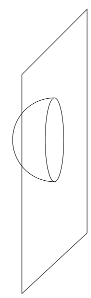


Figure 4: The Wilson loop operator creates a string worldsheet ending on the corresponding loop on the boundary of AdS.

The area as defined above is divergent. The divergence arises from the fact that the string worldsheet is going all the way to the boundary of AdS. If we evaluate the area up to some radial distance $U = r$, we see that for large r it diverges as $r|\mathcal{C}|$, where $|\mathcal{C}|$ is the length of the loop in the field theory [166, 167]. On the other hand, the perturbative computation in the field theory shows that $\langle W \rangle$, for W given by (56), is finite, as it should be since a divergence in the Wilson loop would have implied a mass renormalization of the BPS particle. The apparent discrepancy between the divergence of the area of the minimum surface in AdS and the finiteness of the field theory computation can be reconciled by noting that the appropriate action for the string worldsheet is not the area itself but its Legendre transform with respect to the string coordinates corresponding to θ^I and the radial coordinate u [168]. This is because these string coordinates obey the Neumann boundary conditions rather than the Dirichlet conditions. When the loop is smooth, the Legendre transformation simply subtracts the divergent term $r|\mathcal{C}|$, leaving the resulting action finite.

As an example let us consider a circular Wilson loop. Take $\mathcal{C}$ to be a circle of radius a on the boundary, and let us work in the Poincaré coordinates. We could find the surface that minimizes the area by solving the Euler-Lagrange equations. However, in this case it is easier to use conformal invariance. Note that there is a conformal transformation in the field theory that maps a line to a circle. In the case of the line, the minimum area surface is clearly a plane that intersects the boundary and goes all the way to the horizon (which is just a point on the boundary in the Euclidean case). Using the conformal transformation to map the line to a circle we obtain the minimal surface we want. It is, using the coordinates (17) for AdS_5,

$$\vec{x} = \sqrt{a^2 - z^2}(\vec{e}_1 \cos\theta + \vec{e}_2 \sin\theta), \tag{57}$$

where $\vec{e}_1$, $\vec{e}_2$ are two orthonormal vectors in four dimensions (which define the orientation of the circle) and $0 \leq z \leq a$. We can calculate the area of this surface in AdS, and we get a contribution to the action

$$S \sim \frac{1}{2\pi\alpha'}\mathcal{A} = \frac{R^2}{2\pi\alpha'}\int d\theta \int_\epsilon^a \frac{dza}{z^2} = \frac{R^2}{\alpha'}(\frac{a}{\epsilon} - 1), \tag{58}$$

where we have regularized the area by putting a an IR cutoff at $z = \epsilon$ in AdS, which is equivalent to a UV cutoff in the field theory [66]. Subtracting the divergent term we get

$$\langle W \rangle \sim e^{-S} \sim e^{R^2/\alpha'} = e^{\sqrt{4\pi g_s N}}. \tag{59}$$

This is independent of a as required by conformal invariance.

We could similarly consider a "magnetic" Wilson loop, which is also called a 't Hooft loop [169]. This case is related by electric-magnetic duality to the previous case. Since we identify the electric-magnetic duality with the $SL(2,\mathbb{Z})$ duality of type IIB string theory, we should consider in this case a D-string worldsheet instead of a fundamental string worldsheet. We get the same result as in (59) but with $g_s \to 1/g_s$.

Using (55) it is possible to compute the quark-antiquark potential in the supergravity approximation [167, 166]. In this case we consider a configuration which is invariant under (Euclidean) time translations. We take both particles to have the same scalar charge, which means that the two ends of the string are at the same point in S^5 (one could consider also the more general case with a string ending at different points on S^5 [166]). We put the quark at $x = -L/2$ and the anti-quark at $x = L/2$. Here "quark" means an infinitely massive W-boson connecting the N branes with one brane which is (infinitely) far away. The classical action for a string worldsheet is

$$S = \frac{1}{2\pi\alpha'}\int d\tau d\sigma \sqrt{\det(G_{MN}\partial_\alpha X^M \partial_\beta X^N)}, \tag{60}$$

where G_{MN} is the Euclidean $AdS_5 \times S^5$ metric. Note that the factors of α' cancel out in (60), as they should. Since we are interested in a static configuration we take $\tau = t$, $\sigma = x$, and then the action becomes

$$S = \frac{TR^2}{2\pi}\int_{-L/2}^{L/2} dx \frac{\sqrt{(\partial_x z)^2 + 1}}{z^2}. \tag{61}$$

We need to solve the Euler-Lagrange equations for this action. Since the action does not depend on x explicitly the solution satisfies

$$\frac{1}{z^2\sqrt{(\partial_x z)^2+1}} = \text{constant}. \tag{62}$$

Defining z_0 to be the maximum value of $z(x)$, which by symmetry occurs at $x=0$, we find that the solution is[8]

$$x = z_0 \int_{z/z_0}^{1} \frac{dy y^2}{\sqrt{1-y^4}}, \tag{63}$$

where z_0 is determined by the condition

$$\frac{L}{2} = z_0 \int_0^1 \frac{dy y^2}{\sqrt{1-y^4}} = z_0 \frac{\sqrt{2}\pi^{3/2}}{\Gamma(1/4)^2}. \tag{64}$$

The qualitative form of the solution is shown in figure 5(b). Notice that the string quickly approaches $x = L/2$ for small z (close to the boundary),

$$\frac{L}{2} - x \sim z^3 \ , \qquad z \to 0 \ . \tag{65}$$

Now we compute the total energy of the configuration. We just plug in the solution (63) in (61), subtract the infinity as explained above (which can be interpreted as the energy of two separated massive quarks, as in figure 5(a)), and we find

$$E = V(L) = -\frac{4\pi^2(2g_{YM}^2 N)^{1/2}}{\Gamma(\frac{1}{4})^4 L}. \tag{66}$$

We see that the energy goes as $1/L$, a fact which is determined by conformal invariance. Note that the energy is proportional to $(g_{YM}^2 N)^{1/2}$, as opposed to $g_{YM}^2 N$ which is the perturbative result. This indicates some screening of the charges at strong coupling. The above calculation makes sense for all distances L when $g_s N$ is large, independently of the value of g_s. Some subleading corrections coming from quantum fluctuations of the worldsheet were calculated in [170, 171, 172].

In a similar fashion we could compute the potential between two magnetic monopoles in terms of a D-string worldsheet, and the result will be the same as (66) but with $g_{YM} \to 4\pi/g_{YM}$. One can also calculate the interaction between a magnetic monopole and a quark. In this case the fundamental string (ending on the quark) will attach to the D-string (ending on the

[8] All integrals in this section can be calculated in terms of elliptic or Beta functions.

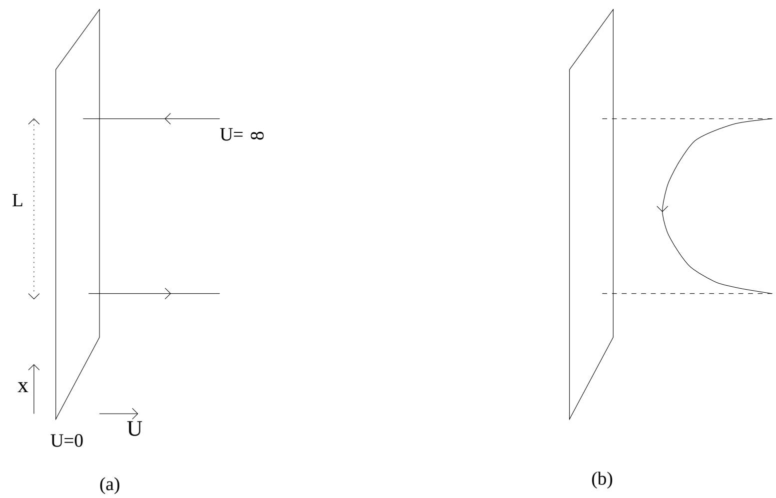


Figure 5: (a) Initial configuration corresponding to two massive quarks before we turn on their coupling to the $U(N)$ gauge theory. (b) Configuration after we consider the coupling to the $U(N)$ gauge theory. This configuration minimizes the action. The quark-antiquark energy is given by the difference of the total length of the strings in (a) and (b).

monopole), and they will connect to form a $(1,1)$ string which will go into the horizon. The resulting potential is a complicated function of g_{YM} [173], but in the limit that g_{YM} is small (but still with $g_{YM}^2 N$ large) we get that the monopole-quark potential is just $1/4$ of the quark-quark potential. This can be understood from the fact that when g is small the D-string is very rigid and the fundamental string will end almost perpendicularly on the D-string. Therefore, the solution for the fundamental string will be half of the solution we had above, leading to a factor of $1/4$ in the potential. Calculations of Wilson loops in the Higgs phase were done in [174].

Another interesting case one can study analytically is a surface near a cusp on $\mathbb{R}^4$. In this case, the perturbative computation in the gauge theory shows a logarithmic divergence with a coefficient depending on the angle at the cusp. The area of the minimum surface also contains a logarithmic divergence depending on the angle [168]. Other aspects of the gravity calculation of Wilson loops were discussed in [175, 176, 177, 178, 179].

5.2 Other branes ending on the boundary

We could also consider other branes that are ending at the boundary [180]. The simplest example would be a zero-brane (i.e. a particle) of mass m.

In Euclidean space a zero-brane describes a one dimensional trajectory in anti-de-Sitter space which ends at two points on the boundary. Therefore, it is associated with the insertion of two local operators at the two points where the trajectory ends. In the supergravity approximation the zero-brane follows a geodesic. Geodesics in the hyperbolic plane (Euclidean AdS) are semicircles. If we compute the action we get

$$S = m \int ds = -2mR \int_{\epsilon}^{a} \frac{a dz}{z\sqrt{a^2 - z^2}}, \tag{67}$$

where we took the distance between the two points at the boundary to be $L = 2a$ and regulated the result. We find a logarithmic divergence when $\epsilon \to 0$, proportional to $\log(\epsilon/a)$. If we subtract the logarithmic divergence we get a residual dependence on a. Naively we might have thought that (as in the previous subsection) the answer had to be independent of a due to conformal invariance. In fact, the dependence on a is very important, since it leads to a result of the form

$$e^{-S} \sim e^{-2mR\log a} \sim \frac{1}{a^{2mR}}, \tag{68}$$

which is precisely the result we expect for the two-point function of an operator of dimension $\Delta = mR$. This is precisely the large mR limit of the formula (14), so we reproduce in the supergravity limit the 2-point function described in section 4. In general, this sort of logarithmic divergence arises when the brane worldvolume is odd dimensional [180], and it implies that the expectation value of the corresponding operator depends on the overall scale. In particular one could consider the "Wilson surfaces" that arise in the six dimensional $\mathcal{N} = (2,0)$ theory. In that case one has to consider a two-brane, with a three dimensional worldvolume, ending on a two dimensional surface on the boundary of AdS_7. Again, one gets a logarithmic term, which is proportional to the rigid string action of the two dimensional surface living on the string in the $\mathcal{N} = (2,0)$ field theory [181, 180].

One can also compute correlation functions involving more than one Wilson loop. To leading order in N this will be just the product of the expectation values of each Wilson loop. On general grounds one expects that the subleading corrections are given by surfaces that end on more than one loop. One limiting case is when the surfaces look similar to the zeroth order surfaces but with additional thin tubes connecting them. These thin tubes are nothing else than massless particles being exchanged between the two string worldsheets [165, 181].

6 Theories at finite temperature

As discussed in section 3, the quantities that can be most successfully compared between gauge theory and string theory are those with some protection from supersymmetry and/or conformal invariance — for instance, dimensions of chiral primary operators. Finite temperature breaks both supersymmetry and conformal invariance, and the insights we gain from examining the $T > 0$ physics will be of a more qualitative nature. They are no less interesting for that: we shall see in section 6.1 how the entropy of near-extremal D3-branes comes out identical to the free field theory prediction up to a factor of a power of 4/3; then in section 6.2 we explain how a phase transition studied by Hawking and Page in the context of quantum gravity is mapped into a confinement-deconfinement transition in the gauge theory.

6.1 Construction

The gravity solution describing the gauge theory at finite temperature can be obtained by starting from the general black three-brane solution and taking the decoupling limit of section 2 keeping the energy density above extremality finite. The resulting metric can be written as

$$\begin{aligned} ds^2 &= R^2\left[u^2(-hdt^2+dx_1^2+dx_2^2+dx_3^2)+\frac{du^2}{hu^2}+d\Omega_5^2\right] \\ h &= 1-\frac{u_0^4}{u^4}\,, \qquad u_0=\pi T. \end{aligned} \tag{69}$$

It will often be useful to Wick rotate by setting $t_E = it$, and use the relation between the finite temperature theory and the Euclidean theory with a compact time direction.

The first computation which indicated that finite-temperature $U(N)$ Yang-Mills theory might be a good description of the microstates of N coincident D3-branes was the calculation of the entropy [182, 183]. On the supergravity side, the entropy of near-extremal D3-branes is just the usual Bekenstein-Hawking result, $S = A/4G_N$, and it is expected to be a reliable guide to the entropy of the gauge theory at large N and large $g_{YM}^2 N$. There is no problem on the gauge theory side in working at large N, but large $g_{YM}^2 N$ at finite temperature is difficult indeed. The analysis of [182] was limited to a free field computation in the field theory, but nevertheless the two results for the entropy agreed up to a factor of a power of 4/3. In the canonical ensemble, where temperature and volume are the independent variables, one identifies the field theory volume with the world-volume of the

D3-branes, and one sets the field theory temperature equal to the Hawking temperature in supergravity. The result is

$$\begin{aligned} F_{SUGRA} &= -\frac{\pi^2}{8} N^2 V T^4, \\ F_{SYM} &= \frac{4}{3} F_{SUGRA} \, . \end{aligned} \tag{70}$$

The supergravity result is at leading order in l_s/R, and it would acquire corrections suppressed by powers of TR if we had considered the full D3-brane metric rather than the near-horizon limit, (69). These corrections do not have an interpretation in the context of CFT because they involve R as an intrinsic scale. Two equivalent methods to evaluate F_{SUGRA} are a) to use $F = E - TS$ together with standard expressions for the Bekenstein-Hawking entropy, the Hawking temperature, and the ADM mass; and b) to consider the gravitational action of the Euclidean solution, with a periodicity in the Euclidean time direction (related to the temperature) which eliminates a conical deficit angle at the horizon.[9]

The $4/3$ factor is a long-standing puzzle into which we still have only qualitative insight. The gauge theory computation was performed at zero 't Hooft coupling, whereas the supergravity is supposed to be valid at strong 't Hooft coupling, and unlike in the 1+1-dimensional case where the entropy is essentially fixed by the central charge, there is no non-renormalization theorem for the coefficient of T^4 in the free energy. Indeed, it was suggested in [184] that the leading term in the $1/N$ expansion of F has the form

$$F = -f(g_{YM}^2 N) \frac{\pi^2}{6} N^2 V T^4, \tag{71}$$

where $f(g_{YM}^2 N)$ is a function which smoothly interpolates between a weak coupling limit of 1 and a strong coupling limit of $3/4$. It was pointed out early [185] that the quartic potential $g_{YM}^2 \mathrm{Tr}[\phi^I, \phi^J]^2$ in the $\mathcal{N} = 4$ Yang-Mills action might be expected to freeze out more and more degrees of freedom as the coupling was increased, which would suggest that $f(g_{YM}^2 N)$ is monotone decreasing. An argument has been given [186], based on the non-renormalization of the two-point function of the stress tensor, that $f(g_{YM}^2 N)$ should remain finite at strong coupling.

[9]The result of [182], $S_{SYM} = (4/3)^{1/4} S_{SUGRA}$, differs superficially from (70), but it is only because the authors worked in the microcanonical ensemble: rather than identifying the Hawking temperature with the field theory temperature, the ADM mass above extremality was identified with the field theory energy.

The leading corrections to the limiting value of $f(g_{YM}^2 N)$ at strong and weak coupling were computed in [184] and [187], respectively. The results are

$$\begin{aligned} f(g_{YM}^2 N) &= 1 - \frac{3}{2\pi^2} g_{YM}^2 N + \ldots \qquad && \text{for small } g_{YM}^2 N, \\ f(g_{YM}^2 N) &= \frac{3}{4} + \frac{45}{32} \frac{\zeta(3)}{(g_{YM}^2 N)^{3/2}} + \ldots \qquad && \text{for large } g_{YM}^2 N. \end{aligned} \tag{72}$$

The weak coupling result is a straightforward although somewhat tedious application of the diagrammatic methods of perturbative finite-temperature field theory. The constant term is from one loop, and the leading correction is from two loops. The strong coupling result follows from considering the leading α' corrections to the supergravity action. The relevant one involves a particular contraction of four powers of the Weyl tensor. It is important now to work with the Euclidean solution, and one restricts attention further to the near-horizon limit. The Weyl curvature comes from the non-compact part of the metric, which is no longer AdS_5 but rather the AdS-Schwarzschild solution which we will discuss in more detail in section 6.2. The action including the α' corrections no longer has the Einstein-Hilbert form, and correspondingly the Bekenstein-Hawking prescription no longer agrees with the free energy computed as βI where I is the Euclidean action. In keeping with the basic prescription for computing Green's functions, where a free energy in field theory is equated (in the appropriate limit) with a supergravity action, the relation $I = \beta F$ is regarded as the correct one. (See [188].) It has been conjectured that the interpolating function $f(g_{YM}^2 N)$ is not smooth, but exhibits some phase transition at a finite value of the 't Hooft coupling. We regard this as an unsettled question. The arguments in [189, 190] seem as yet incomplete. In particular, they rely on analyticity properties of the perturbation expansion which do not seem to be proven for finite temperature field theories.

6.2 Thermal phase transition

The holographic prescription of [23, 24], applied at large N and $g_{YM}^2 N$ where loop and stringy corrections are negligible, involves extremizing the supergravity action subject to particular asymptotic boundary conditions. We can think of this as the saddle point approximation to the path integral over supergravity fields. That path integral is ill-defined because of the non-renormalizable nature of supergravity. String amplitudes (when we can calculate them) render on-shell quantities well-defined. Despite the conceptual difficulties we can use some simple intuition about path integrals to illustrate an important point about the AdS/CFT correspondence: namely, there can be more than one saddle point in the range of integration, and when there is we should sum $e^{-I_{SUGRA}}$ over the classical configurations to obtain the saddle-point approximation to the gauge theory partition function. Multiple classical configurations are possible because of the general feature of boundary value problems in differential equations: there can be multiple solutions to the classical equations satisfying the same asymptotic boundary conditions. The solution which globally minimizes I_{SUGRA} is the one that dominates the path integral.

When there are two or more solutions competing to minimize I_{SUGRA}, there can be a phase transition between them. An example of this was studied in [191] long before the AdS/CFT correspondence, and subsequently resurrected, generalized, and reinterpreted in [24, 68] as a confinement-deconfinement transition in the gauge theory. Since the qualitative features are independent of the dimension, we will restrict our attention to AdS_5. It is worth noting however that if the AdS_5 geometry is part of a string compactification, it doesn't matter what the internal manifold is except insofar as it fixes the cosmological constant, or equivalently the radius R of anti-de Sitter space.

There is an embedding of the Schwarzschild black hole solution into anti-de Sitter space which extremizes the action

$$I = -\frac{1}{16\pi G_5} \int d^5x \sqrt{g} \left(\mathcal{R} + \frac{12}{R^2} \right) . \tag{73}$$

Explicitly, the metric is

$$\begin{aligned} ds^2 &= f dt^2 + \frac{1}{f} dr^2 + r^2 d\Omega_3^2, \\ f &= 1 + \frac{r^2}{R^2} - \frac{\mu}{r^2} . \end{aligned} \tag{74}$$

The radial variable r is restricted to $r \geq r_+$, where r_+ is the largest root of $f = 0$. The Euclidean time is periodically identified, $t \sim t + \beta$, in order to eliminate the conical singularity at $r = r_+$. This requires

$$\beta = \frac{2\pi R^2 r_+}{2r_+^2 + R^2} . \tag{75}$$

Topologically, this space is $S^3 \times B^2$, and the boundary is $S^3 \times S^1$ (which is the relevant space for the field theory on S^3 with finite temperature). We will call this space X_2. Another space with the same boundary which is also a local extremum of (73) is given by the metric in (74) with $\mu = 0$

and again with periodic time. This space, which we will call X_1, is not only metrically distinct from the first (being locally conformally flat), but also topologically $B^4 \times S^1$ rather than $S^3 \times B^2$. Because the S^1 factor is not simply connected, there are two possible spin structures on X_1, corresponding to thermal (anti-periodic) or supersymmetric (periodic) boundary conditions on fermions. In contrast, X_2 is simply connected and hence admits a unique spin structure, corresponding to thermal boundary conditions. For the purpose of computing the twisted partition function, $\mathrm{Tr}(-1)^F e^{-\beta H}$, in a saddle-point approximation, only X_1 contributes. But, X_1 and X_2 make separate saddle-point contributions to the usual thermal partition function, $\mathrm{Tr} e^{-\beta H}$, and the more important one is the one with the smaller Euclidean action.

Actually, both $I(X_1)$ and $I(X_2)$ are infinite, so to compute $I(X_2)-I(X_1)$ a regulation scheme must be adopted. The one used in [68, 184] is to cut off both X_1 and X_2 at a definite coordinate radius $r = R_0$. For X_2, the elimination of the conical deficit angle at the horizon fixes the period of Euclidean time; but for X_1, the period is arbitrary. In order to make the comparison of $I(X_1)$ and $I(X_2)$ meaningful, we fix the period of Euclidean time on X_1 so that the proper circumference of the S_1 at $r = R_0$ is the same as the proper length on X_2 of an orbit of the Killing vector $\partial/\partial t$, also at $r = R_0$. In the limit $R_0 \to \infty$, one finds

$$I(X_2) - I(X_1) = \frac{\pi^2 r_+^3 (R^2 - r_+^2)}{4G_5(2r_+^2 + R^2)} \ , \tag{76}$$

where again r_+ is the largest root of $f = 0$. The fact that (76) (or more precisely its AdS_4 analog) can change its sign was interpreted in [191] as indicating a phase transition between a black hole in AdS and a thermal gas of particles in AdS (which is the natural interpretation of the space X_1). The black hole is the thermodynamically favored state when the horizon radius r_+ exceeds the radius of curvature R of AdS. In the gauge theory we interpret this transition as a confinement-deconfinement transition. Since the theory is conformally invariant, the transition temperature must be proportional to the inverse radius of the space S^3 which the field theory lives on. Similar transitions, and also local thermodynamic instability due to negative specific heats, have been studied in the context of spinning branes and charged black holes in [192, 193, 194, 195, 196, 197, 198]. Most of these works are best understood on the CFT side as explorations of exotic thermal phenomena in finite-temperature gauge theories. Connections with Higgsed states in gauge theory are clearer in [199, 200]. The relevance to confinement is explored in [197]. See also [201, 202, 203, 204] for other interesting contributions to the finite temperature literature.

Deconfinement at high temperature can be characterized by a spontaneous breaking of the center of the gauge group. In our case the gauge group is $SU(N)$ and its center is $\mathbb{Z}_N$. The order parameter for the breaking of the center is the expectation value of the Polyakov (temporal) loop $\langle W(C)\rangle$. The boundary of the spaces X_1, X_2 is $S^3 \times S^1$, and the path C wraps around the circle. An element of the center $g \in \mathbb{Z}_N$ acts on the Polyakov loop by $\langle W(C)\rangle \to g\langle W(C)\rangle$. The expectation value of the Polyakov loop measures the change of the free energy of the system $F_q(T)$ induced by the presence of the external charge q, $\langle W(C)\rangle \sim exp\,(-F_q(T)/T)$. In a confining phase $F_q(T)$ is infinite and therefore $\langle W(C)\rangle = 0$. In the deconfined phase $F_q(T)$ is finite and therefore $\langle W(C)\rangle \neq 0$.

As discussed in section 5, in order to compute $\langle W(C)\rangle$ we have to evaluate the partition function of strings with a worldsheet D that is bounded by the loop C. Consider first the low temperature phase. The relevant space is X_1 which, as discussed above, has the topology $B^4 \times S^1$. The contour C wraps the circle and is not homotopic to zero in X_1. Therefore C is not a boundary of any D, which immediately implies that $\langle W(C)\rangle = 0$. This is the expected behavior at low temperatures (compared to the inverse radius of the S^3), where the center of the gauge group is not broken.

For the high temperature phase the relevant space is X_2, which has the topology $S^3 \times B^2$. The contour C is now a boundary of a string worldsheet $D = B^2$ (times a point in S^3). This seems to be in agreement with the fact that in the high temperature phase $\langle W(C)\rangle \neq 0$ and the center of the gauge group is broken. It was pointed out in [68] that there is a subtlety with this argument, since the center should not be broken in finite volume (S^3), but only in the infinite volume limit ($\mathbb{R}^3$). Indeed, the solution X_2 is not unique and we can add to it an expectation value for the integral of the NS-NS 2-form field B on B^2, with vanishing field strength. This is an angular parameter ψ with period 2π, which contributes $i\psi$ to the string worldsheet action. The string theory partition function includes now an integral over all values of ψ, making $\langle W(C)\rangle = 0$ on S^3. In contrast, on $\mathbb{R}^3$ one integrates over the local fluctuations of ψ but not over its vacuum expectation value. Now $\langle W(C)\rangle \neq 0$ and depends on the value of $\psi \in U(1)$, which may be understood as the dependence on the center $\mathbb{Z}_N$ in the large N limit. Explicit computations of Polyakov loops at finite temperature were done in [205, 6].

In [68] the Euclidean black hole solution (74) was suggested to be holographically dual to a theory related to pure QCD in three dimensions. In the large volume limit the solution corresponds to the $\mathcal{N} = 4$ gauge theory on $\mathbb{R}^3 \times S^1$ with thermal boundary conditions, and when the S^1 is made small (corresponding to high temperature T) the theory at distances larger than

$1/T$ effectively reduces to pure Yang-Mills on $\mathbb{R}^3$. Some of the non-trivial successes of this approach to QCD are summarized in [1].

Acknowledgments

These notes were made by selected pieces of [1]. I thank O. Aharony, S. Gubser, H. Ooguri and Y. Oz for collaborating on writing that review. I also thank the organizers for a very nice school. This work was supported by the Sloan, Packard and Mac Arthur foundations and by the DOE grant DE-FGO2-91ER40654.

References

[1] O. Aharony, S. S. Gubser, J. Maldacena, H. Ooguri and Y. Oz, Phys. Rept. **323**, 183 (2000) [hep-th/9905111].

[2] H. J. Boonstra, B. Peeters and K. Skenderis, Phys. Lett. **B411**, 59 (1997) [hep-th/9706192].

[3] K. Sfetsos and K. Skenderis, Nucl. Phys. **B517**, 179 (1998) [hep-th/9711138].

[4] S. R. Das and S. P. Trivedi, Phys. Lett. **B445**, 142 (1998) [hep-th/9804149].

[5] F. Gonzalez-Rey, I. Park and K. Schalm, Phys. Lett. **B448**, 37 (1999) [hep-th/9811155].

[6] A. Brandhuber, N. Itzhaki, J. Sonnenschein and S. Yankielowicz, Phys. Lett. **B434**, 36 (1998) [hep-th/9803137].

[7] G. 't Hooft, "A Planar Diagram Theory for Strong Interactions," *Nucl. Phys.* **B72** (1974) 461.

[8] A. M. Polyakov, "Quantum Geometry of Bosonic Strings," *Phys. Lett.* **103B** (1981) 207.

[9] J. Maldacena, "The Large N limit of superconformal field theories and supergravity," *Adv. Theor. Math. Phys.* **2** (1998) 231, `hep-th/9711200`.

[10] J. Polchinski, "Dirichlet Branes and Ramond-Ramond charges," *Phys. Rev. Lett.* **75** (1995) 4724–4727, `hep-th/9510017`.

[11] G. 't Hooft, "Magnetic monopoles in unified gauge theories," *Nucl. Phys.* **B79** (1974) 276–284.

[12] A. M. Polyakov, "Particle spectrum in the quantum field theory," *JETP Lett.* **20** (1974) 194–195.

[13] E. Witten, "Bound States of Strings and p-Branes," *Nucl. Phys.* **B460** (1996) 335–350, `hep-th/9510135`.

[14] I. R. Klebanov, "World volume approach to absorption by nondilatonic branes," *Nucl. Phys.* **B496** (1997) 231, `hep-th/9702076`.

[15] S. S. Gubser, I. R. Klebanov, and A. A. Tseytlin, "String theory and classical absorption by three-branes," *Nucl. Phys.* **B499** (1997) 217, `hep-th/9703040`.

[16] S. S. Gubser and I. R. Klebanov, "Absorption by branes and Schwinger terms in the world volume theory," *Phys. Lett.* **B413** (1997) 41–48, `hep-th/9708005`.

[17] A. Strominger and C. Vafa, "Microscopic origin of the Bekenstein-Hawking entropy," *Phys. Lett.* **B379** (1996) 99–104, `hep-th/9601029`.

[18] C. G. Callan and J. M. Maldacena, "D-Brane Approach to Black Hole Quantum Mechanics," *Nucl. Phys.* **B472** (1996) 591–610, `hep-th/9602043`.

[19] G. T. Horowitz and A. Strominger, "Counting States of Near Extremal Black Holes," *Phys. Rev. Lett.* **77** (1996) 2368–2371, `hep-th/9602051`.

[20] S. R. Das and S. D. Mathur, "Comparing decay rates for black holes and D-branes," *Nucl. Phys.* **B478** (1996) 561–576, `hep-th/9606185`.

[21] A. Dhar, G. Mandal, and S. R. Wadia, "Absorption vs. decay of black holes in string theory and T symmetry," *Phys. Lett.* **B388** (1996) 51–59, `hep-th/9605234`.

[22] J. Maldacena and A. Strominger, "Black hole grey body factors and d-brane spectroscopy," *Phys. Rev.* **D55** (1997) 861–870, `hep-th/9609026`.

[23] S. S. Gubser, I. R. Klebanov, and A. M. Polyakov, "Gauge theory correlators from noncritical string theory," *Phys. Lett.* **B428** (1998) 105, `hep-th/9802109`.

[24] E. Witten, "Anti-de Sitter space and holography," *Adv. Theor. Math. Phys.* **2** (1998) 253, `hep-th/9802150`.

[25] A. Salam and E. Sezgin, eds., *Supergravities in Diverse Dimensions*, vol. 1 and 2. North-Holland, Amsterdam, Netherlands, 1989.

[26] M. J. Duff, B. E. W. Nilsson, and C. N. Pope, "Kaluza-Klein Supergravity," *Phys. Rept.* **130** (1986) 1.

[27] E. Witten, "Quantum Field Theory and the Jones Polynomial," *Commun. Math. Phys.* **121** (1989) 351.

[28] T. Banks, W. Fischler, S. H. Shenker, and L. Susskind, "M Theory as a Matrix Model: A Conjecture," *Phys. Rev.* **D55** (1997) 5112–5128, `hep-th/9610043`.

[29] G. 't Hooft, "Dimensional Reduction in Quantum Gravity," `gr-qc/9310026`.

[30] L. Susskind, "The World as a hologram," *J. Math. Phys.* **36** (1995) 6377–6396, `hep-th/9409089`.

[31] J. D. Bekenstein, "Entropy Bounds and Black Hole Remnants," *Phys. Rev.* **D49** (1994) 1912–1921, `gr-qc/9307035`.

[32] P. D. Vecchia, "An Introduction to AdS/CFT correspondence," `hep-th/9903007`.

[33] M. R. Douglas and S. Randjbar-Daemi, "Two lectures on the AdS / CFT correspondence," `hep-th/9902022`.

[34] J. L. Petersen, "Introduction to the Maldacena conjecture on AdS / CFT," `hep-th/9902131`.

[35] I. R. Klebanov, "From three-branes to large N gauge theories," `hep-th/9901018`.

[36] J. Polchinski, "String Theory,". Cambridge University Press (1998).

[37] R. G. Leigh, "Dirac-Born-Infeld Action from Dirichlet Sigma-Model," *Mod. Phys. Lett.* **A4** (1989) 2767.

[38] G. T. Horowitz and A. Strominger, "Black strings and P-branes," *Nucl. Phys.* **B360** (1991) 197–209.

[39] G. W. Gibbons and P. K. Townsend, "Vacuum interpolation in supergravity via super p-branes," *Phys. Rev. Lett.* **71** (1993) 3754–3757, `hep-th/9307049`.

[40] E. Witten, "AdS / CFT correspondence and topological field theory," *JHEP* **12** (1998) 012, `hep-th/9812012`.

[41] N. Seiberg, private communication, February 1999.

[42] S. Elitzur, G. Moore, A. Schwimmer, and N. Seiberg, "Remarks on the Canonical Quantization of the Chern-Simons-Witten Theory," *Nucl. Phys.* **B326** (1989) 108.

[43] C. Fronsdal, "The Dirac Supermultiplet," *Phys. Rev.* **D26** (1982) 1988.

[44] D. Z. Freedman and H. Nicolai, "Multiplet Shortening in $OSp(N, 4)$," *Nucl. Phys.* **B237** (1984) 342.

[45] K. Pilch, P. van Nieuwenhuizen, and P. K. Townsend, "Compactification of $d = 11$ Supergravity on S^4 (Or $11 = 7 + 4$, Too)," *Nucl. Phys.* **B242** (1984) 377.

[46] M. Gunaydin, P. van Nieuwenhuizen, and N. P. Warner, "General construction of the unitary representations of anti-de Sitter superalgebras and the spectrum of the S^4 compactification of eleven dimensional supergravity," *Nucl. Phys.* **B255** (1985) 63.

[47] M. Gunaydin and N. P. Warner, "Unitary supermultiplets of $OSp(8|4,\mathbb{R})$ and the spectrum of the S^7 compactification of eleven dimensional supergravity," *Nucl. Phys.* **B272** (1986) 99.

[48] M. Gunaydin, B. E. W. Nilsson, G. Sierra, and P. K. Townsend, "Singletons and Superstrings," *Phys. Lett.* **176B** (1986) 45.

[49] E. Bergshoeff, A. Salam, E. Sezgin, and Y. Tanii, "Singletons, Higher Spin Massless States and the Supermembrane," *Phys. Lett.* **205B** (1988) 237.

[50] E. Bergshoeff, A. Salam, E. Sezgin, and Y. Tanii, "$N = 8$ Supersingleton Quantum Field Theory," *Nucl. Phys.* **B305** (1988) 497.

[51] E. Bergshoeff, M. J. Duff, C. N. Pope, and E. Sezgin, "Compactifications of the Eleven-Dimensional Supermembrane," *Phys. Lett.* **B224** (1989) 71.

[52] M. R. Douglas, "Branes within branes," `hep-th/9512077`.

[53] C. M. Hull and P. K. Townsend, "Unity of superstring dualities," *Nucl. Phys.* **B438** (1995) 109–137, `hep-th/9410167`.

[54] A. A. Tseytlin, "Selfduality of Born-Infeld action and Dirichlet three-brane of type IIB superstring theory," *Nucl. Phys.* **B469** (1996) 51–67, `hep-th/9602064`.

[55] M. B. Green and M. Gutperle, "Comments on three-branes," *Phys. Lett.* **B377** (1996) 28–35, `hep-th/9602077`.

[56] M. R. Douglas and M. Li, "D-brane realization of N=2 superYang-Mills theory in four- dimensions," `hep-th/9604041`.

[57] N. Seiberg and E. Witten, "Monopoles, duality and chiral symmetry breaking in N=2 supersymmetric QCD," *Nucl. Phys.* **B431** (1994) 484–550, `hep-th/9408099`.

[58] M. Dine and N. Seiberg, "Comments on higher derivative operators in some SUSY field theories," *Phys. Lett.* **B409** (1997) 239–244, `hep-th/9705057`.

[59] M. R. Douglas, D. Kabat, P. Pouliot, and S. H. Shenker, "D-branes and short distances in string theory," *Nucl. Phys.* **B485** (1997) 85–127, hep-th/9608024.

[60] M. R. Douglas and W. Taylor, "Branes in the bulk of Anti-de Sitter space," hep-th/9807225.

[61] S. R. Das, "Holograms of branes in the bulk and acceleration terms in SYM effective action," hep-th/9905037.

[62] S. R. Das, "Brane waves, Yang-Mills theories and causality," *JHEP* **02** (1999) 012, hep-th/9901004.

[63] A. Kehagias, "New type IIB vacua and their F theory interpretation," *Phys. Lett.* **B435** (1998) 337, hep-th/9805131.

[64] S. S. Gubser, "Einstein manifolds and conformal field theories," *Phys. Rev.* **D59** (1999) 025006, hep-th/9807164.

[65] L. J. Romans, "New Compactifications of Chiral $N = 2, d = 10$ Supergravity," *Phys. Lett.* **153B** (1985) 392.

[66] L. Susskind and E. Witten, "The Holographic bound in anti-de Sitter space," hep-th/9805114.

[67] O. Aharony and E. Witten, "Anti-de Sitter Space and the Center of the Gauge Group," *JHEP* **11** (1998) 018, hep-th/9807205.

[68] E. Witten, "Anti-de Sitter space, thermal phase transition, and confinement in gauge theories," *Adv. Theor. Math. Phys.* **2** (1998) 505, hep-th/9803131.

[69] F. Gonzalez-Rey, B. Kulik, I. Y. Park, and M. Rocek, "Selfdual effective action of N=4 superYang-Mills," *Nucl. Phys.* **B544** (1999) 218, hep-th/9810152.

[70] J. Bagger and J. Wess, *Supersymmetry and Supergravity.* Princeton Series in Physics. Princeton University Press, Princeton, 1992.

[71] O. Bergman, "Three Pronged Strings and 1/4 BPS States in $N = 4$ Super Yang-Mills Theory," *Nucl. Phys.* **B525** (1998) 104, hep-th/9712211.

[72] K. Hashimoto, H. Hata, and N. Sasakura, "Three - string junction and BPS saturated solutions in SU(3) supersymmetric Yang-Mills theory," *Phys. Lett.* **B431** (1998) 303–310, hep-th/9803127.

[73] T. Kawano and K. Okuyama, "String network and 1/4 BPS states in N=4 SU(N) supersymmetric Yang-Mills theory," *Phys. Lett.* **B432** (1998) 338–342, `hep-th/9804139`.

[74] O. Bergman and B. Kol, "String Webs and 1/4 BPS Monopoles," *Nucl. Phys.* **B536** (1998) 149, `hep-th/9804160`.

[75] K. Hashimoto, H. Hata, and N. Sasakura, "Multipronged strings and BPS saturated solutions in SU(N) supersymmetric Yang-Mills theory," *Nucl. Phys.* **B535** (1998) 83, `hep-th/9804164`.

[76] K. Lee and P. Yi, "Dyons in N=4 supersymmetric theories and three pronged strings," *Phys. Rev.* **D58** (1998) 066005, `hep-th/9804174`.

[77] N. Sasakura and S. Sugimoto, "M theory description of 1/4 BPS states in N=4 supersymmetric Yang-Mills theory," `hep-th/9811087`.

[78] D. Tong, "A Note on 1/4 BPS states," `hep-th/9902005`.

[79] M. Gunaydin, D. Minic, and M. Zagermann, "Novel supermultiplets of $SU(2,2|4)$ and the AdS_5 / CFT_4 duality," `hep-th/9810226`.

[80] M. Gunaydin and N. Marcus, "The Spectrum of the S^5 Compactification of the Chiral $N = 2, D = 10$ Supergravity and the Unitary Supermultiplets of $U(2,2/4)$.," *Class. Quant. Grav.* **2** (1985) L11.

[81] L. Andrianopoli and S. Ferrara, "K-K Excitations on $AdS_5 \times S^5$ as $N = 4$ 'Primary' Superfields," *Phys. Lett.* **B430** (1998) 248–253, `hep-th/9803171`.

[82] L. Andrianopoli and S. Ferrara, "'Nonchiral' Primary Superfields in the AdS_{d+1} / CFT_d Correspondence," *Lett. Math. Phys.* **46** (1998) 265, `hep-th/9807150`.

[83] S. Ferrara and A. Zaffaroni, "Bulk gauge fields in AdS supergravity and supersingletons," `hep-th/9807090`.

[84] M. Gunaydin, D. Minic, and M. Zagermann, "4-D doubleton conformal theories, CPT and II B string on $\mathrm{AdS}_5 \times S^5$," *Nucl. Phys.* **B534** (1998) 96–120, `hep-th/9806042`.

[85] L. Andrianopoli and S. Ferrara, "On Short and Long $SU(2,2/4)$ Multiplets in the AdS/CFT Correspondence," `hep-th/9812067`.

[86] M. Flato and C. Fronsdal, "One massless particle equals two Dirac singletons: elementary particles in a curved space. 6," *Lett. Math. Phys.* **2** (1978) 421.

[87] C. Fronsdal and M. Flato, "On Dis and Racs," *Phys. Lett.* **97B** (1980) 236.

[88] M. Flato and C. Fronsdal, "Quantum field theory of singletons: the Rac," *J. Math. Phys.* **22** (1981) 1100.

[89] E. Angelopoulos, M. Flato, C. Fronsdal, and D. Sternheimer, "Massless Particles, Conformal Group and de Sitter Universe," *Phys. Rev.* **D23** (1981) 1278.

[90] H. Nicolai and E. Sezgin, "Singleton representations of $OSp(N,4)$," *Phys. Lett.* **143B** (1984) 389.

[91] M. Gunaydin and N. Marcus, "The unitary supermultiplet of $N = 8$ conformal superalgebra involving fields of spin ≤ 2," *Class. Quant. Grav.* **2** (1985) L19.

[92] M. Flato and C. Fronsdal, "Quarks of singletons?," *Phys. Lett.* **B172** (1986) 412.

[93] S. Ferrara and C. Fronsdal, "Conformal Maxwell theory as a singleton field theory on AdS_5, IIB three-branes and duality," *Class. Quant. Grav.* **15** (1998) 2153, `hep-th/9712239`.

[94] S. Ferrara and C. Fronsdal, "Gauge fields and singletons of AdS_{2p+1}," *Lett. Math. Phys.* **46** (1998) 157, `hep-th/9806072`.

[95] S. Ferrara and C. Fronsdal, "Gauge fields as composite boundary excitations," *Phys. Lett.* **B433** (1998) 19–28, `hep-th/9802126`.

[96] V. G. Kac, "Representations of classical Lie superalgebras,". Proceedings, Differential Geometrical Methods In Mathematical Physics.Ii., Berlin 1977, 597-626.

[97] V. K. Dobrev and V. B. Petkova, "Group theoretical approach to extended conformal supersymmetry: function space realizations and invariant differential operators," *Fortschr. Phys.* **35** (1987) 537.

[98] V. K. Dobrev and V. B. Petkova, "All positive energy unitary irreducible representations of extended conformal supersymmetry," *Phys. Lett.* **162B** (1985) 127.

[99] N. Seiberg, "Notes on theories with 16 supercharges," *Nucl. Phys. Proc. Suppl.* **67** (1998) 158, `hep-th/9705117`.

[100] S. Minwalla, "Restrictions imposed by superconformal invariance on quantum field theories," *Adv. Theor. Math. Phys.* **2** (1998) 781–846, `hep-th/9712074`.

[101] S. Ferrara, C. Fronsdal, and A. Zaffaroni, "On N=8 supergravity on AdS_5 and N=4 superconformal Yang- Mills theory," *Nucl. Phys.* **B532** (1998) 153, `hep-th/9802203`.

[102] S. Ferrara, M. A. Lledo, and A. Zaffaroni, "Born-Infeld corrections to D3-brane action in AdS(5) x S(5) and N=4, d = 4 primary superfields," *Phys. Rev.* **D58** (1998) 105029, `hep-th/9805082`.

[103] H. J. Kim, L. J. Romans, and P. van Nieuwenhuizen, "The mass spectrum of chiral $\mathcal{N} = 2$ $d = 10$ supergravity on S^5," *Phys. Rev.* **D32** (1985) 389.

[104] P. Berglund, E. G. Gimon, and D. Minic, "The AdS/CFT correspondence and spectrum generating algebras," `hep-th/9905097`.

[105] O. Aharony, M. Berkooz, D. Kutasov, and N. Seiberg, "Linear Dilatons, NS Five-Branes and Holography," *JHEP* **10** (1998) 004, `hep-th/9808149`.

[106] E. Witten, "Baryons and branes in anti-de Sitter space," *JHEP* **07** (1998) 006, `hep-th/9805112`.

[107] I. R. Klebanov and E. Witten, "Superconformal field theory on three-branes at a Calabi-Yau singularity," *Nucl. Phys.* **B536** (1998) 199, `hep-th/9807080`.

[108] S. S. Gubser and I. R. Klebanov, "Baryons and domain walls in an N=1 superconformal gauge theory," *Phys. Rev.* **D58** (1998) 125025, `hep-th/9808075`.

[109] D. Z. Freedman, S. D. Mathur, A. Matusis, and L. Rastelli, "Correlation functions in the CFT_d / AdS_{d+1} correspondence," `hep-th/9804058`.

[110] G. Chalmers, H. Nastase, K. Schalm, and R. Siebelink, "R Current Correlators in $N = 4$ Super Yang-Mills Theory from Anti-de Sitter Supergravity," *Nucl. Phys.* **B540** (1999) 247, `hep-th/9805105`.

[111] S. Deser and A. Schwimmer, "Geometric classification of conformal anomalies in arbitrary dimensions," *Phys. Lett.* **B309** (1993) 279–284, hep-th/9302047.

[112] M. J. Duff, "Twenty years of the Weyl anomaly," *Class. Quant. Grav.* **11** (1994) 1387–1404, hep-th/9308075.

[113] H. Osborn and A. Petkos, "Implications of conformal invariance in field theories for general dimensions," *Ann. Phys.* **231** (1994) 311–362, hep-th/9307010.

[114] D. Anselmi, M. Grisaru, and A. Johansen, "A Critical Behavior of Anomalous Currents, Electric-Magnetic Universality and CFT in Four-Dimensions," *Nucl. Phys.* **B491** (1997) 221–248, hep-th/9601023.

[115] J. Erdmenger and H. Osborn, "Conserved currents and the energy momentum tensor in conformally invariant theories for general dimensions," *Nucl. Phys.* **B483** (1997) 431–474, hep-th/9605009.

[116] D. Anselmi, D. Z. Freedman, M. T. Grisaru, and A. A. Johansen, "Nonperturbative Formulas for Central Functions of Supersymmetric Gauge Theories," *Nucl. Phys.* **B526** (1998) 543, hep-th/9708042.

[117] M. Henningson and K. Skenderis, "The Holographic Weyl anomaly," *JHEP* **07** (1998) 023, hep-th/9806087.

[118] M. Henningson and K. Skenderis, "Holography and the Weyl anomaly," hep-th/9812032.

[119] V. Balasubramanian and P. Kraus, "A Stress Tensor for Anti-de Sitter Gravity," hep-th/9902121.

[120] W. Mueck and K. S. Viswanathan, "Counterterms for the Dirichlet prescription of the AdS/CFT correspondence," hep-th/9905046.

[121] S. Nojiri and S. D. Odintsov, "Conformal anomaly for dilaton coupled theories from AdS / CFT correspondence," *Phys. Lett.* **B444** (1998) 92, hep-th/9810008.

[122] D. Anselmi and A. Kehagias, "Subleading corrections and central charges in the AdS / CFT correspondence," hep-th/9812092.

[123] O. Aharony, J. Pawelczyk, S. Theisen, and S. Yankielowicz, "A Note on Anomalies in the AdS / CFT Correspondence," hep-th/9901134.

[124] M. Blau, K. S. Narain, and E. Gava, "On subleading contributions to the AdS / CFT trace anomaly," `hep-th/9904179`.

[125] S. Nojiri and S. D. Odintsov, "On the conformal anomaly from higher derivative gravity in AdS / CFT correspondence," `hep-th/9903033`.

[126] G. W. Gibbons and S. W. Hawking, "Action integrals and partition functions in quantum gravity," *Phys. Rev.* **D15** (1977) 2752–2756.

[127] I. Y. Aref'eva and I. V. Volovich, "On large N conformal theories, field theories in anti-de Sitter space and singletons," `hep-th/9803028`.

[128] W. Muck and K. S. Viswanathan, "Conformal field theory correlators from classical scalar field theory on AdS_{d+1}," *Phys. Rev.* **D58** (1998) 041901, `hep-th/9804035`.

[129] H. Liu and A. A. Tseytlin, "D = 4 superYang-Mills, D = 5 gauged supergravity, and D = 4 conformal supergravity," *Nucl. Phys.* **B533** (1998) 88, `hep-th/9804083`.

[130] W. Muck and K. S. Viswanathan, "Conformal field theory correlators from classical field theory on anti-de Sitter space. 2. Vector and spinor fields," *Phys. Rev.* **D58** (1998) 106006, `hep-th/9805145`.

[131] S. N. Solodukhin, "Correlation functions of boundary field theory from bulk Green's functions and phases in the boundary theory," *Nucl. Phys.* **B539** (1999) 403, `hep-th/9806004`.

[132] S. Lee, S. Minwalla, M. Rangamani, and N. Seiberg, "Three point functions of chiral operators in D = 4, N=4 SYM at large N," *Adv. Theor. Math. Phys.* **2** (1999) 697, `hep-th/9806074`.

[133] H. Liu and A. A. Tseytlin, "On four point functions in the CFT / AdS correspondence," *Phys. Rev.* **D59** (1999) 086002, `hep-th/9807097`.

[134] E. D'Hoker, D. Z. Freedman, and W. Skiba, "Field theory tests for correlators in the AdS / CFT correspondence," *Phys. Rev.* **D59** (1999) 045008, `hep-th/9807098`.

[135] D. Z. Freedman, S. D. Mathur, A. Matusis, and L. Rastelli, "Comments on 4 point functions in the CFT / AdS correspondence," `hep-th/9808006`.

[136] E. D'Hoker and D. Z. Freedman, "Gauge boson exchange in AdS_{d+1}," `hep-th/9809179`.

[137] G. Chalmers and K. Schalm, "The Large N_c Limit of Four-Point Functions in $N = 4$ Super Yang-Mills Theory from Anti-de Sitter Supergravity," hep-th/9810051.

[138] W. Muck and K. S. Viswanathan, "The Graviton in the AdS-CFT correspondence: Solution via the Dirichlet boundary value problem," hep-th/9810151.

[139] E. D'Hoker and D. Z. Freedman, "General scalar exchange in AdS_{d+1}," hep-th/9811257.

[140] P. Minces and V. O. Rivelles, "Chern-Simons theories in the AdS / CFT correspondence," hep-th/9902123.

[141] G. Arutyunov and S. Frolov, "Three-Point Green Function of the Stress-Energy Tensor in the AdS / CFT Correspondence," hep-th/9901121.

[142] D. Z. Freedman, K. Johnson, and J. I. Latorre, "Differential regularization and renormalization: A New method of calculation in quantum field theory," *Nucl. Phys.* **B371** (1992) 353–414.

[143] M. Henningson and K. Sfetsos, "Spinors and the AdS / CFT correspondence," *Phys. Lett.* **B431** (1998) 63–68, hep-th/9803251.

[144] A. M. Ghezelbash, K. Kaviani, I. Shahrokh AF Tehran Parvizi, and A. H. Fatollahi, "Interacting spinors - scalars and AdS / CFT correspondence," *Phys. Lett.* **B435** (1998) 291, hep-th/9805162.

[145] G. E. Arutyunov and S. A. Frolov, "On the origin of supergravity boundary terms in the AdS / CFT correspondence," *Nucl. Phys.* **B544** (1999) 576, hep-th/9806216.

[146] G. E. Arutyunov and S. A. Frolov, "Antisymmetric tensor field on AdS(5)," *Phys. Lett.* **B441** (1998) 173, hep-th/9807046.

[147] W. S. l'Yi, "Holographic projection of massive vector fields in AdS / CFT correspondence," hep-th/9808051.

[148] A. Volovich, "Rarita-Schwinger field in the AdS / CFT correspondence," *JHEP* **09** (1998) 022, hep-th/9809009.

[149] W. S. l'Yi, "Generating functionals of correlation functions of p form currents in AdS / CFT correspondence," hep-th/9809132.

[150] W. S. l'Yi, "Correlators of currents corresponding to the massive p form fields in AdS / CFT correspondence," *Phys. Lett.* **B448** (1999) 218, hep-th/9811097.

[151] A. S. Koshelev and O. A. Rytchkov, "Note on the massive Rarita-Schwinger field in the AdS / CFT correspondence," *Phys. Lett.* **B450** (1999) 368, hep-th/9812238.

[152] R. C. Rashkov, "Note on the boundary terms in AdS / CFT correspondence for Rarita-Schwinger field," hep-th/9904098.

[153] A. Polishchuk, "Massive symmetric tensor field on AdS," hep-th/9905048.

[154] D. Z. Freedman, S. S. Gubser, K. Pilch, and N. P. Warner, "Renormalization group flows from holography supersymmetry and a c theorem," hep-th/9904017.

[155] G. Dall'Agata, K. Lechner, and D. Sorokin, "Covariant actions for the bosonic sector of d = 10 IIB supergravity," *Class. Quant. Grav.* **14** (1997) L195–L198, hep-th/9707044.

[156] G. Dall'Agata, K. Lechner, and M. Tonin, "D = 10, N = IIB supergravity: Lorentz invariant actions and duality," *JHEP* **07** (1998) 017, hep-th/9806140.

[157] G. E. Arutyunov and S. A. Frolov, "Quadratic action for Type IIB supergravity on $AdS_5 \times S^5$," hep-th/9811106.

[158] J. H. Schwarz, "Covariant Field Equations of Chiral $N = 2, D = 10$ Supergravity," *Nucl. Phys.* **B226** (1983) 269.

[159] P. S. Howe and P. C. West, "The Complete $N = 2, d = 10$ Supergravity," *Nucl. Phys.* **B238** (1984) 181.

[160] E. D'Hoker, D. Z. Freedman, S. D. Mathur, A. Matusis, and L. Rastelli, "Graviton exchange and complete four point functions in the AdS / CFT correspondence," hep-th/9903196.

[161] E. D'Hoker, D. Z. Freedman, S. D. Mathur, A. Matusis, and L. Rastelli, "Graviton and gauge boson propagators in AdS(d+1)," hep-th/9902042.

[162] H. Liu, "Scattering in anti-de Sitter space and operator product expansion," hep-th/9811152.

[163] G. Chalmers and K. Schalm, "Holographic normal ordering and multiparticle states in the AdS / CFT correspondence," hep-th/9901144.

[164] E. D'Hoker, D. Z. Freedman, and L. Rastelli, "AdS/CFT four point functions: How to succeed at Z integrals without really trying," hep-th/9905049.

[165] D. J. Gross and H. Ooguri, "Aspects of Large N Gauge Theory Dynamics as seen by String Theory," *Phys. Rev.* **D58** (1998) 106002, hep-th/9805129.

[166] J. Maldacena, "Wilson loops in large N field theories," *Phys. Rev. Lett.* **80** (1998) 4859, hep-th/9803002.

[167] S.-J. Rey and J. Yee, "Macroscopic strings as heavy quarks in large N gauge theory and anti-de Sitter supergravity," hep-th/9803001.

[168] N. Drukker, D. J. Gross, and H. Ooguri, "Wilson Loops and Minimal Surfaces," hep-th/9904191.

[169] G. 't Hooft, "A Property of Electric and Magnetic Flux in Nonabelian Gauge Theories," *Nucl. Phys.* **B153** (1979) 141.

[170] S. Forste, D. Ghoshal, and S. Theisen, "Stringy corrections to the Wilson loop in N=4 superYang- Mills theory," hep-th/9903042.

[171] S. Naik, "Improved heavy quark potential at finite temperature from Anti-de Sitter supergravity," hep-th/9904147.

[172] J. Greensite and P. Olesen, "Worldsheet Fluctuations and the Heavy Quark Potential in the AdS/CFT Approach," hep-th/9901057.

[173] J. A. Minahan, "Quark - monopole potentials in large N superYang-Mills," *Adv. Theor. Math. Phys.* **2** (1998) 559, hep-th/9803111.

[174] J. A. Minahan and N. P. Warner, "Quark potentials in the Higgs phase of large N supersymmetric Yang-Mills theories," *JHEP* **06** (1998) 005, hep-th/9805104.

[175] I. I. Kogan and O. A. Solovev, "Gravitationally dressed RG flows and zigzag invariant strings," *Phys. Lett.* **B442** (1998) 136, hep-th/9807223.

[176] I. I. Kogan and O. A. Solovev, "On zigzag invariant strings," hep-th/9901131.

[177] S. Nojiri and S. D. Odintsov, "Running gauge coupling and quark - anti-quark potential from dilatonic gravity," hep-th/9904036.

[178] K. Zarembo, "Wilson loop correlator in the AdS / CFT correspondence," hep-th/9904149.

[179] E. Alvarez, C. Gomez, and T. Ortin, "String representation of Wilson loops," *Nucl. Phys.* **B545** (1999) 217, hep-th/9806075.

[180] C. R. Graham and E. Witten, "Conformal anomaly of submanifold observables in AdS / CFT correspondence," hep-th/9901021.

[181] D. Berenstein, R. Corrado, W. Fischler, and J. Maldacena, "The Operator Product Expansion for Wilson Loops and Surfaces in the Large N Limit," hep-th/9809188.

[182] S. S. Gubser, I. R. Klebanov, and A. W. Peet, "Entropy and temperature of black 3-branes," *Phys. Rev.* **D54** (1996) 3915–3919, hep-th/9602135.

[183] A. Strominger, unpublished notes, November, 1997.

[184] S. S. Gubser, I. R. Klebanov, and A. A. Tseytlin, "Coupling constant dependence in the thermodynamics of N=4 supersymmetric Yang-Mills theory," *Nucl. Phys.* **B534** (1998) 202, hep-th/9805156.

[185] G. T. Horowitz and J. Polchinski, "A Correspondence principle for black holes and strings," *Phys. Rev.* **D55** (1997) 6189–6197, hep-th/9612146.

[186] N. Itzhaki, "A Comment on the entropy of strongly coupled N=4," hep-th/9904035.

[187] A. Fotopoulos and T. R. Taylor, "Comment on two loop free energy in N=4 supersymmetric Yang- Mills theory at finite temperature," hep-th/9811224.

[188] R. M. Wald, "Black hole entropy in Noether charge," *Phys. Rev.* **D48** (1993) 3427–3431, gr-qc/9307038.

[189] M. Li, "Evidence for large N phase transition in N=4 superYang- Mills theory at finite temperature," *JHEP* **03** (1999) 004, hep-th/9807196.

[190] Y. hong Gao and M. Li, "Large N strong / weak coupling phase transition and the correspondence principle," hep-th/9810053.

[191] S. W. Hawking and D. N. Page, "Thermodynamics of black holes in anti-de Sitter space," *Commun. Math. Phys.* **87** (1983) 577.

[192] S. S. Gubser, "Thermodynamics of spinning D3-branes," hep-th/9810225.

[193] K. Landsteiner, "String corrections to the Hawking-Page phase transition," *Mod. Phys. Lett.* **A14** (1999) 379, hep-th/9901143.

[194] R.-G. Cai and K.-S. Soh, "Critical behavior in the rotating D-branes," hep-th/9812121.

[195] M. Cvetic and S. S. Gubser, "Phases of R charged black holes, spinning branes and strongly coupled gauge theories," hep-th/9902195.

[196] A. Chamblin, R. Emparan, C. V. Johnson, and R. C. Myers, "Charged AdS Black Holes and Catastrophic Holography," hep-th/9902170.

[197] M. Cvetic and S. S. Gubser, "Thermodynamic stability and phases of general spinning branes," hep-th/9903132.

[198] M. M. Caldarelli and D. Klemm, "M theory and stringy corrections to Anti-de Sitter black holes and conformal field theories," hep-th/9903078.

[199] P. Kraus, F. Larsen, and S. P. Trivedi, "The Coulomb branch of gauge theory from rotating branes," *JHEP* **03** (1999) 003, hep-th/9811120.

[200] A. A. Tseytlin and S. Yankielowicz, "Free energy of N=4 superYang-Mills in Higgs phase and nonextremal D3-brane interactions," *Nucl. Phys.* **B541** (1999) 145, hep-th/9809032.

[201] D. Birmingham, "Topological black holes in Anti-de Sitter space," *Class. Quant. Grav.* **16** (1999) 1197, hep-th/9808032.

[202] J. Louko and D. Marolf, "Single exterior black holes and the AdS / CFT conjecture," *Phys. Rev.* **D59** (1999) 066002, hep-th/9808081.

[203] S. W. Hawking, C. J. Hunter, and M. M. Taylor-Robinson, "Rotation and the AdS / CFT correspondence," *Phys. Rev.* **D59** (1999) 064005, hep-th/9811056.

[204] A. W. Peet and S. F. Ross, "Microcanonical phases of string theory on $AdS_m \times S^n$," *JHEP* **12** (1998) 020, hep-th/9810200.

[205] S.-J. Rey, S. Theisen, and J. Yee, "Wilson-Polyakov Loop at Finite Temperature in Large N Gauge Theory and Anti-de Sitter Supergravity," *Nucl.Phys.* **B527** (1998) 171–186, `hep-th/9803135`.

CFT & FRW

Erik Verlinde*

Joseph Henry Laboratories, Princeton University, Princeton, USA

Lectures given at the
Spring School on Superstrings and Related Matters
Trieste, 2 – 10 April 2001

LNS027002

*erikv@feynman.princeton.edu

Contents

1 Introduction

The holographic principle is based on the idea that for a given volume V the state of maximal entropy is given by the largest black hole that fits inside V. 't Hooft and Susskind [1] argued on this basis that the microscopic entropy S associated with the volume V should be less than the Bekenstein-Hawking entropy

$$S \leq \frac{A}{4G} \tag{1}$$

of a black hole with horizon area A equal to the surface area of the boundary of V. Here the dependence on Newton's constant G is made explicit, but as usual $\hbar$ and c are set to one.

To shed further light on the holographic principle and the entropy bounds derived from it, we consider a standard cosmology of a closed radiation dominated Friedmann-Robertson-Walker (FRW) universe with general space-time dimension

$$D = n + 1.$$

The metric takes the form

$$ds^2 = -dt^2 + R^2(t) d\Omega_n^2 \tag{2}$$

where $R(t)$ represents the radius of the universe at a given time t and $d\Omega_n^2$ is a short hand notation for the metric on the unit n-sphere S^n. Hence, the spatial section of a $(n+1)$d closed FRW universe is an n-sphere with a finite volume

$$V = \mathrm{Vol}(S^n) R^n.$$

The holographic bound is in its naive form (1) not really applicable to a closed universe, since space has no boundary. Furthermore, the argumentation leading to (1) assumes that it's possible to form a black hole that fills the entire volume. This is not true in a cosmological setting, because the expansion rate H of the universe as well as the given value of the total energy E restrict the maximal size of black hole. As will be discussed in these notes, this will lead to a modified version of the holographic bound.

The radiation in the universe is described by a conformal field theory (CFT) with a very large central charge c. In a finite volume the energy E has a Casimir contribution proportional to c. Due to this Casimir effect, the entropy S is no longer a purely extensive function of E and V. The entropy of a $(1+1)$d CFT is given by the well-known Cardy formula [2]

$$S = 2\pi \sqrt{\frac{c}{6}\left(L_0 - \frac{c}{24}\right)}, \tag{3}$$

where L_0 represents the product ER of the energy and radius, and the shift of $\frac{c}{24}$ is caused by the Casimir effect. In these notes we show that, after making the appropriate identifications for L_0 and c, the same Cardy formula is also valid for CFTs in other dimensions. This is rather surprising, since the standard derivation of the Cardy formula based on modular invariance only appears to work for $n = 1$. By defining the central charge c in terms of the Casimir energy, we are able to argue that the Cardy formula is universally valid. Specifically, we will show that with the appropriate identifications, the entropy S for a $n+1$ dimensional CFT with an AdS-dual is exactly given by (3).

There appears to be a deep and fundamental connection between the holographic principle, the entropy formulas for the CFT, and the FRW equations for a radiation dominated universe. In $n+1$ dimensions the FRW equations are given by

$$H^2 = \frac{16\pi G}{n(n-1)}\frac{E}{V} - \frac{1}{R^2} \tag{4}$$

$$\dot{H} = -\frac{8\pi G}{n-1}\left(\frac{E}{V} + p\right) + \frac{1}{R^2} \tag{5}$$

where $H = \dot{R}/R$ is the Hubble parameter, and the dot denotes as usual differentiation with respect to the time t. The FRW equations are usually written in terms of the energy density $\rho = E/V$, but for the present study it is more convenient to work with the total energy E and entropy S instead of their respective densities ρ and $s = S/V$. Note that the cosmological constant has been put to zero.

Entropy and energy momentum conservation together with the equation of state $p = E/nV$ imply that E/V and p decrease in the usual way like $R^{-(n+1)}$. Hence, the cosmological evolution follows the standard scenario for a closed radiation dominated FRW universe. After the initial Big Bang, the universe expands until it reaches a maximum radius, the universe subsequently re-collapses and ends with a Big Crunch. No surprises happen in this respect.

The fun starts when one compares the holographic entropy bound with the entropy formulas for the CFT. We will show that when the bound is saturated the FRW equations and entropy formulas of the CFT merge together into one set of equation. One easily checks on the back of an envelope that via the substitutions

$$
\begin{aligned}
2\pi L_0 &\Rightarrow \frac{2\pi}{n}ER \\
2\pi\frac{c}{12} &\Rightarrow (n-1)\frac{V}{4GR} \\
S &\Rightarrow (n-1)\frac{HV}{4G}
\end{aligned} \tag{6}
$$

the Cardy formula (3) *exactly* turns into the $n+1$ dimensional Friedmann equation (4). This observation appears as a natural consequence of the holographic principle. In sections 2 and 3 we introduce three cosmological bounds each corresponding to one of the equations in (6) The Cardy formula is presented and derived in section 4. In section 5 we introduce a new cosmological bound, and show that the FRW equations and the entropy formulas are exactly matched when the bound is saturated. In section 6 we present a graphical picture of the entropy bounds and their time evolution.

2 Cosmological entropy bounds

This section is devoted to the description of three cosmological entropy bounds: the Bekenstein bound, the holographic Bekenstein-Hawking bound, and the Hubble bound. The relation with the holographic bound proposed by Fischler-Susskind and Bousso (FSB) will also be clarified.

2.1 The Bekenstein bound

Bekenstein [4] was the first to propose a bound on the entropy of a macroscopic system. He argued that for a system with limited self-gravity, the total entropy S is less or equal than a multiple of the product of the energy and the linear size of the system. In the present context, namely that of a closed radiation dominated FRW universe with radius R, the appropriately normalized Bekenstein bound is

$$S \leq S_B \tag{7}$$

where the Bekenstein entropy S_B is defined by

$$S_B \equiv \frac{2\pi}{n}ER. \tag{8}$$

The bound is most powerful for relatively low energy density or small volumes. This is due to the fact that S_B is super-extensive: under $V \to \lambda V$ and $E \to \lambda E$ it scales like $S_B \to \lambda^{1+1/n} S_B$.

For a radiation dominated universe the Bekenstein entropy is constant throughout the entire evolution, since $E \sim R^{-1}$. Therefore, once the Bekenstein bound is satisfied at one instance, it will remain satisfied at all times as long as the entropy S does not change. The Bekenstein entropy is the most natural generalization of the Virasoro operator $2\pi L_0$ to arbitrary dimensions, as is apparent from (6). Indeed, it is useful to think about S_B not really as an entropy but rather as the energy measured with respect to an appropriately chosen conformal time coordinate.

2.2 The Bekenstein-Hawking bound

The Bekenstein-bound is supposed to hold for systems with limited self-gravity, which means that the gravitational self-energy of the system is small compared to the total energy E. In the current situation this implies, concretely, that the Hubble radius H^{-1} is larger than the radius R of the universe. So the Bekenstein bound is only appropriate in the parameter range $HR \leq 1$. In a strongly self-gravitating universe, that is for $HR \geq 1$, the possibility of black hole formation has to be taken into account, and the entropy bound must be modified accordingly. Here the general philosophy of the holographic principle becomes important.

It follows directly from the Friedmann equation (4) that

$$HR \lesseqgtr 1 \qquad \Leftrightarrow \qquad S_B \lesseqgtr (n-1)\frac{V}{4GR} \tag{9}$$

Therefore, to decide whether a system is strongly or weakly gravitating one should compare the Bekenstein entropy S_B with the quantity

$$S_{BH} \equiv (n-1)\frac{V}{4GR}. \tag{10}$$

When $S_B \leq S_{BH}$ the system is weakly gravitating, while for $S_B \geq S_{BH}$ the self-gravity is strong. We will identify S_{BH} with the holographic Bekenstein-Hawking entropy of a black hole with the size of the universe. S_{BH} indeed grows like an area instead of the volume, and for a closed universe it is the closest one can come to the usual expression $A/4G$.

As will become clear in these notes, the role of S_{BH} is not to serve as a bound on the total entropy, but rather on a sub-extensive component of the entropy that is association with the Casimir energy of the CFT. The relation (6) suggests that the Bekenstein-Hawking entropy is closely related to the central charge c. Indeed, it is well-known from $(1+1)d$ CFT that the central charge characterizes the number of degrees of freedom may be even better than the entropy. This fact will be further explained in sections 5 and 6, when we describe a new cosmological bound on the Casimir energy and its associated entropy.

2.3 The Hubble entropy bound

The Bekenstein entropy S_B is equal to the holographic Bekenstein-Hawking entropy S_{BH} precisely when $HR = 1$. For $HR > 1$ one has $S_B > S_{BH}$ and the Bekenstein bound has to be replaced by a holographic bound. A naive application of the holographic principle would imply that the total entropy S should be bounded by S_{BH}. This turns out to be incorrect, however, since a purely holographic bound assumes the existence of arbitrarily large black holes, and is irreconcilable with a finite homogeneous entropy density.

Following earlier work by Fischler and Susskind [5], it was argued by Easther and Lowe [6], Veneziano [7], Bak and Rey [8], Kaloper and Linde [9], that the maximal entropy inside the universe is produced by black holes of the size of the Hubble horizon, see also [10]. Following the usual holographic arguments one then finds that the total entropy should be less or equal than the Bekenstein-Hawking entropy of a Hubble size black hole times the number N_H of Hubble regions in the universe. The entropy of a Hubble size black hole is roughly $HV_H/4G$, where V_H is the volume of a single Hubble region. Combined with the fact that $N_H = V/V_H$ one obtains an upper bound on the total entropy S given by a multiple of $HV/4G$. The presented arguments of [6, 8, 9, 7] are not sufficient to determine the precise pre-factor, but in the following subsection we will fix the normalization of the bound by using a local version of the Fischler-Susskind-Bousso formulation of the holographic principle. The appropriately normalized entropy bound takes the form

$$S \leq S_H \qquad \text{for} \qquad HR \geq 1 \tag{11}$$

with

$$S_H \equiv (n-1)\frac{HV}{4G}. \tag{12}$$

The Hubble bound is only valid for $HR \geq 1$. In fact, it is easily seen that for $HR \leq 1$ the bound will at some point be violated. For example, when the universe reaches its maximum radius and starts to re-collapse the Hubble constant H vanishes, while the entropy is still non-zero.[1] This should not really come as a surprise, since the Hubble bound was based on the idea that the maximum size of a black hole is equal to the Hubble radius. Clearly, when the radius R of the universe is smaller than the Hubble radius H^{-1} one should reconsider the validity of the bound. In this situation, the self-gravity of the universe is less important, and the appropriate entropy bound is

$$S \leq S_B \qquad \text{for} \qquad HR \leq 1 \tag{13}$$

2.4 The Hubble bound and the FSB prescription

Fischler, Susskind, and subsequently Bousso [12], have proposed an ingenious version of the holographic bound that restricts the entropy flow through contracting light sheets. The FSB-bound works well in many situations, but, so far, no microscopic derivation has been given. Wald and collaborators [13] have shown that the FSB bound follows from local inequalities on the entropy density and the stress energy. The analysis of [13] suggests the existence a local version of the FSB entropy bound, one that does not involve global information about the causal structure of the universe, see also [11]. The idea of to formulate the holographic principle via entropy flow through light sheets also occurred in the work of Jacobson [14], who used it to derive an intriguing relation between the Einstein equations and the first law of thermodynamics. In this subsection, a local FSB bound will be presented that leads to a precisely normalized upper limit on the entropy in terms of the Hubble constant.

According to the original FSB proposal, the entropy flow S through a contracting light sheet is less or equal to $A/4G$, where A is the area of the surface from which the light sheet originates. The following infinitesimal version of this FSB prescription will lead to the Hubble bound. For every $n-1$ dimensional surface at time $t+dt$ with area $A+dA$ one demands that

$$dS \leq \frac{dA}{4G}, \tag{14}$$

where dS denotes the entropy flow through the infinitesimal light sheets originating at the surface at $t+dt$ and extending back to time t, and dA represents the increase in area between t and $t+dt$. For a surface that is kept fixed in co-moving coordinates the area A changes as a result of the

[1] To avoid this problem a different covariant version of the Hubble bound was proposed in [11].

Hubble expansion by an amount

$$dA = (n-1)HA\,dt, \tag{15}$$

where the factor $n-1$ simply follows from the fact that $A \sim R^{n-1}$. Now pick one of the two past light-sheets that originate at the surface: the inward or the outward going. The entropy flow through this light-sheet between t and $t+dt$ is given by the entropy density $s = S/V$ times the infinitesimal volume Adt swept out by the light-sheet. Hence,

$$dS = \frac{S}{V}A\,dt. \tag{16}$$

By inserting this result together with (15) into the infinitesimal FSB bound (14) one finds that the factor Adt cancels on both sides and one is left exactly with the Hubble bound $S \leq S_H$ with the Hubble entropy S_H given in (12). We stress that the relation with the FSB bound was merely used to fix the normalization of the Hubble bound, and should not be seen as a derivation.

3 Time-evolution of the entropy bounds

Let us now return to the three cosmological entropy bounds discussed in section 2. The Friedmann equation (4) can be re-written as an identity that relates the Bekenstein-, the Hubble-, and the Bekenstein-Hawking entropy. One easily verifies that the expressions given in (8), (10), and (12) satisfy the quadratic relation

$$S_H^2 + (S_B - S_{BH})^2 = S_B^2. \tag{17}$$

It is deliberately written in a Pythagorean form, since it suggests a useful graphical picture of the three entropy bounds. By representing each entropy by a line with length equal to its value one finds that due to the quadratic Friedmann relation (17) all three fit nicely together in one diagram, see figure 1. The circular form of the diagram reflects the fact that S_B is constant during the cosmological evolution. Only S_H and S_{BH} depend on time.

Let us introduce a conformal time coordinate via

$$Rd\eta = (n-1)dt \tag{18}$$

and let us compute the η-dependence of S_{BH} and S_H. For S_{BH} this easily follows from: $\dot{S}_{BH} = (n-1)HS_{BH} = (n-1)R^{-1}S_H$. For S_H the calculation is a bit more tedious, but with the help of the FRW equations, the result can eventually be put in the form

$$\begin{aligned}\frac{dS_H}{d\eta} &= S_B - S_{BH}, \\ \frac{dS_{BH}}{d\eta} &= -S_H. \end{aligned} \tag{19}$$

These equations show that the conformal time coordinate η can be identified with the angle η, as already indicated in figure 1. As time evolves the Hubble entropy S_H rotates into the combination $S_B - S_{BH}$ and visa versa. Equation (19) can be integrated to

$$\begin{aligned} S_H &= S_B \sin\eta \\ S_{BH} &= S_B(1-\cos\eta) \end{aligned} \tag{20}$$

The conformal time coordinate η plays the role of the time on a cosmological clock that only goes around once: at $\eta = 0$ time starts with a Big Bang and at $\eta = 2\pi$ it ends with a Big Crunch. Note that η is related to the parameter HR via

$$HR = \cot\frac{\eta}{2} \tag{21}$$

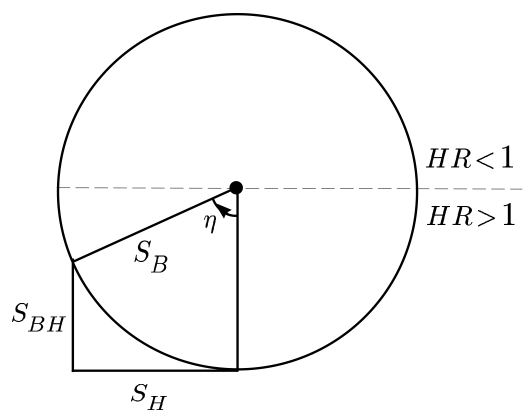


Figure 1: A graphical representation of the Bekenstein entropy S_B, the Hubble entropy S_H and the Bekenstein-Hawking entropy S_{BH}. The angle η corresponds to the conformal time coordinate. The value of each entropy is represented by an actual distance: S_B is constant, while S_H and S_{BH} change with time.

So far we have not yet included the CFT into our discussion. We will see that the entropy of the CFT will 'fill' part of the diagram, and in this way give rise to a special moment in time when the entropy bounds are saturated.

4 Casimir energy and the Cardy formula

We now turn to the discussion of the entropy of the CFT that lives inside the FRW universe. We begin with a study of the finite temperature Casimir energy with the aim to exhibit its relation with the entropy of the CFT. Subsequently a universal Cardy formula will be derived that expresses the entropy in terms of the energy and the Casimir energy, and is valid for all values of the spatial dimension n.

4.1 The Euler relation and Casimir energy

In standard textbooks on cosmology [15, 16] it is usually assumed that the total entropy S and energy E are extensive quantities. This fact is used for example to relate the entropy density s to the energy density ρ and pressure p, via $Ts = \rho + p$. For a thermodynamic system in finite volume V the energy $E(S,V)$, regarded as a function of entropy and volume, is called extensive when it satisfies $E(\lambda S, \lambda V) = \lambda E(S,V)$. Differentiating with respect to λ and putting $\lambda = 1$ leads to the Euler relation[2]

$$E = V\left(\frac{\partial E}{\partial V}\right)_S + S\left(\frac{\partial E}{\partial S}\right)_V \tag{22}$$

The first law of thermodynamics $dE = TdS - pdV$ can now be used to re-express the derivatives via the thermodynamic relations

$$\left(\frac{\partial E}{\partial V}\right)_S = -p, \qquad \left(\frac{\partial E}{\partial S}\right)_V = T. \tag{23}$$

The resulting equation $TS = E + pV$ is equivalent to the previously mentioned relation for the entropy density s.

For a CFT with a large central charge the entropy and energy are not purely extensive. In a finite volume the energy E of a CFT contains a non-extensive Casimir contribution proportional to c. This is well known in (1+1) dimensions where it gives rise to the familiar shift of $c/24$ in the L_0 Virasoro operator. The Casimir energy is the result of finite size effects in the quantum fluctuations of the CFT, and disappears when the volume becomes infinitely large. It therefore leads to sub-extensive contributions to the total energy E. Usually the Casimir effect is discussed at zero temperature [17],

[2] We assume here that there are no other thermodynamic functions like a chemical or electric potential. For a system with a 1st law like $TdS = dE + pdV + \mu dN + \Phi dQ$ the Euler relation reads $TS = E + pV + \mu N + \Phi Q$.

but a similar effect occurs at finite temperature. The value of the Casimir energy will in that case generically depend on the temperature T.

We will now define the Casimir energy as the violation of the Euler identity (22)

$$E_C \equiv n(E + pV - TS) \tag{24}$$

Here we inserted for convenience a factor equal to the spatial dimension n. From the previous discussion it is clear that E_C parameterizes the sub-extensive part of the total energy. The Casimir energy will just as the total energy be a function of the entropy S and the volume V. Under $S \to \lambda S$ and $V \to \lambda V$ it scales with a power of λ that is smaller than one. On general grounds one expects that the first subleading correction to the extensive part of the energy scales like

$$E_C(\lambda S, \lambda V) = \lambda^{1-2/n} E_C(S, V) \tag{25}$$

One possible way to see this is to write the energy as an integral over a local density expressed in the metric and its derivatives. Derivatives scale like $\lambda^{-1/n}$ and because derivatives come generally in pairs, the first subleading terms indeed has two additional factors of $\lambda^{-1/n}$. The total energy $E(S, V)$ may be written as a sum of two terms

$$E(S, V) = E_E(S, V) + \frac{1}{2} E_C(S, V) \tag{26}$$

where the first term E_E denotes the purely extensive part of the energy E and E_C represents the Casimir energy. Again the factor 1/2 has been put in for later convenience. By repeating the steps that lead to the Euler relation one easily verifies the defining equation (24) for the Casimir energy E_C.

4.2 Universality of the Cardy formula and the Bekenstein bound

Conformal invariance implies that the product ER is independent of the volume V, and is only a function of the entropy S. This holds for both terms E_E and E_C in (26). Combined with the known (sub-)extensive behavior of E_E and E_C this leads to the following general expressions

$$E_E = \frac{a}{4\pi R} S^{1+1/n} \qquad\qquad E_C = \frac{b}{2\pi R} S^{1-1/n}$$

where a and b are a priori arbitrary positive coefficients, independent of R and S. The factors of 4π and 2π are put in for convenience. With these expressions, one now easily checks that the entropy S can be written as

$$S = \frac{2\pi R}{\sqrt{ab}}\sqrt{E_C(2E - E_C)}. \tag{27}$$

If we ignore for a moment the normalization, this is exactly the Cardy formula: insert $ER = L_0$ and $E_C R = c/12$, and one recovers (3). It is obviously an interesting question to compute the coefficients a and b for various known conformal invariant field theories. This should be particularly straightforward for free field theories, such as $d = 4$ Maxwell theory and the self-dual tensor theory in $d = 6$. This question is left for future study.

Given the energy E the expression (27) has a maximum value. For all values of E, E_C and R one has the inequality

$$S \leq \frac{2\pi}{\sqrt{ab}} ER$$

This looks exactly like the Bekenstein bound, except that the pre-factor is in general different from the factor $2\pi/n$ used in the previous section. In fact, in the following subsection we will show that for CFTs with an AdS-dual description, the value of the product ab is exactly equal to n^2, so the upper limit is indeed exactly given by the Bekenstein entropy. Although we have no proof of this fact, we believe that the Bekenstein bound is universal. This implies that the product ab for all CFTs in $n+1$ dimensions is larger or equal than n^2. Only then it is guaranteed that the upper limit on the entropy is less or equal than S_B.

The upper limit is reached when the Casimir energy E_C is equal to the total energy E. Formally, when E_C becomes larger that E the entropy S will again decrease. Although in principle this is possible, we believe that in actual examples the Casimir energy E_C is bounded by the total energy E. So, from now on we assume that

$$E_C \leq E \tag{28}$$

In the next subsection we provide further evidence for this inequality.

From now on we will assume that we are dealing with a CFT for which $ab = n^2$. In the next section I will show that this includes all CFTs that have an AdS-dual description.

4.3 The Cardy formula derived from AdS/CFT

Soon after Maldacena's AdS/CFT-correspondence [18] was properly understood [19, 20] it was convincingly argued by Witten [21] that the entropy, energy and temperature of CFT at high temperatures can be identified with the entropy, mass, and Hawking temperature of the AdS black hole previously considered by Hawking and Page [22]. Using this duality relation the following expressions can be derived for the energy and entropy[3] for a $D = n + 1$ dimensional CFT on $R \times S^n$:

$$\begin{aligned} S &= \frac{c}{12}\frac{V}{L^n} \\ E &= \frac{c}{12}\frac{n}{4\pi L}\left(1+\frac{L^2}{R^2}\right)\frac{V}{L^n} \end{aligned} \tag{29}$$

The temperature again follows from the first law of thermodynamics. One finds

$$T = \frac{1}{4\pi L}\left((n+1) + (n-1)\,\frac{L^2}{R^2}\right). \tag{30}$$

The length scale L of the thermal CFT arises in the AdS/CFT correspondence as the curvature radius of the AdS black hole geometry. The expression for the energy clearly exhibits a non-extensive contribution, while also the temperature T contains a corresponding non-intensive term. Inserting the equations (29,30) into (24) yields the following result for the Casimir energy

$$E_C = \frac{c}{12}\frac{n}{2\pi R}\frac{V}{L^{n-1}R}. \tag{31}$$

Now let us come to the Cardy formula. The entropy S, energy E and Casimir energy E_C are expressed in c, L and R. Eliminating c and L leads to a unique expression for S in terms of E, E_C and R. One easily checks that it takes the form of the Cardy formula

$$S = \frac{2\pi R}{n}\sqrt{E_C\,(2E - E_C)} \tag{32}$$

[3]These expressions differ somewhat from the presented formulas in [21] due to the fact that (i) the $D + 1$ dimensional Newton constant has been eliminated using its relation with the central charge, (ii) the coordinates have been re-scaled so that the CFT lives on a sphere with radius equal to the black hole horizon. We will not discuss the AdS perspective in these notes, since the essential physics can be understood without introducing an extra dimension. The discussion of the CFT/FRW cosmology from an AdS perspective will be described elsewhere [3].

In the derivation of these formulas it was assumed that $R \gg L$. One may worry therefore that these formulas are not applicable in the early universe. Fortunately this is not a problem because during an adiabatic expansion both L and R scale in the same way so that R/L is fixed. Hence the formulas are valid provided the (fixed) ratio of the thermal wave-length and the radius R is much smaller than one. Effectively this means, as far as the CFT is concerned, we are in a high temperature regime. We note further that with in this parameter range, the Casimir energy E_C is indeed smaller than the total energy E.

Henceforth, we will assume that the CFT that describes the radiation in the FRW universe will have an entropy given by (32) with the specific normalization of $2\pi/n$. Note that if we take $n = 1$ and make the previously mentioned identifications $ER = L_0$ and $E_C R = c/12$ that this equation exactly coincides with the usual Cardy formula. We will therefore in the following refer to (32) simply as the Cardy formula. To check the precise coefficient of the Cardy formula for a CFT we have made use of the AdS/CFT correspondence. The rest of our discussions in the preceding and in the following sections do not depend on this correspondence.

5 A new cosmological bound

In this section a new cosmological bound will be presented, which is equivalent to the Hubble bound in the strongly gravitating phase, but which unlike the Hubble bound remains valid in the phase of weak self-gravity. When the bound is saturated the FRW equations and the CFT formulas for the entropy and Casimir energy completely coincide.

5.1 A cosmological bound on the Casimir energy

Let us begin by presenting another criterion for distinguishing between a weakly or strongly self-gravitating universe. When the universe goes from the strongly to the weakly self-gravitating phase, or vice-versa, the Bekenstein entropy S_B and the Bekenstein-Hawking entropy S_{BH} are equal in value. Given the radius R, we now define the 'Bekenstein-Hawking' energy E_{BH} as the value of the energy E for which S_B and S_{BH} are exactly equal. This leads to the condition

$$\frac{2\pi}{n} E_{BH} R \equiv (n-1)\frac{V}{4GR}. \tag{33}$$

One may interpret E_{BH} as the energy required to form a black hole with the size of the entire universe. Now, one easily verifies that

$$\begin{array}{rclll} E & \leq & E_{BH} & \text{for} & HR \leq 1 \\ E & \geq & E_{BH} & \text{for} & HR \geq 1. \end{array} \tag{34}$$

Hence, the universe is weakly self-gravitating when the total energy E is less than E_{BH} and strongly gravitating for $E > E_{BH}$.

We are now ready to present a proposal for a new cosmological bound. It is not formulated as a bound on the entropy S, but as a restriction on the Casimir energy E_C. The physical content of the bound is the Casimir energy E_C by itself can not be sufficient to form a universe-size black hole. Concretely, this implies that the Casimir energy E_C is less or equal to the Bekenstein-Hawking energy E_{BH}. Hence, we postulate

$$E_C \leq E_{BH} \tag{35}$$

To put the bound in a more conventional notation one may insert the definition (24) of the Casimir energy together with the defining relation (33) of the Bekenstein-Hawking energy. We leave this to the reader.

The virtues of the new cosmological bound are: (i) it is universally valid and does not break down for a weakly gravitating universe, (ii) in a strongly gravitating universe it is equivalent to the Hubble bound, (iii) it is purely holographic and can be formulated in terms of the Bekenstein-Hawking entropy S_{BH} of a universe-size black hole, (iv) when the bound is saturated the laws of general relativity and quantum field theory converge in a miraculous way, giving a strong indication that they have a common origin in a more fundamental unified theory.

The first point on the list is easily checked because E_C decays like R^{-1} while E_{BH} goes like R^{-n}. Only when the universe re-collapses and returns to the strongly gravitating phase the bound may again become saturated. To be able to proof the other points on the list of advertised virtues, we have to take a closer look to the FRW equations and the CFT formulas for the entropy an entropy.

5.2 A cosmological Cardy formula

To show the equivalence of the new bound with the Hubble bound let us write the Friedmann equation as an expression for the Hubble entropy S_H in terms of the energy E, the radius R and the Bekenstein-Hawking energy

E_{BH}. Here, the latter is used to remove the explicit dependence on Newton's constant G. The resulting expression is unique and takes the form

$$S_H = \frac{2\pi}{n} R \sqrt{E_{BH} \left(2E - E_{BH}\right)} \tag{36}$$

This is exactly the Cardy formula (32), except that the role of the Casimir energy E_C in CFT formula is now replaced by the Bekenstein-Hawking energy E_{BH}. Somehow, miraculously, the Friedmann equation knows about the Cardy formula for the entropy of a CFT!

With the help of (36) is now a straightforward matter to proof that when $HR \geq 1$ the new bound $E_C \leq E_{BH}$ is equivalent to the Hubble bound $S \leq S_H$. First, let us remind that for $HR \geq 1$ the energy E satisfies $E \geq E_{BH}$. Furthermore, we always assume that the Casimir energy E_C is smaller than the total energy E. The entropy S is a monotonically increasing function of E_C as long as $E_C \leq E$. Therefore in the range

$$E_C \leq E_{BH} \leq E \tag{37}$$

the maximum entropy is reached when $E_C = E_{BH}$. In that case the Cardy formula (32) for S exactly turns into the cosmological Cardy formula (36) for S_H. Therefore, we conclude that S_H is indeed the maximum entropy that can be reached when $HR \geq 1$. Note that in the weakly self-gravitating phase, when $E \leq E_{BH}$, the maximum is reached earlier, namely for $E_C = E$. The maximum entropy is in that case given by the bekenstein entropy S_B. The bifurcation of the new bound in two entropy bounds is a direct consequence of the fact that the Hubble bound is written as the square-root of a quadratic expression.

5.3 A limiting temperature

So far we have focussed on the entropy and energy of the CFT and on the first of the two FRW equations, usually referred to as the Friedmann equation. We will now show that also the second FRW equation has a counterpart in the CFT, and will lead to a constraint on the temperature T. Specifically, we will find that the bound on E_C implies that the temperature T in the early universe is bounded from below by

$$T_H \equiv -\frac{\dot{H}}{2\pi H} \tag{38}$$

The minus sign is necessary to get a positive result, since in a radiation dominated universe the expansion always slows down. Further, we assume that we are in the strongly self-gravitating phase with $HR \geq 1$, so that there is no danger of dividing by zero.

The second FRW equation in (5) can now be written as a relation between E_{BH}, S_H and T_H that takes the familiar form

$$E_{BH} = n(E + pV - T_H S_H) \tag{39}$$

This equation has exactly the same form as the defining relation $E_C = n(E + pV - TS)$ for the Casimir energy. In the strongly gravitating phase we have just argued that the bound $E_C \leq E_{BH}$ is equivalent to the Hubble bound $S \leq S_H$. It follows immediately that the temperature T in this phase is bounded from below by T_H. One has

$$T \geq T_H \qquad \text{for} \quad HR \geq 1 \tag{40}$$

When the cosmological bound is saturated all inequalities turn into equalities. The Cardy formula and the defining Euler relation for the Casimir energy in that case exactly match the Friedmann equation for the Hubble constant and the FRW equation for its time derivative.

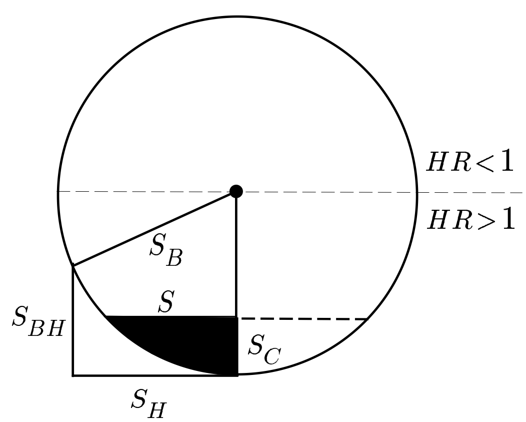


Figure 2: The entropy S and Casimir entropy S_C fill part of the cosmological entropy diagram. The diagram shows: (i) the Bekenstein bound $S \leq S_B$ is valid at all times (ii) the Hubble bound $S \leq S_H$ restricts the allowed range of η in the range $HR > 1$, but is violated for $HR < 1$, (iii) the new bound $S_C \leq S_{BH}$ is equivalent to the Hubble bound for $HR > 1$, and remains valid for $HR < 1$.

6 The entropy bounds revisited

We now return to the cosmological entropy bounds introduced in sections 2 and 3. In particular, we are interested in the way that the entropy of the CFT may be incorporated in the entropy diagram described in section 3. For this purpose it will be useful to introduce a non-extensive component of the entropy that is associated with the Casimir energy.

The cosmological bound $E_C \leq E_{BH}$ can also be formulated as an entropy bound, not on the total entropy, but on a non-extensive part of the entropy that is associated with the Casimir energy. In analogy with the definition of the Bekenstein entropy (8) one can introduce a 'Casimir' entropy defined by

$$S_C \equiv \frac{2\pi}{n} E_C R. \tag{41}$$

For $d = (1+1)$ the Casimir entropy is directly related to the central charge c. One has $S_C = 2\pi\, c/12$. In fact, it is more appropriate to interpret the Casimir entropy S_C as a generalization of the central charge to $n+1$ dimensions than what is usually called the central charge c. Indeed, if one introduces a dimensionless 'Virasoro operator' $\widetilde{L}_0 \equiv \frac{1}{2\pi} S_B$ and a new central charge $\frac{\widetilde{c}}{12} \equiv \frac{1}{2\pi} S_C$, the $n+1$ dimensional entropy formula (32) is exactly identical to (3).

The Casimir entropy S_C is sub-extensive because under $V \to \lambda V$ and $E \to \lambda E$ it goes like $S_C \to \lambda^{1-1/n} S_C$. In fact, it scales like an area! This is a clear indication that the Casimir entropy has something to do with holography. The total entropy S contains extensive as well as sub-extensive contributions. One can show that for $E_C \leq E$ the entropy S satisfies the following inequalities

$$S_C \leq S \leq S_B \tag{42}$$

where both equal signs can only hold simultaneously. The precise relation between S and its super- and sub-extensive counterparts S_B and S_C is determined by the Cardy formula, which can be expressed as

$$S^2 + (S_B - S_C)^2 = S_B^2. \tag{43}$$

This identity has exactly the same form as the relation (17) between the cosmological entropy bounds, except that in (17) the role of the entropy and Casimir entropy are taken over by the Hubble entropy S_H and Bekenstein-Hawking entropy S_{BH}. This fact will be used to incorporate the entropy S and the Casimir entropy S_C in the entropy diagram introduce in section 3.

The cosmological bound on the Casimir energy presented in the section 4 can be formulated as an upper limit on the Casimir entropy S_C. From the definitions of S_C and E_{BH} it follows directly that the bound $E_C \leq E_{BH}$ is equivalent to

$$S_C \leq S_{BH} \tag{44}$$

where we made use of the relation (33) to re-write E_{BH} again in terms of the Bekenstein-Hawking entropy S_{BH}. Thus the bound puts a holographic upper limit on the d.o.f. of the CFT as measured by the Casimir entropy S_C.

In figure 2 we have graphically depicted the quadratic relation between the total entropy S and the Casimir entropy S_C in the same diagram we used to related the cosmological entropy bounds. From this diagram it easy to determine the relation between the new bound and the Hubble bound. One clearly sees that when $HR > 1$ that the two bounds are in fact equivalent. When the new bound is saturated, which means $S_C = S_{BH}$, then the Hubble bound is also saturated, *ie.* $S = S_H$. The converse is not true: there are two moments in the region $HR < 1$ when the $S = S_H$, but $S_C \neq S_{BH}$. In our opinion, this is an indication that the bound on the Casimir energy has a good chance of being a truly fundamental bound.

7 Summary and conclusion

We have used the holographic principle to study the bounds on the entropy in a radiation dominated universe. The radiation has been described by a continuum CFT in the bulk. Surprisingly the CFT appears to know about the holographic entropy bounds, and equally surprising the FRW-equations know about the entropy formulas for the CFT. Our main results are summarized in the following two tables. Table 1. contains an overview of the bounds that hold in the early universe on the temperature, entropy and Casimir energy. In table 2. the Cardy formula for the CFT and the Euler relation for the Casimir energy are matched with the Friedmann equations written in terms of the quantities listed in table 1.

CFT-bound	*FRW-definition*
$T \geq T_H$	$T_H \equiv -\dot{H}/2\pi H$
$S \leq S_H$	$S_H \equiv (n-1)HV/4G$
$E_C \leq E_{BH}$	$E_{BH} \equiv n(n-1)V/8\pi G R^2$

Table 1: Summary of cosmological bounds

CFT-formula	*FRW-equation*
$S = \frac{2\pi R}{n}\sqrt{E_C(2E - E_C)}$	$S_H = \frac{2\pi R}{n}\sqrt{E_{BH}(2E - E_{BH})}$
$E_C \equiv n(E + pV - TS)$	$E_{BH} = n(E + pV - T_H S_H)$

Table 2: Matching of the CFT-formulas with the FRW-equations

The presented relation between the FRW equations and the entropy formulas precisely holds at this transition point, when the holographic bound is saturated or threatens to be violated. The miraculous merging of the CFT and FRW equations strongly indicates that both sets of these equations arise from a single underlying fundamental theory.

The discovered relation between the entropy, Casimir energy and temperature of the CFT and their cosmological counterparts has a very natural explanation from a RS-type brane-world scenario [23] along the lines of [24]. The radiation dominated FRW equations can be obtained by studying a brane with fixed tension in the background of a AdS-black hole. In this description the radius of the universe is identified with the distance of the brane to the center of the black hole. At the Big Bang the brane originates from the past singularity. At some finite radius determined by the energy of the black hole, the brane crosses the horizon. It keeps moving away from the black hole, until it reaches a maximum distance, and then it falls back into the AdS-black hole. The special moment when the brane crosses the horizon precisely corresponds to the moment when the cosmological entropy bounds are saturated. This world-brane perspective on the cosmological bounds for a radiation dominated universe are described in detail in [3].

We have restricted our attention to matter described by a CFT in order to make our discussion as concrete and coherent as possible. Many of the used concepts, however, such as the entropy bounds, the notion of a non-extensive entropy, the matching of the FRW equations, and possibly even the Cardy formula are quite independent of the equation of state of the matter. One point at which the conformal invariance was used is in the diagrammatic representation of the bounds. The diagram is only circular when the energy E goes like R^{-1}. But it is possible that a similar non-circular diagram exists for other kinds of matter. It would be interesting to study other examples in more detail.

Finally, the cosmological constant has been put to zero, since only in that case all of the formulas work so nicely. It is possible to modify the formalism to incorporate a cosmological constant, but the analysis becomes less transparent. In particular, one finds that the Hubble entropy bound needs to be modified by replacing H with the square root of $H^2 - \Lambda/n$. At this moment we have no complete understanding of the case $\Lambda \neq 0$, and postpone its discussion to future work.

References

[1] G. 't Hooft, in *Salamfestschift: a collection of talks*, eds. A.Ali, J.Ellis, S.Randjbar-Daemi (World Scientific 1993), gr-qc/9321026; L. Susskind, J.Math.Phys. **36** (1995) 6337.

[2] J.L. Cardy, Nucl.Phys. **B 270** (1986) 317.

[3] I. Savonije, and E. Verlinde, Phys. Lett. **B507** (2001) 305.

[4] J.D. Bekenstein, Phys.Rev.**D23** (1981) 287, **D49** (1994) 1912; Int.J.Theor.Phys.**28** (1989) 967.

[5] W. Fischler and L. Susskind, hep-th/9806039

[6] R. Easther and D. Lowe, Phys. Rev. Lett. **82**, 4967,(1999), hep-th/9902088.

[7] G. Veneziano, Phys.Lett **B454** (1999), hep-th/9902126; hep-th/9907012.

[8] D. Bak and S.-J. Rey, hep-th/9902173

[9] N. Kaloper and A. Linde, Phys.Rev. **D60** (1999), hep-th/9904120

[10] R. Brustein, gr-qc/9904061; R. Brustein, S. Foffa, and R. Sturani, hep-th/9907032.

[11] R. Brustein, G. Veneziano, Phys.Rev.Lett. **84** (2000) 5695. G. Veneziano, private communication.

[12] R. Bousso, JHEP **07** (1999) 004; JHEP **06** (1999) 028; hep-th/9911002.

[13] E.E. Flanagan, D. Marolf, and R.M. Wald, hep-th/9908070.

[14] T. Jacobson, Phys.Rev.Lett.**75** (1995) 1260, gr-qc/9504004. (1948), 793.

[15] S. Weinberg, *Gravitation and Cosmology*, Wiley, (1972).

[16] E. Kolb, M. Turner, *The Early Universe*, Addison-Wesley (1990)

[17] H.B.G. Casimir, Proc.Kon.Ned.Akad.Wet.**51**,

[18] J. Maldacena, Adv.Theor.Math.Phys.2 (1998) 231, hep-th/9711200.

[19] S. Gubser, I. Klebanov and A. Polyakov, Phys.Lett.B428 (1998) 105, hep-th/9802109;

[20] E. Witten, Adv.Theor.Math.Phys.2 (1998) 253, hep-th/9802150.

[21] E. Witten, Adv.Theor.Math.Phys.2 (1998) 505, hep-th/9803131

[22] S.W. Hawking and D. Page, Commun.Math.Phys. 87 (1983) 577

[23] L. Randall and R. Sundrum, *A Large Mass Hierarchy from a Small Extra Dimension*, hep-ph/9905221; *An Alternative to Compactification*, hep-th/9906064.

[24] S. Gubser, *AdS/CFT and gravity*, hep-th/9912001.

On Confinement and Duality

Matthew J. Strassler*

University of Pennsylvania, Philadelphia, Pennsylvania, USA

Lectures given at the
Spring School on Superstrings and Related Matters
Trieste, 2 – 10 April 2001

LNS027003

*strassler@physics.upenn.edu

Abstract

Confinement in four-dimensional gauge theories is considered from several points of view. General features are discussed, and the mechanism of confinement is investigated. Dualities between field theories, and duality between field theory and string theory, are both put to use.

Contents

1 Introduction to Confinement

One of the most important discoveries of the twentieth century is that our world consists of atoms, of size 10^{-10} meters, made from electrons bound to positively charged nuclei. The size of the atom is set by the uncertainty principle; the electron is nonrelativistic, with a velocity of order α, so the size of the atom is of order $\delta x \sim 1/\delta p \sim (m_e \alpha)^{-1}$, where α is the QED coupling constant. While all experiments to date indicate that the electron itself has a size smaller than 10^{-18} meters, nuclei of atoms have a definite size, of order 10^{-15} meters. They consist of weakly-bound clumps of protons and neutrons. It was learned in the 1950s that the protons and neutrons have a size comparable to the nuclei which contain them. In the 1960s, evidence emerged that nucleons have pointlike constituents, weakly coupled in high-energy scattering processes, but highly relativistic, and therefore strongly bound, inside the proton. By the 1970s the theory of QCD emerged to explain how this strange effect was possible. The QCD interaction is weak in high-energy processes, and grows, through renormalization effects, to become strong in the low-energy processes that bind the quarks in the nucleons. The energy scale Λ_{QCD} at which it becomes strong is a few hundred MeV, corresponding to the size of the nucleon. The pointlike objects in the nucleons are the quarks suggested by Gell-Mann, interacting through the color charge suggested by Greenberg. These quarks are now themselves known to be smaller than 10^{-18} meters. They are also very light; most of the mass of the proton comes from their kinetic energy and from the powerful interactions binding the quarks together.

Yet no one has ever seen a quark, or its fractional electric charge, sitting by itself somewhere. So why should we believe this story? We all know the words: quarks are confined in hadrons — nucleons, pions, etc. — and never come out. But all too often we overlook the subtleties involved in this statement. What actually happens if we send an electron deep into a proton and try to kick a quark away from its two friends? A large amount of energy, in the form of chromoelectric field, appears in the region between the escaping quark and the remaining parts of the proton. Then what? We are familiar with the idea that large electric fields beyond a certain magnitude cannot survive; sufficiently strong fields, with energy densities bigger than $m_e^4 \sim 1$ MeV^4, are able to decay by producing pairs of electrons and positrons, the lightest electrically charged particles. The same holds for chromoelectric fields; when they become sufficiently strong, of order $\Lambda_{QCD}^4 \sim$

(300 MeV)^4, they can pair-produce light quarks and antiquarks. How does this affect the departing quark? Well, as it moves away, the field between it and the other two quarks starts producing pairs. If for example a single pair is created, the new antiquark can end up bound to the escaping quark, and the new quark can end up bound to the other two quarks in the proton, making a new nucleon. Or perhaps multiple pairs will be created, and many quark-antiquark bound states will result. But in any case, the original quark succeeds in its escape. The force between it and the remaining quarks in the proton drops to zero as it moves away. Is this really "confinement"?

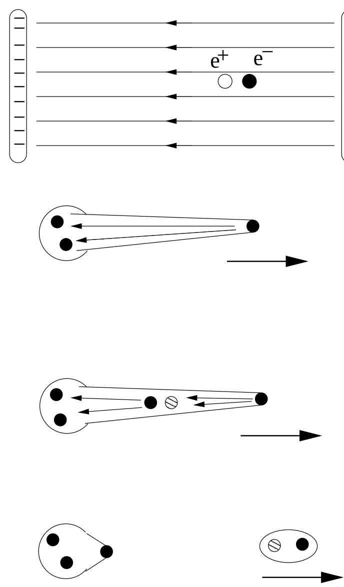


Figure 1: As with pair production of electrons in a strong electric field, pair production occurs as a quark tries to escape from a proton.

Let's contrast this with what might happen in an imaginary world in which all of the quarks had masses much larger than 100 MeV. In fact, let's take *all* of their masses to all be, oh, say, about 1000 GeV. Now the proton is a very heavy object, with mass of order 3000 GeV, and it is now quite a bit smaller than usual, about 10^{-17} meters in size (the factor of ten compared to (1000 GeV)$^{-1}$ comes from the fact that the strong coupling constant is about 1/10 at these energies.) But let's imagine trying to kick a quark out of the proton now. As it rushes away, the chromoelectric field becomes very large, but the energy density, of order $\Lambda_{QCD}^4 \sim$ (300 MeV)^4, is far too low to produce pairs of 1000 GeV quarks. (Notice that for pair-production to be impossible, it is essential that *all* flavors of quarks be heavy; if even *one* type of quark is light, the field will pair-produce it, independent of whether the quarks in the proton are themselves heavy.) So what happens now? Does the quark escape?

No; it cannot — or at least, it is extremely difficult. In this imaginary world, where all the quark masses m_q are very large compared to Λ_{QCD}, the quark is truly imprisoned. The force between the escaping quark and the remains of the proton goes to a constant; as we will discuss further, a "string" or "tube" of chromoelectric flux, of thickness $\Lambda_{QCD}^{-1} \sim 10^{-15}$ meters, and of tension (energy per unit length) Λ_{QCD}^2, connects the two colored objects to one another. Unless the tube becomes very long, of length m_q/Λ_{QCD}^2 (which in this case $\sim 10^{-12}$ meters, many times larger than the proton radius), there is insufficient energy in the chromoelectric field to pair-produce quarks. Even if the string does become this long, there is an exponentially low probability that all of its energy, spread out over 10^{-12} meters, will find itself localized in a region of radius $m_q^{-1} \sim 10^{-18}$ meters, as would be necessary to produce a pair of heavy quarks. So this tube of flux, stable if short, metastable if long but with a exponentially long lifetime, makes it essentially impossible for the kicked quark to escape. Eventually, the constant force from the flux tube will bring it to a stop, and pull it back into its protonic prison. This is true confinement, no doubt about it. The word really means something here.

Notice that it is not just the quarks which can be said to be "confined". *The chromoelectric field emitted by the quarks, rather than spreading out across space as in electromagnetism, is confined into "tubes," or "strings".* This is important, because even when we take the quarks away — say, by taking their masses to infinity — it might still be true that flux is confined, though there are no confined particles. In fact, we will soon see this is a more precise definition of confinement.

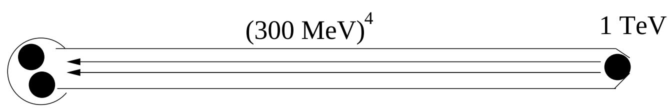


Figure 2: If all quarks were heavy, then flux tubes would break much less readily.

Strict confinement, of flux and of quarks, is thus a property of QCD only when all of its quarks are heavy. [More precisely, it is seen in the limit where the number of flavors N_f of light quarks is much smaller than the number of colors N; the number of flavors need not be strictly zero, because the amplitude for the pair production process that splits flux tubes is of order N_f/N.] QCD with at least one light quark shows only a few remnants of these properties; there are hints of flux tubes between an escaping quark and the proton it leaves behind (though they break very quickly and never become long); and there are hints that some of the bound states of the theory behave as bits of spinning flux tube (though this is a very imprecise statement, and has as its strongest merit that it helped to motivate string theory).

So what is the right way to describe what happens in real-world QCD? We do not live in a truly confining world, and it might have been better for our own conceptual thinking if we had come up with another word for what QCD does to quarks. "Cloaking" or "maximal screening" might have been a better term. What QCD really does is ensure that a quark seeking to be free has a region in its vicinity, of size Λ_{QCD}^{-3} in volume, with chromoelectric energy density that is of order Λ_{QCD}^4. This by itself will cause an antiquark (and its partner quark) to pop out nearby, cloaking — that is, completely screening — the charge of the original quark. Compare this with electrons; in their vicinity there are regions with energy density of order m_e^4, but since the energy density is $(\alpha/r^2)^2$, the size of the region with this energy is too small to pair-produce electrons and positrons by a factor of $\alpha^{3/2}$. Thus to have this cloaking effect we need a strong coupling constant, but it hardly requires something as drastic as the flux tubes and the imprisonment found in worlds with only heavy quarks. (Indeed you might amuse yourselves by considering the possible physical properties of a hypothetical point particle of electric charge greater than $\sqrt{137}e$.)

In these lectures, we are going to explore truly confining gauge theories in some detail. Such theories may indeed exist in nature, but it is important to remember that real-world QCD is not among them.

1.1 Confinement in pure Yang-Mills

How do we even know that true confinement does in fact occur in some theories? This is a long story, and there are many ways to tell it. Let us begin in the middle, by assuming that confinement of flux occurs in pure Yang-Mills (YM) theory.

So instead of QCD, let us discard the quarks, leaving only a gauge boson A_μ in the adjoint of $SU(N)$. The group $SU(N)$ consists of $N \times N$ matrices

$U^{\alpha}_{\bar{\beta}}$ (with row indices α and column indices $\bar{\beta}$.) which are special ($\det U = 1$) and unitary ($U^\dagger = U^{-1}$). The "gluon" field A_μ takes values in the algebra of $SU(N)$,

$$(A_\mu)^{\alpha}_{\bar{\beta}} = A^a_\mu (T^a)^{\alpha}_{\bar{\beta}} \ .$$

Here T^a is a generator of the group $SU(N)$, also an $N \times N$ matrix (normalized to tr $T^a T^b = \frac{1}{2}\delta^{ab}$,) and the group index a runs from 1 to the dimension of $SU(N)$, namely $N^2 - 1$. The theory has the simplest possible Lagrangian; defining $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + i[A_\mu, A_\nu]$ (here F and A are matrices and the brackets indicate a matrix commutator), we write the Lagrangian as

$$\mathcal{L} = -\frac{1}{2g^2} \text{tr } F_{\mu\nu} F^{\mu\nu} \ .$$

This normalization of the field A_μ differs from the one in standard textbooks on perturbation theory. There is good reason for this. We will not be doing perturbation theory. In perturbative calculations, it is more convenient to absorb the $1/g$ into A_μ; then $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + ig[A_\mu, A_\nu]$. The quadratic terms in the Lagrangian are then the free Maxwell equations, and do not depend on g. We may then think of the theory as a set of free fields — simply $(N^2 - 1)$ independent photons — coupled together by interactions of order g. However, in these lectures we will not assume small g, and will rarely expand in powers of g. The normalization chosen here is more profound; it puts the coupling constant in its proper place, multiplying $\hbar$ and therefore determining the size of all quantum effects. Most nonperturbative properties of the theory will involve either $1/g^2$ or e^{-1/g^2}, as we will soon see.

Pure Yang-Mills theory is weakly coupled at high energy, like QCD, and becomes strongly coupled at a scale Λ. More accurately, we can show, using perturbation theory, that it *cannot* become strongly coupled at energies *above* a scale Λ; below this point we simply don't know what it does. The scale Λ can be estimated using one-loop graphs; at this order, the running of the gauge coupling is given by

$$\beta_g = \frac{\partial g}{\partial \ln \mu} = \frac{g^3}{16\pi^2}\left(-\frac{11}{3}N\right)$$

for $SU(N)$. The solution is

$$\frac{8\pi^2}{g^2(\mu)} = \frac{8\pi^2}{g^2(\mu_0)} + \frac{11}{3} N \ln(\mu/\mu_0)$$

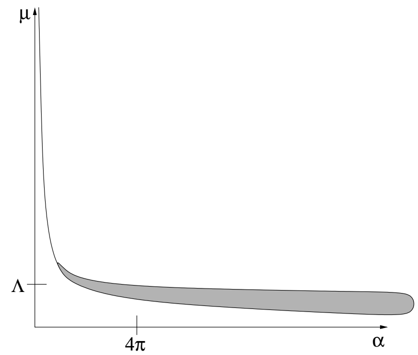


Figure 3: The coupling constant α versus energy scale μ; the one-loop calculation is valid at $\mu \gg \Lambda$ but becomes only approximate at low energy.

where μ_0 is an arbitrary starting point. Thus, the coupling is small above the energy scale $\mu \sim \Lambda$, where

$$\Lambda^{11N/3} \sim \mu_0^{11N/3} \ \exp\left[\frac{-8\pi^2}{g^2(\mu_0)}\right] \ . \tag{1}$$

This is reliable since higher loop graphs and nonperturbative effects are comparatively small above Λ.

As is standard in renormalization, the scale Λ is physical and thus independent of the arbitrary starting point μ_0. Near and below this energy regime, the coupling constant is strong; above it, perturbation theory in g^2 is possible. Also, notice that Λ involves e^{-1/g^2}. All of the really interesting physics in Yang-Mills theory is related to Λ; it is therefore nonperturbative in g^2, and cannot show up at any order in an ordinary Feynman graph expansion.

Now we must consider two more profound claims, which are fully nonperturbative, and are based on a combination of experiment, theoretical reasoning, and both analytic and especially numerical lattice gauge theory. First, the quantum Yang-Mills theory is known to develop a mass gap (that is, it has no massless fields in its spectrum, and instead has a discrete set of states with masses of order Λ) and second, it apparently becomes confining, in the true sense, at the scale Λ. Both of these effects are through strongly-coupled physics not visible semiclassically.

Both statements are strange. The gluons in the above Lagrangian are massless; how can there be no massless particles in the spectrum of the

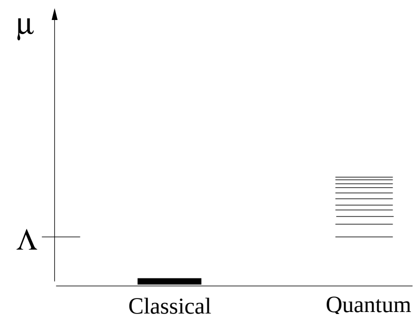


Figure 4: Classically the theory has many massless particles, but the quantum theory has a mass gap and a spectrum of gauge-neutral hadrons.

theory? Well, let us assume that, as in QCD, the effect of the strong interactions will be that we will observe only colorless bound states. What kinds of bound states can we make from gluons? We might say that we can make a bound state of two gluons, or three gluons, or four. But this is surely wrong. The interactions of the theory do not conserve the number of gluons even in perturbation theory; there are terms cubic and quartic in A_μ in the Lagrangian, so one gluon may become two or three, and vice versa. The situation will be worse once the interactions of Yang-Mills become strong. We clearly cannot use "gluon number" as a quantum number describing a state. In fact, the strong coupling dynamics makes it impossible to talk about gluons at low energies. Instead, we have only bound states, whose name "glueballs" is reasonably accurate, in that these gluey states do not really consist of a fixed number of gluons, but rather of a shifting mass of chromoelectric flux lines. There are a large number of these states. Below the scale Λ we might try to write an effective theory of these glueballs. Unlike the gluons, for which mass terms are forbidden (since they have only two polarization states and massive vectors require three), the glueballs include scalars (for which mass terms cannot be forbidden) and vectors with three polarizations (for which mass terms also cannot be forbidden) and similar higher spin particles. Their masses can't be much larger than Λ since that would contradict perturbation theory, but nothing stops them from having masses of order Λ. Essentially, there is a mass gap because there are no symmetries which forbid mass terms for any of the glueballs.

The statement about confinement is also, at first, strange. The theory has only gluons; are they confined? What happens when we try to pull a gluon out of a bound state? Does a flux tube form between it and the other gluons? What does this mean, since the flux tube itself is made from gluons? How is it possible that pair-production of gluons is forbidden? In fact, it is not forbidden, but that is fortunately irrelevant. The statement about confinement has nothing to do with the gluons. The gluons are no more confined in Yang-Mills than light quarks are in real-world QCD; in fact they are even less so, since there is no parameter analogous to the quark mass which when large can make the gluons confined. *"Confinement" means that chromoelectric field is confined; it cannot spread out in space over regions larger than about* Λ^{-1} *in radius.*

One might ask if there is a connection between the mass gap and the confinement of flux. We will return to this issue later.

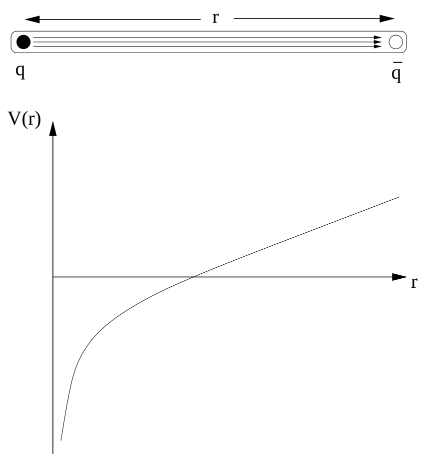


Figure 5: The confined field lines between a heavy quark and antiquark form a tube; the potential energy of the system goes as $1/r$ at distances short compared to Λ^{-1} but becomes linear at larger distances.

Now, how can we detect the presence of the strings which contain the chromoelectric flux? Ideally we would like to find a long and straight flux tube and find its tension (energy per unit length) but we might have trouble convincing one to stay straight long enough to do this measurement. So here we need a new idea. Recall how the heavy quarks of QCD-with-no-light-quarks were truly confined. This suggests that the way to detect confinement of flux in Yang-Mills theory is to put some extremely heavy quarks in it — so heavy that they can't affect the dynamics of the Yang-Mills theory — and see that these quarks are confined! That is, we can compute the quark-antiquark potential $V(r)$ and see that it grows without bound (indicating confinement) and more specifically is linear in r (indicating confinement by flux tube.) Why is the linear potential characteristic of a flux tube? Well, consider Gauss's law. In an unconfined theory, the electric flux is uniformly distributed over a sphere surrounding a charge, and therefore falls off as $1/r^2$. In a confining theory with flux tubes, the flux tube has a fixed cross-sectional area of order Λ^{-2} no matter how long it is; and thus, for any sphere of radius $r \gg \Lambda^{-1}$ surrounding a charge, the flux on the sphere is zero everywhere except in a region of area Λ^{-2} where the flux tube passes through the sphere. From this we conclude that the electric field in that region has a magnitude which is r-independent! In turn, this implies the force that it generates on a test charge is also r-independent, and finally, that the potential between charges grows linearly with r.

So, let us add a charged fermion (or scalar) to the Yang-Mills theory, one whose mass M is so much larger than Λ that it cannot play a role in the strong-coupling physics. Adding a quark ψ we make the Lagrangian

$$\mathcal{L} = -\frac{1}{2g^2}\text{tr } F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\,\psi - M\bar{\psi}\psi \ .$$

The quark ψ is charged under $SU(N)$, but for the moment let us not specify the representation R of $SU(N)$ under which it transforms. Now let us consider the potential $V(r)$ between ψ, placed at one position, and $\bar{\psi}$, placed a distance r away. Since the quarks are very heavy, we can expect that they can be placed at rest and will move only very slowly, allowing us to do this computation. Confinement means that when r is large, a string — a tube of chromoelectric flux — stretches between ψ and $\bar{\psi}$, of constant tension T_R, such that the potential $V(r) = T_R r$ [1]. The force between two such fermions goes to a constant, and never drops off to zero. (That these facts are true in Yang-Mills theory does not follow from any direct theoretical calculation. Highly quantum mechanical in nature, they have only been checked using direct numerical simulation of Yang-Mills theory.)

In the limit where $M \to \infty$, the quarks become completely non-dynamical [1]; they are what we may call "chromoelectrostatic sources", probes which never appear in any loop diagram and thus are purely classical. What remains dynamical is the flux tube. Thus, we didn't really need the quarks as physical particles; using nondynamical chromoelectric sources, we could have detected the confinement of chromoelectric field, which is a property of the Yang-Mills theory without the added quarks. (An equivalent way to make this statement, without introducing the quarks, is to talk about Wilson loops in various representations R; in a confining theory the value of

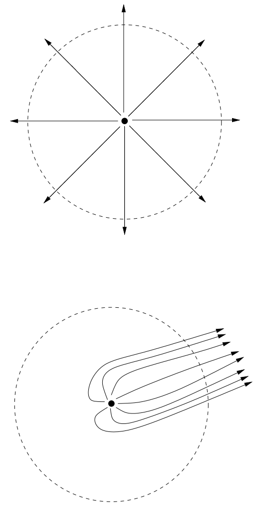


Figure 6: Gauss's law for unconfined and confined flux.

the Wilson loop is proportional to the exponential of minus its area, with proportionality constant T_R [1].)

In general, the string tension, and the corresponding force, between quark and antiquark can depend on the representation R. After all, why not? In particular, for R the adjoint, we already know $T_{adjoint} = 0$: any fermion in the adjoint can combine with a light gluon to make something gauge neutral, so two such fermions will each cloak themselves with a gluon and will feel no long-range force as we pull them apart. So clearly we need to think about how things depend on the representation R. Clearly the map from representations to flux tubes cannot be one-to-one (since both the trivial representation and the adjoint representation have $T_R = 0$.) Lie groups have an infinite number of representations, but the stable flux tubes number at most $\dim C_G$, the dimension of the center of the gauge group. Let us see why this is so.

What is the center of $SU(N)$? A matrix $U^{\alpha}{}_{\bar{\beta}} = e^{2\pi i k/N}\delta^{\alpha}{}_{\bar{\beta}}$, $k = 0, \ldots, N-1$, is an element of $SU(N)$. Being proportional to the identity, it obviously commutes with everything in $SU(N)$; in short, U is in the center $C_{SU(N)}$. The elements of the center are thus labelled by the integer k, which from the definition of U is only determined modulo N, so the labels form the group $\mathbf{Z}_N$, the additive integers mod N. Now consider any representation R. An element ρ of this representation is labelled by a certain number n

of unbarred (upper) and $\bar{n}$ of barred (lower) indices; that is, it takes the form $\rho^{\alpha_1\alpha_2\cdots\alpha_n}_{\bar{\beta}_1\bar{\beta}_2\cdots\bar{\beta}_{\bar{n}}}$. Under a group transformation, each unbarred index is rotated by the matrix U, while each barred index is rotated by $U^\dagger$. Consequently, the transformation of the representation R under the center C_G is by the phase $e^{2\pi ik(n-\bar{n})/N}$, where $n - \bar{n}$ is called the "N-ality" of the representation. The adjoint representation, with one upper and one lower index, is invariant under the center. The fundamental $\mathbf{N}$ representation (one unbarred index) rotates by $e^{2\pi ik/N}$; the antifundamental $\bar{\mathbf{N}}$ rotates by $e^{-2\pi ik/N}$. Both the antisymmetric-tensor and symmetric-tensor representations $\mathbf{N(N \pm 1)/2}$, which have two unbarred indices, rotate by $e^{2\pi i(2k)/N}$. Indeed, all p-upper-index tensors carry charge p under $\mathbf{Z}_N$ — that is, they rotate by $e^{2\pi ipk/N}$ under the k^{th} element of $\mathbf{Z}_N$. In short, the representations R of $SU(N)$ break up into equivalence classes under the center, and can be labelled by their "N-ality" charge p [2, 3]. Note that the conjugate representation of R has "N-ality" $N - p$, since the number of barred and unbarred indices is exchanged.

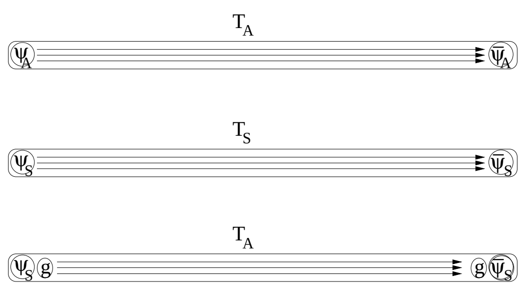


Figure 7: The flux between different quarks, or combinations of quarks and gluons, all with N-ality 2.

Why is this interesting? First consider, for example, adding a heavy quark ψ_A, in the antisymmetric representation, to Yang-Mills theory; the potential between $\bar{\psi}_A$ and ψ_A is $V(r) = T_A r$. Now consider instead adding a heavy quark ψ_S in the symmetric representation; the quark-antiquark potential between $\bar{\psi}_S$ and ψ_S is now $V(r) = T_S r$. Suppose that $T_S > T_A$ in Yang-Mills theory. (This is probably true, but what I'm about to say won't depend on the specific assumption.) Nothing prevents the theory from taking one of its light gluons (remember their number is not conserved so it need not be pair-produced) and putting it very near ψ_S. The combination of the gluon A_μ and the fermion ψ_S looks, from a distance, as though it were a single object. What is its charge? Well, we must consider the group theory of $SU(N)$; what is (adjoint)$\otimes$(symmetric)? It is a direct sum of a number of representations, *all of which have the same "N-ality" as the symmetric representation, namely 2.* Said another way, the product of $(A_\mu)^\alpha_{\bar{\beta}}$ and $\psi_S^{\gamma\delta}$ can lead, no matter how the indices are contracted, only to representations with two more upper indices than lower indices. Among these representations is the antisymmetric representation. (In $SU(3)$, for instance, the symmetric tensor is $\mathbf{6}$, the antisymmetric tensor is $\bar{\mathbf{3}}$, and $\mathbf{8}\otimes\mathbf{6} = \bar{\mathbf{3}}+\mathbf{6}+\overline{\mathbf{15}}+\mathbf{24}$.) But then, since we assumed $T_A < T_S$, there exists a dynamical process by which the theory may lower its energy! By popping a gluon out of the vacuum and putting it near ψ_S, the theory can make ψ_S look more like a fermion in the antisymmetric representation. The same goes for $\bar{\psi}_S$. Then, instead of a string of tension T_S, a string of tension T_A can link these two fermion-gluon combinations. The energy cost is that of making two extra gluons — at most of order Λ — while the energy gain is $(T_S - T_A)r$, which for r sufficiently large always wins. The reverse process will hold if $T_A > T_S$.

More generally, the fact that gluons are in the "N-ality"-zero adjoint representation implies that *the presence of nearby gluons can change one representation to another but only in a way that conserves N-ality.* Thus in Yang-Mills, the representation R of a chromoelectric source is not a conserved quantum number; only its "N-ality" is actually conserved. Consequently, we should expect that for the entire class of representations with the same N-ality charge, there will be only one stable configuration of strings (which might involve one or more tubes — for "N-ality"=2 there might be one tube with two units of flux or two tubes with one unit each.) *The tensions of the stable strings, or combinations of strings, are labelled not by R but by the N-ality p of R.* Charge conjugation symmetry also ensures that $T_p = T_{N-p}$; thus we have of order $N/2$ stable flux tube configurations in $SU(N)$ Yang-Mills theory.

Can we see this in $SU(3)$ Yang-Mills? Yes and no. There is N-ality 0,1, and 2; but $T_0 = 0$ while $T_2 = T_1$, so only one confining string is predicted. The nontrivial statements are then only that, for example, the symmetric $\mathbf{6}$ representation of $SU(3)$ is confined by the same string tension as the antisymmetric tensor, the $\bar{\mathbf{3}}$; this in turn has the same tension as the fundamental $\mathbf{3}$. To have a nontrivial set of strings we must go to $SU(4)$; here the antisymmetric tensor $\mathbf{6}$ should have a tension T_6 different from that of the $\mathbf{4}$ and $\bar{\mathbf{4}}$, T_4. There is still a question as to whether $T_6 < 2T_4$; if not, the flux between two $\mathbf{6}$ fields may be carried by two strings of N-ality 1 rather than a single string of N-ality 2. Theoretical arguments [4] and lattice calculations

[5, 6, 7] support the view that $T_6 < 2T_4$ (and similarly in other theories) so that there really are two independent stable flux tubes, of N-ality 1 and 2 (and again $T_3 = T_1$.)

To summarize, we expect that Yang-Mills theories have stable flux tubes labelled by a charge in the center of the group [2]; for $SU(N)$ this is its N-ality, a charge under the $C_{SU(N)} = \mathbf{Z}_N$ group action. While the gluons are not confined by these strings, any heavy quark with nonzero N-ality will experience a linear potential energy and a constant force which will confine it to an antiquark, or more generally, to some combination of quarks and antiquarks which have the opposite N-ality. (For example, it could combine with $N-1$ other quarks to form a baryon. As another example, a **6** of $SU(4)$ could combine with two $\bar{\mathbf{4}}$ quarks to form an exotic object not found in real-world $SU(3)$ QCD.)

1.2 Confinement in $\mathcal{N} = 1$ Super-Yang-Mills

Let us now consider $\mathcal{N} = 1$ supersymmetric Yang-Mills theory (SYM.) This theory is very interesting in that (1) many of its properties can be exactly or approximately determined, (2) it resembles Yang-Mills theory, in that it has confinement and flux tubes, has a mass gap, and lacks light particles similar to pions, yet (3) it resembles QCD in that it has chiral symmetry breaking and an anomaly which makes a would-be Nambu-Goldstone boson, the η', massive, while (4) it differs from both in that it has multiple isolated, degenerate vacua.

The $SU(N)$ SYM theory is nothing more than $SU(N)$ gauge theory with a vector boson (gluon) A_μ and a massless Majorana spinor (gluino) λ_α, both in the adjoint representation of the gauge group. The Lagrangian is simply

$$\mathcal{L} = \frac{1}{2g^2}\left[\mathrm{tr}\ F_{\mu\nu}F^{\mu\nu} + i\bar{\lambda}\not{D}\,\lambda\right] \tag{2}$$

Pure $\mathcal{N} = 1$ SYM, like pure non-supersymmetric YM, is a confining theory. (Convincing arguments confirming earlier expectations are given in [8, 9].) It will have stable flux tubes, just like YM, despite the presence of the gluinos. The gluino carries the same gauge charge as the gluon, and is neutral under $C_{SU(N)} = \mathbf{Z}_N$. Therefore, like the gluon, it does not break flux tubes carrying $\mathbf{Z}_N$; no flux tube which carries such a charge can end on a $\mathbf{Z}_N$-neutral gluino. (This is in contrast to $SU(3)$ QCD, where the quarks, which carry charge under the $\mathbf{Z}_3$ center, do indeed break the flux tubes.) Thus SYM is a good place to study confining strings.

The theory also has an anomalous $U(1)$ global symmetry, just like QCD. We won't need this, but it is useful for you to know a bit about it. How does this work? Classically, the Lagrangian of the theory has a global symmetry $\lambda \to \lambda e^{i\alpha}$, where α is any real number. However, the path integral of SYM does *not* have this symmetry. There isn't time in these lectures to study anomalies in detail, so let me just quote the classic result: under this rotation, the path integal itself is not invariant unless $2N\alpha$ is a multiple of 2π. (A similar statement applies in QCD with N_f massless flavors of quarks ψ_i and $\tilde\psi_{\tilde j}$; under the global rotation $\psi_i \to \psi_i e^{i\alpha}$, $\tilde\psi_{\tilde i} \to \tilde\psi_{\tilde i} e^{i\alpha}$, the path integral is not invariant unless $2N_f\alpha$ is a multiple of 2π.) Thus the $U(1)$ is a fake; only a discrete $\mathbf{Z}_{2N}$ subgroup of this $U(1)$ is actually a symmetry.

In QCD, with Lagrangian[1]

$$\mathcal{L} = \frac{1}{2g^2}\mathrm{tr}\ F_{\mu\nu}F^{\mu\nu} + \sum_{i=1}^{N_f} i\bar\psi_i \not{D}\,\psi_i + \sum_{\tilde j=1}^{N_f} i\overline{\tilde\psi}_{\tilde j}\not{D}\,\tilde\psi_{\tilde j} - \sum_{i,j} m^{i\tilde j}\tilde\psi_{\tilde j}\psi_i\ ,$$

there is an entire $SU(N_f)$ symmetry for the quarks ψ_i, another $SU(N_f)$ for the antiquarks $\tilde\psi_{\tilde j}$, a $U(1)$ "baryonic" symmetry under which the ψ_i and $\tilde\psi_{\tilde j}$ have opposite charge, and finally the fake "axial" $U(1)$ mentioned above of which only a $\mathbf{Z}_{2N_f}$ is a true symmetry. These symmetries do not all appear at low energy, however. First, the nonzero quark masses $m^{i\tilde j}\tilde\psi_{\tilde j}\psi_i$ break most of the two $SU(N_f)$ symmetries; but the masses are relatively small for the up, down, and strange quarks, so let us imagine for a moment that they are zero, and, forgetting the heavier quarks (which are dynamically less important,) take $N_f = 3$. But even then, for $m^{i\tilde j} = 0$, the vacuum does not show all of the symmetries of the theory. For reasons not entirely understood, a quark-antiquark bilinear operator $\tilde\psi_{\tilde j}\psi_i$ develops a nonzero expectation value[2] proportional to $\delta_{i\tilde j}$, with a magnitude of order $(\Lambda_{QCD})^3$. This quark-antiquark condensate is not invariant under the $SU(N_f)$ symmetries mentioned above;

[1] Note the fermion fields $\psi, \tilde\psi$ written here are not each others' complex conjugates! They are left-handed quarks and left-handed antiquarks; they form two separate sets of two-component Weyl fermions, transforming in the $\mathbf{N}$ and $\bar{\mathbf{N}}$ representations. Mass terms $m^{i\tilde j}\tilde\psi_{\tilde j}\psi_i$ make them into massive four-component Dirac fermions, but without the masses they are independent fields, with independent generation indices $i = 1, \ldots, N_f$ and $\tilde j = 1, \ldots, N_f$.

[2] All of the following statements about chiral symmetry breaking apply at least for small N_f; they are certainly not true for $N_f > (11/2)N$, at which point $SU(N)$ QCD has a positive one-loop beta function and can't possibly be strongly-coupled in the infrared. At what value of N_f they stop being true is not known, although most guesses these days for $N = 3$ range from 5 to 12.

it is only invariant under *simultaneous* rotations of the quarks ψ_i by a matrix U in *their* $SU(N_f)$ and of the antiquarks $\tilde{\psi}_{\tilde{j}}$ by the conjugate matrix $U^\dagger$ in the *other* $SU(N_f)$. These "diagonal" rotations define a group $SU(N_f)_D$, which remains a symmetry of the vacuum. All other $SU(N_f) \times SU(N_f)$ rotations change the vacuum, and thus are not symmetries of it. This is known as "spontaneous chiral symmetry breaking"; the equations of the theory still have an $SU(N_f) \times SU(N_f)$ symmetry, but the vacuum itself, a particular solution of those equations, is invariant only under its $SU(N_f)_D$ subgroup. As both Nambu and Goldstone taught us years ago, this implies, as an automatic consequence, that there are massless particles corresponding to the broken rotations. These are the pions. They tell us that QCD has not one vacuum, but in fact a continuous set of degenerate vacua (if the quarks are strictly massless!) The pions are massive in nature only because the quark masses are in fact not zero, and the $SU(N_f) \times SU(N_f)$ flavor symmetry is only approximate. Note that the baryonic $U(1)$ is unbroken. If the axial $U(1)$ had been a true symmetry, it would have been broken, and we would have expected a Goldstone boson for it, the η', which corresponds to shifts of the phase of the condensate $\langle\tilde{\psi}_{\tilde{j}}\psi_i\delta^{i\tilde{j}}\rangle$. However, the $U(1)$ is a fake; and although the $\mathbf{Z}_{2N_f}$ axial symmetry mentioned above is also spontaneously broken by the condensate to a $\mathbf{Z}_2$ subgroup, only continuous symmetries give continuous sets of degenerate vacua and corresponding massless particles. The η' in fact has a periodic potential, with N_f minima rotated by the $\mathbf{Z}_{2N_f}$ symmetry. In each of these minima the potential has some upward curvature, so the η' has a mass. Note however, that these minima are not actually isolated since they are connected via $SU(N_f) \times SU(N_f)$ rotations.

What happens in SYM? In this case the operator $\lambda\lambda$ develops an expectation value (this is a largely rigorous statement, for which there are many fairly strong proofs; see for example [10].) The $\mathbf{Z}_{2N}$ axial symmetry is broken to $\mathbf{Z}_2$. Because there are no continuous global symmetries, we have no continuous space of vacua. Instead we have N isolated, degenerate vacua, in which

$$\langle\lambda\lambda\rangle \propto \Lambda^3 e^{2\pi i r/N} \ , \quad r = 0, 1, 2, \ldots, N-1 \ .$$

In this theory, the beta function has coefficient $3N$, so the strong-coupling scale satisfies $\Lambda^{3N} = \mu_0^{3N} e^{-8\pi^2/g^2(\mu_0)}$. Notice that the $\mathbf{Z}_{2N}$ symmetry rotates one vacuum into the next, so the N vacua, though distinct from one another, are isomorphic. This guarantees they are degenerate with one another. Again, in each vacuum the $\mathbf{Z}_{2N}$ is broken, but the *space* of N vacua is $\mathbf{Z}_{2N}$ symmetric, and the symmetry rotates one vacuum into the next. The

η' particle in this theory is the phase of $\langle\lambda\lambda\rangle$, and it has a periodic potential, with N degenerate minima. Thus, like QCD, SYM has a fermion bilinear condensate which breaks global symmetries, and it has an η' with a periodic potential; but unlike QCD, and similar to YM, it has no massless or very light particles.

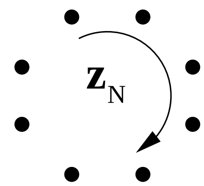


Figure 8: The vacua of $\mathcal{N} = 1$ $SU(N)$ SYM (shown in the complex $\langle\lambda\lambda\rangle$ plane) are rotated by a $\mathbf{Z}_N$ global symmetry.

2 Confinement of Magnetic Flux

Now let us try to understand why and how confinement occurs. In Yang-Mills theory it occurs through a process requiring strong coupling; detailed investigations have revealed no small parameter in which we can do perturbation theory, and no simple calculation that we can perform. From where can we gain some insight? We might ask: where we have seen tubes of confined flux before?

2.1 Superconductors and the Abelian Higgs Model

In Type I superconductors, magnetic flux is excluded from the material. This occurs through the appearance of surface currents, which can exist without energy cost due to the absence of any resistance in the material. These currents generate an exactly-compensating magnetic field which cancels any external magnetic field trying to penetrate the material, and instead produces some additional magnetic field outside. This makes it appear that all external magnetic fields are "expelled" from the superconductor. This famous piece of physics is called the "Meissner effect."

In Type II superconductors, however, the situation is a bit more complicated. Flux can indeed penetrate the superconductor in this case, although only in a very specific way. The material becomes nonsuperconducting in a narrow tube running from one side of the material to another, and the magnetic flux threads that tube. The magnetic field, which would have been free to roam in a normal material, is trapped inside "Abrikosov vortices" [11] traversing the superconductor. These vortices carry one or more quanta of flux; in short, they carry an integer charge, $q \in \mathbf{Z}$. *Superconductors confine magnetic flux into quantized vortices.*

Indeed this looks familiar. We have learned that $SU(N)$ YM and $\mathcal{N}=1$ SYM both confine *electric* flux into tubes which carry a discrete charge in $\mathbf{Z}_N$. This looks similar enough to set off alarm bells. We had better look at this more closely.

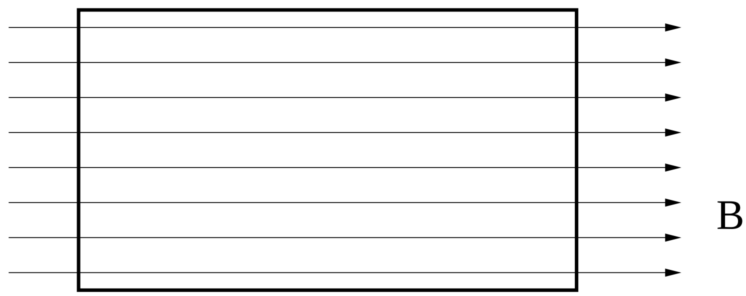


Figure 9: Normal materials can sustain magnetic fields.

How does a superconductor accomplish this? The superconductivity occurs because electrons form Cooper pairs, which are bosons. Let us call the density of these pairs ϕ. Since the pairs carry electric charge 2, ϕ must be complex, and couples to the photon. More specifically, the photon must couple to a conserved current, namely

$$J^\mu = \phi^\dagger \partial^\mu \phi - (\partial^\mu \phi^\dagger)\phi \tag{3}$$

Now suppose that there were a magnetic field attempting to pass through the material. Since the Cooper pairs can flow without resistance, they can respond by creating a compensating current. For instance, suppose we have a long cylinder of material of radius R; let us use cylindrical coordinates r, θ, z. Suppose we attempt to apply a uniform magnetic field $B_z > 0$ along the axis of the cylinder. The Cooper pairs can respond by generating a current J^θ, which can propagate without resistance, at the surface of the cylinder $r = R$. This completely cancels the applied field, reducing the energy density inside the superconductor. It also generates a dipole field outside the cylinder. The field appears to have been "expelled" from the material.

However, the material could also respond in an additional way, and does so in the type II case. In addition to generating a current at $r = R$, it could

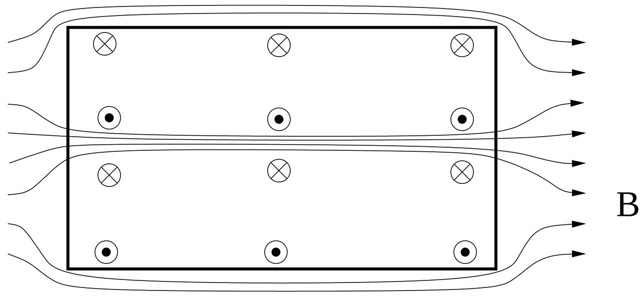


Figure 10: In superconductors, Cooper pair currents (shown into and out of the plane of the paper) are induced, causing the magnetic flux to be expelled or trapped in vortices.

also generate a current in the opposite direction at $r = r_0 \ll R$, deep within the material. This current, like the current in a solenoid, generates a field in the positive B_z direction, all confined within the region $r < r_0$. This is a magnetic flux tube.

What does ϕ do near this flux tube? Consider a circle of radius $r_1 > r_0$. The integral of the magnetic flux inside this circle, $\int_{r<r_1} B_z\ r\ dr\ d\theta$, should be independent of r_1 if flux is indeed confined. On the other hand, it is also equal to $\oint_{r=r_1} d\theta\ A_\theta$. By cylindrical symmetry, A_θ can be only a function of r. From this we learn that A_θ is a constant for large r. But this poses problems. The kinetic terms for ϕ itself surely include $\vec{\nabla}\phi \cdot \vec{\nabla}\phi$, where $\nabla_i = \partial_i + iA_i$, and thus $A_\theta^2|\phi^2|/r^2$. If ϕ is a constant v at infinity, then the integral of such a term in the Hamiltonian density is divergent! So this cannot give a finite energy solution. The only way out is to have $\partial_\theta\phi = -iA_\theta\phi$, which can be accomplished if $\phi(r) = ve^{is\theta}$ at large r, where s a real constant. Furthermore, we can avoid a divergent potential energy only if v is at a potential minimum; and at the minimum $v \neq 0$ (or we would not have superconductivity!) But then single-valuedness of ϕ requires that s is an integer. Therefore this approach only works if $A_\theta = s \in \mathbf{Z}$, and thus if $\int B_z\ rdrd\theta$ is an integral multiple of a fundamental flux quantum.

From Eq. (3), we see that J^θ is now nonzero; as advertised, the flux is of necessity enclosed by a current of Cooper pairs. Furthermore, because the phase of ϕ is winding as we go once around in θ, the radial derivatives of ϕ will be ill-defined at $r = 0$ unless ϕ has a zero there. Thus we have $\phi = ve^{is\theta}f(r)$, where $f(0) = 0$ and $f(r \to \infty) \to 1$, and s an integer. The material becomes nonsuperconducting at the vortex core, paving the way for the magnetic field to pass through unobstructed.

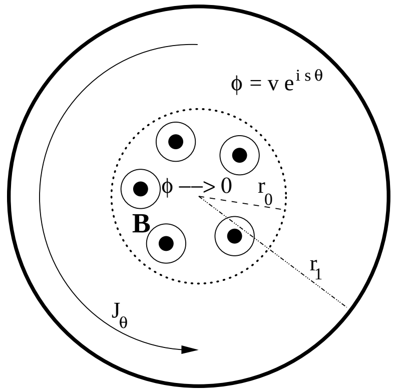


Figure 11: A flux tube of radius r_0; the phase of ϕ winds as one circles the core, in which the magnetic flux is trapped and $|\phi| < v$.

This configuration, with quantized magnetic flux and a zero for ϕ at its center, and a winding of A_θ and a corresponding winding of the phase of ϕ outside its core, is the Abrikosov vortex. Let us consder the topology associated with this vortex. We have a $U(1)$ gauge group, under which ϕ is charged. When the vacuum expectation value of ϕ is nonzero, the $U(1)$ group is broken spontaneously; gauge transformations will rotate the phase of $\langle\phi\rangle$. [However, remember that gauge transformations are not real symmetries! Therefore, unlike the case of sponaneously broken global symmetries, we do not have a continuous set of physically distinct vacua and associated Nambu-Goldstone bosons. Instead we will get a massive photon!] To make a magnetic flux tube, it must be that as we traverse a circle around the flux tube in space, the phase of the field ϕ makes a closed loop inside the $U(1)$ group. We may think of this as a map from a circle in space to a closed loop in the broken gauge group. Such a map may wind s times around the $U(1)$ as we make a single circle in space. In short, the topology of such maps, given by the first homotopy group of $U(1)$, is the group $\pi_1[U(1)] = \mathbf{Z}$. Every element in the group is labelled by an integer, the winding number s.

To round out the story, it is a bit more convenient to look at a slightly different system. Instead of studying superconductors — three-dimensional nonrelativistic systems — I will take us on a quick tour of the relativistic version, the "abelian Higgs model". This model has Nielsen-Olesen vortices [12], magnetic flux tubes very similar to those of Abrikosov.

Let us take a photon — a $U(1)$ gauge field — and a charged scalar field ϕ. The action for ϕ must be invariant under local $U(1)$ rotations $\phi \to \phi e^{i\alpha(x)}$, which can only happen if all derivatives of ϕ are covariant, that is, of the form $D_\mu\phi \equiv (\partial_\mu + iA_\mu)\phi$, where A_μ is the photon vector potential. In particular, the kinetic term for ϕ must be of the form

$$(D_\mu\phi)^\dagger D^\mu\phi \ .$$

There can also be a potential for ϕ, but gauge invariance requires it be a function only of $\phi^\dagger\phi$. In addition we should add the action for the photon. The action is thus of the form

$$-\frac{1}{4g^2}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^\dagger D^\mu\phi - V(\phi^\dagger\phi) \ .$$

The potential V may have its minimum at $\phi^\dagger\phi = 0$. In this case the vacuum of the theory is much like the one we live in; the photon is massless, propagates at maximum speed, and generates a long-range force. Magnetic and electric fields are related by a symmetry; both fall off as $1/r^2$ from magnetic and electric point charges.

However, the potential might instead have its minimum at $\phi^\dagger\phi = |v|^2 \neq 0$. Now the physics is very different. First, the photon is now massive. To see this, consider small fluctuations of electric fields A_μ for fixed $\phi = ve^{i\sigma}$. The Lagrangian for these modes is

$$-\frac{1}{4g^2}F_{\mu\nu}F^{\mu\nu} - |v|^2(A_\mu A^\mu)$$

A massive photon, which can be brought to rest, must have three polarization states ($J_3 = 1, 0, -1$) unlike a photon which has only two, $J_3 = \pm 1$. Where does this extra state come from? It comes from σ, the phase of ϕ! Let us see this; if we write $\phi = ve^{i\sigma(x)}$ the Langrangian density now becomes

$$-\frac{1}{4g^2}F_{\mu\nu}F^{\mu\nu} - |v|^2(\partial_\mu\sigma + A_\mu)^\dagger(\partial^\mu\sigma + A_\mu)$$

from which we see that σ and A_μ mix. We cannot think of them any longer as separate fields, and thus σ and A_μ together form a massive, three-polarization-state spin-one particle. (If we like, we can use a gauge transformation to set $\sigma = 0$ and absorb it into A_μ, but this merely puts the degree of freedom of σ into A_μ. It will not always be useful to do this.) This is the Higgs mechanism, discovered by Anderson (always remember that condensed matter physicists have much to teach us) and then rediscovered by many others independently, including Higgs.

Finally, we still have the magnitude of ϕ. Writing $\phi = v + \delta\phi$, we can quickly see from the Lagrangian that $\delta\phi$ acts as a neutral, massive field. I will leave this as an exercise. This means *the theory has a mass gap!* There are no massless modes and no long-range forces.

Now, what happens to electric fields in this context? Suppose I put an electric charge at the origin. The equation of free electrostatics

$$\nabla^2 A^0 = g^2 \delta(x)$$

whose solution is the usual $1/r$ electrostatic potential, is now modified. The new equation is

$$[\nabla^2 + (gv)^2] A^0 = g^2 \delta(x)$$

The solution to this equation is the Yukawa potential for a massive field with mass $m_\gamma = gv$, $V(r) \propto e^{-m_\gamma r}/r$. The electrostatic field falls off exponentially rapidly at distances larger than the inverse of m_γ. *Electric fields are screened!*.

What about magnetic fields? We cannot expel magnetic fields from an infinite system, but we can make currents, just as in superconductors, from the charged scalar ϕ, and use them to confine magnetic flux. Since the photon is massive, it is energetically preferable for the magnetic field to be localized in tubes where ϕ shrinks to zero and the photon is lighter than m_γ. On the other hand, the presence of the magnetic field in a confined region requires, as we saw, that the phase of ϕ wind an integer number of times around the center of the vortex. Classical solutions to the above equations satisfying these conditions can be found; they are called Nielsen-Olesen vortices. Their tensions can be calculated, and are proportional to $1/g^2$. Thus, *magnetic flux is confined!* The topological analysis that we did for the Abrikosov vortex — that the charges of these vortices is given by the first homotopy group of $U(1)$, the group $\pi_1[U(1)] = \mathbf{Z}$ — goes through here as well, without alteration.

Magnetic flux tubes can arise in other gauge groups as well when they are broken via the Higgs mechanism. If we have a gauge group G broken down to a smaller gauge group H (which might be the identity, as in the example above) we will get magnetic flux tubes if $\pi_1(G/H)$ is not trivial. For example, if we have the group $SU(N)$, and it is broken down to nothing, then there are no flux tubes; $SU(N)$ is simply connected, so all closed curves on it can be shrunk down to nothing, and all of its homotopy groups are trivial. However, if we break $SU(N)$ down to its center $\mathbf{Z}_N$, then since

$\pi_1(G/H) = \pi_0(H)$ if G is simply connected, and since $\pi_0(H)$ is the number of distinct components of H, we have simply $\pi_1(SU(N)/\mathbf{Z}_N) = \mathbf{Z}_N$. Magnetic flux tubes are generated, and they carry a charge in $\mathbf{Z}_N$, the integers modulo N [2]. [As an example, consider $SU(2)$. The matrices diag$(e^{i\alpha}, e^{-i\alpha})$ are in $SU(2)$; for $\alpha = 0$ and π they are in the center. The path from $\alpha = 0$ to $\alpha = 2\pi$ is a closed path in $SU(2)$, but the path from $\alpha = 0$ to $\alpha = \pi$ is not closed. However, in $SU(2)/\mathbf{Z}_2$, the matrices with $\alpha = 0$ and $\alpha = \pi$ are identified, so the second path is also closed and forms the nontrivial element of $\pi_1(SU(2)/\mathbf{Z}_2) = \mathbf{Z}_2$.]

2.2 Electric Sources and Fluxes

Let us review what we learned in the first lecture, but a bit more formally. Consider a pure gauge theory with gauge group G. Suppose we have a source — an infinitely massive, static, electrically charged particle — in a representation R of G. If we surround the source with a large sphere, what characterizes the flux passing through the sphere? If G is $U(1)$, the flux measures the electric charge directly. However, in non-abelian gauge theories the gauge bosons carry charge. Since there may be a number (varying over time) of gauge bosons inside the sphere, the representation under which the charged objects in the sphere transform is not an invariant. But, by definition, the gauge bosons are neutral under the discrete group C_G, the center of G. It follows that the charge of R under the center *is* a conserved quantity, and that the total flux exiting the sphere carries a conserved quantum number under C_G.

Electric sources and fluxes in pure gauge theories carry a conserved C_G quantum number. If the gauge group confines, then the confining electric flux tubes will also carry this quantum number.

If the theory also contains light matter charged under C_G but neutral under a subgroup C_m of C_G, then the above statements are still true with C_G replaced with C_m. For example, if we take $SU(N)$ with light fields in the $\mathbf{N}$ representation, then C_m is just the identity, reflecting the fact that all sources can be screened and all flux tubes break. If we take $SO(10)$ with fields in the $\mathbf{10}$, then the center $\mathbf{Z}_4$ is replaced with spinor-number $\mathbf{Z}_2$. Sources in the $\mathbf{10}$ will be screened and have no flux tube between them, while sources in the $\mathbf{16}$ or $\overline{\mathbf{16}}$ will be confined by a single type of flux tube.

2.3 Magnetic Sources and Fluxes

Before discussing the magnetic case, I review some basic topology. [The presentation which follows is overly naive, though it serves for present purposes. A more rigorous story requires a study of the relevant fiber bundles.] The p-th homotopy group of a manifold $\mathcal{M}$, $\pi_p(\mathcal{M})$, is the group of maps from the p-sphere into $\mathcal{M}$, where we identify maps as equivalent if they are homotopic (can be continuously deformed into one another) in $\mathcal{M}$. All we will need for present purposes are the following examples. Suppose a Lie group G has rank r, so that its maximal abelian subgroup is $U(1)^r$; then

$$\pi_2[G] = \mathbf{1} \quad \Rightarrow \quad \pi_2\left[G/U(1)^r\right] = \pi_1[U(1)^r] = \mathbf{Z} \times \mathbf{Z} \times \cdots \times \mathbf{Z} \equiv [\mathbf{Z}]^r \ . \quad (4)$$

Similarly,

$$\pi_1[G] = \mathbf{1} \quad \Rightarrow \quad \pi_1\left[G/C_G\right] = \pi_0[C_G] = C_G \ . \quad (5)$$

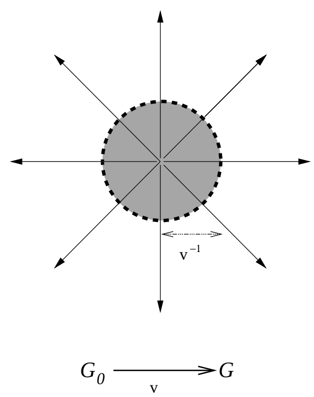


Figure 12: A magnetic monopole soliton of size v^{-1}.

We will need to investigate both monopole solitons and string solitons below. The classic monopole soliton is that of 't Hooft and of Polyakov, which arises in $SU(2)$ broken to $U(1)$; in this case the important topological relation is $\pi_2[SU(2)/U(1)] = \pi_1[U(1)] = \mathbf{Z}$. This leads to a set of monopole solutions carrying integer charge. Note that the stability of, for example, a single monopole which has charge two against decay to two monopoles, each of charge one, is determined not by topology but by dynamics. The situation is similar for the Nielsen-Olesen magnetic flux tube of the abelian Higgs model; here the relevant topological relation is $\pi_1[U(1)] = \mathbf{Z}$. This again leads to solutions with an integer charge, whose stability against decay to minimally charged vortices is determined dynamically.

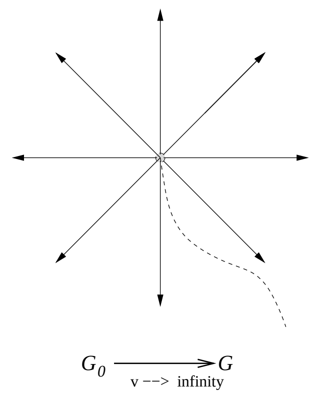


Figure 13: A pointlike Dirac monopole, with its unphysical Dirac string.

More generally, if we have a *simply connected* gauge group G_0 which breaks to a group G at a scale v, there will be solutions to the classical equations in the form of magnetic monopoles carrying a quantum number in $\pi_2[G_0/G]$ (see, for example, [13].) These will have mass [radius] proportional to v $[1/v]$. Now imagine that we take $v \to \infty$. In this limit the gauge group G_0 disappears from the system. The monopoles become pointlike and infinitely massive; their only non-pointlike feature is their (nonphysical) Dirac string, which stems from our having discarded G_0, and which carries a quantum number in $\pi_1[G]$. In short, the solitonic monopoles become fundamental Dirac monopoles in this limit. Note that since $\pi_2[G_0/G] = \pi_1[G]$, the charges carried by the solitonic monopoles and their Dirac monopole remnants are the same. At this point, we can forget about G_0, which is only relevant at infinitely high energies. Since the Dirac monopoles are heavy, we may use them as magnetic sources in a theory with gauge group G.

Let's further suppose that the gauge group G is broken completely at some scale v'. In this case no Dirac strings can exist in the low-energy theory, and so the monopoles allowed previously have seemingly vanished. However, solitonic magnetic flux tubes, carrying charges under $\pi_1[G]$, will be generated; they will have tension [radius] of order v'^2 $[1/v']$. Their $\pi_1[G]$ quantum numbers are precisely the ones they need to confine the $\pi_1[G]$-charged Dirac monopole sources of the high-energy theory. Thus, when G is completely broken, the Dirac monopoles disappear because they are confined by flux tubes.

Magnetic sources and fluxes in pure gauge theories carry a conserved

$\pi_1[G]$ quantum number. If the gauge group is completely broken, then the confining magnetic flux tubes will also carry this quantum number.

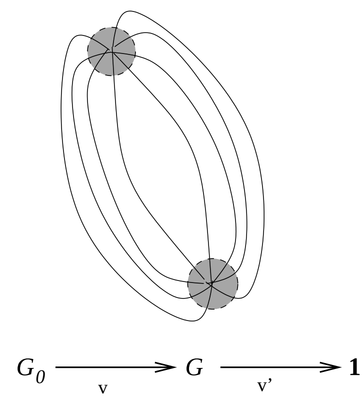


Figure 14: Confined monopole solitons in a theory with flux tubes.

3 Electric-Magnetic Duality?

So let us observe something about $SU(N)$. The electric fluxes of $SU(N)$ are in $C_{SU(N)} = \mathbf{Z}_N$, while its magnetic fluxes are in $\pi_1[SU(N)] = \mathbf{1}$. The electric fluxes of $SU(N)/\mathbf{Z}_N$ are in $C_{SU(N)/\mathbf{Z}_N} = \mathbf{1}$, while its magnetic fluxes are in $\pi_1[SU(N)/\mathbf{Z}_N] = \mathbf{Z}_N$. In fact, more generally, for k a divisor of N, a theory with $SU(N)/\mathbf{Z}_k$ has electric fluxes in $\mathbf{Z}_{(N/k)}$ and magnetic fluxes in $\mathbf{Z}_k$. This electric-magnetic symmetry appears very interesting. What does it mean?

3.1 Duality in Maxwell's theory

The symmetry between electric and magnetic fields in the case of classical electromagnetism is well known. If there are no electric charges present, the Maxwell equations have a symmetry $E \to B$, $B \to -E$. This is physically meaningful, since E and B are both gauge invariant. Without charges, there is no way to say which type of field is which.

Let us be more explicit. Under this transformation, the Bianchi identities $\nabla \times E + \dot{B} = 0, \nabla \cdot B = 0$ are exchanged with the equations of motion $\nabla \times B - \dot{E} = 0, \nabla \cdot E = 0$. Said more covariantly,

$$F_{\mu\nu} \to \tilde{F}_{\mu\nu} = \epsilon_{\mu\nu\rho\sigma} F^{\rho\sigma}$$

and the Bianchi identity $\epsilon^{\mu\nu\rho\sigma}\partial_\rho F_{\mu\nu} = 0$ goes to the equation of motion $\partial^\mu F_{\mu\nu} = 0$.

None of this is particularly obvious if one uses the formalism of potentials, and with good reason. Because of the Bianchi identities, we are free to write $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, which defines A_μ. The symmetry of the equations under $A_\mu(x) \to A_\mu(x) + \partial_\mu \chi(x)$ is the $U(1)$ gauge symmetry — let us call it the "electric" gauge symmetry. Notice it is not a symmetry of anything physical! It is a symmetry of the variables A_μ! The physical quantities — E and B — are gauge invariant, and are trivial under this "symmetry." This is a good thing, because under exchange of E and B, we cannot exchange A_μ with anything. We must introduce a new, and entirely different, vector potential C_μ, with $\tilde{F}_{\mu\nu} = \partial_\mu C_\nu - \partial_\nu C_\mu$. *No local expression will convert A_μ to C_μ.* Furthermore, C_μ has its own $U(1)$ symmetry — let us call it "magnetic" — $C_\mu(x) \to C_\mu(x) + \partial_\mu \rho(x)$. This is just as unphysical as the first $U(1)$. Even if we were to find a transformation from A_μ to C_μ, nonlocal as it would be, *we are free to redefine C_μ through a magnetic $U(1)$ transformation separately from any redefinition of A_μ through an electric $U(1)$ transformation!* There are two $U(1)$ groups here; they are two entirely distinct symmetries of two entirely distinct sets of variables, and both are unphysical. When we say that the Maxwell equations are the equations of a $U(1)$ gauge theory, we are being extremely careless with the truth.

Let's see this as a path integral statement. (I learned the following from Seiberg and Witten's first paper [9].) They start with the free Maxwell theory

$$\int \mathcal{D}A \; e^{-i\int \frac{F^2}{4g^2}} \;\; \delta(\partial \cdot A)$$

(I will suppress indices except where clearly needed.) Notice the gauge fixing term. This expression is just a number. Let us instead write something more useful. Let's introduce a source $J^{\mu\nu}$ for $F_{\mu\nu}$, and write

$$Z[J] = \int \mathcal{D}A \; e^{-i\int \frac{F^2}{4g^2} + \int JF} \;\; \delta(\partial \cdot A) \; .$$

Functional derivatives of $\ln Z$ with respect to J now give the correlation functions of F.

Let us now change variables in this path integral. Up to an overall constant. the path integral can be rewritten as an integral, not over A, but over F. We have to be careful, though, because F is subject to the Bianchi identities, which are exact operator identities. Consider the expression

$$Z[J] = \int \mathcal{D}F \ e^{-i\int \frac{F^2}{4g^2} + \int JF} \ \delta(\epsilon^{\mu\nu\rho\sigma}\partial_\rho F_{\mu\nu}) \ .$$

There's no gauge fixing needed now, but the Bianchi identities must be implemented through a delta function. Let us rewrite this Bianchi identity using a Lagrange multiplier which we will for some unknown reason call C_μ,

$$Z[J] = \int \mathcal{D}F\mathcal{D}C \ e^{-i\int \frac{F^2}{4g^2} + \int JF + \frac{i}{4\pi}\int \epsilon^{\mu\nu\rho\sigma} C_\sigma \partial_\rho F_{\mu\nu}} \ \delta(\partial \cdot C) \ .$$

Notice that the integral over C enforces the Bianchi identity, but since $\epsilon^{\mu\nu\rho\sigma}\partial_\rho\partial_\sigma F_{\mu\nu} = 0$, the Langrange multiplier field C itself has a gauge invariance, which must be fixed by the new delta function. Now let us integrate by parts

$$\epsilon^{\mu\nu\rho\sigma} C_\sigma \partial_\rho F_{\mu\nu} = -\epsilon^{\mu\nu\rho\sigma}\partial_\rho C_\sigma F_{\mu\nu} + \partial^\rho(\cdots) \equiv -\frac{1}{2}\tilde{F}_C F + \partial^\rho(\cdots)$$

where F_C is the field strength of C and $\sim$ represents contraction with an ϵ tensor. We next carry out the integral over F, obtaining

$$Z[J] = e^{-ig^2\int J^2} \int \mathcal{D}C \ e^{-i\int \frac{g^2}{64\pi^2} F_C^2 - \frac{g^2}{4\pi}\int J\tilde{F}_C} \ \delta(\partial \cdot C) \ ,$$

where I have used $\tilde{F}_C^2 = F_C^2$. Thus we recover a free Maxwell theory for C! It looks identical to the original one, except (1) g has been replaced with $\tilde{g} = 4\pi/g$ — weak coupling and strong coupling have been exchanged — (2) the source J, which coupled to F, now couples to $\frac{g^2}{4\pi}\tilde{F}_C$, so the electric field F/g of A_μ is proportional to the magnetic field $\tilde{F}_C/\tilde{g}$ for C_μ, and (3) there is a contact term proportional to J^2 (a typical quantum subtlety which does not affect Green's functions of fields at different points — you may want to experiment with Fourier transforms of Gaussian integrals to see why it is there.) Thus we have found that we can express a single quantum theory (in the form of a generating function for gauge-invariant correlation functions) using two, entirely distinct, integral representations, both of which are nice-looking and well-behaved. One theory, two descriptions, each with its own $U(1)$ gauge (non-)symmetry. This is *duality*.

3.2 The addition of charged fields

Only when we add charges to the theory do we start to learn the distinction between electric and magnetic fields. We know that in nature we have only electric charges, and all symmetry between electric and magnetic charges is lost. And yet — what if there are magnetic monopoles? Could the symmetry be restored?

Yes, and no. If we put both electric and magnetic currents in the classical Maxwell equations, they look beautifully symmetric: the Bianchi identity $\epsilon^{\mu\nu\rho\sigma}\partial_\rho F_{\mu\nu} = J^\sigma_m$ is exchanged with the equation of motion $\partial_\mu F^{\mu\nu} = J^\nu_e$. All seems well. Electric charges have charge e, while magnetic charges are proportional to $1/e$; thus if electrons are weakly coupled, monopoles are strongly interacting with the photon, and vice versa.

So let us return to the question of confinement. We have seen that we can use condensing electric charges to cause electric charge to be screened, and make magnetic flux confined through the Meissner effect. Clearly, there should be a "dual" Meissner effect; *if we have condensing magnetic charges, then magnetic charge will be screened and electric flux will be confined.* In both cases there will be a mass gap in the theory. Thus we now have a guess as to how confinement will occur: if there are some magnetically charged objects around — perhaps composite ones not visible even semiclassically — then their condensation would cause electric flux to be confined via the dual Meissner effect. All we have to do now is write down the equations governing this process, and see that in such a world, electrons are confined by flux tubes...

But there's a problem. The Bianchi identities are now $\nabla \times E + \dot{B} = J_{mag}$, $\nabla \cdot B = q_{mag}$. This means we *cannot* introduce A_μ anymore; the very introduction of the vector potential imposes the Bianchi identities with zero for the right-hand sides. If we want to introduce a *magnetically* charged field, we will have to use C_μ. In this case, the equations for C_μ will look exactly the same as they did before for A_μ, simply relabelled. And that's not good, if we want to see that *electrically* charged particles are confined. Fields for *electrically* charged particles must have kinetic terms defined using covariant derivatives which contain A_μ! We cannot write a local expression for an electron's kinetic terms if we only have C_μ. Even worse, the presence of the electron field ruins the Bianchi identity for C_μ, so we can't really introduce C_μ either. There isn't going to be a local Lagrangian, and there isn't going to be an ordinary, classical analysis. All we have is a mess.

And that's before quantum mechanics. These complications prevent us from repeating the argument for duality using the path integral. Once there are charged fields, we do not know how (as of yet — though see [14]) to write a path integral which converts an electric description of a theory to an magnetic one. (Furthermore, in contrast to $U(1)$ without charged matter, there is in fact little reason to expect that a $U(1)$ theory with charged matter is actually quantum-mechanically dual to an identical theory; it could easily be dual to a nonabelian gauge theory, and/or have multiple dual representations [15].)

We could, of course, forgo the electrically charged particles. Then we would just have a photon coupled to magnetically charged particles; but this would look exactly the same as the superconductor we just considered. That won't help us with Yang-Mills theory, or any other theory with electric confinement that we would like to understand. In such theories, the gluons themselves are chromoelectrically charged, and we can't simply choose to discard all possible chromoelectrically charged objects.

3.3 Duality in pure Yang-Mills?

Can we find a similar duality for the pure Yang-Mills theory? We know that Yang-Mills has the property that it generates electric flux tubes with $\mathbf{Z}_N$ quantum numbers. We might hope that Yang-Mills has an obvious duality to some theory with a gauge group H which has $\pi_1(H) = \mathbf{Z}_N$, so that when H is broken by a condensing field, it generates magnetic flux tubes with $\mathbf{Z}_N$ charges. A natural guess for H would be $SU(N)/\mathbf{Z}_N$. Of course we will need some additional matter — at least a couple of scalar fields — if we are to break this gauge group completely, so the dual description of this theory *can't itself be pure Yang-Mills.* Is there any hope that there exists a dual $SU(N)/\mathbf{Z}_N$ gauge theory of some type, which gives a weakly-coupled (and therefore calculable) dual description analogous to the Meissner effect of confinement in Yang-Mills?

This type of idea, popular briefly in the 1970s, has a few serious problems. First, unlike the case of $U(1)$ gauge theories, the electric and magnetic fields of $SU(N)$ are in the adjoint representation of the gauge group and are not themselves gauge-invariant. This makes the Bianchi identities $\epsilon^{\mu\nu\rho\sigma}D_\rho F^{\mu\nu} = 0$ nonlinear. Secondly, their equations of motion $D_\mu F^{\mu\nu} = 0$ are nonlinear. In both expressions, covariant derivatives appear, which means we always have to write expressions using the vector potential A_μ. This means we cannot simply exchange electric and magnetic fields as we did in the classsical Maxwell equations; the potential appears in the classical equations. At the quantum level, this is equally problematic; the path-integral trick used above for $U(1)$ is useless here, since it required we write the path integral only in terms of F. (Note Halpern [16] showed in the 1970s that it is consistent to write A nonlocally in terms of F inside a path integral, but no one has figured out how to make use of this fact.)

Another complication is that Yang-Mills theory has a running coupling constant. At high energies it is weak. (Any magnetic description therefore will be strongly coupled at these high energies, but we don't mind that, since the original description is weakly coupled, and extremely useful, in this regime.) At low energies, below Λ, it is strong — but how strong? Is it infinite, or merely order 1? This is important, because we are interested in trying to find a dual description of confinement which presumably inverts the coupling constant $g \to 1/g$. Unless the gauge coupling is much larger than one, our dual description will itself have a large coupling (of order one) and we won't be able to use it for a semiclassical description of the physics. In this case the dual magnetic description will be as hard to use as our original, electric one.

Unfortunately, all indications are that the coupling in the region near Λ is closer to $\sqrt{4\pi}$ than to infinity. There is no evidence that the theory at low energies has a weakly-coupled magnetic description, and the dynamics of the theory does not seem to have any small parameters, or large separations of scales, which could make it easier to analyze. The nonperturbative physics of Yang-Mills may just be a hard problem.

We might be stuck. But here's an idea. What happens if we make the gauge coupling g artificially large? Maybe in that limit a dual description can be found, and its description of confinement will be easier to study and to use. And maybe from there we can get back to the Yang-Mills theory that we want to understand.

How could we do this? Well, let's review... why does the coupling become small in the ultraviolet? It does so because the theory is asymptotically free; its beta function is negative, so the coupling becomes smaller and smaller as we go to high energy. We can't avoid this region of small coupling unless we do something drastic...

Well, one drastic thing we can do is put the theory on a lattice. This means there is a shortest distance below which there can be no vibrations; the theory only looks like pure Yang-Mills at much longer scales. The ultraviolet modes are simply removed, so we won't have to worry about the theory becoming weakly coupled at high energy. In fact, we are free to choose the coupling constant $g(a^{-1})$ at the energy a^{-1} corresponding to the lattice spacing a. Instead of choosing it small and allowing the theory to run to strong coupling at low energy, let's just choose $g(a^{-1})$ very large. What happens?

In this case we can do a "strong-coupling expansion". I won't review this here, but the expansion on the lattice in powers of $1/g^2$ can in fact be performed [1], and one sees the existence of confining strings right away. There, we're done. Yang-Mills confines chromoelectric field, and Strassler's lectures are over.

Or does it? The problem is that the theory on the lattice has very different dynamics from that of pure Yang-Mills. If $g(a^{-1})$ is very large, then the confinement scale Λ will be at the same order as $1/a$. This can be seen from Eq. (1) with $\mu_0 = a^{-1}$, using the fact that $e^{-8\pi^2/g^2} \sim 1$ if $g^2 \gg 1$. There will be no separation between the scale of the lattice and the scale of confinement. The mass gap will be at this scale also, so there will be no long-distance physics at all. All of the glueball spectrum will be sensitive to the lattice. Thus the theory is very different from Yang-Mills, in fact. If we change the lattice from a square lattice to a triangular one, we will change the glueball spectrum significantly. So why should the fact that the lattice theory confines convince us that when we take the limit

$$a \to 0,\ g^2 \to 0,\ \Lambda^{11N/3} = a^{-11N/3} e^{-8\pi^2/g^2(a)} \text{ fixed},$$

thereby recovering the pure Yang-Mills theory, that the confinement, the flux tubes, and the mass gap will actually survive? Couldn't there easily be a phase transition at some value of g which would change the physics completely?

It's a serious objection. Indeed, we see here a general approach at work, and its basic advantages and disadvantages. Let's review them. We can't study Yang-Mills directly; it is too hard. But let's *change the theory* in a way that allows us to artificially make a parameter small (in this case $1/g^2$.) By doing so, we permit a new expansion in powers of the small parameter. This gives us a calculational technique in which it may be possible to show that confinement and other nonperturbative properties do actually occur, and explain how and why they arise. That's a great idea; and it works, too! But we changed the theory; it is related continuously to Yang-Mills, but that's all. Let's now try to go back to Yang-Mills itself. The problem is that our small parameter will become large again as we do so, and we have no guarantee that confinement, etc., and especially the explanation for confinement, will survive as we make our way back to our starting point. This is especially true since the dynamics of the theory with the small parameter *depends in* *detail on how we changed the theory.*

In fact, experience shows that in considering a variety of weakly-coupled variations on pure Yang-Mills, one finds (1) all of the reasonable variations confine, tending to confirm that Yang-Mills confines, and (2) each variation has its own, separate explanation as to how confinement happens. The various explanations have a few things in common but their details are different. Is this progress? I leave this as a question for you to decide.

Similar issues arise for $\mathcal{N}=1$ SYM. There are many ways to distort the theory (see for example [17, 3, 9]) so that it becomes easier to study; each shows that the theory confines, although each gives a somewhat different explanation. In the remaining part of these lectures, we will be choosing a couple of these variations, and studying how confinement occurs in these cases. We will embed the $\mathcal{N}=1$ SYM theory into $\mathcal{N}=4$ SYM — the most symmetric of all gauge theories — and use the dualities of $\mathcal{N}=4$ to study the confinement in $\mathcal{N}=1$ (and possibly, if the mathematics is kind, of pure Yang-Mills itself.)

3.4 $\mathcal{N}=4$ Supersymmetric Gauge Theory

We now need to review some properties of $\mathcal{N}=4$ supersymmetric gauge theory. We will take the gauge group to be $SU(N)$ unless otherwise noted. The theory consists of one gauge field, four Majorana fermions, and six real scalars, all in the adjoint representation. It is useful to combine these using the language of $\mathcal{N}=1$ supersymmetry, in which case we have one vector multiplet (the gauge boson A_μ and one Majorana fermion λ) and three chiral multiplets (each with a fermion ψ^s and a complex scalar Φ^s, $s=1,2,3$.)

These fields have the usual gauged kinetic terms, along with additional interactions between the scalars and fermions. I won't write them all here (you can find them in many books and review articles on supersymmetry) and will instead focus on the potential energy for the scalars.

$$V(\Phi^s) = \sum_{a=1}^{dim\ G} |D_a|^2 + \sum_{s=1}^{3} |F_s|^2 \tag{6}$$

where

$$D_a = \left(\sum_{s=1}^{3} [\Phi^{s\dagger}, \Phi^s] \right)_a \tag{7}$$

(here a is an index in the adjoint of G) and

$$F_s = \epsilon_{stu} [\Phi^t, \Phi^u] \ . \tag{8}$$

Supersymmetry requires that $\langle V(\Phi^s)\rangle = 0$, and so all D_a and F_s must vanish separately. The solution to these requirements is that the matrices are all diagonal, namely

$$\langle \Phi^s \rangle = \mathrm{diag}(v_1^s, v_2^s, \cdots, v_N^s) \ . \tag{9}$$

If the v_i^s, thought of as N vectors $\vec{v}_i$, $i = 1, \ldots N$, in a three-dimensional complex space, are all distinct, this breaks G to $U(1)^r$. Since $\pi_2[G/U(1)^r] = [\mathbf{Z}]^r$ [see Eq. (4)] the theory has monopoles carrying r integer charges under $U(1)^r$. (Quantum mechanically, the theory also has dyons, carrying r electric and r magnetic charges (n_e, n_m) [18].)

The space of vacua written in Eq. (9) is not altered by quantum mechanics. In the generic $U(1)^r$ vacuum, each $U(1)$ has no charged matter, and consequently has the usual electric-magnetic duality of the Maxwell equations.

When all v_i^s are zero, the gauge group is unbroken. The theory is conformally invariant. All reasonable Green's functions are power laws. All reasonable operators have a definite, fixed, dimension. The gauge coupling g has an exactly-zero beta function, and does not run. Thus, in contrast to QCD, YM, and $\mathcal{N} = 1$ SYM, the $\mathcal{N} = 4$ SYM theory has a truly dimensionless coupling constant; there is no strong-coupling scale Λ, no dimensional transmutation. We can dial this truly dimensionless g to be whatever we like — it can be small, or it can be large — and it will stay that way at all energy scales. And this nonabelian gauge theory, with lots of charged matter, has a generalization of electric-magnetic duality, suggested first by Montonen and Olive in 1977 [19], in which this coupling constant is inverted.

3.5 Montonen-Olive Duality

Like the pure Maxwell theory, the $\mathcal{N} = 4$ theory has more than one description. There's lots of evidence for this, although it has not been proven directly. Consider this an open challenge.

There is actually an infinite set of alternate descriptions (one has to talk about the θ angle of the theory to obtain them, and I will not have time to cover this here) but the most important one, for our purposes, exchanges electric and magnetic charges. It is generated by a change of variables $\mathbf{S}$ analogous to the one we discussed above for electromagnetism, but whose explicit form remains a mystery. It has the effect

$$\mathbf{S} \ : g \to \frac{4\pi}{g}; q_e \leftrightarrow q_m \ ; \ G \to \tilde{G} \ . \tag{10}$$

S exchanges electric and magnetic charge, inverts the gauge coupling [19], and changes the gauge group [20, 21] from G to its dual group $\tilde{G}$, as defined below.

The group G has a root lattice Γ_G which lies in an $r = \mathrm{rank}(G)$ dimensional vector space. This lattice has a corresponding dual lattice $(\Gamma_G)^*$. It is a theorem that there exists a Lie group whose root lattice $\Gamma_{\tilde{G}}$ equals $(\Gamma_G)^*$ [20]. Here are some examples:

$$\begin{array}{ll} SU(N) \leftrightarrow SU(N)/\mathbf{Z}_N \ ; & SO(2N+1) \leftrightarrow USp(2N) \ ; \\ SO(2N) \leftrightarrow SO(2N) \ ; & Spin(2N) \leftrightarrow SO(2N)/\mathbf{Z}_2 \ . \end{array} \tag{11}$$

Notice that this set of relationships depends on the global structure of the group, not just its Lie algebra; $SO(3)$ (which does not have spin-1/2 representations) is dual to $USp(2) \approx SU(2)$ (which does have spin-1/2 representations.) These details are essential in that they affect the topology of the group, on which Montonen-Olive duality depends.

In particular, there are two topological relations which are of great importance to Montonen-Olive duality. The first is relevant in the generic vacuum, in which G is broken to $U(1)^r$. The electric charges under $U(1)^r$ of the massive electrically charged particles (spin $0, \frac{1}{2}, 1$) lie on the lattice Γ_G. The massive magnetic monopoles (*also* of spin $0, \frac{1}{2}, 1$) have magnetic charges under $U(1)^r$ which lie on the dual lattice $(\Gamma_G)^*$ [20, 21]. Clearly, for the **S** transformation, which exchanges the electrically and magnetically charged fields and the groups G and $\tilde{G}$, to be consistent, it is essential that $\Gamma_{\tilde{G}} = (\Gamma_G)^*$ — which, fortunately, is true.

The second topological relation is the one we will use below. We have seen that the allowed electric and magnetic sources for a gauge theory with adjoint matter (such as $\mathcal{N} = 4$) are characterized by quantum numbers in C_G and $\pi_1(G)$ respectively. Consistency of the **S** transformation would not be possible were these two groups not exchanged under its action. Fortunately, it is a theorem of group theory that [20]

$$\pi_1(G) = C_{\tilde{G}} \ ; \ \pi_1(\tilde{G}) = C_G \ . \tag{12}$$

For example, $\pi_1[SU(N)] = C_{SU(N)/\mathbf{Z}_N} = \mathbf{1}$ while $C_{SU(N)} = \pi_1[SU(N)/\mathbf{Z}_N] = \mathbf{Z}_N$.

Thus, as a consequence of Eq. (12) and the results discussed in our earlier discussions of electric and magnetic fluxes and sources, the allowed magnetic sources of G are the same as the allowed electric sources for $\tilde{G}$, and vice

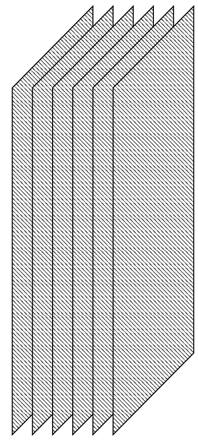


Figure 15: N D3 branes have a $U(N)$ $\mathcal{N} = 4$ SYM on their world volume.

versa. This is a significant piece of evidence in favor of S-duality, and will be essential later on.

Now, this is not the only way to approach $\mathcal{N} = 4$ SYM, as you have already heard in Prof. Maldacena's lectures. As he showed you, the world-volume theory on a stack of N D3 branes of Type IIB string theory has a complicated action, but at low energy it reduces to $\mathcal{N} = 4$ $U(N)$ SYM theory. The extra $U(1)$ decouples, and all of the interesting physics is in the $SU(N)$ part of the theory.

Do we see signs of S-duality in this string construction of $\mathcal{N} = 4$ SYM? We certainly do! Type IIB string theory itself has an S-duality — for which, again, there is tremendous evidence but no proof (see for example [22] and [23].) The duality inverts the string coupling: $g_s \to 1/g_s$. It also changes various extended objects into one another. The theory has (among other things) fundamental strings, Neveu-Schwarz 5-branes, and D1, D3 and D5 branes. (It also has D(-1) and D7 branes but we won't discuss them.) Now, under S-duality, the D1 and F1 (fundamental) strings are exchanged, as are the D5 and NS5 branes. The D3 branes, however, are unchanged. The $\mathcal{N} = 4$ $SU(N)$ SYM theory goes back to itself, except that its coupling constant $g_{YM}^2 = g_s/4\pi$ is inverted — just as we expected! Furthermore, a fundamental string ending on a D3 brane looks like a point electric charge from the perspective of an observer stuck on the D3 brane. A D1 brane ending on a D3 brane looks like a point magnetic charge. Thus S-duality in

Type IIB string theory correctly inverts the $\mathcal{N}=4$ SYM coupling constant, exchanges its electric and magnetic charges, and exchanges the gauge groups of the electric and magnetic descriptions.[3]

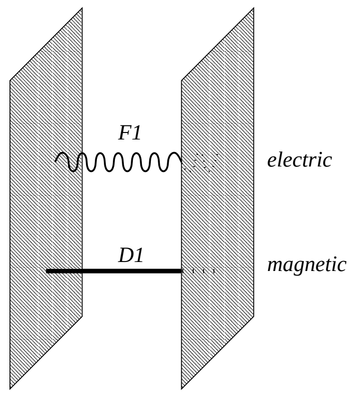


Figure 16: F1 (D1) strings appear as electrically (magnetically) charged particles.

A word of warning about this beautiful structure. Most examples of duality are much more complicated than this! The identification of the dual group is vastly more difficult, and the relations which we have used in arguing that it is $\tilde{G}$ do not work. So don't be fooled into thinking that most of the other known dualities are this elegant. They are both less straightforward and much richer in content. A good example for you to look at is the Seiberg duality of $\mathcal{N}=1$ supersymmetric gauge theories [17, 24, 25], which could actually be relevant in nature. But the example of $\mathcal{N}=4$ duality proves to be a good one for examining confinement in $\mathcal{N}=1$ SYM and pure YM, so we'll stick with it.

4 Breaking $\mathcal{N}=4$ to $\mathcal{N}=1$

It's time to return to our goal of discussing confinement in $\mathcal{N}=1$ SYM theory. Let's try to apply the trick we discussed earlier in the context of the strong-coupling expansion on the lattice. Is there, perhaps, a way to take $\mathcal{N}=1$ SYM, make its coupling artificially large, and do a strong-coupling expansion? The lattice badly breaks supersymmetry, so it won't help us very

[3]Well, almost. Actually, the D3-branes give $U(N)$, whose dual is $U(N)$ again. To remove the $U(1)$ factors, and see the $\mathbf{Z}_N$, is subtle. It is much easier to see that $SO(2N+1)$ is exchanged with $USp(2N)$, so you might try that instead.

much (although it might be worth revisiting this point after recent advances in lattice theory [26].) A different approach would be to put $\mathcal{N}=1$ Yang-Mills theory inside of $\mathcal{N}=4$ Yang-Mills. How might we do this?

We could add to the $\mathcal{N}=1$ SYM theory three chiral multiplets (that's three Majorana fermions and six real scalars) in the adjoint representation of the group, all with a common mass m. We'll also add some additional interactions, so that when m goes to zero the theory has $\mathcal{N}=4$ supersymmetry. We take all scalars to have expectation values less than or of order m (an assumption which will be justified later.)

At energies well above m, the theory is approximately $\mathcal{N}=4$ SYM. Since the masses m are comparatively tiny at these energy scales, the theory will be approximately conformally invariant. The gauge coupling will run very little for energies bigger than m, and for very high energy it goes to a constant g_0. But at energies well below m, the classically massless particles will be those of $\mathcal{N}=1$ SYM. Quantum mechanically, the gauge coupling will run below the scale m, and confinement will presumably occur at some scale $\Lambda < m$.

Thus this $\mathcal{N}=1$ supersymmetric theory — which we will call "$\mathcal{N}=1^*$", for short — interpolates between $\mathcal{N}=4$ SYM and $\mathcal{N}=1$ SYM. As required for our trick, we have kept the basic $\mathcal{N}=1$ SYM infrared dynamics but have changed the ultraviolet behavior of the theory in such a way that we can, if we wish, ensure the coupling constant is always large! In particular, we can simply choose the ultraviolet value of the coupling g_0 much larger than one. Since $g(\mu)\approx g_0$ for $\mu > m$, the coupling constant at $\mu = m$ will also be large — and thus, just below the scale m, we obtain a theory with the matter content of $\mathcal{N}=1$ SYM, but with an artificially large coupling constant. All we have to do now is expand in $1/g_0$. But that's exactly what Montonen-Olive duality allows us to do! The magnetic dual description of this physics will be weakly coupled, with coupling constant $1/g_0 \ll 1$.

But how close will the $\mathcal{N}=1^*$ theory be to $\mathcal{N}=1$ SYM? What properties will they share? It is worth examining the strong coupling scale of the $\mathcal{N}=1^*$ theory. Below the scale m, the coupling constant $g(\mu)$ will run as it does in pure $\mathcal{N}=1$ SYM theory, so the one-loop relation between $g(\mu)$ at the scale $\mu=m$ and the scale Λ reads

$$\Lambda^{3N} = m^{3N} e^{-8\pi^2/g^2(m)} \approx m^{3N} e^{-8\pi^2/g_0^2}$$

Notice that if g_0 is small, $\Lambda \ll m$, but if g_0 is large, as we will want for our strong-coupling expansion, $\Lambda \sim m$. Thus, just as in the lattice strong-

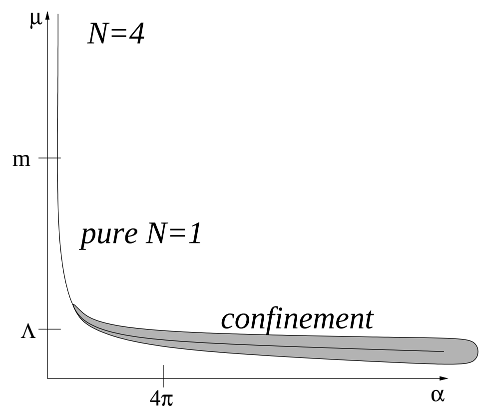


Figure 17: $\mathcal{N}=1^*$ for small g_0.

coupling expansion, there will not be a separation of scales between the new physics (in this case the three massive adjoint multiplets) and the scale of confinement, glueball masses, etc. We will not be doing much better than the lattice case. Our strong-coupling expansion will depend on the details of our the mass scale m. For example, if we give the extra chiral multiplets different masses instead of a common mass m, the glueball spectrum will reflect this change, although there would be no such change at small g_0 where $m \gg \Lambda$. This is the standard limitation; we accept it and move on.

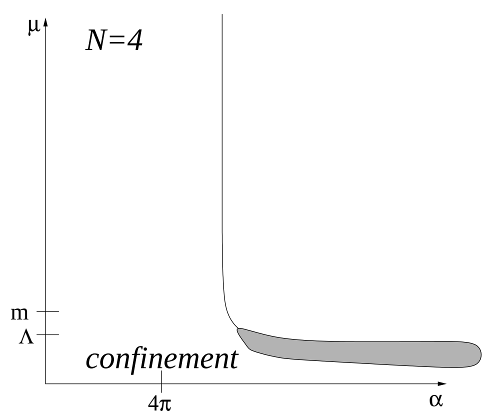


Figure 18: $\mathcal{N}=1^*$ for large g_0.

You might wonder if there is some danger that the massive chiral multiplets will ruin the confinement we want to study. In fact, there is not much to worry about. As we noted earlier, $\mathcal{N}=1$ SYM has confining strings because neither gluons nor gluinos can break these flux tubes; fields in the adjoint representation are neutral under the center of the gauge group C_G. The addition of massive matter in the adjoint representation does not change this; heavy particles would only obstruct confinement by breaking flux tubes, which adjoint matter cannot do. We therefore can expect that $\mathcal{N}=1^*$ should share some qualitative features with pure $\mathcal{N}=1$ SYM: both should have mass gaps and confine flux into tubes carrying a C_G quantum number.

Now let's examine things more closely. Let's first take g_0 very small so we can do a semiclassical analysis. When we break the $\mathcal{N}=4$ supersymmetry by adding masses m for the fields Φ^s, the F_s functions of (8) become

$$F_s = \epsilon_{stu}[\Phi^t, \Phi^u] + m\Phi^s \ , \tag{13}$$

so that $F_s = 0$ implies $\epsilon_{stu}[\Phi^t, \Phi^u] = -m\Phi^s$ [8]. Up to normalization, these are the commutation relations for an $SU(2)$ algebra; thus solutions will take the form

$$\Phi^1 = -imJ_x \ ; \Phi^2 = -imJ_y \ ; \Phi^3 = -imJ_z \ , \tag{14}$$

where J_x, J_y, J_z are $N \times N$ matrices satisfying $[J_x, J_y] = iJ_z$, *etc.*, a representation of $SU(2)$. Each possible gauge-inequivalent choice for the J's gives a separate, isolated vacuum of the classical $\mathcal{N}=1^*$ theory [8].

How does this work, explicitly, in $SU(N)$? We can write the Φ^s as $N \times N$ traceless matrices, so the J_s should be an N-dimensional (generally reducible and possibly trivial) representation of $SU(2)$ [8, 3]. The trivial choice corresponds to $J_i = 0$; clearly if $\Phi^s = 0$ the JJ commutation relations are satisfied. We will call the corresponding vacuum the "unbroken" vacuum, since the $SU(N)$ gauge group is preserved. Another natural choice is to take the J_s in the irreducible spin-$\frac{N-1}{2}$ representation of the $SU(2)$. In this case $SU(N)$ is completely broken (this is left as an exercise); we will call this the "Higgs vacuum". We may also choose the J_s in a reducible representation

$$J_s = \begin{bmatrix} \sigma_s & | & 0 \\ -- & -|- & -- \\ 0 & | & 0 \end{bmatrix} \ ; \tag{15}$$

here the σ_s are the Pauli matrices. In this case $SU(N)$ is partly broken. There are many vacua like this last one, but they will play no role in today's story; we will only need the unbroken vacuum and the Higgs vacuum.

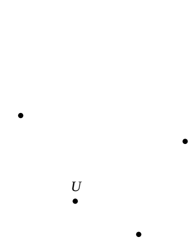


Figure 19: A few of the classical vacua of $\mathcal{N} = 1^*$, including the unbroken (U) and completely Higgesd (H) vacua.

In all of these vacua, the scalar fields are massive, as are most of the fermions. However, in any vacuum with unbroken gauge symmetry, there are both massless gauge bosons and their massless fermionic superpartners. Thus, the Higgs vacuum has a mass gap — there are no massless fields — while the unbroken vacuum has the massless gauge bosons and fermions of an $SU(N)$ $\mathcal{N} = 1$ SYM theory.

As an example, let's take the case of an $SU(2)$ gauge group [8]. This is a rather degenerate one, but it has all the essential features. In this case we need two-by-two matrices which satisfy the above commutation relations; the only solutions are $J_s = 0$ and $J_s = im\sigma_i$. We thus have two classical vacua, one with unbroken $SU(2)$ gauge symmetry, and one in which the $SU(2)$ is completely broken by the Higgs mechanism. (The expectation value for Φ^3 breaks $SU(2)$ to $U(1)$, while the expectation values for Φ^1 and Φ^2 break the remaining $U(1)$.)

In summary, the classical analysis of the $SU(N)$ $\mathcal{N} = 1^*$ theory shows that it has isolated supersymmetric vacua scattered about, with the unbroken (U) vacuum at the origin of field space and the Higgs vacuum (H) at large Φ^s expectation values (of order m) [8, 3]. The Higgs vacuum has a mass gap, while the unbroken vacuum has the matter content of an $SU(N)$ $\mathcal{N} = 1$ SYM theory.

4.1 OM Duality and the Yang-Mills String

The above picture is modified by quantum mechanics. The U vacuum has the matter content of $SU(N)$ $\mathcal{N} = 1$ SYM theory. Remember we are still working at small g_0. We know this theory is asymptotically free, so at an energy scale exponentially small compared to m — more precisely, at an energy $\Lambda \sim me^{-8\pi^2/3Ng_0^2} \ll m$ — the gauge coupling will become strong.

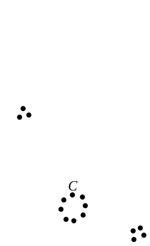


Figure 20: Quantum mechanically, the vacua with unbroken gauge groups split; the U vacuum splits into N, one of which (C) has confinement via magnetic monopole condensation.

Since this scale is so small, the physics at energies of order m cannot affect it. We already know, then, what this theory will do; it will confine, generate a mass gap of order Λ, and it will have not one but N vacua due to the breaking of the $\mathbf{Z}_{2N}$ axial symmetry down to $\mathbf{Z}_2$. As we noted earlier, these vacua are related by this $\mathbf{Z}_{2N}$ symmetry, so we can focus on just one of them.[4] Let's call it the confining (C) vacuum.

By contrast, in the H vacuum the gauge group is completely broken at the scale $m \gg \Lambda$, and there is a mass gap of order m, so there is no way for non-trivial low-energy dynamics to take place. Consequently, the H vacuum remains a single vacuum.

Now let's compare the Higgs vacuum and the confining vacuum. Recall that we took the gauge group to be $SU(N)$. The confining vacuum has a strongly-coupled process of confinement and generation of a mass gap of order $\Lambda \ll m$. We expect the confining electric flux tubes to have tension of order Λ, and for them to carry a $\mathbf{Z}_N$ charge. In the Higgs vacuum, on the other hand, there is a weakly-coupled breaking of the gauge group. We can see classically that a mass gap is generated. But actually the gauge group is not completely broken. The adjoint scalar fields carry no charge under the center of the group, so $SU(N)$ is in fact broken down to its center $\mathbf{Z}_N$! Now, we have already learned that there will be solitonic magnetic flux tubes in any breaking of a gauge theory $G \to H$ if $\pi_1(G/H)$ is nontrivial, and these strings will carry charges in $\pi_1(G/H)$. Here we have $SU(N) \to \mathbf{Z}_N$, and $\pi_1[SU(N)/\mathbf{Z}_N]$ is $\mathbf{Z}_N$. So the Higgs vacuum has *confining magnetic flux tubes*, carrying charge $\mathbf{Z}_N$, as a result of condensation of the electrically

[4] A caution: this symmetry is actually only exact when $m \to \infty$, and is approximate if $m \gg \Lambda$. However, for reasons explained below, the number of vacua cannot change when m varies, so our counting of vacua is correct for any m. The vacua are actually related by shifting the θ angle by $2\pi k$, $k \in \mathbf{Z}$.

charged fields Φ^s. The scale of these flux tubes and of the mass gap is $\sim m$.

This is extremely suggestive. Let us attempt to rewrite this physics using the magnetic description of the theory. Montonen-Olive duality converts g_0 to $1/g_0 \gg 1$... oops. The physics we were just discussing for small g_0 will now be converted to a very strongly coupled description. In such a highly-fluctuating set of variables, we won't know how to calculate anything. Bad move.

So instead, let's first *continously* vary g_0 from small to large, as we had discussed doing earler. Now Λ and m will gradually become of the same order. The classical analysis we performed of the Higgs vacuum will become invalid, as will our semiclassical analysis of the unbroken vacuum. However, we may now appeal to a special property of supersymmetric field theories. Even an $\mathcal{N}=1$ supersymmetric theory has the property that the energy of any field configuration is positive. All supersymmetric vacua have exactly zero energy, and are global minima of the potential. Furthermore, the potential energy is proportional to the square of a complex function, whose zeroes are controlled by complex analysis. These zeroes cannot simply disappear. Even if we change g_0 (which, when combined with the θ angle of the gauge theory, is actually complex) the number of zeroes cannot suddenly change. (This hand-waving argument is vastly improved by consideration of Witten's index [27], discovered around 1980.) This gives us great confidence that even at large g_0, the H vacuum will still exist, with a mass gap and confining magnetic flux tubes, and so will the C vacuum, with its own mass gap and confining electric flux tubes. This is not quite a proof, but the evidence is very strong. (The mathematics of [3] elevates the argument to a near-proof.)

Now, having moved to a theory with $g \gg 1$ which still has the flux tubes of interest, let's apply a strong-coupling expansion by switching over to the magnetic description of the theory, using $SU(N)/\mathbf{Z}_N$ variables whose gauge coupling is $\tilde{g} = 4\pi/g_0$. What happens in the magnetic description? Not only does Montonen-Olive duality invert the gauge coupling, exchange electric and magnetic charge, and switch $SU(N)$ with $SU(N)/\mathbf{Z}_N$, giving a new description in terms of new adjoint gauge, spinor, and scalar fields $\hat{\Phi}^s$, magnetically charged, *it also exchanges the H vacuum with the C vacuum* [8, 3]!

It's important not to get confused, so let's review. In the electric theory, there is an H vacuum, described at small g_0 by simple breaking of a gauge group by condensation of the Φ^s fields. We don't have a good electric de-

		$g_0 << 1$	$g_0 >> 1$
H vacuum	SU(N) description	Higgs effect; $g(\mu) << 1$ for all μ solitons: Z_N magn. flux tubes	Higgs effect; $g(\mu) >> 1$ for all μ
	Dual $SU(N)/Z_N$ description	Confinement; $\tilde{g}(\mu) >> 1$ for all μ	Confinement; $\tilde{g}(\mu) << 1$ for $\mu >> \Lambda$ expect Z_N dual electric flux tubes
C vacuum	SU(N) description	Confinement; $g(\mu) << 1$ for $\mu >> \Lambda$ expect Z_N electric flux tubes	Confinement; $g(\mu) >> 1$ for all μ
	Dual $SU(N)/Z_N$ description	Higgs effect; $\tilde{g}(\mu) >> 1$ for all μ	Higgs effect; $\tilde{g}(\mu) << 1$ for all μ solitons: Z_N dual magn. flux tubes

Figure 21: The Higgs and Coulomb vacua, in the regions of large and small g_0, as described by the two different sets of variables.

scription of it at large g_0, but we know it still exists. We also know there is a C vacuum, and we don't have a good electric description of it even at small g_0, much less at large g_0. Each of these two vacua may also be described using the magnetic variables of the $\mathcal{N} = 4$ theory. In these variables, we do not have any good descriptions when g_0 is small, since $1/g_0$ is big. However, when g_0 is large, and $1/g_0$ is small, we have a good description of the C vacuum (!) which is exactly *isomorphic* to the small-g_0 electric description of the H vacuum at small g_0. And that's what we want: a magnetic description of the C vacuum, valid at $g_0 \gg 1$, which makes it easy to see the confining electric flux tubes of the C vacuum. In this magnetic description of the C vacuum, the electric flux tubes are simply the semiclassical (remember $\tilde{g}_0 \ll 1$) solitonic strings which emerge from the condensation of the scalars $\hat{\Phi}^s$, which are *magnetically* charged and break the *magnetic* gauge group from $SU(N)/\mathbf{Z}_N$ to nothing. These solitons carry $\mathbf{Z}_N = \pi_1[SU(N)/\mathbf{Z}_N]$ charge — which is exactly what we need! Furthermore, we can easily see how the mass gap is generated in this context, just as it is generated classically at small g_0 in the H vacuum.

So we have found our strong-coupling description of confinement, and it is precisely as we originally suggested: it is a non-Abelian generalization of the dual Meissner effect, in which condensation of magnetically charged scalar fields generates a mass gap and confines electric flux. The picture even gives us flux tubes with the correct charges!

Can we go back to $\mathcal{N} = 1$ SYM? No; that would require varying $m \to \infty, g_0 \to 0$, which would make the magnetic description of the C vacuum strongly coupled and unreliable. But by supersymmetry, the physics should not change too much as we vary g_0. We may therefore consider this a near-proof that $\mathcal{N} = 1$ SYM does indeed have a mass gap and confinement. It is a strong argument that the corresponding flux tubes carry $\mathbf{Z}_N$ charges for the flux tubes. However, it is no proof at all that confinement occurs via a simple picture of condensing, weakly-coupled magnetically-charged objects. In fact, it firmly suggests that the magnetic condensation process is *strongly coupled.* This means, for example, that any calculation of the string tension, or even of ratios of tensions of different flux tubes, will be suspect. Qualitatively things look great; but a quantitative tool this is not.

Should we expect this picture to survive to the non-supersymmetric case? Take the theory with $\mathcal{N} = 4$ supersymmetry broken to $\mathcal{N} = 1$, and futher break $\mathcal{N} = 1$ supersymmetry by adding an $SU(N)$ gluino mass $m_\lambda \ll m$. Duality is in fact enough to tell us how to implement this breaking at leading order in m_λ/m. However we don't need to think very hard. We know that the theory has a mass gap, so small supersymmetry-breaking can only change some properties of the massive fields, *without altering the fact that* $SU(N)/\mathbf{Z}_N$ *is completely broken.* The strings, whose existence depends only on this breaking, thus survive for small m_λ. To reach pure YM, however, requires taking m, m_λ all to infinity together as $g_0 \to 0$. It seems probable, given what we know of YM physics, that the strings undergo no transition as these masses are varied. In particular, we may hope that there is no phase transition for the strings between pure $\mathcal{N} = 1$ SYM and pure YM. Note that this conjecture can, and should, be tested numerically on the lattice.

If in fact the strings of $\mathcal{N} = 1$ SYM and of YM are continously related, without a transition as a function of the gluino mass, then the arguments given above for $\mathcal{N} = 1$ SYM extend to YM, establishing a direct link between Montonen-Olive duality of $\mathcal{N} = 4$ gauge theory and the confining $\mathbf{Z}_N$-strings of pure YM theory.

4.2 A gravitational description of confinement

We have used up most of these lectures, and yet still not reached the latest developments. I will give an overview of some recent work with Polchinski [28] which gives a new and remarkable picture of confinement. A somewhat different picture emerged earlier in this context [29], and other pictures were discovered later [30, 31, 32]. The reason for the existence of all of these different pictures is the same as before: each of them represents a distinct modification of the confining theory of interest into a regime where there is a new small parameter, and each therefore agrees that confinement occurs but disagrees on the precise mechanism.

Let me comment on these disagreements. We should abstract a lesson from all this, namely that confinement is a generic property of gauge theories for which there can be many causes. The various causes we are learning about need not be directly relevant for pure YM, or $\mathcal{N}=1$ SYM, which is too bad, since it means that we are not yet learning any quantitative method for computing in such a theory. But it may be that neither of these theories has enough small parameters to permit simple computation. We are not guaranteed that a given physical phenomenon has a perturbative expansion in some parameter, any more than we are guaranteed a similar property for a generic function. It may be that the only way to understand Yang-Mills theory is either to simulate it or solve it exactly. The latter goal is far beyond any mathematical problem ever solved. Simulation may be the end of the line. [Fortunately, in real-world QCD, there are large global symmetries among the *quarks* which are only weakly broken. Expansions around an exactly-globally-symmetric theory in the small symmetry-breaking parameters has allowed many *relations between quantities* in nonperturbative QCD to be predicted. This was essential in the development of the theory of the strong interactions.]

But even if our new descriptions of confinement are less relevant for YM and $\mathcal{N}=1$ SYM (and we already know they are even less relevant for QCD,) they still provide new phenomena for us to think about, ones which could be relevant in yet other contexts. The goal of these lectures is not merely to explore confinement in YM and SYM. It is to show you the variety of phenomena in gauge theories, and encourage you to consider the possibility that confinement occurs elsewhere in nature, perhaps in unexpected ways and in unexpected places.

In particular, the most strange and wonderful of all of the developments of the 1990s has been the discovery that string theory and field theory are not even distinct mathematical entities. In the Maldacena [33] conjecture, sharpened further by Witten [34] and by Gubser, Klebanov and Polyakov [35], there is strong evidence for a new form of duality. We saw earlier that we may take a generating functional and give it multiple integral representations, each of them with a four-dimensional local Lagrangian in its integrand, giving us a local quantum field theory. But it turns out that we may also rewrite this functional as a well-known string theory in 9+1 dimensions, with five of the dimensions compact. Even though Polyakov [36] has argued for years that we should seek a five-dimensional string to describe gauge theories in four-dimensions, it is astonishing that the needed string is one that we already know. (Of course the string theory has its own dualities, so we mustn't limit ourselves to a single set of variables for it either.)

There are many technical problems with this duality. First, we don't know how to write a path integral for string fields. (The usual two-dimensional world-sheet path integral is analogous to a one-dimensional particle world-line path integral, not to the path integral of a four-dimensional field theory. The first is "first-quantization", the second is "second-quantization".) We therefore have no explicit way to write the equating of the field theory and the string theory. Second, the string theory is particularly nasty. The presence of large curvatures and large Ramond-Ramond fields makes the usual techniques of classical string theory invalid. But fortunately there is a limit in which these issues are unimportant, and it is in that limit that we may hope to study new properties of field theory. This is the limit in which the quantum string theory reduces simply to classical supergravity. (Actually this is too restrictive as has been shown very recently [37, 38].) In the remaining time, we will seek to study the $\mathcal{N} = 1^*$ theory in a regime where it is simply described by semiclassical supergravity coupled to strings and to branes.

Both pure YM and the $\mathcal{N} = 1$ SYM theory have two parameters, the QCD scale Λ and the number of colors N. (The coupling $g(\mu)$ runs with scale and is a determined function of μ and Λ; thus it is not an independent parameter.) However, Λ is simply the only scale in the problem, so it is not a dimensionless quantity that it is meaningful to vary. The only other parameter available is N, and it has long been suggested that as $N \to \infty$ gauge theory might simplify, and might even be soluble. The solution to large N gauge theory has remained elusive, however.

By contrast, the $\mathcal{N} = 4$ theory has *two* dimensionless parameters: N

and the high-energy coupling g_0. As Maldacena has shown you, the two parameters play an essential role in the string theoretic description of the $\mathcal{N} = 4$ theory. The coupling $g_0^2/4\pi$ is the string coupling g_s, which when small makes the string theory classical. However, this is not enough, since even the classical theory in background Ramond-Ramond fields is too complicated. When $g_0^2 N/4\pi \equiv \lambda$, the 't Hooft coupling, is large, then the space on which the classical string theory is defined becomes very large, with very low curvature; then the string theory reduces to its low-energy limit, namely type IIB supergravity.

Here we see that the hope of the previous paragraphs, that the large N limit of gauge theory might simplify, appears to be partially realized. At large N we do indeed find a new description, a classical string theory. But only if we simultaneously take λ large do we obtain a well-understood theory, one in which anything can be calculated. At small λ the theory is very complicated. This is unfortunate, because the YM and SYM theories we might want to study do not have a dimensionless parameter corresponding to λ. The gauge coupling runs from small to large, so we are guaranteed that at high energy $\mu \gg \Lambda$ the running $\lambda(\mu)$ will be small (which is not a problem, because we can use field theory perturbation theory in that regime) and that $\lambda(\mu)$ becomes potentially large only near to the energy scale Λ. Unfortunately, there is no evidence that $\lambda \gg 1$ at $\mu \sim \Lambda$. More likely, it is only of order 2π, which (when you check the factors of 2π) is not sufficiently large for gravity to work. In particular, in $\mathcal{N} = 1$ SYM, the scale of confinement and the mass gap is

$$\Lambda \sim \mu e^{-2\pi/3\lambda(\mu)}$$

(in pure YM, replace 3 with 11/3) so the energy scale μ is of the same order as the confining scale when λ is of order 2π. Thus, even if gravity were to actually describe confinement in YM or $\mathcal{N} = 1$ SYM, it could only do so at energy scales extremely close to Λ, corresponding to a ten-dimensional space whose curvature would be large everywhere except (at best) in a small region.

If this is true, then gravity cannot provide a nice description of confining YM or $\mathcal{N} = 1$ SYM. The confinement occuring in these theories can only be studied using the classical but extremely complicated theory of strings in Ramond-Ramond fields and on a highly curved space. This duality is not much better than the electric-magnetic duality we had before. But we can consider our by-now familiar trick; can we find a way to distort YM or

$\mathcal{N} = 1$ SYM in such a way that we can take λ artificially large?

Yes; just as before, let us consider $\mathcal{N} = 1^*$. The $\mathcal{N} = 1^*$ theory has *three* parameters: N, g_0 and m. The first two are those of $\mathcal{N} = 4$ and are the important ones in the ultraviolet. In the infrared, g_0 and m are combined into Λ, leaving N as the only dimensionless constant. As in our earlier discussion, we may take g_0 small but $g_0^2 N$ large; then the ultraviolet theory will be approximately $\mathcal{N} = 4$ SYM *in the supergravity regime!* We can then consider the effect of $m \neq 0$ in the context of supergravity, and see if we can obtain a picture of how confinement occurs. As always, the corresponding picture will be special to this particular deformation of $\mathcal{N} = 1$ SYM — note that $\Lambda \sim m e^{-8\pi^2/3Ng_0^2}$ and m will be of the same order, so as usual our confining scale will not be well-separated from the physics of the massive adjoint chiral multiplets — but we'll accept this limitation and move forward.

4.3 Confinement in the supergravity regime of $\mathcal{N} = 1^*$

This is a long story, and I can't describe it all here. One needs a nice discussion of branes, fluxes, and all the rest. So let me be schematic, and give you a brief but telling overview of what happens in this theory. Needless to say, a significantly more rigorous discussion appears in our paper [28].

The key idea was provided by Rob Myers, in a slightly different context [39]. What he showed was this. Suppose you take a collection of flat Dp branes, forming $p+1$ dimensional Minkowski space $\mathcal{M}^{p+1}$ embedded in 9+1 dimensional flat space. Now subject them to a certain electric field, not an ordinary $F_{\mu\nu} = \partial_{[\mu} A_{\nu]}$ but rather a derivative of an antisymmetric-tensor potential with $p+3$ indices — in short, an electric field with $p+4$ indices. In this background field, the Dp branes link together and expand into a D$(p+2)$ brane, with a $p+3$-dimensional worldvolume in the form of a two-sphere [40] times $\mathcal{M}^{p+1}$.

Myers called this "dielectric branes", and with good reason. Take an atom; it is electrically neutral, but carries a global charge, its atomic number. Now subject it to an electric field. It will polarize, as in a dielectric. It is still electrically neutral, but it locally has electric charge. Also, it still has its atomic number charge, which is unaffected. Here, our N Dp branes carry a charge, the total number N. After they expand into a D$(p+2)$ brane, what do they have? First, the number of Dp branes hasn't changed; that charge remains. Second, the total D$(p+2)$ brane charge hasn't changed; a brane

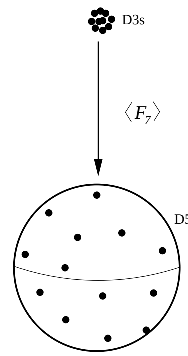


Figure 22: The Myers effect for D3-branes.

in the form of a two-sphere can collapse and disappear, so our D$(p+2)$-brane will vanish if we turn off the electric flux, and there is no net charge associated with it. Still, locally on the two-sphere, there *is* D$(p+2)$ brane charge. Go near to the two-sphere and you can feel it; the other side of the sphere, with cancelling charge, is far away. Thus the Dp branes have expanded into a D$(p+2)$-brane *dipole*! Particles form dipoles by moving apart a certain distance; strings and other branes form dipoles by forming closed surfaces; but the idea is the same.

What's the connection? Take the $\mathcal{N}=4$ theory, described as type IIB string theory on $AdS_5 \times S^5$. Now modify the gauge theory by adding mass terms as in $\mathcal{N}=1^*$. It turns out that the modification of the Lagrangian by the mass operators corresponds, in supergravity, to turning on a background electric field, a tensor with 7 indices. The D3-branes, whose near-horizon geometry formed the $AdS_5 \times S^5$ spacetime, expand, as Myers suggested, into a 5-brane. However, they have two choices (actually many more, but we'll only consider these two for now.) They can expand into a D5-brane. But by S-duality, under which D3-branes are invariant and D5-branes are exchanged with NS5-branes, it must also be possible for the D3-branes to expand into an NS5-brane. Solving the equations, one finds that both of these possibilities are realized. The first corresponds to the Higgs vacuum of $\mathcal{N}=1^*$, the second to the confining vacuum!

What does this do to the supergravity? The full supergravity solution has still not been found. However, we were able to show that there exists a good perturbative expansion in this theory which allows us to demonstrate solutions of the following form: at large AdS radius r, near the boundary, we have $AdS_5 \times S^5$ modified slightly by corrections of order $1/r$ to a power. At a radius of order $m\alpha' N$ these corrections become large. A singularity is avoided, however, by the presence of a D5-brane (or NS5-brane) carrying N units of D3-brane charge. The brane has world-volume S^2 (placed on an equator of the S^5) times $\mathcal{M}^4$ (parallel to the boundary of AdS_5.) Specifically, this prevents the 7-form electric flux from diverging and causing the metric to do the same. Instead, there is a smooth solution (except at the position of the 5-brane, where there is a standard and understood singularity) which rounds off nicely at $r = 0$, without a horizon or singularity at that point. In fact, for $r \ll m\alpha' N$, the spacetime is approximately flat ten-dimensional space.

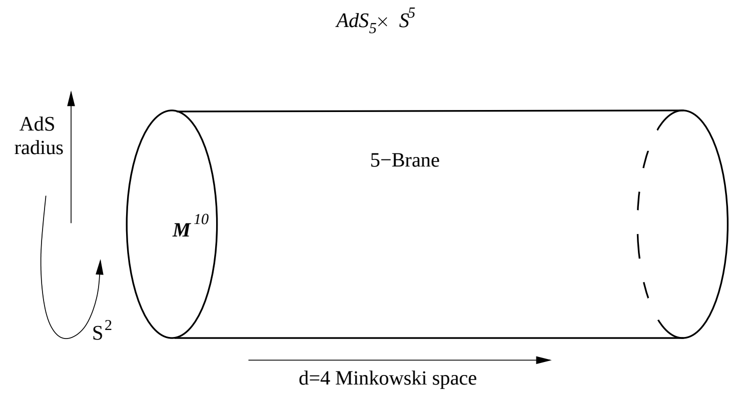


Figure 23: A useful geometrically-reduced representation of a 5-brane of the sort found in the $\mathcal{N} = 1^*$ solution.

What about confinement? Can we see that magnetic flux is confined in the H vacuum and that electric flux is confined in the C vacuum? Indeed we can. D-branes, by definition, are places where strings can end. In particular, F1-strings can end on D3- and D5-branes. But then, by S-duality, D1-branes can end on D3- and NS5-branes. On the other hand, F1-strings cannot end on NS5 branes, nor D1-branes on D5-branes. Another important feature is that D1 branes, and F1-strings, if placed parallel to D3-branes, can dissolve in them. But D1-branes cannot dissolve into D5-branes, nor can F1-strings dissolve into NS5-branes.

All of these facts have physical implications for the $\mathcal{N} = 4$ and $\mathcal{N} = 1^*$ field theories. F1-strings ending on D3-branes look like electrically charged particles; D1-strings look like magnetic monopoles. We can create a pair of oppositely-oriented F1-strings, for example, and move them apart without large energy cost; thus the electric charges are unconfined, as expected in $\mathcal{N} = 4$ SYM. An F1-string placed parallel to and inside a stack of D3-branes corresponds to putting a line of electric flux into the $\mathcal{N} = 4$ theory. The dissolving of this line indicates that electric flux prefers to minimize its energy by expanding to infinity. Thus electric flux is, as expected, unconfined. The same holds for magnetic flux, a dissolving D1-brane.

However, in the $\mathcal{N} = 1^*$ theory the vacua of the theory correspond to 5-branes with D3-brane charge. Now, in the H vacuum, we have a spherical D5-brane, on which D1-branes cannot end! Magnetic charges can no longer appear with finite energy. And suppose we put a D1-brane parallel to and near a D5-brane which also carries D3-brane charge. Here a remarkable thing happens; the D1-brane can only *partially* dissolve. The D3-branes try to make the D1-brane expand, but the D5-brane charge prevents its complete dissolution. We are left with a diffuse, but nonetheless finite-thickness, D1-brane–D5/D3-brane bound state. The magnetic flux corresponding to the D1-brane expands, but only to a tube of fixed size; it is *confined* in this tube. Furthermore, if we attempt to produce a pair of magnetic monopoles in the form of D1-branes ending on this D5/D3-brane composite, we will find instead that they are connected by this diffuse flux tube. The charges, kinematics and dynamics of D-branes tell us that magnetic charge is confined in the Higgs vacuum of $\mathcal{N} = 1^*$!

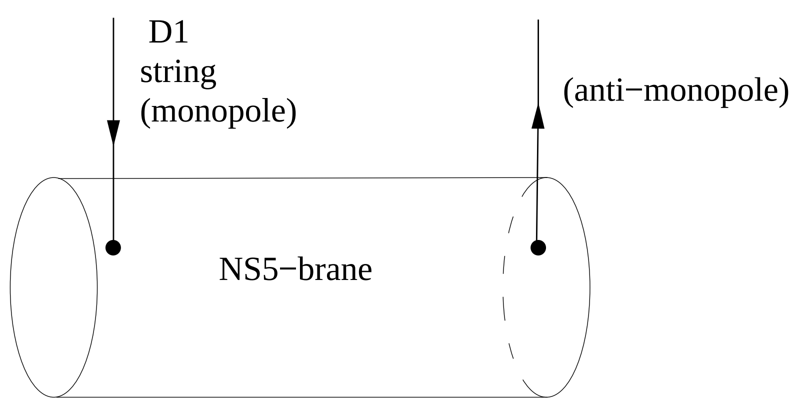


Figure 24: Monopoles (D1 strings) can end on NS5-branes; they are not confined.

The S-dual story holds in the confining vacuum. We can repeat the entire previous paragraph, exchanging D1 with F1, D5 with NS5, and magnetic with electric. The conclusion is also exchanged: the charges, kinematics and dynamics of NS5-branes, D3-branes and fundamental strings tell us that electric charge is confined in the appropriate vacuum of $\mathcal{N} = 1^*$. We have found a new picture for confinement. It occurs through the appearance of an NS5-brane dipole in the 9+1-dimensional spacetime. The dipole prevents flux tubes, in the form of fundamental strings, from dissolving into the D3-branes contained within the dipole, and instead makes them into flux tubes which are fundamental strings bound to the NS5-brane![5]

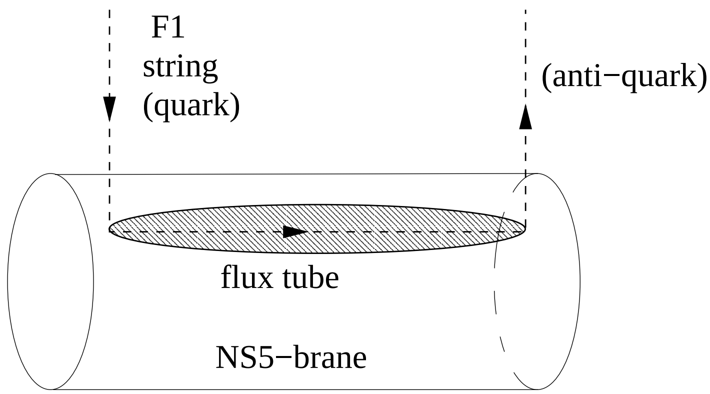


Figure 25: Heavy quarks (fundamental strings) cannot end on NS5-branes; however, there is an NS5-F1 bound state that serves as a flux tube connecting the quark and antiquark.

Of course this is not the end of the story. One should (and can) check that there is a mass gap, that strings carry $\mathbf{Z}_N$ charges, that various expectation values come out correctly, etc. For a few quantities, there are exact results from field theory that are complicated functions of λ and N; comparison with our gravity solution shows precise agreement, even for the numerical coefficients. There is also an exciting new form of duality, which is beyond the scope of these lectures, which takes not $g_0 \to 1/g_0$ but $\lambda \to 1/\lambda$! This is still largely unexplored territory, although it has been discussed further in [41].

It is important to remember that we have not been constructing an analogy. We have not found a new "model" for confinement in field theory. This *is* confinement in field theory. The string theory is just a convenient

[5]In principle it is also possible to break supersymmetry. If the supersymmetry breaking is small the story does not change much. For large supersymmetry breaking, of order m, the technical challenges become greater. It is not known whether reliable computations can be done in that regime, although there are no known obstructions.

description of it; but we are not dealing with a different theory, just an alternative description of the *same* theory. This mechanism for confinement is a new behavior of ordinary, four-dimensional continuum field theory which was not previously known. It is one of several which have been uncovered in the regime of large 't Hooft coupling.

However, as always, this is not confinement in pure $\mathcal{N} = 1$ SYM. To reach that theory, we would have to take the 't Hooft coupling λ small. In that limit, the NS5-brane dipole would shrink in size, its radius becoming of order the string scale. All calculational control would be lost. That's the price we paid for our new picture. Like Moses, we can see the promised land but never quite manage to reach it.

5 Wrap-up

In these lectures I have given you an overview of some of the key ideas underlying confinement as a property of field theory, and now, of string theory as well. This is a tiny fraction of what field theory (and now string theory) is capable of, and we are still uncovering new features on a monthly basis. In fact, most field theories do not have confinement, for reasons entirely different from those of QCD. Many become nontrivial conformal field theories at low energy. Others become composite, weakly-coupled gauge theories (the so-called "free-magnetic phase" [17].) Dualities of many stripes are found everywhere. Ordinary dimensional analysis in string theory is totally wrong in the regime where it looks like weakly-coupled field theory, and ordinary dimensional analysis in field theory is totally wrong in the regime where it looks like weakly-coupled supergravity. There's much more. You are encouraged to stride into the midst of these developments, to search with us for new features of both field theory and string theory (or, better said, of the single theory of which both are a part,) and most importantly, and most difficult, to explain to us what all these dualities really mean, and where they come from. Good hunting.

References

[1] K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).

[2] G. 't Hooft, Nucl. Phys. B **190**, 455 (1981), Nucl. Phys. B **153**, 141 (1979), Nucl. Phys. B **138**, 1 (1978).

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

[4] M. J. Strassler, Nucl. Phys. Proc. Suppl. **73**, 120 (1999) [hep-lat/9810059].

[5] M. Wingate and S. Ohta, Phys. Rev. D **63**, 094502 (2001) [hep-lat/0006016].

[6] B. Lucini and M. Teper, hep-lat/0103027.

[7] B. Lucini and M. Teper, Phys. Lett. B **501**, 128 (2001) [hep-lat/0012025].

[8] C. Vafa and E. Witten, Nucl. Phys. B **431**, 3 (1994) [hep-th/9408074].

[9] N. Seiberg and E. Witten, Nucl. Phys. B **426**, 19 (1994) [Erratum-ibid. B **430**, 485 (1994)] [hep-th/9407087].

[10] D. Finnell and P. Pouliot, Nucl. Phys. B **453**, 225 (1995) [hep-th/9503115].

[11] A. A. Abrikosov, Sov. Phys. JETP **5**, 1174 (1957) [Zh. Eksp. Teor. Fiz. **32**, 1442 (1957)].

[12] H. B. Nielsen and P. Olesen, Nucl. Phys. B **61**, 45 (1973).

[13] S. Coleman *Aspects of Symmetry, Selected Erice lectures* (Cambridge University Press, Cambridge, 1985)

[14] A. Kapustin and M. J. Strassler, JHEP **9904**, 021 (1999) [hep-th/9902033].

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

[16] M. B. Halpern, Phys. Rev. D **19**, 517 (1979).

[17] N. Seiberg, Nucl. Phys. B **435**, 129 (1995) [hep-th/9411149].

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

[19] C. Montonen and D. Olive, Phys. Lett. B **72**, 117 (1977).

[20] P. Goddard, J. Nuyts and D. Olive, Nucl. Phys. B **125**, 1 (1977).

[21] H. Osborn, Phys. Lett. B **83**, 321 (1979).

[22] J. Polchinski *String Theory, Volume II* (Cambridge University Press, Cambridge, 1998)

[23] J. H. Schwarz, Nucl. Phys. Proc. Suppl. **55B**, 1 (1997) [hep-th/9607201].

[24] K. Intriligator and N. Seiberg, Nucl. Phys. Proc. Suppl. **45BC**, 1 (1996) [hep-th/9509066].

[25] M. J. Strassler, Prog. Theor. Phys. Suppl. **131**, 439 (1998) [hep-lat/9803009].

[26] F. Niedermayer, Nucl. Phys. Proc. Suppl. **73**, 105 (1999) [hep-lat/9810026].

[27] E. Witten, Nucl. Phys. B **188**, 513 (1981).

[28] J. Polchinski and M. J. Strassler, hep-th/0003136.

[29] E. Witten, Adv. Theor. Math. Phys. **2**, 505 (1998) [hep-th/9803131].

[30] I. R. Klebanov and M. J. Strassler, JHEP **0008**, 052 (2000) [hep-th/0007191].

[31] J. M. Maldacena and C. Nunez, Phys. Rev. Lett. **86**, 588 (2001) [hep-th/0008001].

[32] C. Vafa, hep-th/0008142.

[33] J. Maldacena, Adv. Theor. Math. Phys. **2**, 231 (1998) [Int. J. Theor. Phys. **38**, 1113 (1998)] [hep-th/9711200].

[34] E. Witten, Adv. Theor. Math. Phys. **2**, 253 (1998) [hep-th/9802150].

[35] S. S. Gubser, I. R. Klebanov and A. M. Polyakov, Phys. Lett. B **428**, 105 (1998) [hep-th/9802109].

[36] A. M. Polyakov, hep-th/9304146.

[37] J. Polchinski and M. J. Strassler, Phys. Rev. Lett. **88**, 031601 (2002) [arXiv:hep-th/0109174].

[38] D. Berenstein, J. Maldacena and H. Nastase, JHEP **0204**, 013 (2002) [arXiv:hep-th/0202021].

[39] R. C. Myers, JHEP **9912**, 022 (1999) [hep-th/9910053].

[40] D. Kabat and W. I. Taylor, Adv. Theor. Math. Phys. **2**, 181 (1998) [hep-th/9711078].

[41] O. Aharony, N. Dorey and S. P. Kumar, JHEP **0006**, 026 (2000) [hep-th/0006008].

Introduction to Little String Theory

David Kutasov*

Department of Physics, University of Chicago, Chicago, USA

Lectures given at the
Spring School on Superstrings and Related Matters
Trieste, 2 – 10 April 2001

LNS027004

*kutasov@theory.uchicago.edu

Abstract

These notes, based on lectures presented at the ICTP Spring School on Superstrings and Related Matters in April 2001, provide an introduction to Little String Theory.

Contents

1 Introduction

Much has been learned over the years by studying string dynamics near various kinds of "impurities." Examples include string propagation on orbifolds [1], where one finds "twisted sectors" corresponding to fundamental strings trapped at the orbifold singularities, and vacua with D-branes which contain localized excitations corresponding to open strings ending on the branes.

In both of these examples, the states localized at the impurity couple to the bulk – *e.g.* two open strings ending on a D-brane can fuse into a closed string that can leave the brane. It is sometimes possible to decouple the physics of the localized modes from bulk dynamics by taking a low energy limit, $E << m_s$, where $m_s = 1/\sqrt{\alpha'}$ is the string scale, associated with the tension of the fundamental string $T = 1/2\pi\alpha'$.

Whenever this limit gives rise to an interacting theory, it corresponds to a local quantum field theory (QFT), such as the non-abelian gauge theories found on branes. This embedding of field theoretic dynamics into string theory led in recent years to many insights into field theory and string theory (see *e.g.* [2, 3] for reviews).

The purpose of these lectures is to describe another class of impurities – Neveu-Schwarz fivebranes [4], or equivalently singularities of Calabi-Yau manifolds and other spaces[1]. One of the striking features of the dynamics of $NS5$-branes is that it can be decoupled from the bulk without taking the low energy limit $\alpha' \to 0$. The decoupled theory of $NS5$-branes is known as Little String Theory[2] (LST). It has the following properties:

(1) The theory is non-local. In particular, upon compactification on tori, LST exhibits T-duality.

(2) It has a Hagedorn density of states at high energies, $\rho(E) \sim E^{\alpha}\exp(\beta_H E)$.

(3) The theory can be defined in six or fewer spacetime dimensions. It has super – Poincare invariant vacua with sixteen or fewer supercharges.

(4) LST is a non-gravitational theory: there is no massless spin two particle in the spectrum.

[1] Orbifolds are examples of such singularities, but in [1] they are in fact resolved by a finite expectation value of a modulus – the B field [5]. We will be interested below in situations where this v.e.v. is zero or at least very small.

[2] A name due to [6].

(5) The theory appears to have well defined off-shell Green functions, unlike (closed) critical string theory, where it is believed that only on-shell observables can be studied.

Note that while properties (1) and (2) are reminiscent of critical string theory, properties (3), (4) and (5) are different in the two cases.

The main purpose of these lectures is to describe in more detail some of the above properties and the techniques that were used to study them. Most of these results were obtained by using holography, and this is the approach that will be followed here. In particular, We will not describe an alternative approach to LST based on a discrete light-cone quantization (DLCQ) of the theory, which utilizes a certain $1+1$ dimensional sigma model [7, 8, 9]. For a review of that approach and LST in general as of mid-1999, see [10].
There are several reasons why I think LST is of some interest. Among them:

(1) In most (compactified) supersymmetric string theories one finds moduli spaces of vacua. For generic values of the moduli the perturbative description is non-singular, but one can often tune the moduli so that a singularity appears somewhere on the compact manifold. The dynamics near the singularity is described by LST. Thus LST is part of the dynamics of rather conventional looking string vacua at special points in the moduli space. Furthermore, when supersymmetry is broken, it is possible that the theory is dynamically driven to such singular points in moduli space.

(2) LST is relevant for the study of strongly coupled gauge theories, which can be realized on $NS5$-branes wrapped around Riemann surfaces or D-branes stretched between fivebranes (see [2] for a review). There are also applications to matrix theory [11], which in fact provided some of the original motivation for the construction of this theory [12, 13].

(3) It was proposed that LST might be phenomenologically relevant for brane world scenarios with a relatively low string scale [14].

More generally, LST appears to be a structure that is intermediate in complexity between local QFT and critical string theory. It has the non-locality and Hagedorn spectrum characteristic of critical string theory, but not the complications associated with gravity. A better understanding of its structure might shed light on string theory, strongly coupled gauge theory (QCD strings), holography and other matters.

The plan of these lectures is as follows. We start in section 2 by describing the limit in which the dynamics of $NS5$-branes decouples from bulk physics. In section 3 we discuss the holographic description of this limit and some of the properties of LST mentioned above. In particular, we exhibit some observables and physical states in the theory.

In section 4 we discuss the high energy thermodynamics of LST. We show that the spectrum has a Hagedorn growth and compute the Hagedorn temperature and the first subleading term in the entropy which shows that the thermodynamics is unstable. In section 5 we introduce and study a class of vacua of LST which can be analyzed in a controlled weak coupling expansion.

Section 6 contains some comments on aspects of LST that we cannot treat in detail due to lack of time, including singularities of Calabi-Yau manifolds which give rise to $d < 6$ dimensional vacua of LST and models with reduced supersymmetry, D-branes in the vicinity of $NS5$-branes, and instabilities in LST. In section 7 we discuss some open problems.

2 The decoupling limit of flat $NS5$-branes

Consider a vacuum of type II string theory which contains N parallel $NS5$-branes[3], which are extended in the directions $(x^1, \cdots, x^5)$ and are localized in $(x^6, \cdots, x^9)$. We will initially take the fivebranes to be at the same point and will later examine the deformations that separate them in the directions $(6, 7, 8, 9)$.

The presence of the fivebranes breaks the Lorenz symmetry:

$$SO(9,1) \to SO(5,1) \times SO(4). \tag{2.1}$$

From the fivebrane worldvolume point of view, $SO(5,1)$ is the Lorenz symmetry, while $SO(4)$ is an internal R-symmetry. The fivebranes also break half of the supersymmetry, reducing the number of unbroken supercharges from thirty two to sixteen. In terms of six dimensional supersymmetry along the fivebranes, IIA fivebranes preserve a chiral $(2,0)$ supersymmetry[4], while IIB fivebranes preserve $(1,1)$ supersymmetry.

Since $NS5$-branes are dynamical objects, like D-branes, one expects to find a rich spectrum of excitations on the branes. To decouple the dynamics

[3] Neveu-Schwarz fivebranes are magnetically charged under the Neveu-Schwarz $B_{\mu\nu}$ field. See *e.g.* [15] for a review of some of their properties.

[4] I.e. two complex supercharges in the **4** of Spin(5,1).

on the fivebranes from the bulk, consider the limit

$$g_s \to 0; \quad \frac{E}{m_s} = \text{fixed}. \tag{2.2}$$

Processes in which modes that live on the fivebranes are emitted into the bulk as closed strings are suppressed in this limit, since the corresponding amplitudes are proportional to g_s and thus go to zero. At the same time, the dynamics on the $NS5$-branes does not become free in this limit. One way to see this is to consider the low energy limit of the resulting theory and to show that it is not free.

Consider first the low energy limit of N $NS5$-branes in type IIB string theory. S-duality relates this to N $D5$-branes; thus the low energy theory is a six dimensional gauge theory with $(1,1)$ supersymmetry and gauge group $U(N)$. The gauge coupling of the theory on the $D5$-branes is

$$\frac{1}{g_D^2} = \frac{m_s^2}{g_s}. \tag{2.3}$$

Using the transformation of g_s and m_s under S-duality one finds that the gauge coupling on the $NS5$-branes is

$$\frac{1}{g_N^2} = m_s^2. \tag{2.4}$$

Thus in the limit (2.2) the gauge coupling remains fixed. Since the gauge theory in question is non-renormalizable, the gauge coupling g_N in fact changes with the scale, approaching zero at long distances and growing at short distances. At energies of order m_s the gauge theory description breaks down and more data needs to be supplied to define the theory. As we will see, there are in fact additional degrees of freedom in the theory at (roughly) that scale, and the full density of states is much larger than that in any local QFT. At any rate, since the dynamics at scales $E \simeq m_s$ is not free, the full theory must be interacting.

Note that the above arguments are only valid for $N > 1$ fivebranes. The low energy theory on a single $NS5$-brane *is* free[5]. Indeed, we will see later that LST is interacting only for $N > 1$.

[5]In the IIA case it contains a self-dual $B_{\mu\nu}$ field, five massless scalars and fermions related to them by $(2,0)$ supersymmetry. In the IIB theory one finds a gauge field, four scalars and fermions, related by $(1,1)$ supersymmetry.

The infrared dynamics of N IIA $NS5$-branes is more involved. One finds in this case a non-trivial IR fixed point with $(2,0)$ superconformal symmetry [16]. To see that something special is happening in the IR imagine separating the fivebranes in the $(6,7,8,9)$ directions. In the IIB theory, one then finds massive states corresponding to D-strings stretched between the fivebranes; their masses go to zero as the fivebranes approach each other. The resulting massless states are the off-diagonal $U(N)$ gauge bosons on the fivebranes.

The analogous process for IIA involves $D2$-branes stretched between the fivebranes. The ends of the $D2$-branes are strings bound to the fivebranes. Their tension goes to zero when the fivebranes coincide [17]. These tensionless strings signal the interacting nature of the low energy limit of the IIA fivebrane theory – the $(2,0)$ superconformal field theory.

Thus, we conclude that the limit (2.2) corresponds to an interacting theory on the $NS5$-branes decoupled from the bulk. What sort of theory is it? Already at the level of the present discussion there are a few hints of non-local/stringy behavior. Let us mention two:

(1) T-duality: Compactify some or all of the dimensions $(1,2,3,4,5)$ on circles. $NS5$-branes are known to transform to themselves under T-duality along their worldvolume. Since the limit (2.2) commutes with T-duality, inversion of the radius of a single circle ($R \to 1/m_s^2 R$) exchanges the IIA and IIB LST's, while inversion of an even number of radii is a symmetry of the theory.

(2) The theory contains strings with tension $T = 1/2\pi\alpha'$, which can be interpreted as fundamental strings bound to the fivebranes. In the IIB case[6], these strings can be constructed in the low energy gauge theory as instanton solutions, which are extended (say) in $(0,1)$ and localized in $(2,3,4,5)$. The tension of these strings is proportional to the instanton action, $1/g_N^2$, which using (2.4) is indeed tension of a fundamental string. Of course, this construction gives rise to long strings, and it is not clear what are the properties of short strings which actually govern the dynamics, but it suggests that LST is a theory of strings. Later we will see further evidence that supports this.

It is instructive to compare the decoupling limit (2.2) with the limits studied in D-brane physics. Usually, to decouple the physics of D-branes from the

[6] A similar construction can be performed in the IIA case.

bulk one considers the low energy limit

$$\frac{E}{m_s} \to 0; \quad g_s = \text{fixed}, \tag{2.5}$$

and the decoupling from the bulk is the standard low energy decoupling of QFT from gravity. In contrast, the limit (2.2) for D-branes gives rise in general to a free theory on the branes, since g_s determines both the open and the closed string couplings.

A limit for N D-branes which is more analogous to (2.2) is

$$N \to \infty; \quad g_s \to 0; \quad \lambda = g_s N = \text{fixed}; \quad \frac{E}{m_s} = \text{fixed}. \tag{2.6}$$

The open string coupling λ is fixed; hence the theory on the D-branes remains interacting. Since $g_s \to 0$, the closed string sector decouples, despite the fact that a low energy limit has not been taken. The resulting theory is an open string theory without closed strings; it has some things in common with LST although there are differences as well.

3 A holographically dual description of LST

The construction described in the previous section is useful for establishing the existence of LST, but it does not provide efficient techniques for studying the theory. To proceed, we will use a holographically dual description proposed in [18] (see also [19, 20]). This duality is a generalization of the AdS/CFT correspondence [3]; it postulates that LST is equivalent to ten dimensional string theory in the background of the fivebranes, in the limit (2.2). In this section we will describe the fivebrane geometry and will briefly discuss the duality of [18].

The metric, dilaton and NS B-field around N $NS5$-branes in type II string theory are [4]:

$$\begin{aligned} ds^2 &= dx_\mu dx^\mu + (1 + \frac{N\alpha'}{r^2}) dx^i dx^i, \\ e^{2\Phi} &= g_s^2 (1 + \frac{N\alpha'}{r^2}), \\ H_{ijk} &= -\epsilon_{ijkl}\partial^l \Phi, \end{aligned} \tag{3.1}$$

where $\mu = 0, 1, 2, \cdots, 5$ are worldvolume coordinates and $i, j, k, l = 6, 7, 8, 9$ are transverse ones. We parameterize the space transverse to the branes by spherical coordinates,

$$dx^i dx^i = dr^2 + r^2 d\Omega_3^2. \tag{3.2}$$

To take the limit (2.2) one must send $r \to 0$ at the same rate as g_s. Defining $r = g_s \exp\sigma$ we have in this limit

$$\begin{aligned} ds^2 &= dx_\mu dx^\mu + N\alpha'(d\sigma^2 + d\Omega_3^2), \\ \Phi &= -\sigma, \end{aligned} \tag{3.3}$$

and we suppress the B-field (3.1). String propagation in this geometry corresponds to an "exact conformal field theory" [4]:

$$\mathbb{R}^{5,1} \times \mathbb{R}_\phi \times SU(2)_N. \tag{3.4}$$

$\mathbb{R}^{5,1}$ is the worldvolume of the fivebranes. $\mathbb{R}_\phi$ is the real line labeled by $\phi = \sqrt{N\alpha'}\sigma$. The dilaton goes like (3.3):

$$\Phi = -\frac{Q}{2}\phi; \quad Q = \frac{2}{\sqrt{N\alpha'}} \ . \tag{3.5}$$

The last factor in (3.4) describes the angular three-sphere in (3.3). The B-field (3.1) is precisely such that the CFT on the three-sphere, whose radius is

$$R_{\text{sphere}} = \sqrt{N\alpha'} \ , \tag{3.6}$$

is described by a level N WZW model. We see that the number of fivebranes N determines the slope of the linear dilaton, Q, and the level of $SU(2)$ current algebra. More precisely, since (3.4) is a background for the superstring, the worldsheet theory contains, in addition to the bosonic coordinates, ten free fermions: ψ^μ, $\mu = 0, 1, 2, \cdots, 5$, the superpartners of x^μ; ψ^i, $i = 3, +, -$, the superpartners of the $SU(2)$ currents J^i; and ψ^ϕ, the superpartner of ϕ. The total level N of the $SU(2)$ current algebra receives a contribution of $N-2$ from the worldsheet bosons, and 2 from the fermions ψ^i, which transform in the adjoint of the total $SU(2)$ current algebra. The total central charge of the worldsheet theory (3.4) is

$$\left(6 + \frac{1}{2} \times 6\right) + \left(1 + \frac{6}{N} + \frac{1}{2}\right) + \left(\frac{3(N-2)}{N} + 3 \times \frac{1}{2}\right) = 15, \tag{3.7}$$

which is the correct value for the superstring.

The background (3.4) is thus expected to be holographically dual to the LST on the fivebranes. We next discuss some features of this duality. First note that while the string coupling (3.5) vanishes far from the fivebranes (*i.e.* as $\phi \to \infty$), it diverges as one approaches the branes ($\phi \to -\infty$, or $r \to 0$ in (3.1)). The $NS5$-branes have the remarkable property that quantum effects near the branes cannot be turned off no matter how small the string coupling is far from the branes [4]. This makes it clear that LST is not a free theory[7], as argued above, but it raises the question whether one can analyze the physics of the string background (3.4), (3.5) perturbatively. We will return to this question below.

As is familiar from the AdS/CFT correspondence, on-shell observables in the "bulk" theory – string theory on (3.4) – correspond to off-shell observables in the "boundary" theory – the LST corresponding to N $NS5$-branes. More precisely, off-shell observables in LST correspond to *non-normalizable* observables in string theory on (3.4), whose wavefunctions are supported near the "boundary" at $\phi \to \infty$. This can be understood as follows (in analogy with the AdS case).

Consider (say) a scalar field Ψ on the manifold (3.4), corresponding to one of the modes of the string. As $\phi \to \infty$, the field behaves as (assuming for simplicity a profile constant on the angular S^3):

$$\Psi(\phi, x^\mu) \sim \sum_k C_k e^{\lambda_k \phi} e^{ik_\mu x^\mu} \tag{3.8}$$

where

$$\lambda_k^2 = k_\mu k^\mu + C. \tag{3.9}$$

C is a constant which depends on the mass of the scalar field. Choosing the positive root of (3.9), we see that the mode (3.8) is non-normalizable and thus the coefficients C_k do not fluctuate – they are not integrated over in the process of integrating over all field configurations in the path integral [21]. Thus, we can think of the C_k as fixed sources. The string partition sum with the fixed boundary conditions (3.8) as $\phi \to \infty$, $Z_{\text{bulk}}(C_k)$, can be

[7]For $N \geq 2$ fivebranes. Note that for $N = 1$, the bosonic $SU(2)$ current algebra has formally a negative level, $N - 2 = -1$, and the construction breaks down. This is usually taken to mean that a single fivebrane does not have a throat region (3.4) associated with it, and the dynamics on it becomes trivial in the limit (2.5).

interpreted as the generating functional of off-shell Green functions in the six dimensional LST via:

$$Z_{\mathrm{bulk}}(C_k) = \langle \exp\left(-\sum_k C_k \Theta(k)\right)\rangle_{LST} \ , \tag{3.10}$$

where $\Theta(k_\mu)$ is the off-shell observable which couples to the source C_k. Qualitatively, (3.10) is natural because modes that are non-normalizable in the "near-horizon" geometry (3.4) are nothing but bulk modes in the full geometry (3.1); they are supported at finite r. Thus, they are not part of the LST but rather are fixed background sources (in the limit (2.2)), which couple to the brane modes via couplings like (3.10).

Similarly, *normalizable modes* in the geometry (3.4) correspond to *states* in LST, since in the full geometry (3.1) they correspond to modes localized on the fivebranes (*i.e.* at $r \to 0$). To illustrate all this, we next give an example each of off-shell observables and states in LST, as described in the holographically dual picture.

3.1 Example 1: Chiral operators in LST

As discussed above, the low energy limit of IIB LST is a $U(N)$ gauge theory with $(1,1)$ supersymmetry. This theory contains four scalar fields in the adjoint of $SU(N)$, X^i, $i = 6,7,8,9$, which parameterize the locations of the N fivebranes in $(6,7,8,9)$. The gauge invariant off-shell operators

$$\mathrm{Tr} X^{i_1} X^{i_2} \cdots X^{i_n}; \quad n = 2,3,4,\cdots N, \tag{3.11}$$

where we only take the completely symmetric and traceless combination in $(i_1, \cdots, i_n)$, are lowest components of short multiplets of supersymmetry. Writing the $SO(4)$ symmetry in (2.1) as

$$SO(4) \simeq SU(2)_L \times SU(2)_R \ , \tag{3.12}$$

the operators (3.11) transform in the spin $(\frac{n}{2}, \frac{n}{2})$ representations. In string theory on (3.4) these chiral operators are described as follows. The $SU(2)_L \times SU(2)_R$ symmetry on (3.12) corresponds to the left and right moving $SU(2)$ symmetries in the $SU(2)_N$ WZW model in (3.4). Physical primaries of this symmetry are $V_{j;m,\bar{m}}$ with the same spin $(2j = 0,1,2,\cdots, N-2)$ under both $SU(2)$'s. $(m, \bar{m})$ are the eigenvalues of $(J_3, \bar{J}_3)$.

The lowest lying observables have the form (in the -1 picture)

$$\xi_{\alpha\beta}\psi^\alpha\bar{\psi}^\beta e^{\beta\phi}e^{ik_\mu x^\mu}V_j, \tag{3.13}$$

where $\alpha, \beta = 0, 1, 2, \cdots 9$ and $\xi_{\alpha\beta}$ is a polarization tensor satisfying the usual physical state conditions. One can show that (3.11) correspond to[8]

$$\mathrm{Tr}X^{i_1}X^{i_2}\cdots X^{i_n} \leftrightarrow (\psi\bar{\psi}V_j)_{j+1}e^{\frac{2j}{\sqrt{N\alpha'}}\phi}, \;\; j+1=\frac{n}{2} \tag{3.14}$$

On the right-hand side of (3.14), ψ stands for the three fermions associated with the $SU(2)$ WZW and the brackets mean that ψ, which has spin 1 under $SU(2)_L$, is coupled with V_j into a spin $j+1$ combination (and similarly for the right movers). Thus, the non-normalizable operators (3.14) transform under $SU(2)_L \times SU(2)_R$ as

$$(j+1, j+1); \;\; 2j = 0, 1, 2, \cdots, N-2, \tag{3.15}$$

in exact agreement with what was found for (3.11) above. Applying the spacetime supercharges gives the other members of the supermultiplets. Thus, the sets of short representations of supersymmetry in LST and in string theory on (3.4) agree.

3.2 Example 2: Normalizable states

A large set of normalizable states is obtained by considering vertex operators of the form

$$V(\phi) \sim e^{(-\frac{Q}{2}+i\lambda)\phi} \tag{3.16}$$

on $\mathbb{R}_\phi$. Recall that the vertex operators are related to the wavefunctions (3.8) by a factor of g_s, which here is a function of ϕ (3.5). Therefore, (3.16) actually corresponds to a wavefunction

$$\Psi(\phi) \sim e^{i\lambda\phi}, \tag{3.17}$$

which is (δ-function) normalizable, and thus gives rise to states in LST. Since λ is arbitrary, there is in fact a continuum of such states. To compute their masses, consider the states (3.13) as an example. The mass shell condition reads:

$$k_\mu k^\mu - \beta(\beta+Q) = 0. \tag{3.18}$$

[8]We set k_μ to zero for simplicity.

Plugging in $\beta = -\frac{Q}{2} + i\lambda$, we find

$$M^2 = \frac{1}{N\alpha'} + \lambda^2. \tag{3.19}$$

Thus, we find a continuum above the gap $m_s/\sqrt{N}$. The gap is given by a natural scale in LST; looking back at (2.4), we see that it is the 't Hooft coupling of the low energy super Yang Mills theory (for IIB fivebranes).

3.3 The strong coupling problem

As we have seen before, the background (3.4) has the property that the string coupling depends on ϕ; it goes to zero as $\phi \to \infty$ and diverges as $\phi \to -\infty$. In this subsection we would like to discuss the physical origin of this behavior and its implications. The strong coupling region $\phi \to -\infty$ corresponds to the vicinity of the brane ($r \to 0$). This is the low energy region in the theory on the branes [19].

The low energy behavior of LST is different for IIA and IIB fivebranes. In the IIB case, the low energy limit is a six dimensional $U(N)$ gauge theory, which is weakly coupled in the IR. Thus, in the limit $\phi \to -\infty$ of the near-horizon geometry, which should be dual to the infrared limit on the brane [3], string theory on (3.4) should reproduce the weakly coupled gauge theory on the branes. Since one does not expect to find two different weakly coupled description of the same physics, the "bulk" description should either be strongly coupled, or exhibit large curvatures (or both). Since in our case the curvature of (3.4) is small, it is natural to find that the string coupling is growing in the infrared region.

In the IIA case the infrared limit of LST is somewhat different. As discussed earlier, one finds in this case a non-trivial superconformal field theory with chiral $(2,0)$ supersymmetry, the $(2,0)$ theory. Thus, it is not obvious that one should run into any strong coupling problems in the dual description.

To see what is going on, recall that type IIA string theory can be thought of as an eleven dimensional theory, M-theory, compactified on a circle of radius R_{11}, which is related to the eleven dimensional Planck scale l_{11}, and the string scale m_s and coupling g_s via

$$m_s R_{11} = \ell_{11}^3 m_s^3 = g_s. \tag{3.20}$$

The eleven dimensional theory contains membranes and fivebranes (the $M2$ and $M5$-branes), which preserve half of the supersymmetry; their tensions are (up to numerical constants) $1/l_{11}^3$ and $1/l_{11}^6$, respectively. The IIA $NS5$-branes are M_5-branes located at points on the circle. Thus, to study them using holography we should construct the background around N coincident $M5$-branes. Taking the limit (2.2), which corresponds to $R_{11}, l_{11} \to 0$ with m_s fixed, one finds the eleven dimensional metric

$$ds^2 = H^{-\frac{1}{3}} \left[dx_\mu dx^\mu + H(dx_{11}^2 + dr^2 + r^2 d\Omega_3^2) \right], \tag{3.21}$$

where

$$H = \sum_{n=-\infty}^{\infty} \frac{N l_{11}^3}{\left[r^2 + (x_{11} - 2\pi n R_{11})^2 \right]^{\frac{3}{2}}} . \tag{3.22}$$

x_{11} is a coordinate on the circle; it is periodic with period $2\pi R_{11}$. In the limit $r \to \infty$, the background (3.21) goes over to (3.4). The radius of the x_{11} circle goes to zero and one finds the linear dilaton behavior discussed above. As $r, x_{11} \to 0$ only one term in the sum over n in (3.22) (say $n = 0$) contributes, and the metric reduces to the near-horizon background of N coincident $M5$-branes in eleven dimensions. This background, $AdS_7 \times S^4$, is known to be dual to the $(2,0)$ superconformal field theory via AdS/CFT [3]. If N is large, it can be studied using eleven dimensional supergravity; otherwise one needs the full M-theory, which is not understood for these backgrounds.

Thus, we see that the growth of the coupling and associated breakdown of string perturbation theory as $\phi \to -\infty$ in the background (3.4) have slightly different origins in the IIA and IIB cases. However, regardless of the origin of this problem, one can ask what is the dual description of LST good for in view of its existence? We have already seen two examples of applications of the formalism. Since off-shell observables correspond to non-normalizable wavefunctions supported in the region $\phi \to \infty$, we can classify the observables of LST by analyzing such wavefunctions; since the coupling is small at large ϕ, perturbative string theory is suitable for this. Also, any normalizable states that are supported in the weakly coupled asymptotic region, like those described in section *3.2*, can be studied using the formalism.

Correlation functions of the observables discussed above are in general difficult to analyze. Since the string coupling goes to zero as $\phi \to \infty$, disturbances on the boundary have to propagate to finite ϕ in order to interact. Thus, to compute correlation functions in LST one needs information about the strong coupling region. E.g. for IIA fivebranes, one has to understand

M-theory in the background (3.21), (3.22) which seems difficult[9].

There are actually some situations in which the strong coupling problem can be avoided. In the next section we describe an example of such a situation, which is in fact of independent interest, the high energy density thermodynamics of LST.

4 High energy thermodynamics of LST

At very high energy density one expects the thermodynamics of fivebranes to be dominated by black brane states. Thus, in this section we will analyze the thermodynamics of near-extremal fivebranes and deduce from it the entropy-energy relation. We will find that the density of states has the Hagedorn behavior

$$\rho(E) \sim E^{\alpha} e^{\beta_H E} \left[1 + O\left(\frac{1}{E}\right) \right]. \tag{4.1}$$

One of our main purposes is to compute β_H and α. This section is based on [23]. For some additional recent work on LST thermodynamics, see [24, 25, 26].

4.1 Thermodynamics of near-extremal fivebranes

The supergravity solution for N coincident near-extremal $NS5$-branes in the string frame is [27]:

$$ds^2 = -\left(1 - \frac{r_0^2}{r^2}\right) dt^2 + \left(1 + \frac{N\alpha'}{r^2}\right) \left(\frac{dr^2}{1 - \frac{r_0^2}{r^2}} + r^2 d\Omega_3^2\right) + dy_5^2, \tag{4.2}$$

$$e^{2\Phi} = g_s^2 \left(1 + \frac{N\alpha'}{r^2}\right). \tag{4.3}$$

$r = r_0$ is the location of the horizon, dy_5^2 denotes the flat metric along the fivebranes, and $d\Omega_3^2$ is the metric on a unit three-sphere, as before. The solution also involves a non-zero NS $B_{\mu\nu}$ field which we suppress. The configuration (4.2), (4.3) has energy per unit volume

$$\frac{E}{V_5} = \frac{1}{(2\pi)^5 \alpha'^3} \left(\frac{N}{g_s^2} + \mu\right), \tag{4.4}$$

[9] For large N and energies much lower than m_s one can use classical eleven dimensional supergravity to compute correlation functions. See [22] for details.

where

$$\mu = \frac{r_0^2}{g_s^2 \alpha'}. \tag{4.5}$$

The first term in (4.4) is the tension of extremal $NS5$-branes and can be ignored for the thermodynamic considerations below – it is a ground state energy. μ measures the energy density above extremality (in string units) and g_s is the asymptotic string coupling, which goes to zero in the decoupling limit.

The near-horizon geometry is obtained by sending $r_0, g_s \to 0$, keeping the energy density μ fixed. Changing coordinates to $r = r_0 \cosh\sigma$ and Wick rotating $t \to it$ to study the thermodynamics, one finds

$$ds^2 = \tanh^2\sigma dt^2 + N\alpha' d\sigma^2 + N\alpha' d\Omega_3^2 + dy_5^2, \tag{4.6}$$

$$e^{2\Phi} = \frac{N}{\mu \cosh^2\sigma}. \tag{4.7}$$

This background corresponds to the worldsheet CFT

$$H_3^+/U(1) \times SU(2)_N \times \mathbb{R}^5, \tag{4.8}$$

where

$$H_3^+ = \frac{SL(2,C)_N}{SU(2)_N} \tag{4.9}$$

is the Euclidean AdS_3 CFT which plays an important role in the AdS-CFT correspondence; the coset $H_3^+/U(1)$, parametrized by (σ, t) in (4.6), is a semi-infinite cigar [28]. The background (4.8) describes the high energy density thermodynamics of fivebranes; it should be compared to (3.4), which is dual to the zero temperature theory.

The absence of a conical singularity at the tip ($\sigma = 0$ in (4.6)) requires the circumference of the cigar to be

$$\beta_H = 2\pi\sqrt{N\alpha'}. \tag{4.10}$$

Thus, Euclidean time lives on a circle of radius $\sqrt{N\alpha'}$, and the temperature of the system is $T_H = 1/\beta_H$. In particular, the temperature is independent of the energy density μ, which determines the value of the string coupling at the tip of the cigar (4.7).

The fact that the temperature is independent of the energy means that the entropy is proportional to the energy (since $\beta = \frac{\partial S}{\partial E}$). Therefore, the free energy is expected to vanish[10],

$$-\beta\mathcal{F} = S - \beta E = 0. \tag{4.11}$$

In general in string theory the free energy is related to the string partition sum via

$$-\beta\mathcal{F} \equiv \log Z(\beta) = Z_{\text{string}}, \tag{4.12}$$

where Z_{string} is the single string partition sum, given by a sum over connected Riemann surfaces [30]. The string path integral should be performed over geometries in which Euclidean time is compactified on a circle of radius $R = \beta/2\pi$ (asymptotically). As mentioned above, for high energies one expects the thermodynamics to be dominated by the black brane geometry (4.2), (4.6) and thus the free energy is proportional to the partition sum of string theory in the background (4.8).

The string partition sum Z_{string} can be expanded as follows:

$$Z_{\text{string}} = e^{-2\Phi_0} Z_0 + Z_1 + e^{2\Phi_0} Z_2 + \cdots, \tag{4.13}$$

where $\exp(\Phi_0)$ is the effective string coupling in the geometry (4.6) and Z_h the genus h partition sum in the background (4.8). Although the string coupling varies along the cigar (see (4.7)), it is bounded from above by its value at the tip,

$$e^{2\Phi_0} = \frac{N}{\mu}. \tag{4.14}$$

Therefore, it is natural to associate (4.14) with the effective coupling in (4.13). We see that the string coupling expansion in the background (4.8) provides an asymptotic expansion of the free energy in powers of $1/\mu$.

The leading term in the free energy (4.12), (4.13) goes like

$$-\beta\mathcal{F} = \frac{\mu}{N} Z_0 \tag{4.15}$$

and corresponds to a free energy that goes like the energy (Z_0 is proportional to the volume of the fivebrane). This term is expected to vanish (see (4.11)),

[10] See [29] for a related discussion in the low energy gravity approximation.

and therefore we conclude that the spherical partition sum in the background (4.8) should vanish. The fact that this is indeed the case follows from the results of [31]; we will not discuss it further here (see [23]).

To compute $1/\mu$ corrections to the free energy we have to examine string loop effects in the background (4.8). We next turn to the one loop correction Z_1 (see (4.13)).

4.2 The leading $1/\mu$ correction to classical thermodynamics

As discussed above, one expects the entropy-energy relation to take the form (4.1)

$$S(E) = \beta_H E + \alpha \log \frac{E}{\Lambda} + O\left(\frac{1}{E}\right), \tag{4.16}$$

where Λ is a dimensionful constant (a UV cutoff) which we will not keep track of below. Consider the canonical partition sum

$$Z(\beta) = \int_0^\infty dE \rho(E) e^{-\beta E}. \tag{4.17}$$

Near the Hagedorn temperature one might expect $Z(\beta)$ to be dominated by the contributions of high energy states;[11] if this is the case, one can replace $\rho(E)$ by (4.1) and find,

$$Z(\beta) \simeq \int dE E^\alpha e^{(\beta_H - \beta)E} \simeq (\beta - \beta_H)^{-\alpha-1}. \tag{4.18}$$

The free energy (4.12) is thus given by

$$\beta \mathcal{F} \simeq (\alpha + 1) \log(\beta - \beta_H). \tag{4.19}$$

The energy computed in the canonical ensemble is

$$E = \frac{\partial(\beta \mathcal{F})}{\partial \beta} \simeq \frac{\alpha + 1}{\beta - \beta_H}; \tag{4.20}$$

thus the free energy (4.19) can be written as

$$-\beta \mathcal{F} \simeq (\alpha + 1) \log E. \tag{4.21}$$

[11] We will see that this assumption is valid slightly *above* the Hagedorn temperature, but is *not* valid slightly below it.

Comparing to the expansion (4.12) – (4.14) we see that the leading term in the free energy arises from the torus (one loop) diagram in the background (4.8), since it scales as μ^0, like Z_1 in (4.13).

The torus partition sum in the background (4.8) is in fact divergent, since it is proportional to the infinite volume of the cigar, associated with the region far from the tip, $\phi \to \infty$. As is standard in other closely related contexts, we will regulate this divergence by requiring that

$$\phi \leq \phi_{UV}. \tag{4.22}$$

In the fivebrane theory, this can be thought of as introducing a UV cutoff. This makes the partition sum finite, but the bulk of the amplitude still comes from the region far from the tip of the cigar. For the purpose of computing this "bulk contribution" one can replace the cigar by a long cylinder with ϕ bounded on one side by the UV cutoff (4.22) and on the other by the location of the tip of the cigar. Combining (3.5) and (4.14) we find that

$$\frac{1}{Q}\log\frac{\mu}{N} \leq \phi \leq \phi_{UV}. \tag{4.23}$$

Thus, the length of the cut-off cylinder is

$$L_\phi = \phi_{UV} - \frac{1}{Q}\log\frac{\mu}{N} = -\frac{1}{Q}\log E + \text{const}. \tag{4.24}$$

Since we are only interested in the energy dependence, we suppress in (4.24) a large energy independent contribution. Any contributions to the torus partition sum from the region near the tip of the cigar can also be lumped into this constant. Note the minus sign in front of $\log E$ in (4.24). The length L_ϕ is of course positive; the minus sign simply means that L_ϕ decreases as E grows.

To recapitulate, for the purpose of calculating the bulk contribution to the torus partition sum, we can replace the background (4.8) by

$$\mathbb{R}_\phi \times S^1 \times SU(2)_N \times \mathbb{R}_5. \tag{4.25}$$

The linear dilaton direction is regulated as in (4.23). The circumference of the S^1 is β_H (4.10).

The background (4.25) is easy to analyze since it is very similar to that describing flat space at finite temperature (see *e.g.* [32, 33, 34]). The bosonic fields on the worldsheet are seven free fields, one of which (Euclidean time) is compact, and a level $N-2$ $SU(2)$ WZW model. The worldsheet fermions are free and decoupled from the bosons; their partition sum, and in particular the sum over spin structures, is the same as in the flat space analysis, which we briefly review next.

Collecting all the contributions to the thermal torus partition sum in the background (4.25) we find,[12]

$$\begin{aligned} Z_1 &= \frac{\beta V_5 L_\phi}{4}\int_F \frac{d^2\tau}{\tau_2}\left(\frac{1}{4\pi^2\alpha'\tau_2}\right)^{7/2}\frac{1}{|\eta(\tau)|^{10}} Z_{N-2}(\tau)\times \\ & \sum_{n,m\in Z}\sum_{\mu,\nu=1}^{4}\delta_\mu U_\mu(n,m)\delta_\nu U_\nu(n,m)\left(\frac{\vartheta_\mu(0,\tau)}{\eta(\tau)}\right)^4\left(\frac{\vartheta_\nu(0,\bar\tau)}{\eta(\bar\tau)}\right)^4 e^{-S_\beta(n,m)}. \end{aligned} \tag{4.26}$$

The modular integral runs over the standard fundamental domain F. Z_{N-2} is the partition sum of level $N-2$ $SU(2)$ WZW [13] (see for example [35]),

$$Z_{N-2}(\tau)=\sum_{m=0}^{N-2}\chi_m^{(N-2)}(q)\chi_m^{(N-2)}(\bar q)=\sum_{m=0}^{N-2}|\chi_m^{(N-2)}(q)|^2, \tag{4.27}$$

where $q=\exp(2\pi i\tau)$ and

$$\chi_m^{(N-2)}(q)=\frac{q^{\frac{(m+1)^2}{4N}}}{\eta(q)^3}\sum_{n\in Z}[1+m+2nN)]q^{n(1+m+Nn)}. \tag{4.28}$$

We note for future reference that Z_{N-2} is real and positive.

μ, ν denote the spin structure for left and right moving worldsheet fermions, respectively. $\delta_\mu=(\pm,-,+,-)$ are signs coming from the usual GSO projections for IIA and IIB superstrings at zero temperature; n, m are winding numbers of Euclidean time around the two non-contractible cycles of the torus. The soliton factor $S_\beta(n,m)$ is given by

$$S_\beta(n,m)=\frac{\beta^2}{4\pi\alpha'\tau_2}(m^2+n^2|\tau|^2-2\tau_1 mn). \tag{4.29}$$

$U_\mu(n,m)$ are additional signs that are associated with finite temperature. Their role is to implement the standard thermal boundary conditions, that

[12] We follow the conventions of [34], which should be consulted for additional details. We also drop the subscript H on β_H, and will reinstate it later.

[13] We choose the A series modular invariant; the D and E series modular invariants can also be studied and correspond to other vacua of LST [18].

spacetime bosons (fermions) are (anti-)periodic around the Euclidean time direction. One can show [34] that this requirement together with modular invariance leads to:

$$\begin{aligned} U_1(n,m) &= \frac{1}{2}\left(-1+(-1)^n+(-1)^m+(-1)^{n+m}\right) \\ U_2(n,m) &= \frac{1}{2}\left(1-(-1)^n+(-1)^m+(-1)^{n+m}\right) \\ U_3(n,m) &= \frac{1}{2}\left(1+(-1)^n+(-1)^m-(-1)^{n+m}\right) \\ U_4(n,m) &= \frac{1}{2}\left(1+(-1)^n-(-1)^m+(-1)^{n+m}\right). \end{aligned} \tag{4.30}$$

The terms with $\mu = 1$ in (4.26) vanish because of the presence of fermionic zero modes for the $(+,+)$ spin structure, or equivalently since $\vartheta_1(0,\tau) = 0$.

The torus partition sum (4.26) can be rewritten in a way that makes it manifest that the coefficient of $\beta V_5 L_\phi/4$ is positive,

$$\begin{aligned} Z_1 &= \frac{\beta V_5 L_\phi}{4}\int_F \frac{d^2\tau}{\tau_2}\left(\frac{1}{4\pi^2\alpha'\tau_2}\right)^{7/2}\frac{1}{|\eta(\tau)|^{18}}Z_{N-2}(\tau)\times \\ &\sum_{n,m\in Z}\left|\sum_{\mu=2}^{4}U_\mu(n,m)\delta_\mu\vartheta_\mu^4(0,\tau)\right|^2 e^{-S_\beta(n,m)}. \end{aligned} \tag{4.31}$$

It is not difficult to check that the integral (4.31) is convergent at $\tau_2 \to \infty$, the only region where a divergence could occur.

To exhibit the interpretation of (4.31) as a sum over the free energies of physical string modes one can proceed as follows [30, 32, 33]. Using the modular invariance of the integrand and the covariance of (n,m), one can extend the integral from the fundamental domain to the strip

$$S: \quad -\frac{1}{2} \le \tau \le \frac{1}{2}; \quad \tau_2 \ge 0, \tag{4.32}$$

while restricting to configurations with $n = 0$ in (4.31). This leads to

$$\begin{aligned} Z_1 &= \frac{\beta V_5 L_\phi}{4}\int_S \frac{d^2\tau}{\tau_2}\left(\frac{1}{4\pi^2\alpha'\tau_2}\right)^{7/2}\frac{1}{|\eta(\tau)|^{18}}Z_{N-2}(\tau)\times \\ &\sum_{m=-\infty}^{\infty}\left|\sum_{\mu=2}^{4}U_\mu(0,m)\delta_\mu\vartheta_\mu^4(0,\tau)\right|^2 e^{-S_\beta(0,m)}. \end{aligned} \tag{4.33}$$

The integral over τ_1 projects on physical states (*i.e.* those with $L_0 = \bar{L}_0$), while τ_2 plays the role of a Schwinger parameter. Because of the Jacobi identity $\vartheta_2^4(0,\tau) - \vartheta_3^4(0,\tau) + \vartheta_4^4(0,\tau) = 0$, and the fact that $U_2(0,m) = (-)^m$, $U_3(0,m) = U_4(0,m) = 1$, the sum over m in (4.33) can be restricted to odd integers. It is not difficult to check in this representation too that the integral over τ_2 is convergent.

We are now ready to determine the parameter α in (4.16), (4.21). Using the relation (4.12) between the free energy $\mathcal{F}$ and the string partition sum, as well as (4.21), we see that Z_1 should be proportional to $\log E$. This is indeed the case in (4.33) since the length L_ϕ goes like $-\log E$ (see (4.24)). Combining these relations we find that

$$\begin{aligned} \alpha + 1 &= -\frac{\beta V_5}{4Q} \int_S \frac{d^2\tau}{\tau_2} \left(\frac{1}{4\pi^2 \alpha' \tau_2} \right)^{7/2} \frac{1}{|\eta(\tau)|^{18}} Z_{N-2}(\tau) \times \\ & \sum_{m=-\infty}^{\infty} \left| \sum_{\mu=2}^{4} U_\mu(0,m) \delta_\mu \vartheta_\mu^4(0,\tau) \right|^2 e^{-S_\beta(0,m)}. \end{aligned} \tag{4.34}$$

We see that $\alpha + 1$ is negative.[14] Physically, it is clear that it is counting the free energy of the perturbative string modes which live in the vicinity of the black brane. An interesting point which was mentioned in [36, 37] is that α is an extensive quantity – it is proportional to the volume of the fivebrane V_5, in contrast, say, to the one particle free energy in critical string theory, where the analogous quantity is of order one.

The integral (4.34) appears in general to be rather formidable and we do not know whether it can be performed exactly. In the remainder of this section we will compute it in the limit $N \to \infty$, where the computation simplifies.

For large N the partition sum corresponding to the three-sphere, $Z_{N-2}(\tau)$, simplifies significantly. Indeed, for $N \gg 1$ (4.27) can be approximated as

$$Z_{N-2}(\tau) = \frac{1}{|\eta(q)|^6} \sum_{p=0}^{\infty} |q|^{\frac{(p+1)^2}{2N}} (p+1)^2. \tag{4.35}$$

[14]Of course, since the r.h.s. of (4.34) is proportional to V_5 which is assumed to be very large, we can neglect the $+1$ on the left-hand side.

Returning to the evaluation of α, (4.34), we have

$$\begin{aligned}\alpha + 1 \quad = \quad & -\frac{\beta V_5}{4Q}\left(\frac{1}{4\pi^2\alpha'}\right)^{7/2}\int_S \frac{d^2\tau}{\tau_2^{9/2}}\left|\frac{1}{\eta(\tau)}\right|^{24} \times \\ & \sum_{m\in 2Z+1}\sum_{p=0}^{\infty} e^{-\frac{(p+1)^2\tau_2}{2N}}(p+1)^2 e^{-\frac{\beta^2 m^2}{4\pi\alpha'\tau_2}}\left|\vartheta_2^4+\vartheta_3^4-\vartheta_4^4\right|^2(0,\tau).\end{aligned} \tag{4.36}$$

At this point it is useful to recall that the inverse temperature β in (4.36) is in fact the Hagedorn temperature of LST, (4.10). In the large N limit, $\beta_H \sim \sqrt{N}$ becomes large (or, equivalently, the Hagedorn temperature is small in string units) and the exponential term in (4.36) suppresses the amplitude, unless τ_2 is large as well (of order N). Therefore, the τ integral in (4.36) is dominated by the large τ_2 region, which corresponds to the free energy of the supergravity modes. To compute the integral we recall the asymptotic forms of the ϑ and η functions at large τ_2 (see *e.g.* [38])

$$\begin{aligned}\vartheta_2(0,\,\tau) \quad &= \quad \sum_{n=-\infty}^{\infty} q^{\frac{1}{2}(n-\frac{1}{2})^2} = 2q^{\frac{1}{8}}(1+q+\dots) \\ \vartheta_3(0,\,\tau) \quad &= \quad \sum_{n=-\infty}^{\infty} q^{\frac{1}{2}n^2} = 1+2q^{\frac{1}{2}}+\dots \\ \vartheta_4(0,\,\tau) \quad &= \quad \sum_{n=-\infty}^{\infty} (-1)^n q^{\frac{1}{2}n^2} = 1-2q^{\frac{1}{2}}+\dots \\ \eta(\tau) \quad &= \quad q^{\frac{1}{24}}\prod_{n=1}^{\infty}(1-q^n) = q^{\frac{1}{24}}+\dots\,.\end{aligned} \tag{4.37}$$

Plugging in (4.36) and using the definition of the modified Bessel function

$$K_\nu(z) = \frac{1}{2}\left(\frac{2}{z}\right)^{\nu}\int_0^\infty t^{\nu-1}e^{-\frac{z^2}{4t}-t}dt, \tag{4.38}$$

we find

$$\begin{aligned}\alpha + 1 \quad = \quad & -\frac{8V_5}{\pi^6(N\alpha')^{5/2}}\sum_{k,p=0}^{\infty}\left(\frac{2\pi(2k+1)^2}{(p+1)^2}\right)^{-7/4}(p+1)^2 \times \\ & K_{-\frac{7}{2}}(\sqrt{2\pi}(p+1)(2k+1)) \simeq -4.08\cdot 10^{-4}V_5(N\alpha')^{-5/2} \equiv -a_1 V_5.\end{aligned} \tag{4.39}$$

Note that, as expected, α is negative. Of course, as is clear from (4.36), we can write $\alpha+1$ as $-a_1 V_5$ with a_1 a positive constant for all N, but in general a_1 receives contributions from massive string modes and is thus given by a complicated modular integral. The large N behavior of a_1 is simpler and is given by (4.39). It should be emphasized that, as mentioned above, the large N result (4.39) comes entirely from the thermodynamics of the supergravity modes in the near-extremal fivebrane background (4.6), (4.7), and thus could have been obtained by a supergravity calculation.

The fact that α goes like $N^{-5/2}$ for large N was found in a different way in [36], by analyzing the deformation of the classical solution (4.6) at one string loop. The analysis described here determines the coefficient of $N^{-5/2}$, and in particular its sign, which is important for the thermodynamics.

In the discussion above, the fivebrane was assumed to be effectively non-compact. It is interesting to study the thermodynamics of fivebranes wrapped around compact manifolds, and in particular the dependence of α on the size and shape of the manifold. As an example of the sort of dependence one can expect, consider compactifying the fivebrane on $(S^1)^5$ where all five circles have the same radius R. It is sufficient to consider the case $R \geq \sqrt{\alpha'}$ since smaller radii give rise to the same physics due to T-duality.

As is standard in string theory, the effect of this is to replace the contribution of the non-compact zero modes on R^5 by the momentum and winding sum on $(S^1)^5$:

$$\frac{V_5}{(4\pi^2\alpha'\tau_2)^{5/2}} \longrightarrow \left(\sum_{l,p\in Z} q^{\frac{\alpha'}{4}\left(\frac{l}{R}+\frac{pR}{\alpha'}\right)^2} \bar{q}^{\frac{\alpha'}{4}\left(\frac{l}{R}-\frac{pR}{\alpha'}\right)^2}\right)^5 . \qquad (4.40)$$

Consider for simplicity the limit $N \to \infty$ discussed above. As mentioned after eq. (4.36), since the Hagedorn temperature is very low, the modular integral is dominated in this case by $\tau_2 \sim N$. If the radius R is much larger than $\sqrt{N\alpha'}$, the sum over momenta on the r.h.s. of (4.40) can be approximated by an integral and gives the same contribution as in the non-compact case (namely the l.h.s. of (4.40)). For $R \sim \sqrt{N\alpha'}$ one has to include a few low lying momentum modes – this is a transition region. For $\sqrt{\alpha'} < R \ll \sqrt{N\alpha'}$ one can neglect all contributions of momentum (and winding) modes, just like one is neglecting the contributions of oscillator states. Thus, we get in this case

$$\alpha + 1 = -\frac{\beta}{2Q}\left(\frac{1}{4\pi^2\alpha'}\right)\int_0^\infty \frac{d\tau_2}{\tau_2^2}\cdot 1024\sum_{k,p=0}^{\infty} e^{-\frac{\beta^2(2k+1)^2}{4\pi\alpha'\tau_2}-\frac{(p+1)^2\tau_2}{2N}} =$$
$$-\frac{256}{\pi}\sum_{k,p=0}^{\infty}\left(\frac{2\pi(2k+1)^2}{(p+1)^2}\right)^{-1/2}(p+1)^2 K_{-1}(\sqrt{2\pi}(p+1)(2k+1)) \simeq -3.693. \quad (4.41)$$

Interestingly, we find that for small fivebranes α is independent of the number of fivebranes N in the $N \to \infty$ limit. Note also that in this case it is important to keep the $+1$ on the l.h.s. of (4.41), since α is of order one.

To summarize, the power α that appears in the high energy density of states (4.1) is negative, and exhibits an interesting dependence on the size of the spatial manifold that the fivebranes are wrapping. For manifolds of size much larger than the characteristic scale of LST, $\sqrt{N\alpha'}$, α is proportional to the volume of the manifold, while for sizes much smaller than this charateristic scale, it saturates at a finite value, which is independent of N (for large N), (4.41). If the density of states (4.1) is due to strings confined to the fivebranes, then these strings belong to a new universality class, with typical configurations not exceeding the size $\sqrt{N\alpha'}$. It would be interesting to understand this universality class better (see also [36]).

4.3 Comments on the near-Hagedorn thermodynamics of LST

The main result of the previous subsections is that the temperature-energy relation has the form (4.20), with α given by (4.36) or for large N by (4.39), (4.41). Since it is negative, the temperature is above the Hagedorn temperature, and the specific heat is negative. This raises two immediate questions:

(1) What is the thermodynamics for temperatures slightly below the Hagedorn temperature?

(2) What is the nature of the instability, reflected by the negative specific heat, above the Hagedorn temperature?

Consider first the behavior well below the Hagedorn temperature, $\beta \gg \beta_H$. In this regime, the thermodynamics is expected to reduce to that corresponding to the extreme IR limit of LST, which is the $(2,0)$ six dimensional SCFT for type IIA LST, or six dimensional $(1,1)$ SYM for IIB. From the point of view of the holographic description, this regime corresponds to the strong coupling region of the near-horizon geometry of the fivebranes, and thus should not be well described by the perturbative theory on the cigar (4.6).

What happens as the temperature approaches T_H from below? One might expect that due to the Hagedorn growth in the density of states (4.1), the high energy part of the spectrum dominates as $\beta \to \beta_H$, and the partition sum becomes better and better approximated by (4.18). What actually happens depends on the value of α, as we discuss next.

Consider first the case of large V_5 ($R \gg \sqrt{N\alpha'}$ in the discussion at the end of section *4.2*). In this case, $|\alpha|$ is large, and the contribution to the partition sum of the high energy part of the spectrum, (4.18), goes rapidly to zero as $\beta \to \beta_H$. The integral over E is dominated by states with moderate energies, whose contribution to the partition sum is analytic at β_H. It is clear that the mean energy remains finite as we approach the Hagedorn temperature from below, and that thermodynamic fluctuations are suppressed (by a factor of the volume V_5). Since the Hagedorn temperature is reached at a finite energy, it corresponds to a phase transition.

As V_5 decreases, α decreases as well, until it reaches the value (4.41). The fluctuations in energy in the canonical ensemble increase with decreasing α. To see that, consider the case $R \ll \sqrt{N\alpha'}$ in the discussion at the end of section *4.2*. Since $-5 < \alpha < -4$ in that case, the expectation values $\langle E^n \rangle$ with $n \geq 4$ in the canonical ensemble diverge as

$$\langle E^n \rangle \sim (\beta - \beta_H)^{-\alpha-n-1}. \tag{4.42}$$

In such situations, one is instructed to pass to the microcanonical ensemble, in which the energy is fixed and the temperature is defined by

$$\beta = \frac{\partial \log \rho}{\partial E} = \beta_H + \frac{\alpha}{E} + \cdots \tag{4.43}$$

where on the r.h.s. we included the first two terms in a perturbative expansion in $1/E$. The perturbative evaluation of β in (4.43) gives a temperature *above* the Hagedorn temperature. This of course does not imply that LST cannot be defined at temperatures below T_H; instead, it means that to study the theory at such temperatures one must compute $S(E)$ to all orders in $1/E$, include non-perturbative corrections, and solve the equation (4.43) to find the energy E corresponding to a particular $\beta > \beta_H$. From the form of the leading terms in $S(E)$ it is clear that the solution of this equation will correspond to finite E. We are led again to the conclusion that the Hagedorn temperature is reached at a finite energy and thus is associated with a phase transition.

Since the study of the non-extremal fivebrane geometry in the previous sections is perturbative in $1/E$, it is not useful for studying the regime $\beta > \beta_H$. Nevertheless, it seems clear that the specific heat is positive there (this is certainly the case for the infrared theory on the fivebranes). Furthermore, since the energy – temperature relation is such that the Hagedorn temperature is reached at a finite energy, we are led to the second question raised in the beginning of this section: what is the nature of the high temperature phase of LST?

The perturbative analysis of the near-extremal fivebrane, which is valid for β slightly below β_H, predicts that the thermodynamics is unstable. Usually, in such situations the instability is associated with a negative mode in the Euclidean path integral (a tachyon). Examples include the instability of flat space at finite temperature in Einstein gravity [39], and the thermal tachyon that appears above the Hagedorn transition in critical string theory. The one loop instability found above leads one to believe that a similar negative mode should appear in LST above the Hagedorn temperature.

In [23] it was shown that there is a natural candidate for this, a mode that lives near the tip of the cigar and is classically massless. It is likely that one loop corrections give a tachyonic correction to the mass of this state above the Hagedorn temperature, but this has not been proven and we will not discuss the detailed properties of this state here.

5 Weakly coupled LST

In the previous section we saw that the high energy thermodynamics of LST can be analyzed reliably using the holographically dual description, since at large energy density the strongly coupled region on $\mathbb{R}_\phi$ is eliminated, and the coupling never exceeds (4.14), a value that can be made arbitrarily small by increasing the energy density. In this section we will describe another situation where something similar happens at zero temperature, by studying the theory away from the origin of its moduli space of vacua. This section is based on [40].

Recall that the theory of N fivebranes contains four massless scalars in the adjoint of $U(N)$, X^i, $i = 6, 7, 8, 9$, parameterizing motions in $(6, 7, 8, 9)$. IIA fivebranes have one more scalar X^{11}, which is compact, but we will not discuss it here. The moduli space of vacua of LST is $\mathbb{R}^{4N}/S_N$ for IIB and

$(\mathbb{R}^4 \times S^1)^N/S_N$ for IIA. The origin corresponds to coincident fivebranes; other points are labeled by relative separations of the fivebranes.

The four scalars X^i can be parametrized by two complex $N \times N$ matrices,

$$\begin{aligned} A &\equiv X^8 + iX^9, \\ B &\equiv X^6 + iX^7. \end{aligned} \tag{5.1}$$

Consider a point on the moduli space where

$$\begin{aligned} \langle A \rangle &= 0, \\ \langle B \rangle &= r_0 \mathrm{diag}(1, e^{\frac{2\pi i}{N}}, e^{\frac{4\pi i}{N}}, \cdots, e^{\frac{2\pi i (N-1)}{N}}) \ . \end{aligned} \tag{5.2}$$

This corresponds to fivebranes symmetrically distributed around a circle of radius r_0 in the $(6,7)$ plane. The gauge invariant characterization of this vacuum is

$$\langle \mathrm{Tr}\, B^N \rangle = r_0^N \tag{5.3}$$

with all other v.e.v.'s of the operators (3.11) set to zero. Since for a single fivebrane the worldvolume dynamics is trivial, in order to get a non-trivial result in the limit (2.2), we have to tune $r_0 \to 0$ as we take the limit. E.g., in the IIB case the masses of D-strings stretched between $NS5$-branes

$$M_W \sim \frac{r_0 m_s^2}{g_s} \tag{5.4}$$

must be kept finite in the limit. This leads one to consider the double scaling limit

$$g_s \to 0; \ r_0 m_s \to 0 \tag{5.5}$$

with M_W/m_s (5.4) held fixed.

Distributing the branes on a circle as in (5.2) breaks the $SO(4)$ R-symmetry

$$SO(4) \to SO(2) \times \mathbb{Z}_N \ . \tag{5.6}$$

We will next show that is also eliminates the strong coupling singularity at $\phi \to -\infty$ discussed above.

The first thing we have to understand is how to describe the vacuum (5.3) in the holographically dual theory. In section *3.1* we found the vertex operators corresponding to the gauge invariant operators (3.11). It is not difficult to see that

$$\mathrm{Tr}\, B^N \leftrightarrow \psi^+\bar{\psi}^+ V_{\frac{N}{2}-1;\frac{N}{2}-1,\frac{N}{2}-1} \exp\left[\frac{2}{\sqrt{N\alpha'}}(\frac{N}{2}-1)\phi\right] . \tag{5.7}$$

Adding the vertex operator (5.7) to the worldsheet action is equivalent, via the prescription (3.10), to adding the operator Tr B^N to the action of LST. In order to turn on a v.e.v. of Tr B^N instead, as in (5.3), we have to use the same vertex operator but replace the charge β in (3.13) by

$$\beta \to -Q - \beta. \tag{5.8}$$

Thus, to describe the vacuum (5.3) we must study the worldsheet Lagrangian

$$\mathcal{L} = \mathcal{L}_0 + \lambda G_{-\frac{1}{2}}\bar{G}_{-\frac{1}{2}}\psi^+\bar{\psi}^+ V_{\frac{N}{2}-1;\frac{N}{2}-1,\frac{N}{2}-1} e^{-\sqrt{\frac{N}{\alpha'}}\phi} + \text{c.c.} \tag{5.9}$$

where we explicitly wrote the worldsheet supercharges which are needed to turn a $(-1,-1)$ picture vertex operator to a $(0,0)$ picture one (the appropriate picture for a term in the worldsheet Lagrangian). λ is a coupling related to r_0. $\mathcal{L}_0$ is the free Lagrangian describing string propagation on (3.4). Since the coupling λ breaks explicitly the $SU(2)_L \times SU(2)_R$ symmetry, it is convenient to analyze its effect by rewriting the background (3.4) as

$$\mathbb{R}^{5,1} \times \mathbb{R}_\phi \times \left(S^1 \times \frac{SU(2)}{U(1)}\right)/\mathbb{Z}_N \tag{5.10}$$

where $SU(2)/U(1)$ is an $N = 2$ minimal model, and S^1 a circle of radius $\sqrt{N\alpha'}$. Denoting the coordinate along the circle by Y, one can show that the interaction in (5.9) can be written as

$$\delta\mathcal{L} = \lambda G_{-\frac{1}{2}}\bar{G}_{-\frac{1}{2}} e^{-\frac{2}{\alpha' Q}(\phi+iY)} + \text{c.c.} \tag{5.11}$$

This interaction is familiar in CFT as the $N = 2$ Liouville interaction. Thus, we find that to describe the vacuum (5.3), we must replace the infinite cylinder $\mathbb{R}_\phi \times S^1$ in (5.10) by the $N = 2$ Liouville model. Note that:

(1) The fact that the interaction (5.9), (5.11) preserves $N = 2$ superconformal invariance is related to the fact that spacetime supersymmetry remains unbroken along the moduli space of LST.

(2) The interaction (5.11) grows as $\phi \to -\infty$. One can show that it resolves the strong coupling singularity discussed in section 3. We will see this directly momentarily.

To study $N = 2$ Liouville theory, it is convenient to use a dual description of this background. It was argued in [40] that $N = 2$ Liouville is equivalent via strong-weak coupling duality on the worldsheet to CFT on the cigar, $H_3^+/U(1)$, which was discussed in section 2. The parameter N which enters the definition of $N = 2$ Liouville (5.11) via Q is mapped under the duality to the level of the underlying $SL(2)$ current algebra.

I will not describe the duality or the evidence for it here[15], but rather will use it to conclude that the vacuum (5.2), (5.3) is dual to

$$\mathbb{R}^{5,1} \times \left(\frac{SL(2)}{U(1)} \times \frac{SU(2)}{U(1)} \right) / \mathbb{Z}_N . \tag{5.12}$$

Note that the unbroken R-symmetry $SO(2) \times \mathbb{Z}_N$ of the vacuum (5.3) is manifest in the description (5.12). The $SO(2)$ symmetry corresponding to rotations in the $(8,9)$ plane is realized as the $U(1)$ translation symmetry around the cigar. The rotation symmetry in the $(6,7)$ plane, which is broken to $\mathbb{Z}_N$ by the v.e.v. of B, corresponds to winding number around the cigar. This quantum number is not conserved, since winding can slip off the tip of the cigar. The $\mathbb{Z}_N$ orbifold in (5.12) leads to a $\mathbb{Z}_N$ remnant of it (since it allows fractional windings $\in \mathbb{Z}/N$).

The radius of the circle on which the fivebranes lie, r_0 in (5.2), is related to the value of the string coupling at the tip of the cigar, g_{cigar}. The precise relation can be determined by noting that D-branes stretched between fivebranes, whose mass is given by (5.4), correspond in (5.12) to D-branes at the tip of the cigar, whose mass is m_s/g_{cigar}. This implies that

$$g_{\text{cigar}} \simeq \frac{m_s}{M_W} . \tag{5.13}$$

Thus, the theory is weakly coupled when $M_W >> m_s$; as M_W decreases, we recover the original strongly coupled theory described holographically by (3.4). As mentioned above, the behavior (5.13) is very reasonable: as $M_W/m_s \to \infty$ the fivebranes become infinitely separated and decouple (recall that the dynamics on a single fivebrane is trivial).

The weakly coupled nature of the theory (5.12) for $M_W >> m_s$ allows one to determine the spectrum in a wide range of energies $0 < E << M_W$,

[15] See [41, 42] for more detailed discussions.

and to compute various off-shell correlation functions of the observables discussed in section 3. Interactions can be turned on gradually by increasing $g_{\rm cigar}$ (5.13). For energies $E \gg M_W$ one expects the weak coupling expansion to break down. Physically, the reason for that is that in this regime the symmetry breaking in (5.2) can be neglected, and the physics is that of coincident fivebranes. All this is very similar to critical string theory, where the string coupling expansion is associated with a large hierarchy of energy scales, m_s/m_p. For $E \sim m_p$ the string coupling expansion breaks down.

Two and three point functions as well as the spectrum of weakly coupled LST were analyzed in [40]. We next illustrate the resulting structure by discussing an example.

Consider the operator Tr $B^N(x)$. The dual vertex operator (5.7) can be written in terms of the background (5.12) as

$$\text{Tr } B^N(x) \leftrightarrow e^{-\varphi-\bar\varphi} e^{ik_\mu x^\mu} V_{j;m,m} \tag{5.14}$$

with $m = N/2$. φ, $\bar\varphi$ are the standard bosonized superconformal ghosts needed for the -1 picture, $V_{j;m,\bar m}$ is a Virasoro primary on the cigar carrying p units of momentum and w units of winding, with

$$m = \frac{1}{2}(p + wN) \ ; \ \ \bar m = -\frac{1}{2}(p - wN) \ . \tag{5.15}$$

In the case (5.14), $p = 0$ while $w = 1$ (*i.e.* $m = \bar m = N/2$). The worldsheet scaling dimension of $V_{j;m,m}$ is

$$\Delta = \bar\Delta = \frac{m^2 - j(j+1)}{N} \tag{5.16}$$

Requiring that (5.14) be physical gives rise to the mass-shell condition

$$\alpha' k_\mu k^\mu = \frac{4}{N}(j - m + 1)(j + m). \tag{5.17}$$

To compute the two point function of Tr $B^N(k_\mu)$ we use the correspondence (3.10):

$$\langle \text{Tr } B^N(k_\mu) \text{Tr } \bar B^N(-k_\mu)\rangle = \langle e^{-\varphi-\bar\varphi} e^{ik_\mu x^\mu} V_{j;m,m} e^{-\varphi-\bar\varphi} e^{-ik_\mu x^\mu} V_{j;-m,-m}\rangle. \tag{5.18}$$

The only non-trivial part of the correlator on the r.h.s. is $\langle VV\rangle$. It was computed in [43]:

$$\langle V_{j;m,\bar m} V_{j;-m,-\bar m}\rangle = N[\nu(N)]^{2j+1} \frac{\Gamma(1 - \frac{2j+1}{N})\Gamma(-2j-1)\Gamma(j-m+1)\Gamma(1+j+\bar m)}{\Gamma(\frac{2j+1}{N})\Gamma(2j+2)\Gamma(-j-m)\Gamma(\bar m - j)} \ . \tag{5.19}$$

where

$$\nu(N) \equiv \frac{1}{\pi} \frac{\Gamma(1+\frac{1}{N})}{\Gamma(1-\frac{1}{N})} . \tag{5.20}$$

The two point function (5.19) has a series of poles; these can be interpreted as contributions of on-shell states in weakly coupled LST, which are created from the vacuum by the operator (5.14). The masses of these states can be computed by using the relation (5.17) between j and $M^2 = -k_\mu k^\mu$. The locations of the poles are given by

$$|m| = j + n; \quad n = 1, 2, 3, \cdots \tag{5.21}$$

These values of m and j belong to the principal discrete series representations of $SL(2)$. The corresponding states can be thought of as bound states that live near the tip of the cigar [44]. Such bound states are to be expected since winding modes around the cigar feel an effective attractive potential towards the tip – their energy decreases as they approach the tip and shrink.

For the particular case (5.14), $m = \bar{m} = N/2$, and the masses of these states are given by

$$\begin{aligned} &\tfrac{\alpha'}{2} M_n^2 = \tfrac{2}{N}(n-1)(N-n) \ , \\ &N + 1 > 2n > 1 \ . \end{aligned} \tag{5.22}$$

The second line in (5.22) comes from a unitarity constraint on j which must be imposed, $-1/2 < j < (N-1)/2$. Note that all the masses squared in (5.22) are non-negative; For $n = 1$ one finds massless states, which correspond to the eigenvalues of the scalar matrix B.
A few comments are in order here:

(1) By analyzing the behavior of the two point function (5.18), (5.19) one can check that the residues of the poles corresponding to the states (5.22) are positive, in agreement with the unitarity of the theory.

(2) In addition to the discrete spectrum given by (5.22), one also has the continuum discussed in section 3 (3.19). One can show that the continuum starts right above the heaviest state (5.22). Thus the spectrum of states that can be created from the vacuum by the operator (5.14) is a finite discrete set, followed by a continuum (similar to the spectrum of bound states and scattering states in quantum mechanics).

(3) It is interesting that the low lying spectrum of states associated with N $NS5$-branes is independent of M_W, or equivalently the radius of the circle on which the fivebranes are placed. This should be contrasted with D-branes, for which masses of open strings stretched between different branes depend on the separation. When the distance between D-branes goes to infinity, states associated with strings stretched between different branes go to infinite mass and decouple. For $NS5$-branes, the masses of low lying states remain finite, and the decoupling is due to the vanishing of the effective coupling (5.13).

(4) In addition to the poles (5.21), which corespond to principal discrete series states near the tip of the cigar, the amplitude (5.19) has poles at $0 \leq 2j+1 \in \mathbb{Z}$, $0 < 2j+1 \in N\mathbb{Z}$. These poles have a different interpretation than (5.21). They are associated with "bulk scattering processes" which can occur anywhere in the infinite throat corresponding to either the $N=2$ Liouville (5.11), or $SL(2)/U(1)$ (5.12) description. This is discussed further in [42].

(5) One can repeat the above discussion for other observables as well. The resulting picture is similar; one always finds a finite set of discrete states which live near the tip of the cigar, followed by a continuum of states which propagate in the semi-infinite throat [40].

(6) Since there is a Hagedorn growth in the number of observables (coming from oscillator states on (5.12)), one finds a Hagedorn density of states in LST. But the exponent β_H (4.1) does not grow like $\sqrt{N}$ as expected from (4.10). Instead one gets $\beta_H \sim 1/m_s$. This is not particularly surprising since (4.10) is the expected behavior for high energies $E >> M_W$, whereas the present analysis is only valid in the intermediate regime $m_s << E << M_W$.

(7) Three point functions of the off-shell observables discussed above can be computed as well using the results of [43]. One finds a similar analytic structure to that exhibited by the two point functions. There are poles associated with external legs going on-shell; their locations correspond again to the spectrum (5.22). The residues of these poles describe the scattering amplitudes of the physical states; they seem to have sensible physical properties. See [40] for details.

6 Other aspects of LST

In this section we would like to briefly list some additional topics in Little String Theory, which were not covered in detail in the lectures due to lack of time.

6.1 Singular Calabi-Yau manifolds and lower dimensional vacua of LST

The theory of N $NS5$-branes discussed in sections 2 – 5 is related to string dynamics on an ALE space $\mathbb{C}^2/\mathbb{Z}_N$, which can be described as the manifold

$$z_1^N + z_2^2 + z_3^2 = \mu \tag{6.1}$$

in $\mathbb{C}^3$. For $\mu = 0$, (6.1) corresponds to a cone; non-zero μ smoothes out the tip of the cone. String propagation on $\mathbb{R}^{5,1} \times \mathbb{C}^2/\mathbb{Z}_N$ is dual [45, 46] to a vacuum with coincident fivebranes. The blowing up parameter μ is related by duality to the distance between the fivebranes. From the perspective of the geometry (6.1), LST describes the dynamics of the modes localized at the singularity, which can be decoupled from the rest of the theory.

This picture can be naturally generalized to a large class of vacua of LST in $d < 6$ dimensions [47]. Consider, for example, string propagation on

$$\mathbb{R}^{3,1} \times \mathcal{M} \ , \tag{6.2}$$

where $\mathcal{M}$ is a Calabi-Yau manifold with an isolated singularity, which looks locally like

$$F(z_1, z_2, z_3, z_4) = 0 \ . \tag{6.3}$$

Here F is a quasi-homogeneous polynomial,

$$F(\lambda^{r_1} z_1, \lambda^{r_2} z_2, \lambda^{r_3} z_3, \lambda^{r_4} z_4) = \lambda F(z_1, z_2, z_3, z_4) \tag{6.4}$$

for some set of charges r_1, r_2, r_3, r_4. Viewed as a hypersurface in $\mathbb{C}^4$, (6.3) describes the vicinity of the singular point $z_1 = z_2 = z_3 = z_4 = 0$.

In analogy to the six dimensional situation (6.1), string theory in the background (6.3) is expected to contain modes localized near the singularity; these modes can be decoupled from the bulk in the same way as in the six dimensional case.

The decoupled dynamics at the singularity (6.3) is holographically dual to string theory in the background

$$\mathbb{R}^{3,1} \times \mathbb{R}_\phi \times \left(S^1 \times LG(F)\right)/\Gamma \ , \tag{6.5}$$

where $LG(F)$ is a Landau-Ginsburg model with the superpotential given by the quasi-homogeneous polynomial $F(z_1, \cdots, z_4)$ defining the singularity (6.3). Γ is a discrete group whose origin is the chiral GSO projection in the vacuum (6.5). As before, $\mathbb{R}_\phi$ is a linear dilaton direction, with the slope Q determined such that the total central charge of (6.5) is fifteen, as appropriate for a critical superstring vacuum. One can show that this implies that

$$\frac{1}{2}Q^2 = \sum_{a=1}^{4} r_a - 1 \ . \tag{6.6}$$

Vacua of the form (6.5) preserve eight supercharges and give rise to $N = 2$ supersymmetric theories in four dimensions.

A simple example is

$$F = z_1^2 + z_2^2 + z_3^3 + z_4^2 \ , \tag{6.7}$$

which corresponds to the conifold. In this case, (6.5) reduces to

$$\mathbb{R}^{3,1} \times \mathbb{R}_\phi \times S^1 \ , \tag{6.8}$$

which is the background holographically dual to string theory on the conifold. Smoothing out the singularity as in (6.1) corresponds to replacing the factor $\mathbb{R}_\phi \times S^1$ in (6.8) by the cigar $SL(2)/U(1)$ (or equivalently $N = 2$ Liouville).

In the same way that the ALE space (6.1) is dual to parallel fivebranes, the background (6.7), (6.8) arises from two orthogonal $NS5$-branes intersecting along $3 + 1$ dimensional Minkowski spacetime.

More generally, if

$$F(z_1, \cdots, z_4) = H(z_1, z_2) + z_3^2 + z_4^2 \ , \tag{6.9}$$

the background (6.3) can be thought of as arising from an $NS5$-brane wrapped around the surface $H(z_1, z_2) = 0$ [48]. An interesting class of examples corresponds to $H(z_1, z_2)$ describing an ADE singularity (*e.g.* $H = z_1^n + z_2^2$ for A_{n-1}), in which case the fivebrane wraps a Seiberg-Witten curve at the corresponding Argyres-Douglas point. For type IIA fivebranes, at low energies the system approaches an interacting four dimensional $N = 2$ SCFT. In the description (6.5), this SCFT corresponds to the strong coupling region in the background (6.5), where the theory is eleven dimensional, and is difficult to study in detail (beyond the supergravity approximation). In [47] it was shown that certain properties of chiral operators which can be studied at weak coupling (such as the R-charges), agree with known results.

The construction described in this subsection can be generalized to other dimensions and more complicated models in four dimensions. For some work in this direction, see [49, 50, 51, 52, 53, 54]. Other vacua of LST in six dimensions with less than maximal supersymmetry were discussed in [55, 56, 57].

6.2 D-branes in the vicinity of $NS5$-branes

D-branes stretched between $NS5$-branes in the weak coupling limit $g_s \to 0$ have been seen in recent years to be very useful for studying the dynamics of a wide class of gauge theories, which are realized as the low energy theories on such branes [2]. In particular, $D4$-branes stretched between parallel adjacent fivebranes realize $N = 2$ SYM and are very useful for embedding Seiberg-Witten theory in string theory [58]. $D4$-branes stretched between orthogonal fivebranes which share $3 + 1$ dimensions, give rise to $N = 1$ SYM and are very useful for studying Seiberg duality in string theory [59].

We have seen above that nearby fivebranes can be described by throat geometries which involve the cigar $SL(2)/U(1)$. Adjacent parallel fivebranes are described by (5.12), while orthogonal fivebranes intersecting on $3 + 1$ dimensional Minkowski spacetime correspond to $\mathbb{R}^{3,1} \times SL(2)/U(1)$. D-branes stretched between the fivebranes correspond in this description to D-branes localized on the cigar and extended in some or all of the non-compact directions. For example, a $D4$-brane stretched between parallel fivebranes corresponds in the geometry (5.12) to a $D3$-brane in $\mathbb{R}^{5,1}$, which is localized on the cigar, and is in one of the familiar boundary states in the $N = 2$ minimal model $SU(2)/U(1)$. One can show that different boundary states in the minimal model correspond to D-branes stretched between different pairs of $NS5$-branes.

Thus, one is led to study D-branes localized on the cigar. It is clear that such D-branes will live near the tip of the cigar, since this is where the string coupling is largest, and thus the energy of the D-branes is smallest. The physics of D-branes living near the tip of the cigar is at present not completely understood, but progress has been made on two closely related problems. In [60], D-branes in Liouville theory have been constructed which can be thought of as being localized in the Liouville direction. It was found that such D-branes are labeled by two integers, and exhibit very interesting properties, such as having a finite number of Virasoro primaries in the open string sector. This construction was generalized in [61] to Euclidean AdS_3, where similar D-branes were found.

Both Liouville theory and AdS_3 are known to share many properties with the cigar and $N = 2$ Liouville CFT's. Thus, it is reasonable to expect that localized D-branes exist on the cigar and $N = 2$ Liouville backgrounds as well. It would be interesting to construct them and use their properties to study gauge dynamics.

Another class of objects that figures in many brane constructions is D-branes ending on $NS5$-branes. In some cases such branes can be studied by analyzing them in the throat region of the fivebranes. For example, consider a D-brane that ends on a stack of N fivebranes. Assuming that the brane extends into the throat of the fivebranes,[16] one can study the behavior of the brane inside the throat region described by (5.10), or for separated fivebranes by (5.12). This analysis was carried out in [62], where many properties of such D-branes that were previously deduced by using spacetime considerations, were verified by using the technology of LST.

6.3 Low dimensional toy models of LST

From the modern perspective, the matrix model description of two dimensional string theory (see *e.g.* [63]) for a review) provides an early example of holography in string theory. Since the "bulk" theory involves in this case a linear dilaton direction, the situation looks like a low dimensional toy model of LST.

Unlike LST, which is difficult to formulate directly (except for the DLCQ construction of [7]), here there is an alternative definition of the theory, which is moreover exactly solvable to all orders in the string coupling expansion (or equivalently, the $1/N$ expansion in the matrix model). This is especially interesting since, like LST, two dimensional string theory is expected to exhibit a Hagedorn growth in the density of states, as in (4.1), with the parameters

[16]This is a non-trivial assumption; it is believed that in some cases D-branes that end on fivebranes do not extend into the throat. The simplest example is D-branes ending on a single fivebrane, which does not have a throat region.

β_H and α known from thermodynamic considerations [64]. Therefore, two dimensional string theory is an interesting toy model of the dynamics of LST in higher dimensions.

In [64], the matrix model description was used to study some properties of the Euclidean black hole solution of two dimensional string theory. In particular, some steps were taken towards developing a description of the states that give rise to the Hagedorn entropy (4.1) directly in the matrix model.

7 Some open problems in LST

While a lot has been achieved, many interesting questions regarding Little String Theory await resolution. Some examples of open problems are:

(1) We have seen in section 4 that LST has at high energies a Hagedorn spectrum of states (4.1). This was established by a thermodynamic analysis; it would be very interesting to exhibit the density of state (4.1) by an explicit counting of states. In the background (3.4) corresponding to coincident fivebranes (and the lower dimensional analogs (6.5)), this is complicated by the fact that string theory in the linear dilaton background is not weakly coupled. The weakly coupled theory described in section 5 does have a Hagedorn spectrum of perturbative states, but the Hagedorn coefficient β_H is smaller than that of the full theory (4.10). As explained in section 5, this is not surprising – most of the states contributing to (4.1) are expected to be non-perturbative. As a first step to counting non-perturbative states in the background (5.12), it would be interesting to enumerate states corresponding to collections of D-branes living near the tip of the cigar, which were briefly discussed in section 6.

(2) It would be interesting to understand the dynamics of D-branes stretched between $NS5$-branes, which correspond to D-branes localized near the tip of the cigar in the dual geometry (5.12). This might be useful for application to gauge dynamics, as well as for providing a direct formulation of LST, independent of holography.

(3) Most of the work on LST concerned spacetime supersymmetric vacua. In the absence of spacetime SUSY one expects to find infrared instabilities such as tachyons, and the system might decay to a more stable

vacuum. It would be interesting to understand the physics associated with supersymmetry breaking and vacuum instabilities in LST. First steps in that direction were recently taken in [65].

Acknowledgements:

I thank O. Aharony, M. Berkooz, S. Elitzur, A. Giveon, V. Kazakov, I. Kostov, O. Pelc, E. Rabinovici, D. Sahakyan, G. Sarkissian, and N. Seiberg for collaboration on the topics presented here. This work is supported in part by DOE grant #DE-FG02-90ER40560. I thank the Weizmann Institute of Science for hospitality during the preparation of these notes.

References

[1] L. J. Dixon, J. A. Harvey, C. Vafa and E. Witten, "Strings On Orbifolds," Nucl. Phys. B **261**, 678 (1985).

[2] A. Giveon and D. Kutasov, "Brane dynamics and gauge theory," Rev. Mod. Phys. **71**, 983 (1999) [hep-th/9802067].

[3] O. Aharony, S. S. Gubser, J. Maldacena, H. Ooguri and Y. Oz, "Large N field theories, string theory and gravity," Phys. Rept. **323**, 183 (2000) [hep-th/9905111].

[4] C. G. Callan, J. A. Harvey and A. Strominger, "Supersymmetric string solitons," hep-th/9112030.

[5] P. S. Aspinwall, "Enhanced gauge symmetries and K3 surfaces," Phys. Lett. B **357**, 329 (1995) [hep-th/9507012].

[6] A. Losev, G. W. Moore and S. L. Shatashvili, "M & m's," Nucl. Phys. B **522**, 105 (1998) [hep-th/9707250].

[7] O. Aharony, M. Berkooz, S. Kachru, N. Seiberg and E. Silverstein, "Matrix description of interacting theories in six dimensions," Adv. Theor. Math. Phys. **1**, 148 (1998) [hep-th/9707079].

[8] E. Witten, "On the conformal field theory of the Higgs branch," JHEP **9707**, 003 (1997) [hep-th/9707093].

[9] O. Aharony, M. Berkooz and N. Seiberg, "Light-cone description of (2,0) superconformal theories in six dimensions," Adv. Theor. Math. Phys. **2**, 119 (1998) [hep-th/9712117].

[10] O. Aharony, "A brief review of little string theories", Class. Quant. Grav. **17** (2000) 929, [hep-th/9911147].

[11] T. Banks, W. Fischler, S. H. Shenker and L. Susskind, "M theory as a matrix model: A conjecture," Phys. Rev. D **55**, 5112 (1997) [hep-th/9610043].

[12] M. Berkooz, M. Rozali and N. Seiberg, "On transverse fivebranes in M(atrix) theory on T**5," Phys. Lett. B **408**, 105 (1997) [hep-th/9704089].

[13] N. Seiberg, "New theories in six dimensions and matrix description of M-theory on T**5 and T**5/Z(2)," Phys. Lett. B **408**, 98 (1997) [hep-th/9705221].

[14] I. Antoniadis, S. Dimopoulos and A. Giveon, "Little string theory at a TeV," JHEP **0105**, 055 (2001) [hep-th/0103033].

[15] J. Polchinski, "String theory", vol. 2, Cambridge University Press, 1998.

[16] E. Witten, "Some comments on string dynamics," hep-th/9507121.

[17] A. Strominger, "Open p-branes," Phys. Lett. B **383**, 44 (1996) [hep-th/9512059].

[18] O. Aharony, M. Berkooz, D. Kutasov and N. Seiberg, "Linear dilatons, NS5-branes and holography," JHEP **9810**, 004 (1998) [hep-th/9808149].

[19] N. Itzhaki, J. M. Maldacena, J. Sonnenschein and S. Yankielowicz, "Supergravity and the large N limit of theories with sixteen supercharges," Phys. Rev. D **58**, 046004 (1998) [hep-th/9802042].

[20] H. J. Boonstra, K. Skenderis and P. K. Townsend, "The domain wall/QFT correspondence," JHEP **9901**, 003 (1999) [hep-th/9807137].

[21] N. Seiberg and S. H. Shenker, "A Note on background (in)dependence," Phys. Rev. D **45**, 4581 (1992) [hep-th/9201017].

[22] S. Minwalla and N. Seiberg, "Comments on the IIA NS5-brane," JHEP **9906**, 007 (1999) [hep-th/9904142].

[23] D. Kutasov and D. A. Sahakyan, "Comments on the thermodynamics of little string theory," JHEP **0102**, 021 (2001) [hep-th/0012258].

[24] M. Rangamani, "Little string thermodynamics," JHEP **0106**, 042 (2001) [hep-th/0104125].

[25] K. Narayan and M. Rangamani, "Hot little string correlators: A view from supergravity," JHEP **0108**, 054 (2001) [hep-th/0107111].

[26] A. Buchel, "On the thermodynamic instability of LST," hep-th/0107102.

[27] G. Horowitz and A. Strominger, "Black Strings and p–Branes", Nucl. Phys. **B360** (1991) 197; J. Maldacena and A. Strominger, "Semiclassical decay of near extremal fivebranes", JHEP **9712** (1997) 008, hep-th/9710014.

[28] E. Witten, "On String Theory and Black Holes", Phys. Rev. **D44** (1991) 314.

[29] G. Gibbons and M. Perry, "The Physics of 2-d Stringy Spacetimes", Int. J. Mod. Phys. **D1** (1992) 335, hep-th/9204090; C. R. Nappi and A. Pasquinucci, "Thermodynamics of Two-Dimensional Black-Holes", Mod. Phys. Lett. **A7** (1992) 3337, gr-qc/9208002.

[30] J. Polchinski, "Evaluation of the One Loop String Path Integral", Commun. Math. Phys. **104** (1986) 37.

[31] D. Kutasov and N. Seiberg, "Noncritical Superstrings," Phys. Lett. B **251**, 67 (1990); D. Kutasov, "Some properties of (non)critical strings," hep-th/9110041.

[32] B. McClain and B. D. B. Roth, "Modular Invariance for Interacting Bosonic String at Finite Temperature", Commun. Math. Phys. **111** (1987) 539.

[33] K. H. O'Brien and C.-I. Tan, "Modular Invariance of the thermopartition function and global phase structure of the heterotic string", Phys. Rev. **D36** (1987) 1184.

[34] J. J. Atick and E. Witten, "The Hagedorn Transition and the Number of Degrees of Freedom of String Theory", Nucl. Phys. **B310** (1988) 291.

[35] P. Di Francesco, P. Mathieu, D. Senechal, "Conformal Field Theory", Springer-Verlag, 1997.

[36] M. Berkooz and M. Rozali, "Near Hagedorn dynamics of NS fivebranes, or a new universality class of coiled strings," JHEP **0005**, 040 (2000) [hep-th/0005047].

[37] T. Harmark and N. A. Obers, "Hagedorn behaviour of little string theory from string corrections to NS5-branes," Phys. Lett. B **485**, 285 (2000) [hep-th/0005021].

[38] J. Polchinski, "String theory", vol. 1, Cambridge University Press, 1998.

[39] D. J. Gross, M. J. Perry and L. G. Yaffe, "Instability of Flat Space at Finite Temperature", Phys. Rev. **D25** (1982) 330.

[40] A. Giveon and D. Kutasov, "Little string theory in a double scaling limit," JHEP **9910**, 034 (1999) [hep-th/9909110]; "Comments on double scaled little string theory," JHEP **0001**, 023 (2000) [hep-th/9911039].

[41] K. Hori and A. Kapustin, "Duality of the fermionic 2d black hole and N = 2 Liouville theory as mirror symmetry," JHEP **0108**, 045 (2001) hep-th/0104202.

[42] A. Giveon and D. Kutasov, "Notes on AdS(3)," hep-th/0106004.

[43] J. Teschner, "On structure constants and fusion rules in the $SL(2,C)/SU(2)$ WZNW model", Nucl. Phys. **B546** (1999) 390, hep-th/9712256; "The Mini-Superspace Limit of the $SL(2,C)/SU(2)$-WZNW Model", Nucl. Phys. **B546** (1999) 369, hep-th/9712258; "Operator product expansion and factorization in the H_3^+-WZNW model", Nucl. Phys. **B571** (2000) 555, hep-th/9906215.

[44] R. Dijkgraaf, H. Verlinde and E. Verlinde, "String propagation in a black hole geometry," Nucl. Phys. B **371**, 269 (1992).

[45] H. Ooguri and C. Vafa, "Two-Dimensional Black Hole and Singularities of CY Manifolds," Nucl. Phys. B **463**, 55 (1996) hep-th/9511164.

[46] D. Kutasov, "Orbifolds and Solitons," Phys. Lett. B **383**, 48 (1996) hep-th/9512145.

[47] A. Giveon, D. Kutasov and O. Pelc, "Holography for non-critical superstrings," JHEP **9910**, 035 (1999) hep-th/9907178.

[48] A. Klemm, W. Lerche, P. Mayr, C. Vafa and N. P. Warner, "Self-Dual Strings and N=2 Supersymmetric Field Theory," Nucl. Phys. B **477**, 746 (1996) [hep-th/9604034].

[49] O. Pelc, "Holography, singularities on orbifolds and 4D N = 2 SQCD," JHEP **0003**, 012 (2000) [hep-th/0001054].

[50] T. Eguchi and Y. Sugawara, "Modular invariance in superstring on Calabi-Yau n-fold with A-D-E singularity," Nucl. Phys. B **577**, 3 (2000) [hep-th/0002100].

[51] S. Mizoguchi, "Modular invariant critical superstrings on four-dimensional Minkowski space x two-dimensional black hole," JHEP **0004**, 014 (2000) [hep-th/0003053]; "Noncompact Gepner models for type II strings on a conifold and an ALE instanton," hep-th/0009240.

[52] S. Yamaguchi, "Gepner-like description of a string theory on a non-compact singular Calabi-Yau manifold," Nucl. Phys. B **594**, 190 (2001) [hep-th/0007069].

[53] M. Naka and M. Nozaki, "Singular Calabi-Yau manifolds and ADE classification of CFTs," Nucl. Phys. B **599**, 334 (2001) [hep-th/0010002].

[54] K. Ohta and T. Yokono, "Linear dilaton background and fully localized intersecting five-branes," Phys. Rev. D **63**, 105011 (2001) [hep-th/0012030].

[55] D. E. Diaconescu and J. Gomis, "Neveu-Schwarz five-branes at orbifold singularities and holography," Nucl. Phys. B **548**, 258 (1999) [hep-th/9810132].

[56] D. Kutasov, F. Larsen and R. G. Leigh, "String theory in magnetic monopole backgrounds," Nucl. Phys. B **550**, 183 (1999) [hep-th/9812027].

[57] M. Gremm and A. Kapustin, "Heterotic little string theories and holography," JHEP **9911**, 018 (1999) [hep-th/9907210].

[58] E. Witten, "Solutions of four-dimensional field theories via M-theory," Nucl. Phys. B **500**, 3 (1997) [hep-th/9703166].

[59] S. Elitzur, A. Giveon and D. Kutasov, "Branes and N = 1 duality in string theory," Phys. Lett. B **400**, 269 (1997) [hep-th/9702014].

[60] A. B. Zamolodchikov and A. B. Zamolodchikov, "Liouville field theory on a pseudosphere," hep-th/0101152.

[61] A. Giveon, D. Kutasov and A. Schwimmer, "Comments on D-branes in AdS(3)," Nucl. Phys. B **615**, 133 (2001) [hep-th/0106005].

[62] S. Elitzur, A. Giveon, D. Kutasov, E. Rabinovici and G. Sarkissian, "D-branes in the background of NS fivebranes," JHEP **0008**, 046 (2000) [hep-th/0005052].

[63] P. Ginsparg and G. W. Moore, "Lectures On 2-D Gravity And 2-D String Theory," hep-th/9304011.

[64] V. Kazakov, I. K. Kostov and D. Kutasov, "A matrix model for the two-dimensional black hole," hep-th/0101011.

[65] J. A. Harvey, D. Kutasov, E. J. Martinec and G. Moore, "Localized tachyons and RG flows," hep-th/0111154.

Lectures on D-Branes on Calabi-Yau Manifolds

Michael R. Douglas*

Department of Physics and Astronomy, Rutgers University, Piscataway, USA
and
I.H.E.S., Le Bois-Marie, Bures-sur-Yvette, France

Lectures given at the
Spring School on Superstrings and Related Matters
Trieste, 2 – 10 April 2001

LNS027005

*mrd@physics.rutgers.edu

Abstract

This is a write-up of lectures given at the 2001 Spring School at the Abdus Salam ICTP.

Contents

1 Introduction

Dirichlet branes play many important roles in the modern discussion of superstring compactification and duality. They provide a very general way to embed gauge theories into string theory, which has led to remarkable physical conjectures such as M(atrix) theory and AdS/CFT. They have also led to remarkably detailed connections between physics and mathematics, such as a rederivation of the ADHM construction of instantons.

In these lectures, we will give some introduction to the problem of finding BPS D-branes in weakly coupled type II string compactification on Calabi-Yau manifolds. This problem is prototypical for the case of $\mathcal{N} = 1$ supersymmetry in four dimensions, and as such has received a lot of study, especially in the special case of non-compact Calabi-Yaus such as resolved orbifolds. Almost all of the ideas apply to compact Calabi-Yaus as well, and although these examples are not as well understood at present, no fundamental barrier has been found to progress in this direction, and this could lead to a much more complete understanding of $\mathcal{N} = 1$ compactification than we now have. This is the general direction our lectures will head in (though we won't get very far).

Rather than start from geometry of Calabi-Yau, we start from general principles which apply to all D-brane problems, then specialize to the case of $\mathcal{N} = 1$ supersymmetry on the world-volume, and finally to special features of the Calabi-Yau case.

As with most discussions of string compactification, unless one restricts attention to very special regions in moduli space where the underlying conformal field theory is exactly solvable, which we do not want to do, making a proper discussion involves a good deal of mathematics which is unfamiliar to most physicists. In these lectures, we will discuss a bit of the theory of holomorphic bundles – since the gauge fields on B-type BPS branes live in such bundles, this is obviously relevant – and that of coherent sheaves, a generalization which describes singular limits of bundles.

We will not start with this however, but rather with the mathematics which underlies quiver theories, namely homological algebra and category theory. This is even less familiar, but turns out to be very well motivated in these problems. Let us start out by giving the basic dictionary, to explain why this is.

A category is defined mathematically as a set of *objects*, and for each pair of objects A and B a set of *morphisms* from A to B. There is a multiplication law: given a morphism from A to B and one from B to C, the product is a morphism from A to C.

This data satisfies certain axioms, which are exactly the general properties we want for each open string boundary condition to be an object, and each morphism to be an open string. The multiplication law is just the operator product expansion between open string vertex operators.

There is much more structure in the physical problem than the open string spectrum of course. One needs to distinguish "matter" and "gauge" vertex operators; one needs to interpret the superpotential and D-flatness conditions; and so on. It turns out that essentially all of the structure of the low energy world-volume theory has known mathematical counterparts. We will explain some of this as we go along.

What saves us from having to do this completely abstractly is that all of this structure is already visible in quiver gauge theories. Indeed, the mathematical notion of "quiver" was defined as a particularly simple source of algebras and categories, long before any physical applications emerged. Physically, these theories contain the minimal structure required to discuss the following problem: one takes a basic finite set of "generating objects" and all morphisms between these, and tries to form all the BPS branes as bound states.

This project leads to many further questions. How many branes do we need, and how do we find such a set? Suppose we start with a brane defined using a different construction: can we find some canonical way to decompose this brane into these fundamental constituents? Are there natural symmetries of the spectrum which map between different generating sets? As we discuss, paths in Calabi-Yau moduli space are associated with monodromies which should produce such symmetries. Of course we also need to discuss marginal stability and the variation of the spectrum in this context. Finally, is this a useful way to consider the problem? Can we compute the spectrum, moduli spaces and so forth this way?

Although we will not have time to go into it deeply, there is a a powerful underlying concept which simplifies all of this further discussion, the derived category. In physical terms, this is a structure derived from the original category of boundary conditions, which describes arbitrary bound states of branes and antibranes, and keeps track of everything which does not depend on the precise identification of which are branes and which are antibranes.

The derived category enters at many points in the discussion. For example, it turns out that for string size Calabi-Yau, not all branes are bundles or coherent sheaves; some are more general bound states of branes and antibranes which correspond to "non-classical" objects in the derived category. As a second example, the relationship between quiver theories describing different sets of generating branes can be understood in terms of "Fourier-Mukai transformations," which can be simply formulated as acting on the derived category, and which reduce for concrete quiver theories to "Seiberg dualities" between these theories. Although we will not be able to get into details of this (many of which are presently under investigation), the concepts we discuss should serve as a good introduction to these directions.

2 Topological considerations

D-branes carry Ramond-Ramond charge and this is the most obvious topological classification we can make. If we consider a brane B wrapping an arbitrary cycle Σ in the internal space M, and which looks like a D-particle in $3+1$ dimensions, it will carry electric and magnetic charges under the $3+1$ $U(1)^r$ gauge group of the bulk theory. Although the considerations we discuss now are more general, let us assume M is a CY_3. Then, if we start with IIa theory, we will have $r = b_{1,1}+1$ coming from odd rank potentials, while if we start with IIb it will have $r = b_{2,1}+1$.

A simpler topological invariant which can be derived from these charges is the "intersection form," whose simplest physical definition is as the integer appearing in the Dirac-Schwinger-Zwanziger charge quantization condition

$$\langle B_1, B_2 \rangle = e_1 \cdot m_2 - e_2 \cdot m_1 \tag{1}$$

in an appropriate basis. In the case of 3-branes, this is entirely geometric and counts the signed intersection number of the two cycles Σ_1 and Σ_2. Poincaré duality then tells us that there exists a basis for $H_3(M,\mathbb{Z})$ which makes this form unimodular (determinant 1). More generally it would pair $H_p(M,\mathbb{Z})$ with $H_{n-p}(M,\mathbb{Z})$ to produce a unimodular form; this is sometimes referred to as a "perfect pairing."

For $2p$-branes, computing the intersection form in this definition requires using the general formula for RR charges of branes carrying gauge fields,

$$\int C \wedge \mathrm{Tr} e^F \wedge \sqrt{\hat{A}(M)}.$$

The DSZ term then becomes

$$\langle B_1, B_2 \rangle = \int \mathrm{Tr} e_1^F \ \mathrm{Tr} e^{-F_2} \hat{A}(M). \tag{2}$$

This is exactly the index of the Dirac operator $\not{D}$ for bifundamental fermions coupled to the gauge field $-A_1 \otimes 1 + 1 \otimes A_2$. This is no coincidence but can be derived from stringy considerations. The basic point is that $\langle B_1, B_2 \rangle$ can be computed from a string annulus diagram with all boundary conditions taken to be Ramond, leading in the closed string channel to the part of the RR closed string exchange proportional to the Levi-Civita symbol ϵ, and in the open string channel to the index

$$\langle B_1, B_2 \rangle = \mathrm{Tr}_{B_1,B_2}(-1)^F$$

which of course is equal to (2).

Thus, the intersection form also counts the massless fermion content of the combined world-volume theory of the B_1 and B_2 branes. Let the number of fermions with charges $(-1,+1)$ under the $U(1)$ gauge groups of the two branes and four dimensional left and right chiralities be n_L and n_R, then $n_L - n_R = \langle B_1, B_2 \rangle$. We could also write all the fermions as left chirality of course by complex conjugating the right chirality.

This quantity is also the natural definition of intersection form in K theory. We will not talk specifically about K theory very much as it will be subsumed in the derived category framework we will develop later. The simplest argument for the relevance of K theory to topological classification of branes however is short and well worth keeping in mind. It is simply that a brane B should be identified topologically with anything one can get by adding another brane X, its antibrane $\bar{X}$, and performing any continuous variations on this configuration. This can be expressed mathematically by a simple construction: let a class in the K theory of "branes" be a pair of branes (E, F) subject to the relation $(E, F) \cong (E \oplus X, F \oplus X)$ for all X. This uses very little structure of the branes and indeed the objects under discussion could be almost anything to make this definition; we just need to know how to take direct sums such as $E \oplus X$ and decide when two direct sums produce the same object. Now we know many branes in the large volume limit, namely those which wrap the entire space M and carry arbitrary vector bundles on M, and it is plausible that these already carry all the topological charges, leading to the classification by K theory of vector bundles.

At this writing, it has not really been proven that this is the right classification. The issue is the torsion part, classes $[X]$ which satisfy $n[X] \cong 0$ for some finite n. The rest of the K theory agrees with cohomology and thus the RR charge considerations we started with, but one might imagine that the true answer in some string theory example might not distinguish these torsion classes, or might make more distinctions, presumably associated with singularities which would not be governed by the large volume Yang-Mills equations. Nevertheless the classification by K theory of vector bundles works in all examples studied to date and seems to fit well with our general understanding of string theory.

To get a simple example with torsion, consider a Calabi-Yau M with $\pi_1(M) = G$ some finite group. One might well expect that a string wound around a nontrivial element of $\pi_1(M)$ would be topologically stable, and use this to construct new topological classes of D1-branes. It can be shown that $H^2(M,\mathbb{Z}) \cong H_1(M,\mathbb{Z}) = G/[G,G]$ the abelianization of G and that this appears in $K^0(M)$, so if G is abelian this works. (The story if G is not abelian is less clear).

Any such M can be obtained as a quotient of a simply connected $\tilde{M}$ by a free action of a symmetry group G, so this particular type of torsion can probably be understood by close study of the theory of branes on $\tilde{M}$. We refer to [6, 7] for examples of this. It is not known whether other types of torsion which cannot be understood this way exist in $K(M)$ for M Calabi-Yau; of course other K groups appropriate to type I theories, orientifolding, H fields and so on will generally have other types of torsion.

We move on however and assume that M is a simply connected CY_3, in which case one can prove that $K^*(M) \cong H^*(M,\mathbb{Z})$, and all of the topological information is summarized in the intersection form. All of our considerations would still hold in the presence of torsion, we would just have further conserved quantum numbers which we would not be making explicit.

2.1 Noncompact manifolds

We need to generalize the previous discussion to handle noncompact manifolds such as the local orbifolds we will discuss below. Although all of the same definitions of intersection form can be used, Poincaré duality takes a different form: it relates the homology $H_p(M,\mathbb{Z})$ to the homology with compact support $H_{n-p}(M,\mathbb{Z})$, and provides a perfect pairing between these.

We will use this below, but we still would like a way to decide whether we have a complete basis for the charge lattice. We will be most interested in BPS branes of finite energy which wrap cycles of compact support, so we want a pairing purely in this sector.

We will do this below by using the pairing provided by the index (2) on the compact manifold of interest. This is also an intersection form in a mathematical sense but it should be distinguished from (1) as it is not symmetric or antisymmetric. Physically, we are dropping certain degrees of freedom (fermions partner to normal deformations of the brane) in this definition. However it still serves our purpose as if this pairing is unimodular, one has a complete basis for the K theory of the compact space. It will also turn out that this index will provide much more information about the fermion spectrum than we would get if we restricted attention to (1).

3 Constraints on brane world-volume theories

A good way to study the dynamics of a collection of branes is to derive their effective world-volume theory, which includes only modes which are visible at the energy scales of interest. For analyzing the vacuum structure this means only modes which can become massless somewhere on the moduli space.

We restrict attention in these lectures to weakly coupled type II theory, i.e. we define our effective Lagrangians using only sphere and disk world-sheets, and treat the resulting world-volume theories as classical, solving equations of motion to find vacua.

For D-branes, we can think of this world-volume theory as derived by starting with a complete open string field theory, and then keeping only the potentially massless modes. We could of course define it using world-sheet conformal field theory instead, and we would only be keeping boundary couplings which can become relevant or marginal.

There are some obvious constraints which are already visible at this level. First, the world-volume theory has only a trivial dependence on the number of dimensions in which the brane extends in the flat Minkowski dimensions. If we derive it for branes filling these dimensions, the lower dimensional cases can be defined by trivial Kaluza-Klein reduction, taking fields to be constant in the dimensions we reduce. Components of the vector potential will become coordinates in these dimensions.

The generating branes will each have gauge group at least $U(1)$ and it is natural to restrict attention to those with gauge group $U(1)$, because a world-volume theory with a larger unbroken gauge symmetry (in weakly coupled type II theory it must be $U(N)$) does not really describe a single brane. If we consider D-particles, such a theory will have moduli which enable us to split the brane into N constituents at different positions in Minkowski space, and it is better to consider these as generating branes instead. Thus we define a *simple* configuration of a gauge theory to be one which breaks the gauge symmetry to be $U(1)$ and take the B_i to be simple. Furthermore, we only need to classify the simple branes, as the others are just direct sums of these configurations.

A theory with N_i copies of the elementary brane B_i will then have gauge group

$$G = \prod_i U(N_i),$$

by the usual Chan-Paton arguments. We also know that the matter fields will all transform in the adjoint or bifundamentals of the gauge groups, and the action (computed from the disk world-sheet) can be written as a single trace.

The matter content is constrained by the index theorem arguments discussed above. Let n_{ij} be the number of massless fermions with charge $(\bar{N}_i, N_j)$; we then have

$$n_{ij} - n_{ji} = \langle B_i, B_j \rangle.$$

3.1 Constraints from $\mathcal{N} = 1$ supersymmetry

To go any farther we need some statement about the spectrum of scalar fields. We will now make the important simplifying assumption that the combined world-volume theory has $\mathcal{N} = 1$ supersymmetry in $d = 4$ (so, four supercharges). This is the supersymmetry preserved by a BPS brane in CY_3 compactification and thus this assumption would seem very natural in the problem of classifying BPS branes. On further reflection, however, it is not at all obvious, because typically the generating branes B_i will each preserve a different $\mathcal{N} = 1$ subalgebra contained in the bulk $\mathcal{N} = 2$ supersymmetry, and thus the combination will break all supersymmetry. This does not invalidate the assumption because we can still claim that this configuration, which in particular has zero vev for any open string modes associated to pairs of branes, has spontaneously broken an underlying $\mathcal{N} = 1$ supersymmetry visible in the ground state. This is what we will implicitly be claiming and will verify in examples, but eventually we will find that this picture cannot always be taken literally, and will have to generalize this assumption. Nevertheless it is close enough to the truth to justify devoting a good deal of attention to the particularities of the $\mathcal{N} = 1$ case.

In an $\mathcal{N} = 1$ supersymmetric Lagrangian, the massless fermion content determines the massless field content. Massless fields can be in either chiral or vector multiplets. A bifundamental must be a chiral multiplet, while an adjoint could be either, but we know that each set of N_i generating branes will have exactly the gauge bosons of $U(N_i)$ and no more, so knowing just the integers n_{ij}, the number of massless fermions of each chirality, completely determines the field content. (If one explicitly derives the world-volume theory from world-sheet considerations, there is no difficulty in identifying the fermions which are the gauginos: their vertex operator is just the spectral flow operator, as we discuss later.) The index however is not enough information.

This data can be conveniently summarized in a "quiver diagram," in which we denote each gauge group with a node, and each bifundamental chiral multiplet with an arrow. Thus we obtain an oriented graph with n_{ij} links between nodes i and j, each representing a chiral matter field $X^a_{i,j}$.

This graph summarizes an infinite set of supersymmetric gauge theories, distinguished by the choice of a rank N_i for each gauge group. Let $V_i \cong \mathbb{C}^{N_i}$ carry the fundamental representation of the group $U(N_i)$, then $X^a_{i,j}$ is in $V_i^* \otimes V_j$.

This is a lot of information already and the remaining data we need to know to find supersymmetric vacua is the superpotential W, the D-flatness conditions, and some qualitative information about the Kähler potential, say that it is nonsingular on the moduli space.

The superpotential is a gauge invariant function of chiral superfields which can be written as a single trace. It can be written as a sum of monomials, each of which could be denoted by a closed loop in the quiver. A supersymmetric vacuum must satisfy the F-flatness conditions

$$0 = F_a = \frac{\partial W}{\partial X^a}, \tag{3}$$

a matrix equation for each chiral superfield X^a.

The D-flatness conditions are largely determined by the spectrum and gauge representations, but there is one further input: each $U(N_i)$ factor in the gauge group (so each node) can come with a single real parameter, the Fayet-Iliopoulos term ζ_i. A supersymmetric vacuum must then satisfy the D-flatness conditions,

$$0 = D_i = \sum_{a=1}^{n_{ij}} (X^a)^\dagger X^a - \sum_{a=1}^{n_{ji}} X^a (X^a)^\dagger - \zeta_i \mathbf{1}. \tag{4}$$

Supersymmetric vacua are gauge equivalence classes of solutions of these two sets of equations, and we will be interested in the general problem of finding such vacua which break the gauge symmetry to $U(1)$.

There is a variation on the D-flatness condition which is relevant for our D-brane problems coming from the fact that D-brane world-volume theories have an additional inhomogeneous $\mathcal{N} = 1$ supersymmetry, the shift $\delta\chi = \epsilon'$ of the decoupled gaugino in the diagonal $U(1)$ factor of the gauge group. We need to allow for vacua which break the linearly realized supersymmetry but preserve some combination of the two as well.

Supersymmetry breaking by D terms shows up in an inhomogeneous transformation law for the gaugino. Adding to this the overall inhomogeneous supersymmetry, we have

$$\delta\chi^a = D^a\epsilon + \epsilon',$$

and we see that these more general supersymmetric vacua can be found by the prescription of solving the D-flatness conditions with an overall constant shift $\zeta_a \to \zeta_a + \xi$ of all of the FI terms.

3.2 Finding supersymmetric vacua

An effective way to think about this problem is to first classify solutions of F-flatness modulo complex gauge equivalence, and then check which of these solutions can solve D-flatness as well. The complexified gauge group is

$$G_{\mathbb{C}} = \prod_i GL(N_i, \mathbb{C}),$$

and it acts on a bifundamental as

$$X_{i,j} \to g_i^{-1} X_{i,j} g_j. \tag{5}$$

For general (nonunitary) g, this is a symmetry of the holomorphic part of the theory (the F-flatness conditions in particular) but not of the D-flatness conditions. Thus, any solution of F-flatness in fact comes with an entire $G_{\mathbb{C}}$-orbit of solutions, and in this first stage of the problem it is not natural to distinguish the points on a given orbit. One can then try to find a g_i in (5) which solves (4).

A major advantage of this two-step procedure is that the second step is very well understood mathematically, and indeed we will be able to quote a general theorem which tells us precisely when solutions of D-flatness do and do not exist. As motivation, let us review two very familiar examples.

First, consider $U(1)$ theory with n chiral multiplets z^i of charge $+1$. The D-flatness condition is

$$\sum_i |z^i|^2 = \zeta. \tag{6}$$

Clearly there are three cases: for $\zeta > 0$ there are solutions whose moduli space is $\mathbb{CP}^{n-1}$; for $\zeta = 0$ there is a unique solution $z = 0$, while for $\zeta < 0$ there are no solutions. This exhibits the fact that the moduli space of solutions of D-flatness will depend on the specific values of the FI parameters and can even disappear. It also illustrates the fact that not every $G_{\mathbb{C}}$-orbit will contain a solution of D-flatness. In the first case, the orbit $z^i = 0$ cannot solve (6) while all the others can; in the second case the situation is reversed, while in the final case of course none will.

Second, consider $U(N)$ theory with a single adjoint chiral superfield X. In this case one cannot usefully introduce an FI term, so the D-flatness condition is

$$[X^\dagger, X] = 0.$$

A matrix satisfying this equation is referred to as "normal" and it can be diagonalized; the moduli space is the space of sets of N eigenvalues x_i (the ordering does not matter thanks to a remaining S_N discrete subgroup of the gauge symmetry), so is $\mathbb{C}^N/S_N$.

Let us again compare with the $G_{\mathbb{C}}$ orbits. This includes the normal matrices but also matrices which cannot be diagonalized, such as

$$\begin{pmatrix} x_1 & 1 \\ 0 & x_2 \end{pmatrix}$$

to give the simplest example. Such matrices cannot solve the D-flatness conditions and thus our previous result for the moduli space was correct, but we still can ask: what distinguishes these $G_{\mathbb{C}}$ orbits from the ones which can solve D-flatness?

There is an answer to this question which is fairly well known by physicists but only applies to the case of zero FI terms. It is that the moduli space of solutions of D-flatness is parameterized by a complete set of independent holomorphic gauge invariant polynomials formed from the original fields. In the matrix example, these could be taken to be $\mathrm{Tr}X^k$ for $1 \leq k \leq N$. Clearly these suffice to distinguish different sets of eigenvalues, but they are not a good system of coordinates to describe all matrices up to gauge equivalence, as any non-diagonalizable matrix will have the same invariants as some diagonalizable matrix. On the other hand, the non-diagonalizable matrices never solve the D-flatness conditions, so these are a good system of coordinates to describe these solutions.

This connection between invariants and solutions of D-flatness is very general and provides a very satisfactory description of the moduli space, but only for the case of zero FI terms. For example, there is no obvious way to adapt it to the first problem as this theory admits no holomorphic gauge invariant observables.

There is a more general solution, which will turn out to have a fairly clear picture in terms of the physics of branes, but explaining this will require some additional formalism.

4 Pure quiver theories

Let us spend some time discussing the case with no superpotential first. There is quite a bit to say, and making careful definitions here will in fact carry us much of the way to the final results.

We can make the main points by considering the theory of two elementary branes B_1 and B_2 whose open string spectrum contains q bifundamental chiral multiplets X^i. This is described by the diagram

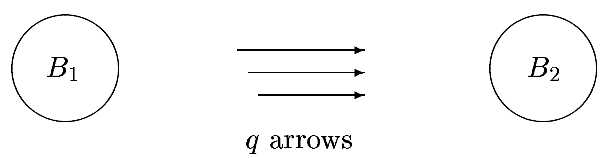


Fig.1. $U(N_1) \times U(N_2)$ quiver

Almost all of what we will say in this section generalizes to the general quiver theory with $W = 0$. In this case we will refer to matter fields between nodes i and j as $X_{i,j}$. Some of this generalizes directly to arbitrary superpotential, and we will say so when it does.

Let us consider a configuration with $N_1B_1 + N_2B_2$ elementary branes, and suppose that all such configurations are described by this $U(N_1)\times U(N_2)$ world-volume gauge theory. Each configuration of the q bifundamental chiral multiplets, modulo complex gauge equivalence, will provide a physically distinct bound state, if it solves the D-flatness conditions. Now even before we consider D-flatness, it is clear that each such configuration represents at most one physical state. Of course it might not be a single bound state; this will be the case only if the gauge symmetry is broken to $U(1)$.

Let us refer to a configuration as an "object" (or holomorphic object) in the category of quiver representations. One which breaks the complex gauge symmetry to $GL(1)$ is a simple object.

We now ask whether this quiver theory contains any simple objects, and if so what is the dimension of their moduli space. Let us refer to such a theory using the notation $(\vec{N})$ or $(N_1\ N_2)$; the moduli space of simple configurations will be $\mathcal{M}(\vec{N})$, while the vector $\vec{N}$ will be called the "charge" of the object. The two elementary branes are $(1\ 0)$ and $(0\ 1)$; let us denote these charge vectors as e_1 and e_2.

There is an obvious guess for the dimension of this moduli space, obtained by counting matter fields minus the number of broken gauge symmetries, and assuming that the resulting object is simple:

$$\dim \mathcal{M}(\vec{N}) = qN_1N_2 - N_1^2 - N_2^2 + 1.$$

One would expect that if this "expected dimension" $\dim \mathcal{M} \geq 0$, there will exist simple configurations and that their moduli space will have this dimension, while if $\dim \mathcal{M} < 0$ there will not exist simple configurations. This is true, although the mathematical proof of this fact is surprisingly complicated. A similar result holds for general quivers with $W = 0$, quoted in the appendix to [14].

We can also write the result in terms of a "Cartan matrix":

$$\dim \mathcal{M}(\vec{N}) = 1 - \tfrac{1}{2}\begin{pmatrix} N_1 & N_2 \end{pmatrix}\begin{pmatrix} 2 & -q \\ -q & 2 \end{pmatrix}\begin{pmatrix} N_1 \\ N_2 \end{pmatrix} \tag{7}$$

$$\equiv 1 - \tfrac{1}{2}(N|N). \tag{8}$$

In this language, we can distinguish three types of vector $\vec{N}$:

- Real roots with $(N|N) = 2$, which are "rigid" configurations with no moduli.

- Imaginary roots with $(N|N) \leq 0$, which are configurations with moduli.

- Vectors with $(N|N) > 2$ which do not correspond to simple configurations.

Using results which are probably familiar from Lie algebra theory, we can see that the general nature of the solutions depends on q in the following way:

- For $q = 0, 1$ there are a finite number of real roots and no imaginary roots, so this is the finite case.
- For $q = 2$, the real roots are $(n+1\ n)$ and $(n\ n+1)$, and the imaginary roots are $(n\ n)$. This is the affine case.
- For $q \geq 3$, the hyperbolic case, there are infinitely many roots of both types.

This last case is probably less familiar although as the reader may have guessed, all three cases indeed admit a relation to Kac-Moody algebras. Let us discuss it a bit more. The symmetries of a root system are generated by Weyl reflections, which act as follows:

$$r_i : \vec{N} \to \vec{N} - 2\frac{(N|e_i)}{(e_i|e_i)}e_i.$$

In the hyperbolic and affine cases, these reflections generate infinite discrete groups (which are cyclic in this simple case).

The condition $(N|N) \leq 0$ defines a region in the $(N_1\ N_2)$ plane in which all the points are imaginary roots. It contains infinitely many copies of a fundamental region, defined by the condition $r_i(\vec{N}) \geq \vec{N}$.

The real roots can all be obtained by Weyl reflection from the elementary roots, and thus form an infinite series which for $q = 3$ starts $(0\ 1), (1\ 3), (3\ 8), \ldots$.

Thus even for these very simple theories, the spectrum of branes has quite a bit of structure. One can even define explicit operations corresponding to the Weyl reflections which take one object into another. This is also discussed in the appendix to [14], and underlies the more complicated Seiberg dualities discussed in [4, 18, 9]. This structure is known to also be present in the quantum mechanical treatment of such branes for the finite and affine cases, and very likely is for the hyperbolic case.

We will say more about this structure and the nature of the Weyl reflections below, after introducing some more formalism.

4.1 Bound states and Ext

A question of primary interest for us is the following: given two objects A and A', when can they form a bound state?

That this question is nontrivial can be seen by considering the example of bound states of (0 1) and (1 n). Starting with (1 0), one can add successive (0 1)'s until one reaches (1 q). However (1 $q+1$) is clearly not a simple object as there are not enough matter fields to break $U(q+1)$ gauge symmetry. We would like a rule which tells us when this can happen, ideally depending only on the charges N and N'.

Clearly the answer to this question can be found by studying the $U(N_1 + N_1') \times U(N_2 + N_2')$ gauge theory which describes the combination of their constituents. Its matter will decompose in a block diagonal way:

$$X^a = \begin{pmatrix} X^a & \rho^a \\ \psi^a & (X')^a \end{pmatrix}. \tag{9}$$

If we can turn on matter fields ρ^a or ψ^a in a way which breaks the total gauge symmetry back to $U(1)$, these two objects will form a bound state. In other words, if after using all possible gauge symmetry to set components of ρ or ψ to zero, we are left with any nonzero components, we will find a bound state.

Let us consider only turning on ρ^a as we can then repeat the discussion, exchanging the two objects, to treat ψ^a. The gauge symmetries (5) can also be decomposed in block diagonal form, and the relevant parameters which act on X and can modify ρ are

$$g_1 = \begin{pmatrix} 1 & \epsilon_1 \\ 0 & 1 \end{pmatrix}; \qquad g_2^{-1} = \begin{pmatrix} 1 & -\epsilon_2 \\ 0 & 1 \end{pmatrix}.$$

The resulting gauge action is then

$$\delta\rho^a = X^a \epsilon_1 - \epsilon_2 (X')^a. \tag{10}$$

Note that the g_i are finite (not infinitesimal) complexified gauge transformations; in this sense this is not a linearized result but is exact. This also generalizes in the obvious way to any matter field $X_{i,j}$ and its off-diagonal part $\rho_{i,j}$ in any quiver.

The result can be seen much more quickly in terms of the following diagram:

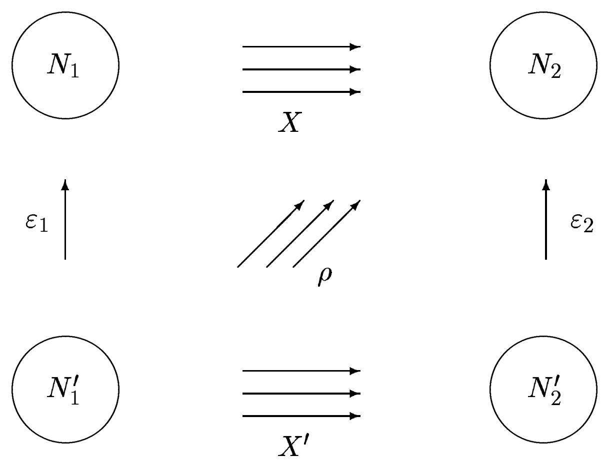


Fig.2. Combining two bound states

The vertical lines represent the off-diagonal degrees of freedom, and sensible (gauge invariant) products must respect the graphical structure.

Thus matter variations which cannot be gauged away are variations ρ^a which cannot be obtained from (10). These form a linear space which is denoted by

$$\mathrm{Ext}(A', A).$$

We are also interested in gauge transformations which are unbroken in the combined configuration, i.e. solutions to the equation $\delta\rho^a = 0$. These form another linear space which is denoted by

$$\mathrm{Hom}(A', A).$$

Note that some of the ranks N_i or N_i' might vanish in a particular example. The definitions still make sense if we just omit those gauge transformations or matter multiplets for which one of the ranks is zero. Often in these cases, the equation (10) will also degenerate; this is fine. One should check that this is clear for the simple examples

$$\dim \mathrm{Hom}(B_i, B_j) = \delta_{i,j} \tag{11}$$

$$\dim \mathrm{Ext}(B_i, B_j) = q\delta_{i,1}\delta_{j,2}. \tag{12}$$

All of these definitions generalize in a direct way to arbitrary quivers with $W = 0$. We will generalize them in the next sections to certain quiver theories with superpotentials as well, and they are quite important in all of the subsequent discussion.

The spaces Hom and Ext are examples of the spaces of morphisms associated with pairs of objects, which as discussed in the introduction define a category. The multiplication law is just the obvious composition of arrows and multiplication of the matrices associated with each arrow. This structure provides an obvious generalization of representations of groups, and it is in this spirit that quivers were first introduced in mathematics [20].

One can regard the equation (10) as defining a linear operator acting on the space of parameters ϵ_i and producing a configuration in the space of matter fields ρ^a, whose matrix elements depend on the configuration X and X'. Let us denote this operator as D, we then have

$$\mathrm{Hom}(A', A) = \ker D; \qquad \mathrm{Ext}(A', A) = \text{coker } D.$$

and we are studying a cohomology problem. This formalizes the observation that, although the dimensions of these two spaces depend on the specific configuration, they can only change in a paired way, with an element disappearing or appearing on both sides, corresponding to the Higgsing or unHiggsing of an off-diagonal gauge boson.

This also shows that it is easy to compute the difference in the dimensions of these two spaces, since we can just do it for $X = 0$, by again counting multiplets. This number is called the relative Euler character; it is

$$\chi(A', A) = \dim \mathrm{Hom}(A', A) - \dim \mathrm{Ext}(A', A) \tag{13}$$

$$= \begin{pmatrix} N_1' & N_2' \end{pmatrix} \begin{pmatrix} 1 & -q \\ 0 & 1 \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \end{pmatrix}. \tag{14}$$

It contains the information both of the intersection form and of the Cartan matrix,

$$\langle A|B\rangle = \chi(A, B) - \chi(B, A) \tag{15}$$

$$(A|B) = \chi(A, B) + \chi(B, A). \tag{16}$$

Although this does not directly determine either $\dim \mathrm{Hom}(A, B)$ or $\dim \mathrm{Ext}(A, B)$, it will give a lower bound for one of them, so given only the charges of the two objects we can often prove that one of dim Hom or dim Ext is nonzero. It is harder to prove that one of these is zero, although in typical examples where A and B are simple objects, one of these dimensions will in fact be zero.

This provides a procedure to decide whether two objects A and B can form a bound state, and a formal construct which is associated with each way of forming a bound state. There is a larger structure which this fits into: any $\mathrm{Ext}(A, B)$ and bound state E produced by turning it on will have an associated exact sequence

$$0 \longrightarrow B \xrightarrow{f} E \xrightarrow{g} A \longrightarrow 0. \tag{17}$$

By an exact sequence one means first of all that $g \cdot f = 0$ defined by composing these pairs of matrices. Furthermore, the terms $0 \longrightarrow B$ and $A \longrightarrow 0$ at the beginning and end of this sequence indicate that the map f must be injective (with no kernel), and g must be surjective (with no cokernel). In other words, E must incorporate all constituents of A and B, with nothing left over.

The object B is called a subobject of E, while A is a quotient object. Physically the Hom's represent the possibility of seeing that these two objects are contained in E by bringing up either one "next to it;" an enhanced complex gauge symmetry appears.

A simple example is provided by the bound states of the two elementary branes $B_1 = (1\ 0)$ and $B_2 = (0\ 1)$. Bound states with charge $(1\ 1)$ exist with a moduli space of dimension $q - 1$. Call one of these E; it will fit into the exact sequence

$$0 \longrightarrow (0\ 1) \xrightarrow{f} E \xrightarrow{g} (1\ 0) \longrightarrow 0$$

where $f \in \mathrm{Hom}(B_2, E)$ and $g \in \mathrm{Hom}(E, B_1)$ are easy to write down using our general definitions (exercise).

We can even write a "triangle" which completes the structure as follows: given $\phi \in \mathrm{Ext}(A, B)$, we have

$$B \xrightarrow{f} E \xrightarrow{g} A \xrightarrow{\phi} B \tag{18}$$

where $f \cdot \phi \cdot g$ is an $\mathrm{Ext}(E, E)$ we can vary in the resulting bound state. We will better define and use this structure later. It is present in general theories of D-branes on Calabi-Yau (and probably even more generally).

So far we have only talked about branes, and not their antibranes. The exact sequence (17) can also be interpreted as describing certain processes involving antibranes, namely the inverse to the bound state formation we just described. For example, given that A and B can form the bound state E, we might expect that E and $\bar{B}$ could partially annihilate to produce A, and also that $E + \bar{A} \to B$, and indeed these are all valid readings of (17). This does not get us too far on the problem of describing bound states of branes with antibranes, however, as all three of the objects involved are each made only from elementary branes or only from their antibranes. In general we will need to talk about bound states of elementary branes with antibranes, but this will require formalism we discuss later.

5 D-flatness and stability

In this section we will explain the general result promised earlier on D-flatness conditions; it applies to general quiver theories with or without superpotentials. The structure we described in the previous section will play an essential role.

As we saw, the question of whether an object really corresponds to a physical brane, i.e. solves the D-flatness conditions, can depend on the FI terms. This is how marginal stability will appear in the physical theories of branes on CY. Conversely, one might imagine that if by varying an FI term, an object becomes physically unstable, it will have to decay into constituents described by an exact sequence of the type we just discussed.

We can study the question of whether, given an $\mathrm{Ext}(A', A)$ and exact sequence (17), these two branes can actually form a physical bound state, by again considering the direct sum $U(N_1 + N_1') \times U(N_2 + N_2')$ gauge theory. Now we want to write down the D-term part of the potential $V = \frac{1}{2}\mathrm{tr}\, D^2$ for the mode ρ. Taking D from (4), we have

$$D_1 = -\sum_{a=1}^{q} X^a (X^a)^\dagger - \zeta_1 \tag{19}$$

$$D_2 = \sum_{a=1}^{q} (X^a)^\dagger X^a - \zeta_2. \tag{20}$$

We first note that there is a rather trivial but necessary condition for solving the D-flatness conditions, obtained by taking the trace of both and adding: one finds

$$0 = \sum_i (N_i + N_i')\zeta_i \equiv (\vec{N} + \vec{N}') \cdot \vec{\zeta}. \tag{21}$$

In this form it clearly holds for arbitrary quiver theories. We will further assume that $\vec{N} \cdot \vec{\zeta} = \vec{N'} \cdot \vec{\zeta} = 0$, so that both initial and final states solve the D-flatness conditions with the same FI terms. We will see below that this is the special case where all three branes preserve the same supersymmetry. The same analysis can be made without this assumption; see [26].

We now compute the quadratic term in V for the mode ρ, assuming that $D_i = 0$ before we turn it on. Substituting (9) into (19) we have

$$D_1 = \begin{pmatrix} \rho\rho^\dagger & \rho(X')^\dagger \\ X'\rho^\dagger & 0 \end{pmatrix}.$$

Using $X'X'^\dagger = \zeta_1$, and taking the trace, the quadratic term in $D_1^2/2$ becomes $N_1\zeta_1$. Adding similar contributions from each gauge group, the total mass squared for the ρ mode is

$$m_\rho^2 = \sum_i N_i \zeta_i.$$

If this is positive, the potential prevents us from turning on ρ to make a bound state, while if it is negative, ρ is tachyonic near the origin and this combination is in fact unstable to forming the bound state.

Since the vacuum energy is always bounded below in this supersymmetric theory, the process of tachyon condensation is guaranteed to stop, resulting in a bound state. Since one has the exact potential, one can be much more precise and prove that a solution of the D-flatness conditions (in the $G_{\mathbb{C}}$ orbit of the object E) can exist only if we have (21) and

$$\sum_i N_i' \zeta_i > 0. \tag{22}$$

If it exists, it will be unique.

The mathematical theorem [23] is even stronger than this and asserts the converse. Suppose we are interested in the particular object E and we want to know whether it is stable or not, i.e. whether it will decay into anything. The theorem states that the configuration E can solve D-flatness if and only if (21) is satisfied and if (22) is satisfied for every subobject of E. In other words, if the mode ρ A and A' becomes massive, not only is this process of bound state formation prevented, the product becomes unstable even if there are potentially other ways of forming it from brane pairs, and even if these other pairs would have led to tachyonic modes. Such a subobject is known as a destabilizing subobject.

This theorem is proven, and the general study of this type of problem (known as symplectic quotient), uses the methods of geometric invariant theory. In fact this type of necessary and sufficient condition was already known as stability, an amusing example in which mathematical and physical nomenclature actually coincide in meaning. The particular version defined here is known as θ-stability; other forms will appear below.

The theorem is not hard to prove, but we give only the general idea here. A general strategy for finding a solution of the D-flatness condition on a given orbit is to take the potential as a function of the group element g parameterizing the orbit,

$$V = \sum \operatorname{tr}(g^{-1}Xgg^{\dagger}X^{\dagger}g^{-1\dagger} - \zeta)^2$$

and minimize it by gradient descent. The simple form of the potential makes it possible to show that a minimum will be reached, but it is not guaranteed that the minimum will be on the orbit; it could be a limit of points on the orbit which is not on the orbit. This is illustrated by the second example above (the adjoint chiral field) and the nonnormal matrix. Its gauge orbit includes the matrices

$$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \begin{pmatrix} x_1 & 1 \\ 0 & x_2 \end{pmatrix} \begin{pmatrix} \lambda^{-1} & 0 \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} x_1 & \lambda^2 \\ 0 & x_2 \end{pmatrix}$$

and one sees that the minimum of $\operatorname{tr}[X, X^{\dagger}]^2$ will be achieved as $\lambda \to 0$, a limit point not on the orbit but on a stable orbit with the same value of the invariants $\operatorname{tr} X^k$. One can show that whenever this happens, there is a similar one-parameter subgroup for which taking the limit decomposes the original unstable object into two objects, the subobject and quotient object (here 1-dimensional matrix configurations), and that conversely whenever the condition (22) is violated, a destabilizing one parameter subgroup can be constructed from it. Thus one obtains necessary and sufficient conditions for a solution to exist.

In our D-brane problems, the relation (21) typically will not be satisfied by any of the D-brane charges N, N' or $N + N'$. However it can be restored by taking advantage of the possibility mentioned earlier of making an overall shift ξ of the FI terms. This turns (21) into

$$0 = (\vec{N} + \vec{N}') \cdot (\vec{\zeta_i} + \xi \vec{e}) \tag{23}$$

where $\vec{e}$ is the vector with components $e_i = 1$. This can be solved for ξ.

One must then satisfy (22) with respect to the shifted FI terms. The resulting stability condition, with ξ eliminated using (23), is

$$(\vec{N}' \cdot \vec{\zeta})(\vec{e} \cdot \vec{N}) - (\vec{N}' \cdot \vec{e})(\vec{\zeta} \cdot \vec{N}) > 0. \tag{24}$$

This condition does not depend on $e \cdot \vec{\zeta}$ or on the overall scale of $\vec{\zeta}$. In the particular case of two nodes, only the ordering of the FI terms enters the final condition,

$$\mathrm{sgn}\,(\zeta_2 - \zeta_1)(N_1 N_2' - N_1' N_2) > 0.$$

Using methods we will not describe here, one can show that given two nodes, all simple bound states are stable on one side of the line $\zeta_1 = \zeta_2$, while on the other side only the two elementary branes and their antibranes are stable. This is very analogous to marginal stability in $\mathcal{N} = 2$ supersymmetric gauge theory, and in fact one can formulate that BPS spectrum in terms of representations of affine quivers [19].

5.1 D-brane stability near orbifold points

For quivers with more than two nodes, the condition (24) will have nontrivial dependence on the FI terms, leading to a complicated structure with an infinite number of lines of marginal stability. It is worth looking at this in detail, because this turns out to the the exact result for the spectrum in the neighborhood of orbifold points, and is a good illustration of the general structure.

A detailed analysis is rather involved (see [14]), but the basic idea can be illustrated by the following result: for any simple object, there is a line going into the point $\zeta = 0$ (which will be the orbifold point in our later examples) on which it is "most likely to be stable" (in many cases one can easily prove that it is). The idea is that we have a necessary condition for $\mathrm{Hom}(E', E) = 0$, namely $\chi(E', E) \leq 0$. Although there could still be $\mathrm{Hom}(E', E)$'s which cancel out of χ, this is not generic. Thus, if we could show that for every candidate destabilizing subobject E', i.e. one for which (24) fails, we had $\chi(E', E) \leq 0$, we would have good evidence for stability, while if this condition fails, we know that the object E is unstable.

This can be arranged by choosing $\vec{\zeta}$ so that

$$\vec{N}' \cdot \vec{\zeta} = \chi(\vec{N}', \vec{N}),$$

i.e. $\zeta_i = \chi_{ij} N_j$ in an obvious notation. Given this choice, the opposite of (24) becomes

$$\chi(\vec{N}', \vec{N}) \leq \chi(\vec{N}, \vec{N}) \frac{\vec{e} \cdot \vec{N}'}{\vec{e} \cdot \vec{N}}.$$

Now by considerations in the previous sections, we know that $\chi(\vec{N},\vec{N}) \leq 1$ for simple objects, and $\vec{e}\cdot\vec{N}' < \vec{e}\cdot\vec{N}$ for subobjects, so on this line $\chi(E',E) \leq 0$ follows.

On the other hand, every object has some elementary brane as subobject which will destabilize it by taking its FI term negative, so for every object there will also be a line on which it is unstable. These two lines must be separated by lines of marginal stability, which may be associated with the elementary brane we just discussed, or may be associated with larger subobjects.

Furthermore, one can easily check that the line for decay into a given subobject (say one of the elementary branes) is different for objects with different charge. This will also be clear from more conventional (BPS charge) marginal stability considerations, so there will be infinitely many lines of marginal stability in these problems, and a rather intricate structure.

6 Orbifold quiver theories

The simplest source of physical theories of D-branes on non-flat spaces is the orbifold construction. This is described in many places including [11].

We start with D3-branes extending in 3 + 1 Minkowski space and at points in an internal space $\mathbb{C}^n$, and choose an orbifold group Γ, an action of Γ on $\mathbb{C}^n$ denoted $r(g)$, and another N-dimensional representation of Γ acting on the "Chan-Paton factors" and denoted $\gamma(g)$. One then takes the world-volume $U(N)$ gauge theory and applies the projection

$$\gamma(g)^{-1}\phi\gamma(g) = r(g)\phi \tag{25}$$

to all the fields, where $r(g)$ acts on ϕ in the appropriate representation (scalar for the gauge fields, vector for the coordinates, spinor for the spinors). For $r(g)$ preserving supersymmetry one can also take the action directly on the superfields, which is what we will do.

The representation γ can be written as a direct sum over irreducibles,

$$\gamma = \oplus_i N_i r_i,$$

and one finds that the gauge group of the resulting theory is

$$G = \otimes_i U(N_i)$$

and the chiral matter spectrum is that of a quiver theory with

$$n_{i,j} = [r \otimes r_i^* \otimes r_j]$$

where $[R]$ denotes the number of times the trivial representation r_0 appears in R, e.g. $[r_i] = \delta_{i,0}$.

There will also be a superpotential and FI terms. If the theory arises from a $\mathbb{C}^3$ orbifold, the superpotential is the projection of that of $\mathcal{N} = 4$ $U(N)$ super Yang-Mills,

$$W_{N=4} = \operatorname{tr} X^1[X^2, X^3].$$

We will confine ourselves to the example of $\Gamma = \mathbb{Z}_k$ with action

$$X^m \to \exp \frac{2\pi i a_m}{k} X^m,$$

in which case there are k irreps and the result of the projection is

$$\begin{aligned} W &= \textstyle\sum_{m=1}^{3} \sum_{i=0}^{k-1} \\ &\quad \operatorname{tr} X^1_{i,i+a_1} X^2_{i+a_1,i+a_1+a_2} X^3_{i+a_1+a_2,i} \\ - &\quad \operatorname{tr} X^1_{i,i+a_1} X^3_{i+a_1,i+a_1+a_3} X^2_{i+a_1+a_3,i}. \end{aligned} \tag{26}$$

There are k FI terms which are fairly general, satisfying the relation

$$\sum_i \zeta_i = 0$$

and in certain examples additional relations. They can be derived from world-sheet considerations of couplings to twist sector moduli and their physical interpretation is as blowup modes for the orbifold singularity; they can be considered as real coordinates for the complexified Kähler moduli space in the neighborhood of the orbifold point.

These quiver theories can be used to describe a large number of D-branes anywhere near the orbifold point, along the lines we just described, as bound states of arbitrary numbers of elementary branes, which are usually called fractional branes in this context. We might even be able to describe all branes, if these branes span the charge lattice.

We will focus particularly on the simplest case of $\mathbb{C}^3/\mathbb{Z}_3$, which has the quiver diagram

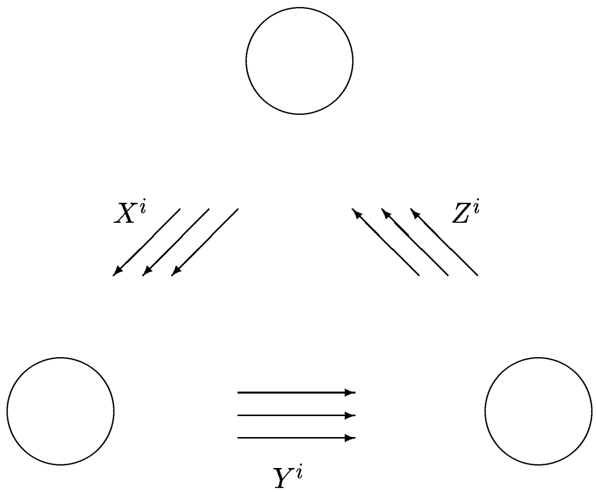


Fig.3. The $\mathbb{C}^3/\mathbb{Z}_3$ quiver

and two independent FI terms related in some way to the single complexified Kähler modulus. For simplicity, denote the three groups of chiral multiplets as X^i, Y^i and Z^i, then the superpotential is

$$W = \epsilon_{ijk} \mathrm{tr}\, Z^i Y^j X^k.$$

The formal orbifold construction can be generalized to higher dimensional complex space. Later we will see that in fact D-branes in Gepner models can also be understood using this construction, essentially because they can be defined as Landau-Ginzburg orbifolds. The case we will consider is $\mathbb{C}^5/\mathbb{Z}_k$, and we will argue that these theories can be derived as the orbifold projection of a $U(N)$ gauge theory with chiral multiplets X^i in the 5 of a global $SU(5)$, chiral multiplets $Y_{[ij]}$ in the $\bar{1}0$, and superpotential

$$W = \mathrm{tr}\, X^i X^j Y_{[ij]} + \ldots.$$

where ... indicates higher order terms which we will pretty much ignore in these lectures. Applying the projection leads to a superpotential much like (26).

Both of these superpotentials are cubic, and impose commutativity conditions on the matrices X^i,

$$X^m_{i,i+a_m} X^n_{i+a_m,i+a_m+a_n} = X^n_{i,i+a_n} X^m_{i+a_n,i+a_n+a_m}. \tag{27}$$

In fact they have a much more important common property: they express the condition that a certain D operator squares to zero.

6.1 $D^2 = 0$ superpotentials and higher Ext groups

We now generalize our previous discussion of bound states and the moduli space to these quiver theories with superpotential.

The formula (10) for a gauge transformation, and the definition of $\mathrm{Hom}(A, B)$ as the kernel of this operator D, go over unchanged.

However, the massless matter variations are different, because some of these can be lifted by the superpotential. We need to supplement the condition that a matter variation be in the cokernel of D with an additional condition that it satisfy the superpotential constraint. For the quadratic constraints (27), we can write this as

$$\begin{aligned} 0 = & X^m_{i,i+a_m} \rho^n_{i+a_m,i+a_m+a_n} \\ & + \rho^m_{i,i+a_m} (X')^n_{i+a_m,i+a_m+a_n} \\ & - X^n_{i,i+a_n} \rho^m_{i+a_n,i+a_n+a_m} \\ & - \rho^n_{i,i+a_n} (X')^m_{i+a_n,i+a_n+a_m}. \end{aligned} \tag{28}$$

Note that this is again linear in ρ and in the configuration variables X and X'. Thus we can define a *new* operator D which takes a variation ρ and maps it to the right hand side of (28). Gauge invariance requires $D^2 = 0$ and in terms of this D, the massless off-diagonal matter variations are now the ρ in $\ker D/\mathrm{Im}\, D$.

Thus we regain the interpretation of $\mathrm{Ext}(A', A)$ as the cohomology of the operator D. One sees now however that to write a formula like (13), we will have to keep track of the cokernel of (28), because now matter variations can pair up with these and become massless. We thus define this space as

$$\mathrm{Ext}^2(A', A)$$

and rename the previous space $\mathrm{Ext}^1(A', A)$ (often it is still called Ext without the superscript).

One naturally wants to know what is the physical meaning of Ext^2. From their construction, these are superpotential constraints which are not actually used in the matter configuration, and do not lift matter fields. This is related to one of the major difficulties in working with general $\mathcal{N} = 1$ theories, namely the fact that the dimension of moduli space need not be constant; it can consist of several branches with various dimensions, because of the relatively unconstrained form of the superpotential and the fact that generic cubic terms will produce such a structure. Indeed the naive estimate for the dimension of moduli space in an $\mathcal{N} = 1$ theory is that it is always zero, because there are as many equations in $W' = 0$ as there are unknowns. This naive estimate can fail because of unused and redundant superpotential constraints, and this is reflected in the fact that the generalization of (13) is not going to involve just $\dim \mathrm{Hom}$ and $\dim \mathrm{Ext}$ but additional terms which can be different on different branches of moduli space.

This story could clearly be repeated again by defining a similar operator acting on new fields (not present in the original Lagrangian) $\rho^{[ab]}$ with two indices, and so forth up to n index fields (one will need to continue to antisymmetrize indices to keep $D^2 = 0$). Should we do this?

The answer appears to be yes, for specific reasons which will appear shortly, and because in general it is useful to keep track of redundancies between the superpotential relations. Suppose we found that a two index element of $\mathrm{coker}\, D$ was not in fact in $\ker D$, so that it drops out of Ext^2 (being paired with a three-index field). This means that some of the unused superpotential constraints become redundant when multiplied by additional powers of the fields X. The presence of such redundancies in solving equations was one of the main motivations for introducing the homological algebra techniques we are now discussing.

The whole setup can be defined at once by introducing a Grassmann algebra with n generators,

$$e_m e_n + e_n e_m = 0,$$

and defining

$$D_X = X^n e_n$$

where X^n is the direct sum of the various matrices $X^n_{i,i+a_n}$ acting on the direct sum of Hilbert spaces associated to all the nodes. It is easy to see that $D^2 = 0$ then precisely reproduces (28).

The operator D of our previous discussion is then $D_X - D_{X'}$ (the relative sign is not important) where D_X acts on the sum

$$\epsilon + \rho^a e_a + \rho^{[ab]} e_a e_b + \ldots$$

on the left, and D'_X acts on it on the right.

The same argument as before suffices to compute the relative Euler character, now defined as

$$\chi(A, B) \quad = \textstyle\sum_i (-1)^i h^1 \quad \equiv \displaystyle\sum_i (-1)^i \dim \mathrm{Ext}^i(A, B) \tag{29}$$

where we let $\mathrm{Ext}^0(A, B) \equiv \mathrm{Hom}(A, B)$. For the configuration $X = X' = 0$, every field ϵ and ρ contributes with the sign $(-1)^i$, leading to a formula bilinear in the charges N_i of the two representations.

In practice, it is often useful to assume that certain of the X or Y links are zero, and also drop the fields ρ which cannot appear under this assumption, to get a definition of χ with stronger consequences. We already followed this policy in the $\mathbb{C}^5$ case in not considering the constraints $\partial W / \partial X = 0$; these can be automatically solved by setting $Y = 0$, leading to the equations we kept. Example of such simplifications are to assume that some nodes are not present, or that all links which cross a certain "line" on the quiver (say all links from i to $j < i$) are zero, in which case all fields ρ which cross this line can also be dropped.

A simplified formula which will be of relevance below is to consider the $\mathbb{C}^3/\mathbb{Z}_3$ quiver with the $X^a_{3,1} = 0$. In this case one obtains

$$\chi(A, B) = \vec{N}_A \cdot \begin{pmatrix} 1 & -3 & 3 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix} \cdot N_B. \tag{30}$$

This corresponds to the "naive" dimension one obtains by taking the number of matter fields minus unbroken gauge symmetries and superpotential constraints, and is indeed the correct dimension (it turns out that $h^2 = 0$) for all cases except the D0-brane, (1 1 1).

This case is instructive. The result $\chi(D0, D0) = 0$ corresponds to a moduli space dimension for the D0 of 1. Of course we left out the $X_{3,1}$ links and this could increase the dimension; however these satisfy superpotential constraints of their own and one finds that they provide one extra modulus, leading to the incorrect answer 2.

The resolution of this is that in fact $h^2 = 1$ in this case. The general solution of the constraints $0 = X^i_{1,2} X^j_{2,3} \epsilon_{ijk}$ is to take $X^i_{1,2} = \lambda X^i_{2,3}$, for a general complex number λ. This is two conditions, and comparing to the three superpotential constraints, we see there is one redundancy. Thus, the result $\chi = 1 = h_0 - h_1 + h_2$ predicts $h_1 = 2$. Similarly $X^i_{3,1} = \lambda' X^i_{2,3}$ solves the remaining constraints, leading to the correct moduli space dimension 3 for an object moving on $\mathbb{C}^3/\mathbb{Z}_3$.

This discussion was of course overkill for such a simple example, but the point is that it illustrates the general (non-trivial) relation between χ and the naive dimension formula, and the true dimension of moduli space. In general, moduli spaces in $\mathcal{N} = 1$ theories can have branches with different dimensions; this will happen because the h^n with $n \geq 2$ can differ on different branches.

We could also extend the definition in the $\mathbb{C}^5$ case to include the $\partial W/\partial X = 0$ superpotential constraints as well, by defining

$$D = e_a X^a + \epsilon^{abcde} e_a e_b e_c Y_{[de]}.$$

Both the forms $\rho^{[a_1 \ldots a_k]} e_{a_1} \ldots e_{a_k}$, and the forms representing Ext^k, are in many ways like differential forms. In particular they have a wedge product which obeys the same formal rules. These rules can be axiomatized in the structure of an "abelian category" as described in Gelfand and Manin [22] and many other textbooks.

The field ρ with n indices (the "top form") is special as it is a sum of adjoints and thus one can take its trace to get the analog of an integral. This trace provides a bilinear form on $\mathrm{Ext}^k(A, B) \times \mathrm{Ext}^{n-k}(B, A)$. If we work in the original orbifold quiver category (i.e. not setting any links to zero), one can see that this form is non-degenerate and thus the dimensions of these two spaces will be the same. This is a quiver analog of "Serre duality" (on which more later).

The D0 problem again provides an example. There, the number of extra moduli from the link $X_{3,1}$ was equal to h^2. This is an example of the duality $\dim \mathrm{Ext}^1(D0, D0) = \dim \mathrm{Ext}^2(D0, D0)$ valid for the orbifold quiver category; the extra Ext^1 is the dual of the h^2 we derived earlier.

Having defined the higher Ext groups, we remark that the relation between Ext^1 and bound states, its further relations to exact sequences and triangles, and the role of these in solving D-flatness conditions, all go through in exactly the same way as before.

7 Some large volume considerations

We more or less have all the tools we need to find the spectrum of BPS branes in simple orbifolds, anywhere near the orbifold point. Indeed it is quite easy to analyze the states which can be formed from arbitrary numbers of two fractional branes, and one might do some exercises of this sort to get familiar with the ideas.

The description does not look much like the way one is used to describing D-branes at large volume, but on a deeper level the two limits can be made part of the same formalism. The key to this is the "decoupling statement" which asserts that all of the holomorphic structure of the branes must be independent of Kähler moduli. This will imply that quiver objects and all the information in the F-flatness conditions must in some sense be the same as the set of holomorphic bundles or suitable generalizations of these.

To see this, we will need to review some of the theory of D-branes in "large volume," i.e. when stringy corrections are negligible. We assume some familiarity with the material in Polchinski's TASI lectures [27].

The problem of finding supersymmetric embeddings of branes governed by the supersymmetrized Nambu-Born-Infeld Lagrangian has been much studied and the general equations governing these are known. In the $\alpha' \to 0$ limit (we will be a bit more general below), and if no other background fields are turned on, the general answer is that a supersymmetric brane must embed into a calibrated submanifold, and the gauge fields must preserve supersymmetry in the usual Yang-Mills sense, except that one can use an inhomogeneous supersymmetry as well: an unbroken supersymmetry is given by two spinor parameters (ϵ, ϵ') which must satisfy

$$0 = \delta\chi = \Gamma^{IJ} F_{IJ} \epsilon + \epsilon'. \tag{31}$$

The inhomogeneous supersymmetry simply corresponds to shifts of the non-interacting gaugino (or its diagonal component in $U(N)$ theory). Such a supersymmetry is guaranteed to be present because the D-brane spontaneously broke half of the supersymmetry of the bulk theory. However, we see that which half is preserved can depend on the world-volume background fields, and in general will be some linear combination of the ϵ and ϵ' bulk supersymmetries.

7.1 Calibrated geometry

A calibration is a p-form λ which is closed, $d\lambda = 0$, and which provides a lower bound for the volume: for any p-dimensional linear subspace V of the tangent space at any point, one has

$$\lambda|_V \leq (\text{volume})_V \tag{32}$$

considered as an equation between oriented p-forms.

Given a calibration, one has an easy way to find minimal volume manifolds: a calibrated submanifold, defined as a submanifold Σ whose tangent bundle saturates (32) at each point, is necessarily minimal volume. This is because Stokes' theorem tells us that $\int_\Sigma \lambda$ will be preserved under any continuous variation of the submanifold Σ, while the bound (32) then tells us that volume can only increase under these variations.

Examples in flat space $\mathbb{C}^n$ are not hard to find and check. They include the Kähler form

$$\omega = \frac{1}{i}\sum_{I=1}^{n} dz^I \wedge d\bar{z}^I,$$

and the various real parts of the holomorphic n-form,

$$\Omega_\theta \equiv \Re e^{i\theta}\Omega$$

where

$$\Omega = dz^1 \wedge dz^2 \wedge \cdots \wedge dz^n.$$

In general, calibrations arise from covariantly constant spinors and thus are closely associated with supersymmetry. Suppose we have a covariantly constant spinor ϵ; then we claim that

$$\lambda = \epsilon^+ \Gamma^{(p)} \epsilon$$

is a calibration, where

$$\Gamma^{(p)} = i^{(p(p-1)/2} \prod_I \Gamma_I dx^I$$

and we choose p to make the product non-zero.

To see this, consider a linear p-dimensional subspace V of the cotangent space to a point, with orthonormal basis e^I, and consider the operator

$$\Gamma_V = i^{(p(p-1)/2} \prod_I \Gamma_I e^I.$$

This operator satisfies the equation

$$(1-\Gamma_V)^2 = 2(1-\Gamma_V)$$

and thus

$$\epsilon^+(1-\Gamma_V)\epsilon \geq 0$$

which implies the bound.

This argument also shows that branes wrapped on calibrated submanifolds preserve supersymmetry: taking V to be a tangent space to the brane, we have $\Gamma_V = \Gamma_D$, so saturating the bound implies $\epsilon = \Gamma_D\epsilon$.

7.2 B type calibrated submanifolds

Each of the types of calibrated manifold has its own distinctive geometry. The most intensively studied case is the Calabi-Yau manifolds, with $SU(n)$ holonomy, for which there are two types of calibration.

The first (called "B-type" for reasons given later) is with respect to the Kähler form, or powers of the Kähler form. It is not hard to see that these are holomorphic submanifolds, defined by an holomorphic map from a p-dimensional complex manifold into space, or else as the zero set of $n-p$ complex equations in space. This also includes branes which embed in the entire CY manifold, and the brane which embeds in a point. All of the methods of algebraic geometry which were so useful in analyzing the geometry of Calabi-Yau manifolds will be just as useful in analyzing these branes.

Gauge field backgrounds which preserve supersymmetry (admit solutions to (31)) can be found by slightly generalizing a traditional argument used in heterotic string compactification (as given in GSW); they are the hermitian Yang-Mills connections. The argument is simply to use the Kähler structure of the manifold to rewrite the Dirac algebra $\{\Gamma^I, \Gamma^J\} = 2g^{IJ}$ as an algebra of fermionic creation and annihilation operators, which naturally act on the space of (antiholomorphic) $(0,p)$-forms. Explicitly, using complex coordinates z^I and $\bar{z}^{\bar{I}}$,

$$\Gamma^{\bar{I}} \rightarrow d\bar{z}^{\bar{I}} \tag{33}$$

$$g_{I\bar{J}}\Gamma^I \rightarrow i_{\bar{J}} \qquad s.t.\{i_{\bar{J}}, d\bar{z}^{\bar{I}} = \delta^{\bar{I}}_{\bar{J}}. \tag{34}$$

The two spinor representations of $SO(2n)$ then reduce to the direct sum of even or odd differential forms. Taking for ϵ and ϵ' the zero form, (31) becomes

$$F^{(2,0)} = F_{IJ} dz^I \wedge dz^J = 0 \quad (35)$$

$$F^{(0,2)} = F_{\bar{I}\bar{J}} d\bar{z}^{\bar{I}} \wedge d\bar{z}^{\bar{J}} = 0 \quad (36)$$

and finally, for $F^{(1,1)} = F_{I\bar{J}} dz^I \wedge d\bar{z}^{\bar{J}}$, we have

$$F^{(1,1)} \wedge \omega^{n-1} = c\omega^n$$

where ω is the Kähler form and c is an arbitrary constant.

7.2.1 BPS central charge

The constant c determines which $\mathcal{N} = 1$ subalgebra of the bulk $\mathcal{N} = 2$ supersymmetry is unbroken. The $\mathcal{N} = 2$ supersymmetry algebra admits a moduli space of $\mathcal{N} = 1$ subalgebras parameterized by a phase θ, defined by

$$0 = \mathrm{Im}\, e^{i\theta} Q^\alpha.$$

In terms of the parameters ϵ and ϵ' we can write

$$0 = \mathrm{Im}\, e^{i\theta}(i\epsilon + \epsilon')$$

which if we insert the solution gives

$$0 = \mathrm{Im}\, e^{i\theta}(\omega + i l_s^4 F) \wedge \omega^2 \quad (37)$$

This is only an $l_s \to 0$ estimate and one can easily do better. Another way to determine the unbroken $\mathcal{N} = 1$ is computing the BPS central charge of the brane; its phase will be $e^{i\theta}$. The BPS central charge of a D-brane is determined by its Ramond-Ramond charges, which are those of a source

$$\int_\Sigma C \wedge \mathrm{Tr} e^{B-F}. \quad (38)$$

Assuming that in the large volume limit a pure D$2p$-brane has central charge $\int (-iV)^p$ leads to the expression

$$Z = \int_\Sigma e^{-F+B+i\omega} = \sum_p \frac{1}{(d-p)!} ch_p (B + iJ)^{d-p} \quad (39)$$

where the Chern character ch_p is the $2p$-form in the expansion of $\mathrm{Tr} e^F$.

These considerations lead to a formula whose leading $l_s \to 0$ limit is (37). They can be confirmed microscopically from an analysis using the supersymmetrized Nambu-Born-Infeld action, leading to an equation derived by Mariño, Minasian, Moore and Strominger (the MMMS equation) [25]. They take into account all powerlike corrections in l_s, but not effects due to world-sheet instantons.

7.2.2 The Donaldson-Uhlenbeck-Yau theorem

There is a standard mathematical approach to solving such equations, based on what is called the Hitchin-Kobayashi correspondence. For hermitian Yang-Mills this is encapsulated in the theorems of Donaldson and Uhlenbeck-Yau, but the approach is in fact more general and applies to the MMMS equations and indeed in a sense we will describe to the general case of string scale Calabi-Yaus.

We start by assuming we can find a solution to the equation $F^{0,2} = 0$. This is the integrability condition for the antiholomorphic part of the connection,

$$[\bar{\partial} + \bar{A}, \bar{\partial} + \bar{A}] = 0,$$

and thus one can locally trivialize the bundle by some holomorphic transformation which acts on sections as $\psi \to g(z)\psi$. Globally, the bundle need not be trivial, but all of the transition functions will be holomorphic. Thus each solution to this equation (up to complex gauge equivalence) oorresponds to a holomorphic bundle. This is a notion which can be defined purely in terms of the complex structure of the manifold, so this part of the problem does not depend on the Kähler moduli of the CY. A hermitian connection will then automatically solve $F^{2,0} = 0$.

This leaves the equation on $F^{(1,1)}$, and the DUY theorems then state necessary and sufficient conditions that there exist a particular connection in this orbit of the complexified gauge group which solves this equation.

First of all, one knows that the first Chern class $c_1 = \int \mathrm{Tr} F \wedge J \wedge J$ is a topological invariant, so it is computable just knowing the holomorphic bundle. Integrating the equation shows that the constant c from above is determined in terms of c_1.

We then define the slope of a bundle E, $\mu(E)$, as the ratio

$$\mu(E) = \frac{c_1(E)}{rk(E)} = \frac{1}{rk(E)} \int \mathrm{Tr} F \wedge J^{n-1}.$$

A holomorphic bundle E is then μ-stable if, for all subbundles E', we have

$$\mu(E') < \mu(E). \qquad (40)$$

The DUY theorems then state that an irreducible (simple) solution to $\omega F^{1,1} = c$ will exist if and only if E is μ-stable. Both the equation and this condition depend on the Kähler form and it can be seen in examples that if $b^{(1,1)} > 1$

(so there is a choice of Kähler form) this dependence is nontrivial; there are "walls" in Kähler moduli space on which the list of stable bundles changes.

Note the very close parallel between this condition and the θ-stability condition we discussed that controls the solvability of D-flatness conditions. Both can be understood using the ideas of geometric invariant theory and in fact the DUY theorem is proven by the same idea we discussed earlier of flowing within a complexified gauge orbit to a minimum of a potential, here the Yang-Mills action. Again the stability condition is what guarantees that the minimum is still on the original orbit.

7.3 A type calibrated submanifolds

The other possibility is to calibrate with respect to one of the n-forms Ω_θ. These are A-type or special Lagrangian (sL) submanifolds. One usually wants to keep track of θ and distinguish A_θ or sL_θ submanifolds, because θ determines the unbroken $\mathcal{N} = 1$ supersymmetry just as in our previous discussion.

The reason for the name special Lagrangian is that one can show (by an easy local argument) that the calibration condition is equivalent to the pair of conditions

$$\omega|_\Sigma = 0 \tag{41}$$

$$\mathrm{Im}\, e^{i\theta}\Omega_\Sigma = 0. \tag{42}$$

For A branes in other than two dimensions, a supersymmetric gauge connection must satisfy $F = 0$, i.e. it is a Wilson line. The case of two dimensions is special and is better understood by bringing in the formalism of hyperkähler geometry, which we will not do here.

One of the basic results about these branes is that the moduli space of a smooth special Lagrangian Σ has real dimension $b_1(\Sigma)$. Combining this with the moduli of the flat gauge connection, the D-brane moduli space has complex dimension $b_1(\Sigma)$.

The BPS central charge of an A brane Σ is simply

$$Z = \int_\Sigma \Omega. \tag{43}$$

7.4 Mirror symmetry and comparison of the two pictures

Mirror symmetry will exchange the A and B branes. The basic physics behind this is a realization of mirror symmetry proposed by Strominger, Yau and Zaslow. The idea is to describe the CY_3 as a T^3 fibration, and then interpret mirror symmetry as T-duality along the fibers. This exchanges the D0-brane with a distinguished D3-brane with topology T^3; since $b_1(T^3) = 3$ one can rederive the original Calabi-Yau as the moduli space of this D3-brane on the mirror. Similarly the T-duality will exchange every B brane on M, with its moduli space and all other physics, with a corresponding A brane on W.

The special Lagrangian picture has some advantages and disadvantages over the holomorphic brane picture. Its main disadvantage at present is that it is much less well understood mathematically, but this situation may improve.

Sometimes in string theory, one finds that duality can exchange a description in which some observable receives quantum corrections, with a dual description in which the corresponding observable is always equal to its classical value. In this case one is usually much better off using the description which is always classical for detailed computation, while the other description retains its primary importance only in limits in which it becomes classical.

Indeed mirror symmetry is the example *par excellence*, as the observables relating to the special geometry of the $\mathcal{N} = 2$ compactified theory, namely the prepotential and BPS central charges, are classically exact in the IIb compactification, in which the BPS central charges are just (43), the periods of the holomorphic three-form, which depend only on complex structure and are independent of Kähler moduli. This is in contrast to the IIa compactification in which the BPS central charges are independent of complex structure and depend only on Kähler moduli; although they take simple values in the large volume limit, in general they receive world-sheet instanton corrections. Physically, mirror symmetry is usually used as a tool for summing these instanton corrections to derive exact prepotentials and solve $\mathcal{N} = 2$ string compactifications.

7.5 The decoupling statement

We should ask the same question, whether there is any dual picture in which the world-volume theory can be determined purely classically, in our D-brane context. It turns out to have a pretty answer [8]: of the two semi-independent parts of our $\mathcal{N} = 1$ world-volume Lagrangian, the holomorphic part and the D-flatness conditions, one of them will receive quantum corrections while the other will not. In the B brane picture, one can see that the holomorphic structure is exact at large volume, while the D-flatness conditions receive instanton corrections. In the A brane picture, it is the other way around.

There are various arguments for this. On one side, we can consider the topologically twisted theory of the open strings ending on the D-branes. This model only depends on a subset of the CY moduli, the complex structure in the B model and the Kähler structure in the A model, and since B model observables do not depend on the volume they must be exact at large volume. One can show that this topological theory can be used to compute the holomorphic structure and superpotential of the world-volume theories.

This leaves the questions of how the Kähler moduli affect B branes, and how complex structure moduli affect A branes. These moduli directly control the BPS central charges of the branes and the most striking physics resulting from this is the variation of the BPS spectrum expressed in lines of marginal stability: as we vary the moduli, a brane can decay into constituents, or new bound states can form. As we discussed, this type of behavior is controlled by the D-flatness conditions, leading to the idea that these moduli couple only through the FI terms. Since A brane central charges are exact at large volume (they are the D-particles of IIb theory), this more or less requires that the A brane D flatness conditions are exact at large volume.

This is explained from a world-sheet point of view in [15]. A simple space-time argument for this can be made by using the decomposition of the moduli of the $\mathcal{N}=2$ bulk theory under the $\mathcal{N}=1$ supersymmetry of the D-branes. This argument is dimension dependent and it is most convenient to make it for the $3+1$ world volume theories we have been discussing so far. In this case, B branes naturally live in IIb theory, while A branes naturally live in IIa theory. Let us consider the case of IIb theory; then the $\mathcal{N}=2$ vector multiplets contain complex structure moduli, while the hypermultiplets contain Kähler moduli and partner RR scalars. Under $\mathcal{N}=1$ supersymmetry, the vector must decompose into a vector and a chiral multiplet, so this chiral multiplet is purely NS-NS and is just a complex structure modulus. The hypermultiplet decomposes into two chiral multiplets and one can check that each of these contains one real NS-NS Kähler modulus and one real partner RR scalar. (The simplest case to check is the related type I theory which keeps only one of these chiral multiplets).

Now, since the $\mathcal{N}=1$ superpotential is holomorphic, if it depends on a chiral multiplet, it must depend on both of its real components. This is fine for the complex structure multiplet, but if it depends on a Kähler modulus, this means it will depend on an RR scalar, call it C_i. Now perturbative string amplitudes very generally do not produce nonderivative couplings to the RR fields; for each RR scalar there is an exact symmetry $\delta C_i = \epsilon_i$. This contradiction implies that the superpotential is independent of Kähler moduli (reproducing our previous claim).

Fayet-Iliopoulos terms are real however and in fact naturally depend only on one real component of a chiral multiplet. The generic coupling of this type to a world-volume vector superfield V is

$$\int d^4\theta \ (\phi + \phi^+ - V)^2.$$

This is gauge invariant under $\delta V = \Sigma + \Sigma^+$ and $\delta\phi = \Sigma$. This shifts the real part of ϕ, but a world-volume gauge transformation cannot act on the bulk fields: thus this must be an exact symmetry of the bulk theory. This will be true only if the real part of ϕ is a RR scalar, and this allows FI terms to be controlled by Kähler moduli but not complex moduli.

The strongest test of the decoupling statement would of course be to simply derive the F or D-flatness conditions in the appropriate brane world-volume theories and check that they are the same in the cases they are supposed to be. Let us consider this problem in an example. Since the F flatness conditions are primary, one should start with these, and thus with the B brane picture.

8 Introduction to the $\mathbb{C}^3/\mathbb{Z}_3$ orbifold

We start by reviewing the basic picture of strings compactified on this space. Defining closed strings on this orbifold leads to an $\mathcal{N} = 2$ theory with marginal operators in the twisted sectors corresponding to a single complexified Kähler modulus. The geometric interpretation of turning these on should correspond to some operation which fixes the complex structure but introduces an element of $H^{1,1}$.

Such an operation is known mathematically and is called a blow-up. The idea is that one can take any point in an n-dimensional complex manifold and replace it by a $\mathbb{CP}^{n-1}$ (henceforth just called $\mathbb{P}^{n-1}$) parameterizing the various tangent vectors to the original point. This can be made precise by the following equations in local coordinates z^i: to blow up $z = 0$, introduce a $\mathbb{P}^{n-1}$ with homogeneous coordinates $w^i \cong \lambda w^i$, $1 \leq i \leq n$, and impose the equations

$$z^i w^j = z^j w^i \qquad \forall i, j.$$

Away from $z = 0$ one can solve for w, while at $z = 0$ they are unconstrained.

Since we define our $\mathbb{Z}^3$ orbifold as $z^i \cong \omega z^i$ with $\omega = e^{2\pi i/3}$, this operation is well defined on the orbifold and replaces the singularity with a $\mathbb{P}^2$. One can also check that the holomorphic three-form has no zeroes, so the result is a Calabi-Yau. In fact it is (the total space of) the line bundle $M = \mathcal{O}_{\mathbb{P}^2}(-3)$.

Thus finite energy D-branes must wrap cycles in the $\mathbb{P}^2$. Now $H^p(\mathbb{P}^2, \mathbb{Z}) \cong \mathbb{Z}$ for $p = 0, 2, 4$ and integrating RR potentials over these three cycles leads to three conserved RR charges (agreeing with the orbifold) which we can call D0, D2 and D4 charge.

8.1 Line bundles on $\mathbb{P}^2$

Let us discuss holomorphic branes on large volume $\mathbb{P}^2$ a bit (much more detail can be found in [14], and in fact the mathematical classification of stable sheaves is completely known in this case). The simplest ones are the line bundles, which we will denote $\mathcal{O}(n)$. These can be defined for $n \geq 0$ as the bundles which admit sections which are degree n polynomials in the homogeneous coordinates w^i. For $n < 0$, one can either talk about sections with poles, or define $\mathcal{O}(-n)$ as the dual object to $\mathcal{O}(n)$ such that multiplication of sections from the two produces a function. All of these are stable and correspond to a D4-brane with n units of D2 charge turned on.

From (38), the RR charge of $\mathcal{O}(n)$ is given by the Chern character $\mathrm{ch}(\mathcal{O}(n)) = e^{nJ}$, where $J = c_1(\mathcal{O}(1))$ is the generator of $H^2(\mathbb{P}^n, \mathbb{Z})$ (the unit of magnetic flux), and we use conventions where the $\sqrt{\hat{A}}$ term has been factored out. The successive terms in the expansion in J (up to $o(J^2)$) are the D4, D2 and D0 charges. Thus for $n \neq 0$ these objects also carry D0 charge.

It turns out that these objects already provide a basis for the K theory of $\mathbb{P}^2$ (and the K theory with compact support of M). Since M is not compact, we cannot directly check this from the intersection form, but as suggested in section 2 we will instead look at the index of the Dirac operator on $\mathbb{P}^2$. In the Kähler context this becomes the index of the $\bar{\partial}$ operator, which is also known as the relative Euler character:

$$\chi(E, F) \equiv \sum_p (-1)^p \dim H^{0,p}(M, E^* \otimes F).$$

The index theorem in this case reduces to the Grothendieck-Riemann-Roch formula, which is (2) with $\hat{A}(M) = Td(M)$.

We now proceed to compute $\chi(\mathcal{O}(m), \mathcal{O}(n)) = \chi(\mathcal{O}, \mathcal{O}(n-m)) \equiv \chi(\mathcal{O}(n-m))$ on $\mathbb{P}^N$. For those who want to do this directly from the index theorem (it is not too hard), we quote $Td(\mathbb{P}^N) = (J/(1-e^{-J}))^{N+1}$.

A shortcut to this is to compute $\chi(\mathcal{O}(n))$ for large n, by explicitly computing $\dim H^0(\mathcal{O}(n))$ and appealing to the following vanishing theorem ([21], p. 159),

Theorem B. Let M be a compact complex manifold and $L \longrightarrow M$ a positive line bundle (i.e. there exists a connection such that $F_{i\bar{j}}$ is everywhere positive; for $\mathbb{P}^n$ we just need $c_1 > 0$). Then for any holomorphic vector bundle E, there exists μ_0 such that

$$H^{0,q}(M, L^\mu \otimes E) = 0 \qquad \forall q > 0, \mu > \mu_0.$$

to conclude that $\chi(\mathcal{O}(n)) = \dim H^0(\mathcal{O}(n))$. But we also know from (2) that χ is polynomial in n, so computing it at large n determines it for all n.

Since the sections of $\mathcal{O}_{\mathbb{P}^N}(n)$ are degree n polynomials in $N+1$ homogeneous variables, we can conclude that

$$\chi(\mathcal{O}_{\mathbb{P}^N}(n)) = \frac{1}{N!}\prod_{i=1}^{N}(n+i).$$

In particular, the three line bundles $\mathcal{O}, \mathcal{O}(1)$ and $\mathcal{O}(2)$ have

$$\chi(\mathcal{O}(m), \mathcal{O}(n)) = \begin{pmatrix} 1 & 3 & 6 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \tag{44}$$

which is a unimodular matrix, and thus this set can be used as a basis. For example, the K theory class of the D0 is determined by solving $N_0 + N_1 e^J + N_2 e^{2J} = J^2 + o(J^3)$ to be

$$[\mathcal{O}_z] = [\mathcal{O}(-1)] - 2[\mathcal{O}] + [\mathcal{O}(1)].$$

We have given the D0 at the point z its mathematical name $\mathcal{O}_z$, the structure sheaf of the point z.

Readers with any familiarity with D-branes on orbifolds will immediately recognize that the three fractional branes B_i of the $\mathbb{C}^3/\mathbb{Z}_3$ orbifold cannot be these, as these satisfy

$$[\mathcal{O}_z] = [B_1] + [B_2] + [B_3]. \tag{45}$$

So what are the B_i ?

8.2 Identifying the fractional branes

This problem was first solved by mirror symmetry techniques [10]. The idea [8] to take expressions for the BPS central charges computed using mirror symmetry and explicitly compare them between the large volume Dp-brane basis and a basis at the orbifold point.

We need to discuss the structure of the stringy Kähler moduli space of M, i.e. the complex structure moduli space of its mirror. It is a Riemann sphere with three singularities. One is the large volume limit near which the BPS charges are (39). The second is the orbifold point, described by our quiver theory. Finally there is a third singularity called "conifold point" at which one of the BPS central charges vanishes. Directly continuing this to large volume, one finds that the corresponding brane is $\mathcal{O}$, the "pure" D4.

As usual in $\mathcal{N} = 2$ theories, the simplest attribute of a singularity is the monodromy it induces on the charges. In large volume this is $B \to B + 1$ which takes $\mathcal{O}(n) \to \mathcal{O}(n + 1)$. Around the conifold point it is determined by the usual considerations involving a massless particle, while the orbifold point has an associated $\mathbb{Z}_3$ monodromy which permutes the fractional branes and the FI terms of the quiver theory.

The results which suffice to determine the identification in this case are that $\mathcal{O}$ is one of the fractional branes (tested by continuing its period to the conifold point), and the $\mathbb{Z}_3$ monodromy expressed in the large volume basis using mirror symmetry.

We did not quote the final identification of the fractional branes as it turns out that there is a simpler way to do this.

9 The McKay correspondence

An independent way to identify the fractional branes follows what is called the "generalized McKay correspondence" in mathematics [28, 5], which we summarize. It can be physically motivated [17] and agrees with the mirror symmetry prediction wherever this has been tested.

The idea is that it is relatively easy to geometrically identify a dual set of "fractional branes" which fill the noncompact space M, and to find their intersection form with the original fractional branes. This data then turns out to determine the original fractional branes.

An extended fractional brane can be taken as a D9-brane filling $\mathbb{C}^3/\Gamma$. Now the orbifold projection (25) acts on the spatial coordinates as well; for the Yang-Mills connection it is

$$\gamma^{-1}(g)A_i(z)\gamma(g) = r_i^j A_j(g(z)). \tag{46}$$

This is a twisted boundary condition and its interpretation is rather clear, at least far from the singularity. It means that scalar matter in the fundamental, i.e. a section of the associated bundle, must transform as

$$\gamma\phi(z) = \phi(g(z)). \tag{47}$$

A particularly simple case is to take γ to be the regular representation, in which case we can consider $\phi(z)$ as a vector-valued field indexed by an element of Γ, so (47) becomes

$$\phi_{gh}(z) = \phi_h(g(z)). \tag{48}$$

This bundle is referred to as the "tautological bundle" over the quotient space. It can be decomposed as a direct sum over bundles R_i associated to irreps γ which if Γ is abelian are line bundles; these are the tautological line bundles.

Both types of fractional brane are labelled by a choice of group representation, and we can write a quiver theory summarizing the massless fermion content of any combination of these, again associating each brane to a quiver node. Let R_i be the D9 node corresponding to r_i and S^j be the D3 node corresponding to r_j, i.e. the original fractional branes.

The spectrum of (3, 9)-strings between a pair (R_i, S^j) is also determined by the orbifold projection. In fact, massless fermions with such boundary conditions (Dirichlet-Neumann boundary conditions in all the transverse dimensions) transform like scalars in $\mathbb{C}^3$, so this projection acts as

$$\gamma_3^{-1}(g)\chi\gamma_9(g) = \chi \tag{49}$$

so we have $n_{ij} = \delta_i^j$ such fermions in each sector. As in section 2, this implies that the intersection form between the two types of branes should be

$$\langle R_i, S^j\rangle = \delta_i^j. \tag{50}$$

This is the natural Poincaré duality on our noncompact space M, between $K(M)$ and the K theory of bundles with compact support $K_c(M)$ (meaning bundles over compact submanifolds) and we see that it indeed gives a perfect pairing.

This relation can then be used to determine the K theory classes of the S_j (and, given more formalism, even identify them as specific holomorphic objects). We need to know the intersection form for the R_i in an explicit basis to make this definition concrete. For example, if we have

$$\langle R_i, R_j \rangle \equiv (I^{-1})_{ij}, \tag{51}$$

then we can write

$$S^j = I^{ij} R_i \tag{52}$$

for which

$$\langle S^j, S^k \rangle = I^{jk}. \tag{53}$$

In terms of the K theory classes, (52) becomes

$$[S^j] = I^{ij}[R_i], \tag{54}$$

a simple explicit formula for the K theory classes of the fractional branes given those of the tautological line bundles.

In practice, we will restrict the bundles R_i from the total space M to the exceptional divisor, and then use $\chi(R_i, R_j)$ on this space as our intersection form in these formulas.

9.1 The $\mathbb{C}^3/\mathbb{Z}_3$ example

We have done almost all the work already if we can convince ourselves. that the R_i for $i = 1, 2, 3$ are in fact the bundles $O(i-1)$ of our previous discussion. This can also be seen fairly directly from the quiver diagram using the fact that the D0 is the object (1 1 1).

The moduli space of this theory is then the space M itself (as seen by the D0). If we are just interested in the $\mathbb{P}^2$, we can solve a simpler problem obtained by setting one set of the links to zero, say $Z^i = 0$. One already sees a $\mathbb{P}^2$ for the moduli space of (say) X^i (with appropriate signs of the FI terms; this went into the choice of which link to set to zero) and it is easy to check that the constraints $Y^{[i}X^{j]} = 0$ then determine $Y^i = X^i$ up to complex gauge equivalence.

The three branes R_i are then distinguished by which node S_i their associated link ψ_i is charged under. The line bundle interpretation of R_i is then determined by the transformation properties or the allowed values of ψ_i, which must be a section of the associated bundle. We can identify these by comparing the gauge invariant observables constructed from each, as we know that both X^i and Y^i correspond to the homogeneous coordinates. Only the relative transformation properties are defined; one can define one of the bundles to be $\mathcal{O}$.

Looking at the figure,

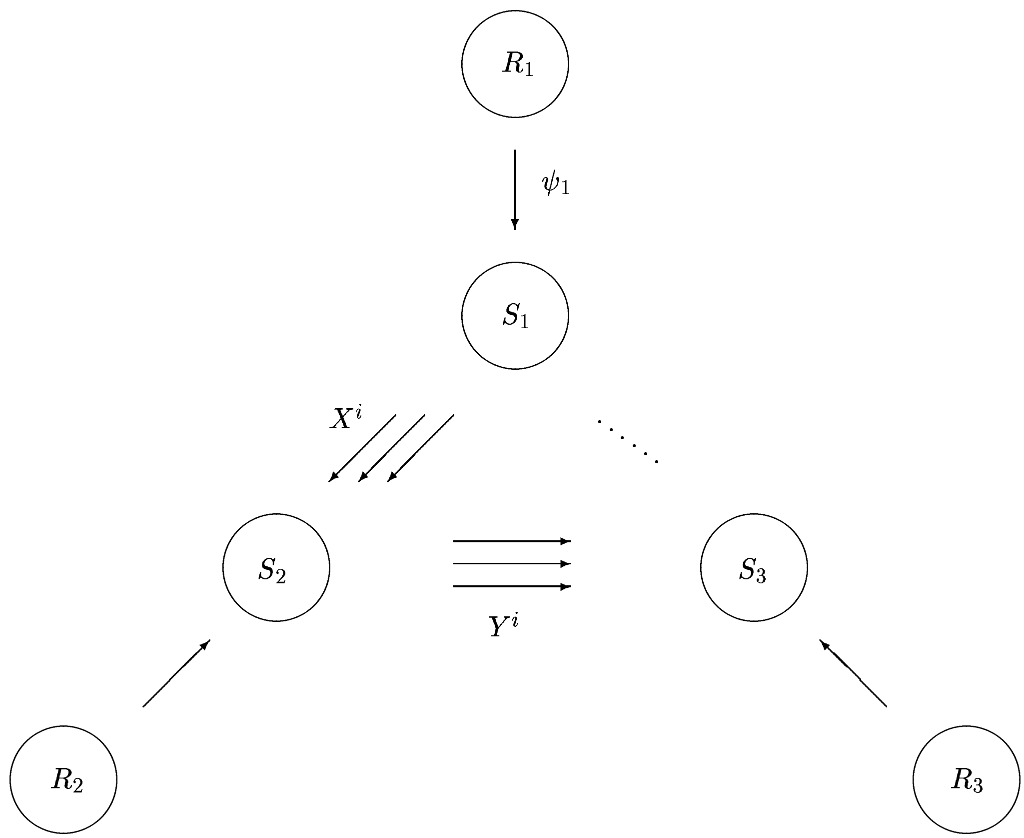


Fig.4. Dual bases

we see that gauge invariant combinations involving these variables are $\psi_1 X^i Y^i \sim \psi_2 X^i \sim \psi_3$. This implies that gauge invariant sections must look like $\psi_1 \sim 1$, $\psi_2 \sim x$, and $\psi_3 \sim x^2$, which establishes the claim.

Thus $\chi(R_i, R_j)$ is given by (44). We need to invert it; this is easy if we realize that it can also be thought of as multiplication of functions $a+bz+cz^2$ by a formal power series $(1-z)^{-3}$, so its inverse is $(1-z)^3$ or

$$\chi(S_i, S_j) = \begin{pmatrix} 1 & -3 & 3 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix}. \tag{55}$$

We can now implement (54) to find that

$$[S_3] = [R_1] = [\mathcal{O}] \tag{56}$$
$$[S_2] = [R_2] - 3[R_1] = e^J - 3 \tag{57}$$
$$[S_1] = [R_3] - 3[R_2] + 3[R_1] \tag{58}$$
$$= e^{2J} - 3e^J + 3 = e^{-J} + o(J^3). \tag{59}$$

We already knew that one of the fractional branes was $\mathcal{O}$ (this seems to be a very general result). The third relation is compatible with

$$S_1 = \mathcal{O}(-1)$$

and indeed a large volume monodromy avoiding the conifold point could clearly turn this into $\mathcal{O}$, so this is consistent with expectations.

The identity of the second brane may not be as obvious but there is a natural exact sequence which this formula suggests, and a much more developed framework in which this can be seen to be necessary. It is

$$0 \longrightarrow \bar{S}_2 \longrightarrow \mathcal{O}^3 \xrightarrow{f} \mathcal{O}(1) \longrightarrow 0. \tag{60}$$

The obvious map f takes a vector of three functions ψ_i (a section of $\mathcal{O}^3$) and produces $z^i\psi_i$. Thus a section of S_2 is a set of functions satisfying $z^i\psi_i = 0$. In fact one can see that this is the cotangent bundle twisted by (tensored with) a line bundle,

$$S_2 = \bar{\Omega}_{\mathbb{P}^2}(1) \equiv \bar{\Omega}_{\mathbb{P}^2} \otimes \mathcal{O}(1).$$

This is seen perhaps more easily in its dual form

$$0 \longrightarrow \mathcal{O} \longrightarrow \mathcal{O}(1)^3 \longrightarrow T_{\mathbb{P}^2} \longrightarrow 0$$

which says that a tangent vector to $\mathbb{P}^2$ can be written as

$$v^i_j z^j \frac{\partial}{\partial z^i} + \lambda z^i \frac{\partial}{\partial z^i}$$

where the choice of λ (a section of $\mathcal{O}$) drops out.

Note that S_2 is not a coherent sheaf but an "antibrane" to a coherent sheaf, with D4 charge -2. That such a thing would be necessary was already clear from (45) and does not seem so remarkable at first, but it is a sign that we are going to have to start thinking harder about bound states of branes and antibranes than we have done so far.

10 Moduli spaces of coherent sheaves on $\mathbb{P}^2$

Having identified the fractional branes, we are at a point where we can make nontrivial comparisons between large volume and the orbifold point. Recall that we have an extremely strong prediction: the holomorphic objects and their moduli spaces should be literally the same in these two limits.

The simplest thing to compare is the intersection form, or $\chi(E,F) - \chi(F,E)$. This is supposed to count massless fermions between pairs of branes, and the claim is that as we vary the Kähler moduli, the massless fermions must carry over unchanged between the two limits.

There is a subtlety in this interpretation of (30), however, because we added fields ρ^{ab} which were not present in the original quiver theory in making our definition. This interpretation is still correct however because we did not count the fermions in the links Z^c which we set to zero to simplify the analysis. These contribute exactly as the term we added, and in fact these two fields are "dual" in a clear sense.

Now $\chi(E,F)$ is (55) at large volume and (30) at the orbifold point, so indeed the intersection forms agree; in fact we have an even stronger statement that $\chi(E,F)$ itself agrees. The simplest argument that this had to happen is that the dimensions of moduli spaces must agree in the two limits, and this and the intersection form is enough information to reconstruct χ. In fact this is a very small part of the equivalence between the two descriptions: all objects, all moduli spaces, and all morphisms must agree as well.

In fact the equivalence between representations of the $\mathbb{C}^3/\mathbb{Z}_3$ quiver (with the Z link set to zero) and a large subset of sheaves on $\mathbb{P}^2$ had been observed in [14] and follows from a mathematical theorem due to Beilinson.

Without describing this in all detail, let us show how a quiver representation E can correspond to a sheaf $S(E)$. If all of the matrices in the quiver representation were zero, we know the correspondence from the above, we have

$$S_0(E) = \oplus N_i S_i = \sum_i V_i \otimes S_i$$

using notations established in section 3.

To incorporate the quiver configuration, we first note that there are natural maps $\hat{e}^a : S_i \longrightarrow S_{i+1}$ for $1 \leq a \leq 3$. For $i = 2$ this is clear from (60), and these are just the maps

$$\hat{e}^a(\psi_i) = \psi_a.$$

For $i = 1$, one needs to check that $\mathcal{O}(-1) \cong \Lambda^2\Omega(1)$; in other words a section of $\mathcal{O}(-1)$ can be written as a vector of three functions $\psi_{[ij]}$ satisfying $z^i\psi_{[ij]} = 0$. Then one has $\hat{e}^a(\psi_{[ij]})_k = \psi_{ak}$.

Because of the antisymmetrization in $\psi_{[ij]}$, these maps satisfy the relations

$$\hat{e}^a\hat{e}^b + \hat{e}^b\hat{e}^a = 0. \qquad (61)$$

Using these, the quiver configuration $X^a_{i,j}$ can be used to construct a natural operator on $S_0(E)$,

$$\hat{D} = \sum_a X^a \hat{e}^a.$$

From (61) and (27), one sees that $\hat{D}^2 = 0$, so the operator D has a cohomology, which is a sheaf. This is the sheaf which corresponds to the original quiver configuration. One can show that this relation is one to one; furthermore the construction can be reversed and used to show equivalences between all morphisms as well.

This provides a very detailed equivalence for many objects in the large volume limit, and at the orbifold point, and is the sense in which the quiver theory really does know about all the geometry of coherent sheaves at large volume. However, on reflection one can see that the set of quiver representations and the set of coherent sheaves on $\mathbb{P}^2$ cannot be literally identical. The simplest counterexample is the D2-brane. This also has a simple representation as an exact sequence,

$$0 \longrightarrow \mathcal{O}(-1) \xrightarrow{f} \mathcal{O} \longrightarrow \mathcal{O}_\Sigma \longrightarrow 0.$$

Here Σ is a $\mathbb{P}^1$ contained in $\mathbb{P}^2$ and $\mathcal{O}_\Sigma$ is its structure sheaf, the D2. The map f is just linear, $f = a_i z^i$, and Σ is the curve $f = 0$ in $\mathbb{P}^2$, leaving $\mathcal{O}_\Sigma$ as its cokernel.

The main point we want to make about this is that the D2 is a bound state of a fractional brane with a fractional antibrane, i.e. it is $(-1\ 0\ 1)$. Since we have a complete basis, there is no other way to make it. Physically, in fact, such a bound state cannot exist at the orbifold point, as its BPS central charge would have vanished. Thus there is no contradiction, but we see that the quiver representation framework as we have defined it so far cannot describe all holomorphic objects at all points in moduli space.

If one's goal is just to describe each brane separately, there are easy ways around this problem. For example, the D2 with a flux turned on, $\mathcal{O}_\Sigma(1)$, can be described with fractional branes – it is $(0\ 1\ 2)$ – and from large volume considerations we know this has the same moduli space as $\mathcal{O}_\Sigma$. In this sense, the quiver does give an adequate description of all sheaves, and is used for this purpose in their classification.

However, one would prefer to have a uniform description of all the branes. Worse, it looks like we have found a contradiction to the decoupling statement. However the agreement between the objects that do exist in both limits is so precise, one feels that there must be some way to extend it to these cases as well. We will eventually find that this can be done, by using the derived category.

11 Flow of gradings

As one explores the relationship between the quiver description of holomorphic branes and the coherent sheaf description, more subtle differences start to appear.

Let us consider the line bundles. Although $\chi(\mathcal{O}(m), \mathcal{O}(n))$ agrees between the two descriptions, the groups $\mathrm{Ext}^{p*}(\mathcal{O}(m), \mathcal{O}(n))$ as defined by the quiver theory do not always agree with the groups $H^p(M, \mathcal{O}(n-m))$ defined by sheaf theory. (We distinguish the quiver groups with a "*" because there is also a definition of $\mathrm{Ext}^p(A, B)$ for sheaves, which is what one would use in a more general discussion. It is equal to $H^p(M, A * \otimes B)$ if A and B are bundles.)

This is fairly clear without detailed analysis as the groups $\mathrm{Ext}^{p*}(\mathcal{O}(m), \mathcal{O}(n))$ do not in fact depend only on $n-m$. Let us look at some specific examples. One can check that $\mathcal{O}(-3) = (6\ 3\ 1)$, $\mathcal{O}(-2) = (3\ 1\ 0)$, $\mathcal{O}(-1) = (1\ 0\ 0)$, $\mathcal{O} = (0\ 0\ 1)$, $\mathcal{O}(1) = (0\ 1\ 3)$, and $\mathcal{O}(2) = (1\ 3\ 6)$. Then one can compute

$A \to B$	Ext^p	Ext^{p*}
$\mathcal{O}(-2) \longrightarrow \mathcal{O}(-1)$	3 0 0	3 0 0
$\mathcal{O}(0) \longrightarrow \mathcal{O}(1)$	3 0 0	3 0 0
$\mathcal{O}(0) \longrightarrow \mathcal{O}(-1)$	0 0 0	0 0 0
$\mathcal{O}(0) \longrightarrow \mathcal{O}(-2)$	0 0 0	0 0 0
$\mathcal{O}(-2) \longrightarrow \mathcal{O}$	6 0 0	0 0 6
$\mathcal{O}(-1) \longrightarrow \mathcal{O}(1)$	6 0 0	0 0 6
$\mathcal{O}(-1) \longrightarrow \mathcal{O}$	3 0 0	0 0 3
$\mathcal{O}(0) \longrightarrow \mathcal{O}(-3)$	0 0 1	1 0 0 .

The last of these is computed using Serre duality, which states that $H^p(\mathbb{P}^2, E) \cong H^{2-p}(\mathbb{P}^2, E^* \otimes O(-3))$.

These results show a simple pattern and can be summarized by the following rule. Let us introduce a notation where we put the gradings of the morphisms into the objects, as so:

$$\mathrm{Ext}^n(A, B) \equiv \mathrm{Hom}(A, B[n]) \equiv \mathrm{Hom}(A[m], B[m+n])$$

Then, to go from Ext to Ext*, all the branes $\mathcal{O}(n)$ with $n \geq 0$ become $\mathcal{O}(n)[1]$, while all the branes $\mathcal{O}(n)$ with $n < 0$ become $\mathcal{O}(n)[-1]$.

The explicit shifts $+1$ and -1 are determined by the following rule. Let the BPS central charges of a brane E at large volume be $Z(E;1)$ and at the orbifold point be $Z(E;2)$ (we could compare any two points in stringy Kähler moduli space). We define the "grade" of the brane E at a point x as

$$\varphi(E;x) = \frac{1}{\pi}\mathrm{Im}\,\log Z(E;x). \tag{62}$$

The "flow of grading" $E \to E[\Delta\varphi]$ from 1 to 2 is then determined as

$$\Delta\varphi = \varphi(E;2) - \varphi(E;1).$$

In this formula, the branches of the logarithm are determined by analytic continuation; defining this in (62) requires some additional discussion.

In the example, all of these branes have Z_2 real and positive, while $Z_1 = (-n + iV)^2$ is approximately real and negative, but with positive imaginary part for $n < 0$ and negative imaginary part for $n \geq 0$ (the case $n = 0$ is actually $n > 0$ as we are taking a path with decreasing V and $B > 0$ to avoid the conifold point). Furthermore the central charges for the two types of branes stay on their respective sides of the origin, positive or negative imaginary part, which leads to the rule we cited.

Although this rule may seem as if it was pulled out of a hat, it not only describes these explicit results but can be justified by a combination of physical and mathematical arguments.

The physical interpretation of the grading q of a morphism $\mathrm{Ext}^q(A, B)$ is that it is in fact the world-sheet $U(1)$ charge of the corresponding (bosonic) open string. Its most direct physical consequence is to determine the mass squared of the boson, which by conformal field theory arguments must be

$$m^2 = \frac{1}{2}(q-1)$$

in string units.

This leads to a very direct conformal field theory generalization of the type of stability argument we gave in our discussion of D-flatness conditions, and the necessary and sufficient conditions found there [15]. When we bring two branes (or a brane-antibrane pair) A and B together, we need to check the gradings of all of the morphisms $\mathrm{Ext}^q(A, B)$. If any satisfy $q < 1$, then bound state formation may be possible. Conversely, one can show that an object E goes unstable if the grading q of an $\mathrm{Ext}^q(A, B)$ between any of its quotient and subobjects goes above 1. This is because one of the Hom's in the triangle (18) would have its degree become negative, but negative $U(1)$ charges for chiral operators in unitary SCFT are not allowed, a contradiction which can only be resolved by the decay of E.

12 Antibranes and the derived category

The considerations of variation of central charge we made above have an even more striking consequence: namely, the distinction between "branes" and "antibranes" is not universal but in fact depends on where one is in Kähler moduli space.

This is not to say that there is any ambiguity in claiming that a particular $\bar{B}$ is "the" antibrane which annihilates B. However this is the only case in which there is no ambiguity. In other cases, the only clear distinction between brane and antibrane is the relative phase of the BPS central charge: for branes these are aligned or roughly aligned, while for brane and antibrane they are antialigned or roughly antialigned.

However, the central charges vary drastically as we move around the moduli space. Let us illustrate this with our example of $\mathbb{C}_3/\mathbb{Z}_3$. We saw that Z changes sign for $\mathcal{O}(n)$ as we go from large volume to orbifold point, so in some sense these branes become antibranes. The other fractional brane S_2 of course started out as an antibrane (meaning negative D4 charge) at large volume and indeed the phase of its central charge does not change during the flow, $\Delta\varphi = 0$, so it stays an antibrane. Thus we have a consistent picture in which all three fractional branes are simultaneously "branes" at the orbifold point. However, they are brane-antibrane pairs at large volume, and indeed using flow of gradings the bosons in $S_1 \longrightarrow S_2 \longrightarrow S_3$ can all be seen to be standard brane-antibrane tachyons which one expects at large volume.

Having realized that this distinction is so fluid, we now see that any description of all of the holomorphic objects which could make sense everywhere in moduli space had better treat branes and antibranes on a very equal footing, and indeed allow continuous evolution between them. It would seem rather hard to imagine such a thing, but as it turns out such a formalism already exists in mathematics, the formalism of the derived category. Indeed the observation that this should be relevant in describing D-branes on Calabi-Yau goes back to Kontsevich's homological mirror symmetry proposal of 1993 [24], so in some sense this part of the story predates D-branes!

It is possible to motivate (and in some sense "derive") the derived category as a systematic extension of the framework of topological open string theory to allow the BRST operator to have a general matrix (Chan-Paton) structure, as is done in [15, 3]. We will not get into this here but instead just describe the resulting formalism as it will appear in our primary application, that of finding the spectrum of BPS branes.

One starts with an abelian category of the sort we have been implicitly describing, of coherent sheaves, quiver representations, or whatever. One can think of the exact sequences (17) and the associated triangles (18) as the primary structure of interest. Since the last arrow in (18) was an Ext^1, however, we write it as

$$B \longrightarrow E \longrightarrow A \longrightarrow B[1].$$

Furthermore, since $\mathrm{Hom}(X, Y) \cong \mathrm{Hom}(X[n], Y[n])$, this sequence can be continued in both directions indefinitely:

$$\longrightarrow B \longrightarrow E \longrightarrow A \longrightarrow B[1] \longrightarrow E[1] \longrightarrow \ldots . \tag{63}$$

This is called a "distinguished triangle" and it plays the role of the exact sequence in the derived category. Note that it contains less information, however: the exact sequence picked out one object as special (the one in the middle), while the distinguished triangle does not. This is an advantage and a disadvantage for our purposes. It is an advantage because it unifies the various processes we discussed before of brane-brane and brane-antibrane bound state formation. It is also a disadvantage because one does not know which of the objects is the bound state and which are the constituents. This shows up mathematically in the statement that one cannot define a notion of "subobject;" indeed every morphism $B \longrightarrow E$ can be completed to a distinguished triangle (63) for some A (called the "cone" of the morphism).

However, this is the universal structure which remains invariant under variations of Kähler moduli. The precise statement of the theorem of Beilinson we referred to, and many similar results on Calabi-Yau monodromies and Fourier-Mukai transforms, is that the natural equivalences and monodromy actions in general do not take sheaves to sheaves, or any other known subclass of holomorphic objects into itself, but instead act on the derived category.

Variation of the Kähler moduli has only two effects on this structure. First, it induces the flow of gradings we discussed. This preserves the only essential constraint on the gradings of the morphisms in (63), namely that they sum to 1, and the two ideas fit very naturally together. Second, it changes the stability of objects, in some way generalizing the orbifold point and large volume phenomena we discussed. We now turn to this.

13 Π-stability

The construction of the derived category now gives a precise meaning to the decoupling statement, at least on the holomorphic side. On the other hand, the flow of gradings we discussed only depended on BPS central charges, and since these are geometric in the A picture, if we could base our discussion of stability only on these, we would have effectively implemented it on the other side.

Let us say a bit more about this. We have a good (though still somewhat abstract) description of the set of all possible F-flat configurations for orbifolds and Calabi-Yaus with a Gepner model realization, as the derived category of representations of a McKay quiver. We could go on to try to formulate and solve analogous stringy D-flatness conditions. However, the discussion we gave of how to find $\mathcal{N} = 1$ supersymmetric vacua suggests a simpler strategy. The procedure we ended up with was to find F flat configurations or objects, but then instead of solving the D flatness conditions, we instead found a necessary and sufficient criterion for such a solution to exist, the θ-stability condition. Furthermore, the DUY theorem shows that the problem of describing BPS branes at large volume can be stated in precisely the same paradigm, we first find holomorphic bundles or objects and then check their stability. Finally, we have now discussed the sense which the holomorphic objects in the B picture are the same in these two limits and indeed everywhere in Kähler moduli space.

All this suggests that we rephrase the problem. Instead, we will try to find a stringy version of the stability condition, which reduces to the conditions we already saw in the large volume and orbifold limits.

Such a condition can be found and is called "Π-stability." We will just state it without the detailed definitions and arguments, which can mostly be found in [13, 15, 1].

We start with a simplified version of Π-stability which was proposed in [13] and is adequate for problems not involving both branes and antibranes. It is essentially to replace the stability conditions (24) at the orbifold point and (40) with the single condition that E is stable if for every subobject E' of E,

$$\varphi(E') < \varphi(E). \tag{64}$$

All of the dependence on Kähler moduli is contained in (62).

This is good when one can define subobject, but there is no concept of subobject in the derived category. Furthermore, comparison of the definitions of subobject in large volume and at the orbifold point shows that they are different (for example, $\mathcal{O}$ is a subobject of $\mathcal{O}(-3)$ at the orbifold point). Thus we must get by without it.

A refined version of Π-stability which can treat this problem was proposed in [15, 1]. One has to start with a list of stable objects, which might be found at large volume or at the orbifold point using the previous definitions. The stability condition is then the following: two stable objects A and B cannot participate in morphisms of negative degree. Taking into account the definition of flow of gradings, this is essentially equivalent to (64), but the difference comes when we cross a line on which this condition is violated.

One can check that the definition of φ in terms of the phases of central charges means that if one of the morphisms in (63) has grade 0, the others must be integral. If the three objects involved were stable, one will be grade 0 and the other will be grade 1. The rule is then that the object between the 0's decays. This is always the heaviest of the three objects, so physically there is no doubt that this is the correct rule.

Conversely, if a morphism between two stable objects drops in degree below 1, the third element of the corresponding distinguished triangle (the "cone") becomes stable. This corresponds to a massless boson becoming tachyonic.

13.1 Examples

A number of examples of these rules are worked out in [15, 16, 1]. Another simple example can already be understood at large volume, namely the decay of a high degree D$n-2$-brane. Consider a compact CY_3 and the following exact sequence:

$$0 \longrightarrow \mathcal{O}(-\frac{N}{2}) \stackrel{f}{\longrightarrow} \mathcal{O}(\frac{N}{2}) \longrightarrow \mathcal{O}_\Sigma \longrightarrow 0. \tag{65}$$

The map f is a polynomial of degree N and generically vanishes on a nonsingular hypersurface of degree N, i.e. a brane with D4 charge N. The total charge is $e^{NJ/2} - e^{-NJ/2}$ and one sees that this brane has zero D2 charge but D0 charge of order N^3.

According to the central charge formula (39), such a brane has central charge

$$Z = -3NV^2 + \frac{N^3}{4}.$$

Although it is large at large volume, $Z = 0$ at $V = N/\sqrt{12}$ in string units. If N is large, this is clearly a nonsingular point in moduli space, so the brane must decay before reaching it. Furthermore, world-sheet instantons are clearly unimportant at this scale, so there is no loophole in this.

The natural exact sequence which might govern the decay of this brane is just (65) as we don't know of any others in general. The map f, as a Hom, has degree 0 in large volume (this is the brane-antibrane tachyon of $m^2 = -1/2$). However, as we decrease V, the central charges $(-N/2+iV)^3$ and $(N/2+iV)^3$ will vary in precisely the same way we described above for the line bundles $\mathcal{O}(n)$ in $\mathbb{C}^3/\mathbb{Z}_3$, and with the same effect: the morphism f will increase in degree until it reaches 1, at which point the D4 will decay. This will happen when these two central charges antialign (so the brane and antibrane charges align), i.e.

$$\begin{aligned} 0 &= (-N/2+iV)^3 + (N/2+iV)^3 \qquad (66)\\ &= (iV)^3 + 3iV\left(\tfrac{N}{2}\right)^2, \qquad (67)\end{aligned}$$

which is at $V = \sqrt{3}N/2$, long before the problematic point.

The same thing happens in the known examples with instanton corrections; the "mysterious brane" in the quintic Gepner model, and even the D2 on $\mathbb{C}^3/\mathbb{Z}^3$, are very similar examples.

14 Parting words

In these lectures, we have given an introduction to a framework for studying and classifying BPS branes in string theory compactified on Calabi-Yau manifolds which, although not complete, has achieved a definite form in which concrete problems can be solved. This work has also shed new light on the structure of $\mathcal{N} = 1$ supersymmetry and provides new methods for studying $\mathcal{N} = 1$ theories, inspired both by the physics of branes and by modern mathematics.

Developments continue in the various directions we discussed; let us mention a few. First, one should be able to get a much better picture of the geometry of string-scale Calabi-Yaus from the behavior of the spectrum of stable BPS branes, and particularly from the D0-brane, in the spirit of "D-geometry" [11, 12]. A good example of this is that the connections between topologically distinct Calabi-Yaus which were visible from the linear sigma model and toric geometry [29, 2] could be rederived by seeing how D0-brane moduli spaces change under variation of stability. Second, one should be able to completely understand Seiberg duality, at least as an equivalence between classical moduli spaces of dual $\mathcal{N} = 1$ effective field theories, along the lines of [4, 18, 9], as examples of Fourier-Mukai transforms, which are the general symmetries of the category of branes on a Calabi-Yau, and thus fit them into a larger, stringy framework.

We feel that a very promising longer term direction for this work is to extend it to describe $\mathcal{N} = 1$ compactifications with branes which make sense as quantum theories (i.e. cancel anomalies), either in type I theory, in type II orientifold theories, or eventually in more general contexts and with quantum corrections, perhaps by using dualities with these constructions. Given our experience with duality, it does not seem unreasonable to hope that a classical moduli space of D-brane configurations could be equivalent to the exact quantum moduli space of some (perhaps very different looking) dual theory. We might look at these constructions as providing "solvable" $\mathcal{N} = 1$ models, somewhat analogous to the role Calabi-Yau compactification has played in the study of $\mathcal{N} = 2$ models.

It seems to us that systematic exploration of $\mathcal{N} = 1$ vacua is the central problem for string/M theory in the coming years, and it seems to us that this is the first framework which gives any usable description at all of complete nontrivial moduli spaces of $\mathcal{N} = 1$ supersymmetric theories arising from string theory, so it will be exciting to see if these ideas can be extended in these ways.

Acknowledgments. These ideas were developed in discussions with many people, and particularly in collaborations with P. Aspinwall, D.-E. Diaconescu, B. Fiol and C. Römelsberger.

This research was supported in part by DOE grant DE-FG05-90ER40559.

References

[1] P. S. Aspinwall and M. R. Douglas, *D-Brane Stability and Monodromy,* hep-th/0110071.

[2] P. Aspinwall, B. R. Greene and D. R. Morrison, Nucl. Phys. B420 (1994) 184-242; hep-th/9311042.

[3] P. S. Aspinwall and A. Lawrence, *Derived Categories and Zero-Brane Stability*, JHEP 0108 (2001) 004, hep-th/0104147.

[4] C. E. Beasley, M. R. Plesser, *Toric Duality Is Seiberg Duality*, JHEP 0112 (2001) 001, hep-th/0109053.

[5] T. Bridgeland, A. King, and M. Reid, *Mukai implies McKay,* math.AG/9908027.

[6] I. Brunner and J. Distler, *Torsion D-Branes in Nongeometrical Phases*, hep-th/0102018.

[7] I. Brunner, J. Distler, and R. Mahajan *Return of the Torsion D-Branes*, hep-th/0106262.

[8] I. Brunner, M. R. Douglas, A. Lawrence, and C. Römelsberger, *D-branes on the quintic,* JHEP 08(2000)015, hep-th/9906200.

[9] F. Cachazo, B. Fiol, K. Intriligator, S. Katz and C. Vafa, *A Geometric Unification of Dualities*, hep-th/0110028.

[10] D.-E. Diaconescu and J. Gomis, *Fractional Branes and Boundary States in Orbifold Theories*, JHEP 0010 (2000) 001, hep-th/9906242.

[11] M. R. Douglas, *Two Lectures on D-Geometry and Noncommutative Geometry,* in *Nonperturbative Aspects of Strings, Branes and Supersymmetry,* World Scientific (1999), hep-th/9901146.

[12] M. R. Douglas, *Topics in D-geometry,* Class. Quant. Grav. **17** (2000) 1057, hep-th/9910170.

[13] M. R. Douglas, B. Fiol, and C. Römelsberger, *Stability and BPS branes,* hep-th/0002037.

[14] M. R. Douglas, B. Fiol, and C. Römelsberger, *The spectrum of BPS branes on a noncompact Calabi–Yau,* hep-th/0003263.

[15] M. R. Douglas, *D-branes, Categories and $N = 1$ Supersymmetry,* J. Math. Phys. 42 (2001) 7, 2818-2843, hep-th/0011017.

[16] M. R. Douglas, *D-Branes and N=1 Supersymmetry,* to appear in the proceedings of Strings 2001, Mumbai, India, hep-th/0105014.

[17] M. R. Douglas and D.-E. Diaconescu, *D-branes on Stringy Calabi-Yau Manifolds,* hep-th/0006224.

[18] B. Feng, A. Hanany, Y.-H. He and A. M. Uranga, *Toric Duality as Seiberg Duality and Brane Diamonds*, hep-th/0109063.

[19] B. Fiol, *The BPS Spectrum of N=2 SU(N) SYM and Parton Branes*, hep-th/0012079.

[20] P. Gabriel, *Unzerlegbare Darstellungen I*, Manuscr. Math. 6 (1972) 71-103.

[21] P. Griffiths and J. Harris, *Principles of Algebraic Geometry*, Wiley (1978).

[22] S. I. Gelfand and Yu. I. Manin, *Homological Algebra*, Springer 1999.

[23] A. D. King, *Moduli of Representations of Finite Dimensional Algebras*, Quart. J. Math. Oxford (2), 45 (1994), 515-530.

[24] M. Kontsevich, *Homological Algebra of Mirror Symmetry,* Proceedings of the 1994 ICM, alg-geom/9411018.

[25] M. Marino, R. Minasian, G. Moore and A. Strominger, *Nonlinear Instantons from Supersymmetric p-Branes,* JHEP 0001 (2000) 005, hep-th/9911206.

[26] R. Minasian and A. Tomasiello, *Variations on stability,* hep-th/0104041.

[27] J. Polchinski, *TASI Lectures on D-branes,* hep-th/9611050.

[28] M. Reid, *La correspondance de McKay,* Séminaire Bourbaki (novembre 1999), no. 867, math.AG/9911165.

[29] E. Witten, *Phases of N=2 theories in two dimensions,* Nucl. Phys. **B403** (1993) 159, hep-th/9301042.

Noncommutative Solitons

Rajesh Gopakumar*

Harish-Chandra Research Institute, Jhusi, Allahabad, India

Lectures given at the
Spring School on Superstrings and Related Matters
Trieste, 2 – 10 April 2001

LNS027006

*gopakumr@mri.ernet.in

Abstract

These are pedagogical lectures on solitons in noncommutative field theories delivered at the Spring School, March'01.

Contents

1 Introduction

Though noncommutative field theories have been explored for several years, a resurgence of interest in it was sparked off after it was realised that they arise very naturally as limits of string theory in certain background fields [1]. It became more plausible (at least to string theorists) that these nonlocal deformations of usual quantum field theories are consistent theories in themselves. This led to a detailed exploration of many of their classical and quantum properties. I will elaborate further on the string theory context in the next section.

One of the consequences of this exploration was the discovery of novel classical solutions in noncommutative field theories [2]. Since then much work has been done in exploring many of their novel properties. My lectures focussed on some specific aspects of these noncommutative solitons. They primarily reflect the topics that I have worked on and are not intended to be a survey of the large amount of work on this topic. Some reviews that give a more comprehensive list of references are [3],[4].

2 The context

Here we will try to provide the context in which the study of noncommutative field theories and their classical solutions assume importance.

The Importance of Open Strings: The understanding of the role of open strings in string theory via D-branes, has proven to be a development of overwhelming importance. This understanding was instrumental in correctly counting black hole microstates, one of the dramatic successes of string theory. One of the surprises was the manner in which a purely gravitational phenomenon, like black hole entropy, was described in terms of open strings, which at least classically don't contain closed string excitations like the graviton.

This connection between open and closed strings was sought to be further exploited in the Matrix theory proposal for a DLCQ description of M-Theory. But it's most striking manifestation was the AdS/CFT duality of Maldacena relating large N gauge theories to purely closed string theories. This conjecture is a reflection of an underlying duality between open and closed strings which is yet to be completely understood.

Decoupling limits: In the AdS/CFT duality, one takes a certain scaling limit of open string theories living on D-branes in which only the massless gauge theory modes survive and are described by a (super) Yang-Mills lagrangian. The massive open string states are effectively decoupled by taking the string scale to infinity. This scaling limit of open string theories is conjectured to describe pure closed strings propagating in the near horizon geometry of the D-branes. The fact that one can gain nontrivial information from studying a simple field theory limit of string theory has led one to examine more closely the various decoupling limits of string theory. (Cf. Kutasov's lectures.) Taking decoupling limits of different sorts also help one to focus more sharply on various aspects of string theory. The idea is to get a limit which is easier to analyse than the full theory, but which nevertheless retains enough of the complexity.

Noncommutativity and String Field Theory: In parallel with these developments, and at first sight unrelated to it, is an ambitious program initiated by Sen which attempts, among other things, to understand closed strings in terms of open strings. The idea is to use the formulation of open string interactions in terms of a cubic string field theory as a complete description of string theory. This formulation relies on a representation of open string interactions which consists of gluing them in a fundamentally noncommutative way [5]. This defines an associative but noncommutative product of string fields in terms of which the string field action is expressed. D-branes are nontrivial classical solutions of this action while closed strings could arise as some kind of quantum excitations.

Since noncommutativity is thus in some sense intrinsic to string theory (and not just a property of some backgrounds) and perhaps plays a crucial role in understanding the notions that replace classical geometry, it is worthwhile to try and understand it better.

However, when one takes the conventional field theory limit of open string theory, the remnant of the noncommutativity is the somewhat trivial matrix algebra of the Chan-Paton indices. It does not involve the noncommutativity that comes from the extended nature of the open string.

Noncommutative Field theories: One might therefore ask if there is a limit of string theory which has the relative simplicity of keeping only a field theoretic number of degrees of freedom and yet displays the extended nature of strings. In particular, it should capture some of the nontrivial noncommutativity of open string interactions. It turns out that the answer is yes. One can obtain a nonlocal deformation of field theories by taking a decoupling limit of open strings in a large magnetic field [1], [6]. The massive string modes decouple leaving a kind of elastic dipole object.

These resulting noncommutative field theories will be the main topic of these lectures.

Noncommutative solitons: More specifically, we will study the classical limit of these noncommutative field theories and find finite energy soliton solutions that have no counterpart in local field theories. Among the nice features of these solitons is that they are fairly universal and more or less insensitive to the details of the theory. They exhibit various novel features like nonabelian enhancement of symmetry when they are coincident.

In fact, these solitons are really the D-branes of string theory manifested in a field theory. This is somewhat surprising as it does not happen that you can find D-branes as finite energy excitations in a conventional field theory limit of string theory. The simplicity of noncommutative solitons implies that one can study many properties of D-branes very explicitly in this context.

Therefore the motivation for studying these solitons will be to use them as a simple set of probes of stringy behaviour in a well controlled manner. Much of the applications have been in the context of issues of tachyon condensation in open string theory. We can however also use these solitons to probe issues of how D-branes see space time, for instance.

Finally, the field theoretic aspects of these solitons are interesting in themselves and might perhaps have applications in very different contexts such as in the Quantum Hall effect.

3 Strings in a large magnetic field

As a prelude to studying strings in a large magnetic field, let us look at point particles in a large magnetic field.

The action for (nonrelativistic) point particles reads as

$$S = \int dt \left(\frac{1}{2} m \dot{x}_\mu \dot{x}^\mu + e B_{\mu\nu} x^\mu \dot{x}^\nu \right). \tag{3.1}$$

The conjugate momentum Π_μ to x^μ is

$$\Pi_\mu = m \dot{x}_\mu + e B_{\mu\nu} x^\nu. \tag{3.2}$$

In the limit where the energy $\omega \ll \frac{e|B|}{m}$, the canonical commutation relations become simply

$$[x^\mu, x^\nu] = i (B^{-1})^{\mu\nu} \frac{m}{e}. \tag{3.3}$$

Thus at energies much less than the cyclotron frequency $\frac{e|B|}{m}$, when one is in the lowest Landau level, one effectively has noncommuting coordinates. This is why the physics of the quantum hall effect displays some features of noncommutativity.

Now write the action for an open string in a constant magnetic field. We assume that the open string ends on a p brane in some of whose worldvolume directions the magnetic field is switched on.

$$S = \frac{1}{4\pi\alpha'}\int_{\Sigma} d^2\sigma \left(g_{\mu\nu}\partial_\alpha X^\mu \partial^\alpha X^\nu - 2\pi i\alpha' B_{\mu\nu}\epsilon^{\alpha\beta}\partial_\alpha X^\mu \partial_\beta X^\nu\right). \qquad (3.4)$$

The additional term involving B is really a boundary term which couples to the charges at the end of the open string like a constant magnetic field.

It leads to boundary conditions in the directions along the brane which are mixed.

$$\left(g_{\mu\nu}\partial_n X^\nu + 2\pi i\alpha' B_{\mu\nu}\partial_t X^\nu\right)|_{\partial\Sigma} = 0. \qquad (3.5)$$

One can write down the Green's functions on the disc worldsheet with these boundary conditions. What we will need is the particular case when the X's are at on the boundary of the disc (parametrised by τ).

$$< X^\mu(\tau)X^\nu(\tau') >= -\alpha' G^{\mu\nu}\ln(\tau-\tau')^2 + \frac{i}{2}\Theta^{\mu\nu}\epsilon(\tau-\tau'). \qquad (3.6)$$

Here

$$\begin{aligned} G^{\mu\nu} &= \left(\frac{1}{g+2\pi\alpha' B}g\frac{1}{g-2\pi\alpha' B}\right)^{\mu\nu} \\ \Theta^{\mu\nu} &= -(2\pi\alpha')^2\left(\frac{1}{g+2\pi\alpha' B}B\frac{1}{g-2\pi\alpha' B}\right)^{\mu\nu} \end{aligned} \qquad (3.7)$$

are usually called the open string metric and the noncommutativity parameter [6]. The open string metric is what determines the mass shell condition for open string states. Θ is called the noncommutativity parameter since the above OPE essentially implies that

$$[X^\mu(\tau), X^\nu(\tau)] = i\Theta^{\mu\nu}. \qquad (3.8)$$

Note that Θ has dimensions of length^2.

There is one more ingredient, namely that the effective coupling of open string modes is also rescaled by a factor that depends on the magnetic field. We will not need the exact expression until later.

The noncommutativity parameter leads to an extra term in the OPE of open string vertex operators $e^{ik\cdot X}$:

$$e^{ik_1\cdot X}(\tau)e^{ik_2\cdot X}(\tau') \sim (\tau-\tau')^{2\alpha' G^{\mu\nu}k_{1\mu}k_{2\nu}} e^{-i\frac{1}{2}\Theta^{\mu\nu}k_{1\mu}k_{2\nu}} e^{i(k_1+k_2)\cdot X}(\tau') + \dots \qquad (3.9)$$

The additional term $e^{-i\frac{1}{2}\Theta^{\mu\nu}k_{1\mu}k_{2\nu}}$ can be understood in position space as giving a nonlocal interaction which is expressed in terms of the Moyal product.

$$(f \star g)(x) = e^{i\frac{1}{2}\Theta^{\mu\nu}\partial_\mu\partial'_\nu} f(x)g(x')|_{x=x'}. \tag{3.10}$$

In general, there will be such a phase factor for all vertex operators implying that the effect of the magnetic field on the effective action for open string modes in spacetime is completely captured by replacing all local products by the Moyal products, if we additionally remember to make all metric contractions with the open string metric (note that it is the open string metric that appears in (3.9) in the anomalous dimension of the vertex operators).

We can now take the equivalent of the limit of a large magnetic field, namely take $\alpha'|B| \gg 1$. Here $|B|^2 = B_{\mu\nu}B_{\rho\sigma}g^{\mu\rho}g^{\nu\sigma}$. We will in addition demand that this limit is taken keeping the open string metric $G^{\mu\nu}$ and $\Theta^{\mu\nu}$ finite. This requires taking the string scale to infinity ($\alpha' \to 0$). In the absence of the magnetic field this would mean decoupling all the massive string modes giving a field theory of the zero mode (if we keep the coupling constant finite).

With the magnetic field, as we have seen the only effect is to replace local products with the moyal product. The terms involving massive modes (both open and closed) then decouple for the same reason as in the case without a magnetic field. The lowest open string modes then interact via a nonlocal deformation of ordinary field theory.

4 Scalar noncommutative solitons

With these motivations we will start our study of semiclassical noncommutative field theories. This section closely follows the discussion in [2]. The simplest example is a theory of a single scalar field in $2+1$ dimensions with noncommutativity in the two spatial directions. Though this does not necessarily arise as any decoupling limit of string theory, the classical solutions we find are generic to noncommutative field theories.

We will parametrize the spatial R^2 by complex coordinates $z, \bar{z}$. The energy functional for static configurations is

$$E = \frac{1}{g^2}\int d^2z\, (\partial_z\phi\partial_{\bar{z}}\phi + V(\phi)_\star)\,, \tag{4.1}$$

where $d^2z = dxdy$. Fields in the action are multiplied using the Moyal star product (which reads in complex form as),

$$(f \star g)\,(z,\bar{z}) = e^{\frac{\theta}{2}(\partial_z\partial_{\bar{z}'} - \partial_{z'}\partial_{\bar{z}})} f(z,\bar{z})g(z',\bar{z}')|_{z=z'}. \tag{4.2}$$

Note that since $\int f \star g = \int fg$, the Moyal product drops out of the quadratic term in the action.

Before we look for classical solutions to this action, let us recall that the scalar theory without noncommutativity does not have any lump solutions. This is actually true for any bounded potential in spatial dimension greater than one, and follows from a simple scaling argument of Derrick [7]. If $\phi_0(x)$ be an extremum of the energy functional (4.1) (with $\theta = 0$), then consider the energy of the field configurations $\phi_\lambda(x) = \phi_0(\lambda x)$.

$$\begin{aligned} E(\lambda) &= \frac{1}{g^2}\int d^D x\left(\frac{1}{2}(\partial\phi_0(\lambda x))^2 + V(\phi_0(\lambda x))\right) \\ &= \frac{1}{g^2}\int d^D x\left(\frac{1}{2}\lambda^{2-D}(\partial\phi_0(x))^2 + \lambda^{-D}V(\phi_0(x))\right). \end{aligned} \tag{4.3}$$

Since $\phi_0(x)$ is an extremum, we require $\frac{\partial E(\lambda)}{\partial\lambda}|_{\lambda=1} = 0$. that is,

$$\int d^D x\left(\frac{1}{2}(D-2)(\partial\phi_0(x))^2 + DV(\phi_0(x))\right) = 0.$$

For spatial dimension $D \geq 2$, for a potential bounded from below by zero, the only way this can be true is for the kinetic and the potential terms to separately vanish. There are therefore no nontrivial configurations. Note that this argument fails once one includes higher derivative terms.

We now seek finite energy (localized) solitons of (4.1) for nonzero θ. Since no solutions exist for $\theta = 0$ (4.3), we begin our search in the limit of large noncommutativity, $\theta \to \infty$. It is useful to non-dimensionalize the coordinates $z \to z\sqrt{\theta}$, $\bar{z} \to \bar{z}\sqrt{\theta}$. As a result, the $\star$ product will henceforth have no θ; i.e. it will be given by (4.2) with $\theta = 1$. Written in rescaled coordinates, the dependence on θ in the energy is entirely in front of the potential term:

$$E = \frac{1}{g^2}\int d^2 z\left(\frac{1}{2}(\partial\phi)^2 + \theta V(\phi)_\star\right) \tag{4.4}$$

In the limit $\theta \to \infty$, with V held fixed, the kinetic term in (4.4) is negligible in comparison to $V(\phi)$, at least for field configurations varying over sizes of order one in our new coordinates.

Our considerations apply to generic potentials $V(\phi)$, but we will, for definiteness, mostly discuss those of polynomial form

$$V(\phi) = \frac{1}{2}m^2\phi^2 + \sum_{j=3}^{r}\frac{b_j}{j}\phi^j. \tag{4.5}$$

4.1 Solutions in the $\theta = \infty$ limit

After neglecting the kinetic term, the energy

$$E = \frac{\theta}{g^2} \int d^2 z V(\phi)_\star, \tag{4.6}$$

is extremised by solving the equation

$$\left(\frac{\partial V}{\partial \phi} \right)_\star = 0. \tag{4.7}$$

For instance, for a cubic potential one has to solve an equation of the form

$$m^2 \phi + b_3 \phi \star \phi = 0. \tag{4.8}$$

If $V(\phi)$ were the potential in a commutative scalar field theory, the only solutions to (4.7) would be the constant configurations

$$\phi = \lambda_i, \tag{4.9}$$

where $\lambda_i \in \{\lambda_1, \lambda_2, \cdots, \lambda_k\}$ are the various real extrema of the function $V(x)$. The derivatives in the definition of the star product allow for more interesting solutions of (4.7).

In order to find all solutions of (4.7) we will exploit the connection between Moyal products and quantization. Given a C^∞ function $f(q,p)$ on R^2 (thought of as the phase space of a one-dimensional particle), there is a prescription which uniquely assigns to it an operator $\widehat{f}(\widehat{q}, \widehat{p})$, acting on the corresponding single particle quantum mechanical Hilbert space, $\mathcal{H}$. It is convenient for our purposes to choose the Weyl or symmetric ordering prescription

$$\widehat{f}(\widehat{q}, \widehat{p}) = \frac{1}{(2\pi)^2} \int d^2 k \widetilde{f}(k) e^{-i(k_q \widehat{q} + k_p \widehat{p})}, \tag{4.10}$$

where

$$\widetilde{f}(k) = \int d^2 x e^{i(k_q q + k_p p)} f(q,p), \tag{4.11}$$

and

$$[\widehat{q}, \widehat{p}] = i. \tag{4.12}$$

With this prescription, it may be verified that

$$\frac{1}{2\pi}\int dpdq f(q,p) = \mathrm{Tr}_{\mathcal{H}}\widehat{f}, \tag{4.13}$$

and that the Moyal product of functions is isomorphic to ordinary operator multiplication

$$\widehat{f}\cdot\widehat{g} = \widehat{f\star g}. \tag{4.14}$$

In order to solve any algebraic equation involving the star product, it is thus sufficient to determine all operator solutions to the equation in $\mathcal{H}$. The functions on phase space corresponding to each of these operators may then be read off from (4.10). We will now employ this procedure to find all solutions of (4.7).

It is easy to see that $\widehat{\phi} = \lambda_i P$ is a solution to $V'(\widehat{\phi}) = 0$, if P is an arbitrary projection operator on some subspace of $\mathcal{H}$ and if λ_i is an extremum of $V(x)$. The energy of this solution is, using (4.13),

$$E = \frac{2\pi\theta}{g^2}\mathrm{Tr}V(\widehat{\phi}) = \frac{2\pi\theta}{g^2}V(\lambda_i)\mathrm{Tr}P. \tag{4.15}$$

Thus the energy is finite if P is projector onto a finite dimensional subspace of $\mathcal{H}$.

In fact, you can convince yourself that the most general solution to (4.7) takes the form

$$\widehat{\phi} = \sum_j a_j P_j \tag{4.16}$$

where $\{P_j\}$ are mutually orthogonal projection operators onto one dimensional subspaces,

$$P_iP_j = \delta_{ij}P_j; \qquad Tr_{\mathcal{H}}P_i = 1, \tag{4.17}$$

with a_j taking values in the set $\{\lambda_i\}$ of real extrema of $V(x)$.

From now on we will restrict ourselves to a potential with one nontrivial minimum λ other than the one at the origin.

We have a huge infinity of solutions of the form λP. To see what they mean, let us translate them into position space. It will be convenient for this purpose to choose a particular basis in $\mathcal{H}$. Let $|n\rangle$ represent the energy eigenstates of the one dimensional harmonic oscillator whose creation and annihilation operators are defined by

$$a = \frac{\widehat{q}+i\widehat{p}}{\sqrt{2}}; \quad a^\dagger = \frac{\widehat{q}-i\widehat{p}}{\sqrt{2}}. \tag{4.18}$$

Note that $a|n\rangle = \sqrt{n}|n-1\rangle$ and $a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle$. Any operator may be written as a linear combination of the basis operators $|m\rangle\langle n|$'s, which, in turn, may be expressed in terms of a and $a^\dagger$ as

$$|m\rangle\langle n| =: \frac{a^{\dagger m}}{\sqrt{m!}} e^{-a^\dagger a} \frac{a^n}{\sqrt{n!}} : \tag{4.19}$$

where double dots denote normal ordering. We will first describe operators of the form (4.16) that correspond to radially symmetric functions in space. As $a^\dagger a \approx \frac{r^2}{2}$, operators corresponding to radially symmetric wavefunctions are functions of $a^\dagger a$. From (4.19), the only such operators are linear combinations of the diagonal projection operators $|n\rangle\langle n| = \frac{1}{n!} : a^{\dagger n} e^{-a^\dagger a} a^n :$. Hence all radially symmetric solutions of (4.7) correspond to operators of the form $\widehat{\phi} = \lambda \sum a_n |n\rangle\langle n|$, where the numbers a_n can take values 0 or 1.

It is not difficult to translate these operators back to position space [2]. One finds

$$|n\rangle\langle n| = \frac{1}{(2\pi)} \int d^2k e^{\frac{-k^2}{4}} L_n(\frac{k^2}{2}) e^{-i\left(k_{\bar{z}} a + k_z a^\dagger\right)} \tag{4.20}$$

where $L_n(x)$ is the n^{th} Laguerre polynomial. The field $\phi_n(x, y)$ that corresponds to the operator $\widehat{\phi}_n = |n\rangle\langle n|$ is, therefore,

$$\phi_n(r^2 = x^2 + y^2) = \frac{1}{(2\pi)} \int d^2k e^{\frac{-k^2}{4}} L_n(\frac{k^2}{2}) e^{-ik.x} = 2(-1)^n e^{-r^2} L_n(2r^2). \tag{4.21}$$

Note that $\phi_0(r^2)$ is the simple gaussian $2e^{-r^2}$. In summary, (4.7) has an infinite number of real radial solutions, given by

$$\sum_{n=0}^{\infty} a_n \phi_n(r^2) \tag{4.22}$$

where $\phi_n(r^2)$ is given by (4.21) and each a_n takes values either 0 or 1. These solutions will have finite energy if only a finite number of the a_n are nonzero, as is evident from (4.15).

We also see from (4.15) that the action at $\theta = \infty$ has a large symmetry $\widehat{\phi} \to U\widehat{\phi}U^\dagger$, where U is any unitary operator acting on $\mathcal{H}$. This $U(\infty)$ global symmetry generates new nonradially symmetric solutions out of the radially symmetric ones. The most general projection operator $\widehat{\phi} = \lambda P$, of rank k, is unitarily related to a projection operator which is diagonal (in the

SHO basis), that is of the form $\lambda(\sum_{i=0}^{k-1}|i><i|)$. And the corresponding solutions are all degenerate in energy. In fact, their energy $E = \frac{2\pi k\theta}{g^2}V(\lambda)$ is k times the energy of the minimal energy soliton $k = 1$. This suggests an interpretation as k solitons which will become clearer as we proceed.

It is remarkable that the energy of the soliton is completely insensitive to the value of the scalar potential at any point except $\phi = \lambda$. Thus the mass of the soliton is unchanged if the height of the barrier in $V(\phi)$ (between $\phi = \lambda$ and $\phi = 0$) is taken to infinity while $V(\lambda)$ is kept fixed. This is true even though $\phi_0(r)$, the solitonic field configuration corresponding to $\lambda|0\rangle\langle 0|$, decreases continuously from $\phi = 2\lambda$ at $r = 0$ to $\phi = 0$ at $r = \infty$! It is also striking that the form of the solutions themselves are remarkably universal too, more or less independent of the details of the potential.

4.2 Stability and Moduli Space at $\theta = \infty$

Because of the $U(\infty)$ symmetry it suffices to examine the stability of radial solutions of the form

$$\phi(r^2) = \lambda\sum_{n=0}^{k-1}\phi_n(r^2) \tag{4.23}$$

to small fluctuations. Since any $U(\infty)$ rotation does not change the energy of our solution (4.23) it is sufficient to study the stability to radially symmetric fluctuations. These are most conveniently parameterized as deformations of the eigenvalues. The energy for an arbitrary radially symmetric state $\phi(r^2) = \sum_{n=0}^{\infty} c_n\phi_n(r^2)$ is

$$E = \frac{2\pi\theta}{g^2}\sum_{n=0}^{\infty} V(c_n).$$

The solutions with $c_n \in \{\lambda, 0\}$ are manifestly local minima of E, as λ and 0 are minima of the function $V(x)$. Thus the solution of the form (4.23) (and all solutions unitarily related to it) are stable to small fluctuations. (If any of the c_n took the value of a local maximum of $V(x)$, then it is equally easy to see that while the corresponding $\phi(r^2)$ would be a solution to (4.7) it is not stable to small radial fluctuations.)

The stability of the gaussian soliton $\lambda\phi_0(r^2)$ may qualitatively be understood as follows. Since $\lambda\phi_0(r^2) = 2\lambda e^{-r^2}$ is a Gaussian of height 2λ, far away from the origin, $\phi_0(x) = 0$, but near $x = 0$, it is in the vicinity of the second vacuum. In other words, the static solution corresponds to a bubble of the "false" vacuum. The area of the bubble is of order one (or θ in our original coordinates), the non-commutativity scale. In a commutative theory such a bubble would decay by shrinking to zero size. Noncommutativity prevents the bubble from shrinking to a spatial size smaller than $\sqrt{\theta}$. In order to decay, ϕ_0 actually has to scale to zero - but that process involves going over the hump in the potential and so is classically forbidden.

The $U(\infty)$ symmetry of (4.7) results in there being an infinite number of zero modes for a given solution with energy $2\pi k V(\lambda)$. This infinite dimensional moduli space can be mathematically characterised as follows. The rank k hermitian projection operators on $\mathcal{H}$ (or equivalently, the k-dimensional hyperplanes in $\mathcal{H}$) form a manifold known as the Grassmannian $\mathrm{Gr}(k, \mathcal{H})$, which can also be described as the coset space

$$\frac{\mathrm{U}(\infty)}{\mathrm{U}(k) \times \mathrm{U}(\infty - k)}, \tag{4.24}$$

where $\mathrm{U}(\infty)$ acts on the entire space, while $\mathrm{U}(\infty - k)$ acts only on the orthogonal complement of a k-dimensional hyperplane.

5 Scalar solitons at finite θ

So far, by working at infinite θ, we have found an infinite number of solutions. This is because we have neglected the kinetic energy. As we will now see, the kinetic energy breaks the $U(\infty)$ symmetry that the potential term possessed. Including it in a systematic expansion in powers of $\frac{1}{\theta}$, we will find that most of the solutions no longer remain. However, we will find to leading order in $\frac{1}{\theta}$, that there is an interesting finite dimensional (approximate) moduli space. These will in some limits correspond to separated gaussian solitons. Apparent singularities in this moduli space are resolved in a very stringy way. The discussion in this section is largely based on [8].

The kinetic term can also be written in terms of operators if we use the Weyl-Moyal correspondence for derivatives

$$\partial_z \to -\frac{1}{\sqrt{\theta}}[a^\dagger, \cdot]. \tag{5.1}$$

The energy functional then reads as

$$E = \frac{2\pi}{g^2}\mathrm{Tr}_{\mathcal{H}}\left([a, \widehat{\phi}][\widehat{\phi}, a^\dagger] + \theta V(\widehat{\phi})\right). \tag{5.2}$$

This no longer has the symmetry under $\widehat{\phi} \to U\widehat{\phi}U^\dagger$.

5.1 The expansion in $\frac{1}{\theta}$

If m^2 is a typical mass scale of the theory one can define a perturbation expansion in $1/(\theta m^2)$ for the energy and solutions of the equations of motion of (5.2),

$$\begin{aligned} \widehat{\phi} &= \widehat{\phi}_0 + \frac{1}{\theta m^2}\widehat{\phi}_1 + \cdots , \\ E &= \theta m^2 E_0 + E_1 + \frac{1}{\theta m^2} E_2 + \cdots , \end{aligned} \tag{5.3}$$

where $\widehat{\phi}_0 = \lambda P$ is a solution at infinite θ. The first correction to the energy is just the kinetic term:

$$E_1[\widehat{\phi}_0] = 2\pi \operatorname{Tr}[a, \widehat{\phi}_0][\widehat{\phi}_0, a^\dagger]. \tag{5.4}$$

Due to the fact that $V'(\widehat{\phi}_0) = 0$, E_1 is independent of the correction $\widehat{\phi}_1$.

A reasonable guess would be that a minimum of the kinetic energy is achieved only by rotationally symmetric solutions. However, the story is not so simple. It turns out that there are non-rotationally symmetric minima of E_1 [8]. Indeed, the full moduli space $\mathcal{M}_k$ has an interesting structure, large enough to allow non-trivial dynamics. This unexpected fact is a consequence, not of any symmetry possessed by E_1, but rather of a Bogomolnyi-like bound that it satisfies:

$$E_1[\widehat{\phi}_0] = 2\pi\lambda^2 \operatorname{Tr}[a, P][P, a^\dagger] = 2\pi\lambda^2 \operatorname{Tr}\Big(P + 2F(P)^\dagger F(P)\Big) \geq 2\pi\lambda^2 k \tag{5.5}$$

where

$$F(P) \equiv (1 - P)aP. \tag{5.6}$$

The bound is saturated when $F(P) = 0$, in other words when the image of P is an invariant subspace of the operator a. The projection operators satisfying this condition define a finite dimensional subspace $\mathcal{M}_k$ of the space of all projection operators. The field configurations corresponding to these projection operators will have a natural interpretation in terms of separated solitons.

Starting with the simplest case of $k = 1$, it is clear that any 1-dimensional invariant subspace of a must be spanned by an eigenstate of that operator, i.e. by a coherent state. We'll use an unnormalised version of coherent states defined by $|z\rangle = \exp a^\dagger z|0\rangle$ obeying $a|z\rangle = z|z\rangle$. The corresponding projector $\widehat{\phi}_z = |z\rangle\langle z|$ is a gaussian soliton localised around z when viewed in position space.

For higher k the one can similarly construct invariant subspaces spanned by k different coherent states $|z_i\rangle$. We can think of this as k solitons, each with independent collective coordinate z_i. (Indeed, if the z_i are far from each other, then the respective coherent states are nearly orthogonal and the corresponding field configuration is approximately the sum of distant Gaussian solitons.) And the moduli space is, at least naively, the k-fold symmetric product of the single-soliton moduli space, $\mathrm{Sym}^k(\mathbb{C}) \equiv \mathbb{C}^k/S_k$ (symmetric because permuting the z_i leaves the configuration unchanged; the solitons are like identical particles).

Examining the potential singularities when the solitons come together will give us some intuition about how these solitons see space time. Consider the case $k = 2$ after factoring out the centre of mass degrees of freedom. The description in terms of coherent states $\{|z\rangle, |-z\rangle\}$ becomes bad when $z \to 0$. But this is the fault of our description. A basis which has a smooth limit as $z \to 0$ is $\{|z\rangle, \frac{|z\rangle - |-z\rangle}{z-(-z)}\}$ which approaches $\{|0\rangle, a^\dagger|0\rangle = |1\rangle\}$. Thus there is no singularity when two solitons come together – one ends up in the radially symmetric configuration $|0\rangle\langle 0| + |1\rangle\langle 1|$. One can study the metric on the k soliton moduli space and find that it is Kahler. An explicit form of the Kahler potential can also be given.

The situation becomes more interesting in higher dimensions. For instance, consider noncommutativity in four spatial directions. Then the Moyal-Weyl correspondence maps the fields in four dimensions to operators in the hilbert space of a particle in $2d$. Again the SHO basis spanned by the $2d$ oscillators $a_1^\dagger, a_2^\dagger$ is the useful one to work in.

As before, projection operators (in this $2d$ hilbert space), are solutions to the equations of motion at $\theta = \infty$. Inclusion of the kinetic energy at leading order in $\frac{1}{\theta}$ leads to a lifting of the degernacy. Nevertheless, a bogomolnyi bound similar to (5.5) implies that there is a finite dimensional moduli space parametrised by projection operators satisfying $Pa_rP = a_rP$ $(r = 1, 2)$.

Again, such projectors can be parametrised by the subspaces spanned by $\{|\vec{z}_i\rangle\}$ $(i = 1 \dots k)$. The moduli space is naively $\mathrm{Sym}^k(\mathbb{C}^2)$. What happens at the coincidence locus is very interesting. When two solitons come together,

$$\lim_{\vec{z}\to\vec{0}} \mathrm{span}\{|\vec{z}\rangle, |-\vec{z}\rangle\} = \mathrm{span}\left\{|0,0\rangle, \vec{\gamma}\cdot\vec{a}^\dagger|0,0\rangle\right\}, \qquad \text{where } \vec{\gamma} = \lim_{\vec{z}\to\vec{0}} \frac{\vec{z}}{|\vec{z}|}. \tag{5.7}$$

Thus the "origin" of the relative moduli space is not a single point, but rather a $\mathbb{P}^1$ parametrized by the complex direction $\vec{\gamma}$ along which the two solitons came together. This is in contrast to the $d = 1$ case. This is also exactly how string theory resolves the $\mathbb{C}^2/\mathbb{Z}_2$ singularity. Namely, by introducing an S^2 at the singular point. Here we see that the geometry seen by the noncommutative algebra of projection operators is very different from that seen by functions. Somehow, this is perhaps a hint of how noncommutative structures in string theory will modify our notion of geometry. Before closing this discussion, it should be mentioned that going to higher than 4 spatial dimensions introduces even more weird behaviour. The moduli spaces seen by the solitons are not even smooth – they are spaces known to mathematicians as Hilbert schemes.

The moduli space $\mathcal{M}_k$ is also useful in constructing solitons on quotient spaces. For example, in two dimensions one can write down stable solitons on $\mathbb{R}^2/\mathbb{Z}_k$, the cylinder and torus. One small surprise is that stable noncommutative solitons do not exist when the area of the torus is smaller than $2\pi\theta$. The torus becomes too small for the solitons to fit on.

6 Noncommutative solitons as D-branes

In this section we provide a brief sketch of how noncommutative solitons in the scalar theories show up as D-branes in studies of tachyon condensation [9], [10].

In the bosonic string theory there are D-branes of all dimensions which are however unstable – they have a real tachyon on their world volume. In particular, the space filling $D25$ brane is unstable and reflects the instability of the bosonic open string theory in 26 dimensions. Ashoke Sen has made a series of definite conjectures [11],[12] about the fate of the tachyon. Firstly, the vacuum that the tachyon rolls down to, is expected to contain no open strings. Secondly, the difference in energy per unit volume of this vacuum to the original unstable one is expected to be equal to the tension of the $D-25$ brane. Thirdly, the lower dimensional D-branes are solitonic excitations of the tachyon potential.

To make the connection with the noncommutative scalar field theories we have studied thus far, we consider the effective action for the tachyon field, obtained by integrating out the massive string fields. It is expected to take the form

$$S = \frac{C}{g_s}\int d^{26}x\sqrt{g}\left(\frac{1}{2}f(T)g^{\mu\nu}\partial_\mu T\partial_\nu T - V(T) + \cdots\right). \qquad (6.1)$$

Here $V(T)$ is a general potential having an unstable extremum at $T = T_0$ (the unstable vacuum) and a minimum chosen to be $V(0) = 0$. The constant $C = g_s T_{25}$ is independent of g_s. Sen's conjecture then requires $V(T_0) = 1$. The terms that are omitted are higher derivative terms and terms involving the massless modes.

Let us now turn on a B field in some of the spatial directions of the theory. In the presence of a B-field, as mentioned in section 3, the action is modified to

$$S = \frac{C}{G_s} \int d^{26}x \sqrt{G} \left(\frac{1}{2} f(T) G^{\mu\nu} \partial_\mu T \partial_\nu T - V(T) + \cdots \right)_\star . \qquad (6.2)$$

Here it is understood that there is noncommutativity only in the directions where the B field has been turned on. Now, the advantage of taking the limit of a large B-field is that derivative terms can be neglected. The solitons of this theory are precisely the noncommutative solitons we constructed earlier. According to Sen's conjecture, these should be the D-branes of the bosonic string theory.

The simplest NC soliton solution takes the form of $T = T_0 \phi_0(r^2)$. Where $\phi_0(r^2)$ is the gaussian localised in two of the noncommutative directions. This would be a codimension two object and a candidate $D23$ brane. It's energy is given by $\frac{2\pi\theta C}{G_s} V(T_0) \int d^{24}x \sqrt{G}$. We can now use Sen's conjecture which implies, in our convention, that $V(T_0) = 1$. In terms of this, using the dictionary between open and closed string quantities, it is then possible to verify that the energy density of the above solution is exactly that of the $D23$ brane. It is lucky that the only information needed to obtain the energy of the noncommutative soliton is the value of V at the extremum T_0 which is one piece of the potential which we have some information about from Sen's conjecture. Using noncommutativity in additional spatial directions, it is also possible to obtain branes of all even codimensions as noncommutative solitons, together with the right tension.

Moreover, by considering a projector of rank k one obtains multiple solitons which have the interpretation as multiple D-branes. The fact that their energy is k times that of a single soliton is a reflection of the fact that, in classical open string theory, D-branes exert no force on each other.

The structure of multiple noncommutative solitons now gives a nice realisation of the nonabelian spectrum of fluctuations around coincident D-branes. This essentially follows from the fact that a projector like $P_k = \sum_{i=0}^{k-1} |i\rangle\langle i|$ leaves an unbroken $U(k)$ group. The reader is referred to [10] for details.

The noncommutative solitons also exhibit the instability of the corresponding D-branes. Since the solitons correspond to an extremum of $V(T)$ which is a maximum, one finds for a rank k soliton, a tachyonic mode which transforms in the adjoint of $U(k)$. It's mass can again be compared with that on the $D23$ brane and one finds agreement.

Thus the identification of D-branes in the bosonic, as also in the type II theories, with noncommutative solitons provides a potentially powerful tool to study many properties of D-branes in a easily controlled manner. In the next section we will see another application of this philosophy.

7 Noncommutative solitons in gauge theories

We will now consider noncommutative Yang-Mills theories in which there are unstable solitons which have no counterpart in the commutative theory. The simplest theory to study will again be one in $(2+1)$ dimensions. The soliton in this system will have an interpretation as a $D0$ brane localised on a noncommutative $D2$ brane worldvolume. The process of the $D0$ brane dissolving into the $D2$ brane can be explicitly studied in the NCYM theory. There is a quartic tachyon potential which allows one to follow the condensation of the tachyon.

7.1 Noncommutative Yang-Mills

The noncommutative version of Yang-Mills theory (we will restrict ourselves to the $U(1)$ version for simplicity) can be written as

$$S = -\frac{1}{4g_{YM}^2}\int G^{\mu\rho}G^{\nu\sigma}\widehat{F}_{\mu\nu}\star\widehat{F}_{\rho\sigma}, \tag{7.1}$$

where

$$\widehat{F}_{\mu\nu} = \partial_\mu\widehat{A}_\nu - \partial_\nu\widehat{A}_\mu + i(\widehat{A}_\mu\star\widehat{A}_\nu - \widehat{A}_\nu\star\widehat{A}_\mu) \tag{7.2}$$

is the noncommutative field strength, expressed in terms of $\widehat{A}$, the noncommutative gauge field. This theory has noncommutative gauge transformations

$$\delta\widehat{A}_\mu = \partial_\mu\epsilon + i(\epsilon\star\widehat{A}_\mu - \widehat{A}_\mu\star\epsilon). \tag{7.3}$$

Note the similarity to nonabelian gauge theories and the fact that the $U(1)$ noncommutative theory is not free.

This Lagrangian can be written in terms of operators using the Weyl-Moyal correspondence. In the (2+1) dimensional case, it is useful to define

$$C_z = C = -i\widehat{A}_z + \frac{a^\dagger}{\sqrt{\theta}}, \qquad C_{\bar{z}} = \bar{C}. \tag{7.4}$$

These shifted gauge fields have the nice property that they transform covariantly under gauge transformations $\delta C = i[\epsilon, C]$ (just like covariant derivatives).

The action takes the simple form (fixing temporal gauge $A_0 = 0$)

$$-S = \frac{2\pi\theta}{g_{YM}^2} \int dt \mathrm{Tr}\Bigg[-\partial_t \bar{C} \partial_t C + \left([C, \bar{C}] + \frac{1}{\theta} \right)^2 \Bigg]. \tag{7.5}$$

The field strength component $F_{z\bar{z}}$ is proportional to the combination $([C, \bar{C}] + \frac{1}{\theta})$. We have "shifted away" the spatial derivative terms and obtained a matrix model-like action.

7.2 Flux lumps

The equation of motion for (7.5), for static solutions, is simply

$$[C, [C, \bar{C}]] = 0. \tag{7.6}$$

In the gauge that we are working in, we should also ensure that the Gauss' law constraint is satisfied

$$[\bar{C}, \partial_t C] + [C, \partial_t \bar{C}] = 0 \tag{7.7}$$

obtained by varying (7.5) with respect to A_0 prior to gauge fixing, and then setting $A_0 = 0$. A trivial solution to both is $C = a^\dagger$. This corresponds to the vacuum with zero gauge field. We would like to find nontrivial solutions.

A class of such solutions labelled by a positive integer $m > 0$ were found in [13] and [14].

$$C = C_0 \equiv (S^\dagger)^m a^\dagger S^m. \tag{7.8}$$

where the shift operator $S = \sum_{i=0}^{\infty} |i\rangle\langle i+1|$. S obeys

$$\begin{aligned} &SS^\dagger = 1, \quad S^\dagger S = 1 - P_0 \\ &S^m (S^\dagger)^m = 1, (S^\dagger)^m S^m = 1 - P_{m-1} \end{aligned} \tag{7.9}$$

where $P_{m-1} = \sum_{i=0}^{m-1} |i\rangle\langle i|$ is the rank m projector. One can actually can easily see that there is a generalisation of (7.8) where one adds $\sum_{a=0}^{m-1} c^a |a\rangle\langle a|$ to C_0. The c_a are arbitrary complex numbers. The matrix C_0 then takes the block form

$$C_0 = \begin{pmatrix} c_0 & 0 \\ 0 & a^\dagger \end{pmatrix} \tag{7.10}$$

where c_0 is an $m \times m$ diagonal matrix with eigenvalues c^a.

The flux operator $-iF_{z\bar{z}} = F_0$ evaluated on C_0 is given by

$$\theta F_0 = 1 + \theta[C_0, \bar{C}_0] = 1 + \theta(S^\dagger)^m[a^\dagger, a]S^m = P_{m-1}. \tag{7.11}$$

For an arbitrary configuration, the normalized integral of the flux over the z plane may be rewritten as a trace over the operator F

$$c_1 = \frac{1}{2\pi}\int F = \theta \mathrm{Tr} F; \tag{7.12}$$

from (7.11) and (7.12) C_0 carries m units of flux. Since $S^m|a\rangle = 0 = \langle a|(S^\dagger)^m$ $(a = 0 \ldots m-1)$

$$[C_0, F_0] = 0, \tag{7.13}$$

and C_0 is a static solution to the equation of motion (7.6). Its energy is

$$E = \frac{2\pi\theta}{2g_{YM}^2}\mathrm{Tr}F_0^2 = \frac{m\pi}{g_{YM}^2\theta}. \tag{7.14}$$

(If one has a nontrivial open string metric $G = G_{z\bar{z}}$, then there is an additional factor of G in the denominator.) Thus the solution we have found is finite energy with $m > 0$ units of flux.

The solution can be generalised to the supersymmetric version of the noncommutative gauge theory. The fermions are not excited in the solution. The additional scalars corresponding to transverse fluctuations can take arbitrary diagonal values. See [15] for details (also [16]), [17]).

7.3 Fluctuation spectrum

The spectrum of quadratic fluctuations about the solution C_0 is easy to compute. It is useful to parametrise the fluctuation δC in $C = C_0 + \delta C$ as

$$\delta C = \begin{pmatrix} A & W \\ \bar{T} & D \end{pmatrix}. \tag{7.15}$$

We then expand (7.5) to quadratic order in the fields A, W, T, D and diagonalize this quadratic form. We find that A is a massless field, while the spectrum of D is exactly that of the vector about the vacuum – these can be identified with the bulk modes on the D2-brane.

The interesting case is that of the offdiagonal fluctuations. One set of linear combinations of the of the W and T are pure gauge, and are set to zero by the Gauss Law constraint. The other set form a tower of states.

Each energy level of the tower has m complex fields in the fundamental of $U(m)$. The base of the tower is tachyonic with the m complex modes $\langle a|T|m\rangle$ $(a = 0 \ldots m-1)$ of $m^2 = -\frac{1}{\theta}$. Thus the solution is unstable as one might intuitively have guessed from the fact that flux prefers to be spread out to reduce energy.

The rest of the tower has a harmonic oscillator spectrum, with energies

$$m_k^2 = \frac{(2k+1)}{\theta}, \quad k = 1, 2, \ldots \tag{7.16}$$

All these modes can be identified with the modes of 0-2 strings.

Again, the transverse scalars and fermions in the supersymmetric case can also be taken into account.

7.4 Comparison with string theory

Since the $(2+1)$ dimensional noncommutative Yang-Mills theory describes the worldvolume theory of a $D2$ brane in the presence of a large B field, we expect a flux solution to be a zero brane. We can make precise tests of this hypothesis.

Firstly, one can compare the energy of a localised zero brane on the two brane with (7.14) obtained from the gauge theory. We will determine the difference between the energy of this configuration and one in which the 0-brane is completely dissolved in the 2-brane. For this purpose we work in commutative variables. Let the constant value of F be equal to B after the 0-brane has dissolved into the 2-brane. The energy of this dissolved state is

$$\begin{aligned} E &= \frac{1}{g_{str}(2\pi)^2(\alpha')^{\frac{3}{2}}} \int d^2x \sqrt{\det\left(g + 2\pi\alpha' B\right)} \\ &= \frac{1}{(2\pi)^2 g_{str}(\alpha')^{\frac{3}{2}}} \int d^2x \sqrt{g^2 + (2\pi\alpha' B)^2}. \end{aligned} \tag{7.17}$$

In the limit of large B field (7.17) may be expanded as

$$E = \frac{1}{g_{str}\sqrt{\alpha'}} \frac{1}{2\pi} \int d^2x B \left(1 + \frac{1}{2}\frac{g^2}{(2\pi\alpha' B)^2} + \ldots\right). \tag{7.18}$$

Removing a unit of D0-brane charge from the constant value of the background F field on the brane, $(\frac{1}{2\pi}\int d^2x \Delta F = -1)$ lowers the energy of the 2-brane by

$$\Delta E = \frac{1}{g_{str}\sqrt{\alpha'}} \left(1 - \frac{1}{2}\frac{g^2}{(2\pi\alpha' B)^2}\right). \tag{7.19}$$

Thus

$$E_{bind} = E(D0) - \Delta E = \frac{1}{g_{str}\sqrt{\alpha'}} - \Delta E = \frac{1}{2g_{str}\sqrt{\alpha'}}\frac{g^2}{(2\pi\alpha' B)^2}. \qquad (7.20)$$

On using the dictionary between open and closed string quantities,

$$\begin{aligned} \theta &= \frac{1}{B} \\ G &= \frac{(2\pi\alpha' B)^2}{g} \\ g_{YM}^2 &= \frac{g_{str} 2\pi(\alpha')^{\frac{1}{2}} B}{g}. \end{aligned} \qquad (7.21)$$

one finds (7.20) is precisely the same as (7.14).

In fact, the spectra of fluctuations we found in the previous subsection can be exactly matched to string theory as well. The free 0-2 conformal field theory has a hagedorn spectrum of stringy states. The moding of the $0-2$ oscillators is shifted by

$$\nu = 1 - \frac{1}{\pi b}, \qquad b = \frac{2\pi\alpha' B}{g} \qquad (7.22)$$

in the scaling limit of a large B-field. Most of the oscillator states have masses of order the string scale and thus decouple. There is however a single tower of massive string states generated by the oscillator $\alpha_{-1+\nu}$ whose energy spacing $\frac{1}{\pi\alpha' b} = \frac{1}{2G\theta}$, is the spacing of the states we found in the gauge theory. A careful analysis reveals a tachyon of exactly the expected mass, as well as the massive tower $m_k^2 = \frac{(2k+1)}{2\pi\alpha' b}$ which matches with (7.16).

7.5 Tachyon condensation

The 0-2 system we have studied in this paper also has a world-volume tachyon, and can be regarded as a toy laboratory for the more difficult and interesting $D\,\bar{D}$ system. In the 0-2 context there is a small parameter, namely the ratio of the string scale to the noncommutativity scale, which can be used to control the analysis. A similar logic was used in [18] to study tachyon condensation of a large number of $D0$ branes in a $D2$ brane.

Let us consider the case of a single flux $m = 1$. The initial state $C = (S^\dagger)a^\dagger S$ decays to the final state $C = a^\dagger$ on exciting the tachyonic mode $T = C_{1,0}$. Note that the tachyonic mode and the nonzero matrix elements in the initial and final state are all of the form $C_{i+1,i}$. One might thus suspect that it is possible to set all C matrix elements not of this form to zero through the entire process of tachyon condensation. This is indeed the case[1], as (7.5) admits a consistent truncation to these modes.

We can then easily expand the action out in terms of the fluctuations T and the $2-2$ modes $C_{i+1,i} = D_{i,i-1}$ (in the notation of (7.15)). Since the gauge theory action is quartic in the fields, the potential for T and $D's$ will also be quartic. In fact, it takes the relatively simple form (setting $\theta = 1$ temporarily for convenience: we can reinstate it at the end using dimensional analysis)

$$\begin{aligned} V &= \frac{\pi}{g_{YM}^2}\Bigg[\left[|T|^2 - 1\right]^2 + \left[|T|^2 - |D_{1,0}+1|^2 + 1\right]^2 \\ &\quad + \sum_{i=1}\left[|D_{i,i-1}+\sqrt{i}|^2 - |D_{i+1,i}+\sqrt{i+1}|^2 + 1\right]^2\Bigg]. \end{aligned} \tag{7.23}$$

An unstable extremum of (7.23) that corresponds to an undissolved 0-brane on the 2-brane is

$$T = D_{i,i-1} = 0. \tag{7.24}$$

It decays into the stable extremum

$$|T| = 1, \qquad D_{i,i-1} = \sqrt{i+1} - \sqrt{i}. \tag{7.25}$$

It is easy to see that (7.25) corresponds to $C = a^\dagger$, the 2-brane vacuum.

It is possible to integrate out the fields $D_{i,i-1}$ and obtain the potential V as a function of the tachyon alone. Minimizing V w.r.t $D_{i,i-1}$ we find

$$|D_{i,i-1} + \sqrt{i}|^2 = |T|^2 + i \tag{7.26}$$

which sets all except the first term (7.23) to zero. Restoring θ, the potential thus takes the simple form[2]

$$V = \frac{\pi\theta}{g_{YM}^2}\left[|T|^2 - \frac{1}{\theta}\right]^2. \tag{7.27}$$

[1]In order to demonstrate this we assign the the fields C_{ij} and C^*_{km} 'angular momentum' quantum numbers $i-j$ and $m-k$ respectively. With this assignment the potential conserves angular momentum. All terms in the potential are the product of an equal number of C and a C^* fields. Angular momentum conservation prohibits linear coupling of 'other fields' to C (C^*) fields of angular momentum 1 (-1).

[2]A quartic tachyon potential was also obtained in [18], (see equation 2.8) using scattering calculations in string theory, strengthening our identification of the fluctuation modes of subsection 2.3 with 0-2 strings.

Thus we can accurately study tachyon condensation in this decoupling limit of string theory.

But as a caveat we should note that the 2-2 string modes in the CFT after tachyon condensation include all the 0-0, 0-2, 2-0, and 2-2 strings of the CFT prior to tachyon condensation. Thus, in the process of tachyon condensation, 0-0 and 0-2 modes are absorbed into the 2-2 continuum. In this respect tachyon condensation in the 0-2 system appears qualitatively different from tachyon condensation in a $p\,\bar{p}$ system. In the latter case there appears to be no continuum for the $p-\bar{p}$ modes to disappear into. Restated, the decay of the 0-brane into 'nothing' in the 0-2 system is not mysterious once the 0-brane is constructed as a soliton on the 2-brane.

8 Conclusion

We have tried to give a flavour of the physics that can be captured by the relatively elementary classical solutions of noncommutative field theories. We have seen in different contexts how these solitons are really simple manifestations of D-branes, possessing many of their important features. Though they have been primarily studied in the context of tachyon condensation, we saw that they can also shed some light on the resolution of singularities in spacetime by D-brane probes. In addition to other applications in string theory it is important at this stage to explore their presence in other systems with a strong magnetic field like the quantum hall effect.

Acknowledgments. I must thank S. Minwalla, A. Strominger, M. Aganagic, M. Headrick and M. Spradlin – my collaborators, with whom much of the work presented here was done.

References

[1] A. Connes, M. Douglas and A. Schwarz, "Noncommutative Geometry and Matrix Theory: Compactification on Tori," hep-th/9711162, JHEP **02** (1998) 003.

[2] R. Gopakumar, S. Minwalla and A. Strominger, "Noncommutative solitons," JHEP **05** (2000) 020, [hep-th/0003160].

[3] J. A. Harvey, "Komaba lectures on noncommutative solitons and D-branes," arXiv:hep-th/0102076.

[4] M. R. Douglas and N. A. Nekrasov, "Noncommutative field theory," arXiv:hep-th/0106048.

[5] E. Witten, "Noncommutative Geometry and String Field Theory," Nucl. Phys. **B268** (1986) 253.

[6] N. Seiberg and E. Witten, "String Theory and Noncommutative Geometry," hep-th/9908142, JHEP **09** (1999) 032.

[7] G. Derrick, "Comments on Nonlinear wave equations as models for Elementary Particles," J. Math. Phys **5** 1252 (1964). S. Coleman, "Aspects of Symmetry", Chapters 6 and 7.

[8] R. Gopakumar, M. Headrick and M. Spradlin, "On noncommutative multi-solitons," hep-th/0103256.

[9] K. Dasgupta, S. Mukhi and G. Rajesh, "Noncommutative tachyons," JHEP **0006**, 022 (2000) [hep-th/0005006].

[10] J. A. Harvey, P. Kraus, F. Larsen and E. J. Martinec, "D-branes and strings as non-commutative solitons," hep-th/0005031.

[11] A. Sen, "Descent Relations Among Bosonic D-branes," Int. J. Mod. Phys. **A14** (1999) 4061, [hep-th/9902105].

[12] A. Sen, "Universality of the Tachyon Potential," JHEP **12** (1999) 027, [hep-th/9911116].

[13] A. P. Polychronakos, "Flux tube solutions in noncommutative gauge theories," Phys. Lett. **B495** (2000) 407, [hep-th/0007043].

[14] D. Bak, "Exact multi-vortex solutions in noncommutative Abelian-Higgs theory," hep-th/0008204.

[15] M. Aganagic, R. Gopakumar, S. Minwalla and A. Strominger, "Unstable solitons in noncommutative gauge theory," [hep-th/0009142].

[16] J. A. Harvey, P. Kraus and F. Larsen, JHEP **0012**, 024 (2000) [arXiv:hep-th/0010060].

[17] D. J. Gross and N. A. Nekrasov, JHEP **0103**, 044 (2001) [arXiv:hep-th/0010090].

[18] E. Gava, K. S. Narain and M. H. Sarmadi, "On the bound states of p- and (p+2)-branes," Nucl. Phys. **B504**, 214 (1997) [hep-th/9704006].