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

*mald@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 .............................................. 5

2 The correspondence ........................................ 9
   2.1 Brane probes and multicenter solutions .................. 16
   2.2 The field ↔ operator correspondence .................... 17
   2.3 Holography .................................................. 20

3 Tests of the AdS/CFT correspondence ......................... 22
   3.1 The spectrum of chiral primary operators ............... 24
      3.1.1 The field theory spectrum ............................ 24
      3.1.2 The string theory spectrum and the matching ....... 30
   3.2 Matching of correlation functions and anomalies ........ 32

4 Correlation functions ......................................... 35
   4.1 Two-point functions ......................................... 37
   4.2 Three-point functions ....................................... 40
   4.3 Four-point functions ....................................... 44

5 Wilson loops .................................................. 46
   5.1 Wilson loops and minimum surfaces ....................... 46
   5.2 Other branes ending on the boundary ...................... 51

6 Theories at finite temperature ................................ 53
   6.1 Construction ................................................ 53
   6.2 Thermal phase transition .................................. 55

References ....................................................... 60
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 \( m^2 \sim T J^2 + \text{const.} \). 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 $x^\mu \rightarrow x^\mu / x^2$). 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 \(g_s N\) 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 \(g_s N\) 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

\[
S = S_{\text{bulk}} + S_{\text{brane}} + S_{\text{int}}.
\]

\( S_{\text{bulk}} \) 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 \( S_{\text{brane}} \) 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, \( S_{\text{int}} \) 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

\[
S_{\text{bulk}} \sim \frac{1}{2\kappa^2} \int \sqrt{g} R \sim \int (\partial h)^2 + \kappa(\partial h)^3 h + \cdots,
\]

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 \( S_{\text{int}} \) is proportional to positive powers of \( \kappa \). If we
take the low energy limit, all interaction terms proportional to $\kappa$ 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

$$ds^2 = f^{-1/2}(-dt^2 + dx_1^2 + dx_2^2 + dx_3^2) + f^{1/2}(dr^2 + r^2 d\Omega_5^2),$$

$$F_5 = (1 + *)dtdx_1 dx_2 dx_3 df^{-1},$$

$$f = 1 + \frac{R^4}{r^4}, \quad R^4 = 4\pi g_s \alpha'^2 N.$$  \hspace{1cm} (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

$$E = f^{-1/4} E_p.$$  \hspace{1cm} (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 $\omega$ 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

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

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 descrip-
tion, 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 $N = 4$ 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
$\sqrt{\alpha'} E_p \sim \text{fixed}$. For small $\alpha'$ (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 $r/\alpha'$ fixed.
It is then convenient to define a new variable $U \equiv r/\alpha'$, so that the metric
becomes

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

This can also be seen by considering a D3 brane sitting at $\tilde{r}$. 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 $\tilde{r} = 0$ and the one at $\tilde{r}$,
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 $\phi^I$ in the field theory are conformally coupled, so there is a term of the form $\int d^4x \text{Tr}(\phi^2)\mathcal{R}$ 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 $S^5$,

$$\int_{S^5} F_5 = N. \quad (7)$$

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

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

where we have also included the relationship of the $\theta$ angle to the expectation value of the RR scalar $\chi$. 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 $\alpha'$, 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 $\alpha'$ is a parameter, but not $\alpha'$ and the radius of curvature independently. Thus, $\alpha'$ 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 $ds^2 = R^2 d\hat{s}^2$, and rewriting everything in terms of $\hat{g}$. 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 $N$. $\alpha'$ 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

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

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,

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

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 ($N = 1, 2, \ldots$) 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 $\alpha'$). 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 $\alpha'$ 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

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

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],

$$S = - \frac{1}{(2\pi)^3 g_s \alpha'^2} \int d^4 x f^{-1} \left[ \sqrt{-\det(\eta_{\alpha\beta} + f \partial_{\alpha} r \partial_{\beta} r + r^2 f g_{ij} \partial_{\alpha} \theta^i \partial_{\beta} \theta^j + 2\pi \alpha' \sqrt{f} F_{\alpha\beta})} - 1 \right],$$

$$f = \frac{4\pi g_s \alpha'^2 N}{r^4},$$

(12)

where $\theta^i$ are angular coordinates on the 5-sphere. We can easily check that if we define a new coordinate $U = r / \alpha'$, then all the $\alpha'$ 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 $\partial U$ 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 ↔ 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
ing 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 $AdS$
$\phi(\vec{x}, z)|_{z=0} = \phi_0(\vec{x})$, in
the coordinate system

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

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

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

(13)

where the left-hand side is the generating function of correlation functions
in the field theory, i.e. $\phi_0$ is an arbitrary function and we can calculate
correlation functions of $\mathcal{O}$ by taking functional derivatives with respect to
$\phi_0$ 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 $\phi$ has the value
$\phi_0$ on the boundary of $AdS$. Notice that $\phi_0$ is a function of the four
variables parametrizing the boundary of $AdS_5$.

A formula like (13) is valid in general, for any field $\phi$. 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
$\phi$ 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 $z^\Delta$ for small $z$ (close to the boundary of $AdS$), where

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

(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

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

(15)
and eventually we would take the limit where $\epsilon \rightarrow 0$. Since $\phi$ is dimensionless, we see that $\phi_0$ has dimensions of $[\text{length}]^{\Delta-d}$ which implies, through the left-hand side of (13), that the associated operator $\mathcal{O}$ has dimension $\Delta$ (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 $\Delta$ [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 $\phi_0$. Each differentiation brings down an insertion $\mathcal{O}$, which sends a $\phi$ 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.](image)

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 $S_{\text{max}} = \text{Area}/4G_N$ [31], where the area is that of the boundary of the region. Suppose that we had a state with more entropy than $S_{\text{max}}$, 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

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

(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

\[
d s^2 = R^2 \frac{-dt^2 + dx^2 + dz^2}{z^2},
\]

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 \( \lambda \) in the \( \vec{x} \) directions, it will also have an extent \( \lambda \) in the \( z \) direction. This can be seen from (17) by eliminating \( \lambda \) through the change of variables \( x \to \lambda x \), \( z \to \lambda z \). This implies that a cutoff at

\[ z \sim \delta \]

(18)
corresponds to a UV cutoff in the field theory at distances \( \delta \), with no factors of \( R \) (\( \delta \) 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 \( \delta \). We get

\[ S \sim N^2 \delta^{-3}, \]

(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

\[
\frac{\text{Area}}{4G_N} = \frac{V_{S^3} R^3 \delta^{-3}}{4G_N} \sim N^2 \delta^{-3}.
\]

(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 $\lambda = g_s^2 N$ 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 $\lambda$ 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 $\lambda$ and for large $\lambda$, 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 $\tau$. 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 $\lambda$. 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 $Q^A_\alpha$ (and their complex conjugates $\bar{Q}_{\dot{\alpha}A}$), where $\alpha$ 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 4 representation). They obey the algebra

$$
\{Q^A_\alpha, \bar{Q}_{\dot{\alpha}B}\} = 2(\sigma^\mu)_{\alpha\dot{\alpha}} P_\mu \delta^A_B,
$$

$$
\{Q^A_\alpha, Q^B_\beta\} = \bar{Q}_{\dot{\alpha}A}, \bar{Q}_{\dot{\beta}B}\} = 0,
$$

(21)

where $\sigma^i (i = 1,2,3)$ 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 $A_\mu$ ($\mu$ is a vector index of the $SO(3,1)$ Lorentz group), four complex Weyl fermions $\lambda_{\alpha A}$ (in the 4 of $SU(4)_R$), and six real scalars $\phi^I$ (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

$$
[Q^A_\alpha, \phi^I] \sim \lambda_{\alpha B},
$$

$$
\{Q^A_\alpha, \lambda_{\beta B}\} \sim (\sigma^{\mu\nu})_{\alpha\beta} F_{\mu\nu} + \epsilon_{\alpha\beta}[\phi^I, \phi^J],
$$

$$
\{Q^A_\alpha, \bar{\lambda}_{\dot{\beta} B}\} \sim (\sigma^{\mu})_{\alpha\dot{\beta}} D_\mu \phi^I,
$$

$$
[Q^A_\alpha, A_\mu] \sim (\sigma^\mu)_{\alpha\dot{\beta}} \bar{\lambda}_{\dot{\beta} B},
$$

(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, $D_\mu$ is the covariant derivative, the field strength $F_{\mu\nu} \equiv [D_\mu, D_\nu]$, and we have suppressed the $SU(4)$ Clebsch-Gordan coefficients corresponding to the products $4 \times 6 \rightarrow \bar{4}$, $4 \times 4 \rightarrow 1 + 15$ and $4 \times 4 \rightarrow 6$ 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 $\Phi^i$ in the adjoint representation (in the $3_{2/3}$ 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 $W \propto \epsilon_{ijk} \text{Tr}(\Phi^i \Phi^j \Phi^k)$. The interactions of the theory include a scalar potential
proportional to $\sum_{I,J} \text{Tr}(\phi^I, \phi^J)^2$, such that the moduli space of the theory is the space of commuting matrices $\phi^I$ ($I = 1, \cdots, 6$).

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 $K_\mu$, and acting on it with the operators $Q$ and $P_\mu$. 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 $\psi$ has helicity $\lambda$) 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 $\phi^I, \lambda_A, \lambda^A$ and $F_{\mu\nu} = \partial_{[\mu} A_{\nu]}$). These operators form an ultrashort 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 $O^{I_1 I_2 \cdots I_n} \equiv \text{Tr}(\phi^{I_1} \phi^{I_2} \cdots \phi^{I_n})$. First we can ask if this operator can be written as $\{Q, \psi\}$ for any field $\psi$. In the SUSY algebra (22) only commutators of $\phi^I$'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 $[O^{I_1 I_2 \cdots I_n}] = n$. 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 \(6\)'s is given by \(6 \times 6 \rightarrow 1 + 20'\). The field in the \(1\) representation is \(\text{Tr}(\phi^I \phi^I)\), for which \([Q_A^I, \text{Tr}(\phi^I \phi^I)] \sim C^{AIB} \text{Tr}(\lambda_{aB} \phi^J)\) where \(C^{AIB}\) is a Clebsch-Gordan coefficient for \(4 \times 6 \rightarrow 4\). The right-hand side is in the \(4\) representation, which is the most general representation that can appear in the product \(4 \times 1\), so we find no null vectors at this level. On the other hand, if we look at the symmetric traceless product \(\text{Tr}(\phi^I \phi^I) \equiv \text{Tr}(\phi^I \phi^I) - \frac{1}{6} \delta^{IJ} \text{Tr}(\phi^K \phi^K)\) in the \(20'\) representation, we find that \([Q_A^I, \text{Tr}(\phi^I \phi^I)] \sim \text{Tr}(\lambda_{aB} \phi^K)\) with the right-hand side being in the \(20\) representation (appearing in \(4 \times 6 \rightarrow 4 + 20\), while the left-hand side could in principle be in the \(4 \times 20' \rightarrow 20 + 60\). Since the \(60\) does not appear on the right-hand side (it is a “null vector”) we identify that the representation built on the \(20'\) 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 \(6\)'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 \(O_n\) 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 \(3\)'s), but for a particular choice of the \(\mathcal{N} = 1\) subalgebra not all the operators \(O_n\) 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 \( n = 2, 3, \ldots, N \), for which the indices \( I \) 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 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 \( Q^4 \) 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

\[
\epsilon^{\alpha \beta} \{ Q_{\alpha}, [Q_{\beta}, \mathcal{O}_n] \} \sim \epsilon^{\alpha \beta} \text{Tr}(\lambda_\alpha \lambda_\beta \phi^{J_1} \cdots \phi^{J_n-2}) + \text{Tr}([\phi^{K_1}, \phi^{K_2}] \phi^{L_1} \cdots \phi^{L_n-1}).
\]

Using an \( \mathcal{N} = 1 \) subalgebra some of these operators may be written as the lowest components of the chiral superfields \( \text{Tr}(W^{a_1}_{\alpha} \phi^{J_1} \cdots \phi^{J_n-2}) \). 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

\[
\{ Q_{\{\alpha}, [Q_{\beta}, \mathcal{O}_n] \} \sim \text{Tr}(\sigma_{\mu \nu}^{\alpha \beta} F_{\mu \nu} \phi^{J_1} \cdots \phi^{J_n-1}) + \text{Tr}(\lambda_\alpha \lambda_\beta \phi^{K_1} \cdots \phi^{K_n-2}).
\]

\[(23)\]
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

$$\{Q_\alpha, \{\bar{Q}_\dot{\alpha}, \mathcal{O}_n\}\} \sim \text{Tr}(\lambda_\alpha \lambda_\dot{\alpha} B \phi^J_1 \cdots \phi^J_{n-2}) + (\sigma^\alpha)_{\dot{\alpha} \dot{\beta}} \text{Tr}((D_\mu \phi^J) \phi^{K_1} \cdots \phi^{K_{n-1}})$$

(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 $\text{Tr}(F_{\mu\nu}^4 \phi^{I_1} \cdots \phi^{I_{n-2}}) + \cdots$.

- A real scalar field in the $(2, n-4, 2)$ representation, given by $Q^2 \bar{Q}^2 \mathcal{O}_n$, of the form $\epsilon_{\alpha\beta} \epsilon^{\dot{\alpha}\dot{\beta}} \text{Tr}(\lambda_\alpha \lambda_\dot{\alpha} B \phi^J_1 \cdots \phi^J_{n-4}) + \cdots$.

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

- 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 $\text{Tr}(D_{\mu} \phi^J D_{\nu} \phi^K \phi^{I_1} \cdots \phi^{I_{n-2}}) + \cdots$.

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 $Q^4 \mathcal{O}_n$, of the form $\text{Tr}(F_{\mu\nu} \phi^{I_1} \cdots \phi^{I_{n-4}}) + \cdots$. Note that fields with more than four $F_{\mu\nu}$'s or more than eight $\lambda$'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 $\frac{1}{8\gamma^2} (\text{of the form } \text{Tr}(F_{\mu\nu}^2) + \cdots)$ and whose imaginary part is the Lagrangian density coupling to $\theta$ (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 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 $\Delta$ 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 $\tau$ 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 $Y^k(y) \sim y^I_1 y^I_2 \cdots y^I_6$ where the $y^I_I$ for $I = 1, 2, \cdots, 6$ and with $\sum_{I=1}^{6}(y^I_I)^2 = 1$, 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 $m_R^2 = k(k+4)/R^2$. 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 $SU(\mathcal{N})$ gauge theory and not a $U(\mathcal{N})$ 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 $U(\mathcal{N})$ 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 $n = 2, 3, \ldots, N$. 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 $\text{Tr}(\phi^I \phi^I)$, 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
representation while the fermions of the opposite chirality $\bar{\chi}_3^4$ are in the 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,

$$\langle J_\mu^a(x) J_\nu^b(y) J_\rho^c(z) \rangle = -\frac{N^2 - 1}{32\pi^3} i d^{abc} \frac{\text{Tr} [\gamma_5 \gamma_\mu (\vec{k} - \vec{\beta}) \gamma_\nu (\vec{\beta} - \vec{k}) \gamma_\rho (\vec{k} - \vec{\beta})]}{(x - y)^4 (y - z)^4 (z - x)^4},$$

(26)

where $d^{abc} = 2 \text{Tr}(T^a \{T^b, T^c\})$ 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 $F_{\mu \nu}^a$,

$$(D^\mu J_\mu)^a = \frac{N^2 - 1}{384 \pi^3} i d^{abc} e^{\mu \nu \rho \sigma} F_{\mu \nu}^b F_{\rho \sigma}^c.$$

(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 $A_\mu^a$ which couple, as explained above, to the $SU(4)_R$ global currents $J_\mu^a$ 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

$$\frac{i N^2}{96 \pi^3} \int_{AdS_5} d^5 x (d^{abc} e^{\mu \nu \lambda \rho \sigma} A_\mu^a \partial_\nu A_\lambda^b \partial_\rho A_\sigma^c + \cdots).$$

(28)

This term is gauge invariant up to total derivatives, which means that if we take a gauge transformation $A_\mu^a \rightarrow A_\mu^a + (D_\mu \Lambda)^a$ for which $\Lambda$ does not vanish on the boundary of $AdS_5$, the action will change by a boundary term of the form

$$-\frac{i N^2}{384 \pi^3} \int_{\partial AdS_5} d^4 x e^{\mu \nu \rho \sigma} d^{abc} \Lambda^a F_{\mu \nu}^b F_{\rho \sigma}^c.$$

(29)

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

**Large N Field Theories and Gravity**
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

$$\langle g^{\mu\nu} T_{\mu\nu} \rangle = -a E_4 - c I_4,$$

where

$$E_4 = \frac{1}{16\pi^2} (\mathcal{R}_{\mu\nu,\rho\sigma}^2 - 4 \mathcal{R}_{\mu\nu}^2 + \mathcal{R}^2),$$

$$I_4 = \frac{1}{16\pi^2} (\mathcal{R}_{\mu\nu,\rho\sigma}^2 - 2 \mathcal{R}_{\mu\nu}^2 + \frac{1}{3} \mathcal{R}^2),$$

where $\mathcal{R}_{\mu\nu,\rho\sigma}$ is the curvature tensor, $\mathcal{R}_{\mu\nu} \equiv \mathcal{R}_{\mu\nu}^\rho$ 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 $e^{-I_{\text{SUGRA}}}$, where $I_{\text{SUGRA}}$ 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 $\alpha'$ corrections that cure the divergences of supergravity, and also all the loop corrections, which are controlled essentially by the gravitational coupling $\kappa \sim g_s \alpha'^2$. 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

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

(32)

where $W$ is the generating functional for connected Green’s functions in the gauge theory. At finite temperature, $W = \beta F$ where $\beta$ 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 $I_{\text{SUGRA}} \simeq \beta F$. 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 $\phi_0$ 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

$$W_{\text{gauge}}[\phi_0] = -\log \left< e^{\int d^4 x \phi(x) \mathcal{O}(x)} \right>_{CFT} \simeq \text{extremum } I_{SUGRA}[\phi] .$$

(33)

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

Note that in (33) we have not attempted to be prescient about inserting factors of $\epsilon$. 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 $\phi$ 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 $\phi_0$. 
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 $\phi_0$. 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

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

(34)

where $\eta$ 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

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

(35)

provides manifest Euclidean symmetry on the directions parametrized by $\bar{x}$. 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 $\phi = e^{i\bar{p} \cdot \bar{x}} Z(pz)$, where the function $Z(u)$ satisfies the radial equation

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

$^6$The results may be analytically continued to give the correlation functions of the field theory on Minkowskian $\mathbb{R}^{1,4}$, 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 $I_\nu$ and $K_\nu$ are Bessel functions and
\[
\Delta = 2 + \sqrt{4 + m^2 R^2}.
\] (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 $\phi(\vec{x}, z) = \phi_0(\vec{x}) = e^{i p \cdot \vec{x}}$ at $z = \epsilon$, we find the bulk-to-boundary propagator
\[
\phi(\vec{x}, z) = K_{\mu}(\vec{x}, z) = \frac{(p \epsilon)^2 K_{\Delta-2}(p \epsilon)}{(p \epsilon)^2 K_{\Delta-2}(p \epsilon)} e^{i p \cdot \vec{x}}.
\] (38)

To compute a two-point function of the operator $\mathcal{O}$ for which $\phi_0$ is a source, we write
\[
\langle \mathcal{O}(\vec{y}) \mathcal{O}(\vec{x}) \rangle = \left. \frac{\partial^2 W}{\partial \lambda_1 \partial \lambda_2} \left[ \phi_0 = \lambda_1 e^{i \theta \cdot \vec{x}} + \lambda_2 e^{i \varphi \cdot \vec{x}} \right] \right|_{\lambda_1 = \lambda_2 = 0}
\]
\[
= (\text{leading analytic terms in } (\epsilon p)^2)
\]
\[
- \eta e^{2 \Delta - 8} (2 \Delta - 4) \frac{\Gamma(3 - \Delta)}{\Gamma(\Delta - 1)} \delta^4(p + q) \left( \frac{p}{2} \right)^{2\Delta - 4}
\] (39)
\[
+ (\text{higher order terms in } (\epsilon p)^2),
\]
\[
\langle \mathcal{O}(\vec{x}) \mathcal{O}(\vec{y}) \rangle = \eta e^{2 \Delta - 8} \frac{2 \Delta - 4}{\Delta} \frac{\Gamma(\Delta + 1)}{\pi^2 \Gamma(\Delta - 2)} \frac{1}{|\vec{x} - \vec{y}|^{2\Delta}}.
\]

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 $\epsilon$. 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 $\Delta$. 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 $\Delta$ by analytic continuation and for positive integer $\Delta$ 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 $\int d^4x \phi(\vec{x}, z = \epsilon) O(\vec{x})$ 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 $\epsilon^\Delta$ into the definition of $O$. 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 $AdS_3$), 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 $\Delta \sim (g^{2}_{YM}N)^{1/4}$. 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 $g^{2}_{YM}N$ 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 $AdS_3$) 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 representa-
tion 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 cor-
respondence, the momentum representation seems uniformly more conve-
nient). The boundary behavior of position space bulk-to-boundary propaga-
tors is specified in a slightly more subtle way: following [109] we require

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

(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 $K_\Delta$ solving the equation of motion
and satisfying (40) is

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

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 \( \epsilon \), 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 \( \epsilon \), 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 \( J_\mu \) 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 \( \phi_1, \phi_2, \) and \( \phi_3 \), of the form

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

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

$$\langle \mathcal{O}_1(\vec{x}_1)\mathcal{O}_2(\vec{x}_2)\mathcal{O}_3(\vec{x}_3) \rangle = -\lambda \int d^5x \sqrt{g} K_{\Delta_1}(x; \vec{x}_1)K_{\Delta_2}(x; \vec{x}_2)K_{\Delta_3}(x; \vec{x}_3)$$

$$\times \frac{\lambda a_1}{\left| x_1 - x_2 \right|^{\Delta_1 + \Delta_2 - \Delta_3} \left| x_1 - \vec{x}_3 \right|^{\Delta_1 + \Delta_3 - \Delta_2} \left| x_2 - \vec{x}_3 \right|^{\Delta_2 + \Delta_3 - \Delta_1}}, \quad (43)$$

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 $K_i$, and the cubic vertex to a factor of $\lambda$. The position of the vertex is integrated over $AdS_5$.

The result [109] is

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

(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 $\text{Tr} \phi_1^{(j_1)} \phi_2^{(j_2)} \ldots \phi_{j_l}$ 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

$$O^I = C_{J_1 \ldots J_l}^I \text{Tr} \phi^{(J_1 \ldots \phi^{(l)}}$$

(45)

follow from an action of the form

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

(46)

Derivative couplings of the form $\bar{\phi} \partial \phi \bar{\phi}$ 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 \( \Box, \Box, \Box, \ldots \). $C_{J_1 \ldots J_l}^I$ is a basis transformation matrix, chosen so that $C_{J_1 \ldots J_l} C_{J_1 \ldots J_l}^J = \delta^{J}$. 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 $\int d^5x \tilde{s}_I O^I$, and the normalization ambiguity is represented by the “leg factors” $w^I$. It is the combination $\tilde{s} = w^I \tilde{s}^I$ rather than $\tilde{s}$ 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 $G_{I_1 I_2 I_3}$. To get rid of factors of $w^I$, we introduce operators $O^I = \tilde{w}^I O^I$. One can choose $\tilde{w}^I$ so that a two-point function computation along the lines of section 4.1 leads to

$$\langle O^{I_1}(\bar{x}) O^{I_2}(0) \rangle = \frac{\tilde{s}_{I_1} \tilde{s}_{I_2}}{\bar{x}^{2\Delta_1}}.$$  

(47)

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

$$\langle O^{I_1}(\bar{x}_1) O^{I_2}(\bar{x}_2) O^{I_3}(\bar{x}_3) \rangle = \frac{1}{N |\bar{x}_1 - \bar{x}_2|^{\Delta_1 + \Delta_2 - \Delta_3} |\bar{x}_1 - \bar{x}_3|^{\Delta_1 + \Delta_3 - \Delta_2} |\bar{x}_2 - \bar{x}_3|^{\Delta_2 + \Delta_3 - \Delta_1}} \left\langle C_{I_1}^I C_{I_2}^I C_{I_3}^I \right\rangle,$$  

(48)

where we have defined

$$\left\langle C_{I_1}^I C_{I_2}^I C_{I_3}^I \right\rangle = C_{J_1 \ldots J_l K_1 \ldots K_j}^I C_{J_1 \ldots J_l L_1 \ldots L_k}^I C_{K_1 \ldots K_j L_1 \ldots L_k}^I.$$  

(49)
Remarkably, (48) is the same result one obtains from free field theory by Wick contracting all the $\phi^J$ 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

$$
D_{\Delta_1 \Delta_2 \Delta_3 \Delta_4}(\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4) = \int d^4\vec{x} \sqrt{g} \prod_{i=1}^{4} \tilde{K}_{\Delta_i}(\vec{x}, z; \vec{x}_i),
$$

$$
\tilde{K}_{\Delta}(\vec{x}, z; \vec{y}) = \left(\frac{z}{z^2 + (\vec{x} - \vec{y})^2}\right)^\Delta.
$$

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

$$
s = \frac{1}{2} \frac{\vec{x}_1^2 \vec{x}_2^2 \vec{x}_3^2 \vec{x}_4^2}{\vec{x}_1^2 \vec{x}_2^2 \vec{x}_3^2 + \vec{x}_4^2 \vec{x}_3^2}, \quad t = \frac{\vec{x}_1^2 \vec{x}_2^2 \vec{x}_3^2 \vec{x}_4^2}{\vec{x}_1^2 \vec{x}_2^2 \vec{x}_3^2 + \vec{x}_4^2 \vec{x}_3^2}.
$$

One can write the connected four-point function as

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

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

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

at least as an asymptotic series, and hopefully even with a finite radius of convergence for $\bar{x}_{13}$ and $\bar{x}_{24}$. 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 descendant 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 $\Delta_n$ and $\Delta_m$ shows up to leading order as a term proportional to a logarithm of some combination of the separations $\bar{x}_{ij}$. 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
\[
\frac{1}{(x_{12})^n} \log \left( \frac{x_{13}}{x_{12}} \right).
\]
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
\[
W(C) = \text{Tr} \left[ P \exp \left( i \oint_C A \right) \right]
\]
(54)
depends on a loop \( C \) 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 \( \langle W(C) \rangle \) 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 \( e^{-TV} \) where \( E \) is the lowest possible energy of the quark-anti-quark configuration. Thus, we have
\[
\langle W \rangle \sim e^{-TV(L)} ,
\]
(55)
where \( V(L) \) is the quark anti-quark potential. For large \( N \) and large \( g_Y^2 \), 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 $N$ 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

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

where $x^\mu(\tau)$ is any parametrization of the loop and $\theta^I$ ($I = 1, \cdots, 6$) 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 "$e^{i\theta}$" in front of $\theta^I$.

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 $\sqrt{x^2}$. 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 $C$ 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 $r|C|$, where $|C|$ 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 $\theta^i$ 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 $r|C|$, leaving the resulting action finite.

As an example let us consider a circular Wilson loop. Take $C$ 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 $AdS_5$,

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

(57)

where $\bar{e}_1$, $\bar{e}_2$ 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 get a contribution to the action

$$S \sim \frac{1}{2\pi \alpha'} A = \frac{R^2}{2\pi \alpha'} \int d\theta \int_\epsilon^a \frac{dz a}{z^2} = \frac{R^2}{\alpha'} \left( \frac{a}{\epsilon} - 1 \right),$$

(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

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

(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 $g_s \rightarrow 1/g_s$.

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

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

(60)

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

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

(61)
We need to solve the Euler-Lagrange equations for this action. Since the action does not depend on \( x \) explicitly the solution satisfies
\[
\frac{1}{z^2 \sqrt{(\partial_z z)^2 + 1}} = \text{constant.} \tag{62}
\]
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
\[
\frac{L}{2} = z_0 \int_{0}^{1} \frac{dy y^2}{\sqrt{1 - y^4}} = z_0 \sqrt{2 \pi^{3/2}} \Gamma(1/4)^2. \tag{64}
\]
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),
\[
\frac{L}{2} - x \sim z^3, \quad z \to 0. \tag{65}
\]
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
\[
E = V(L) = \frac{4 \pi^2 (2g_{YM}^2 N)^{1/2}}{\Gamma(1/4)^4 L}. \tag{66}
\]
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.
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 $m$. 

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).
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

\[ S = m \int ds = -2mR \int_\epsilon^0 \frac{dz}{z \sqrt{a^2 - z^2}} \]  \hspace{1cm} (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 \( \epsilon \rightarrow 0 \), 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

\[ e^{-S} \sim e^{-2mR \log a} \sim \frac{1}{a^{2mR}}, \]  \hspace{1cm} (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

\[
ds^2 = R^2 \left[ u^2 (-h dt^2 + dx_1^2 + dx_2^2 + dx_3^2) + \frac{du^2}{hu^2} + d\Omega_5^2 \right]
\]

\[ h = 1 - \frac{u_0^4}{u^4}, \quad u_0 = \pi T. \]  

(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

\[ F_{SUGRA} = -\frac{\pi^2}{8} N^2 VT^4, \]
\[ F_{SYM} = \frac{4}{3} F_{SUGRA}. \]

The supergravity result is at leading order in \( \ell_s/R \), 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

\[ F = -f(g_{YM}^2 N) \frac{\pi^2}{6} N^2 VT^4, \]

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 \( g_{YM}^2 \text{Tr}[\phi^4] \) 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

\[
\begin{align*}
  f(g_{YM}^2 N) &= 1 - \frac{3}{2\pi^2} g_{YM}^2 N + \ldots & \text{for small } g_{YM}^2 N, \\
  f(g_{YM}^2 N) &= \frac{3}{4} + \frac{45}{32} \zeta(3) (g_{YM}^2 N)^{3/2} + \ldots & \text{for large } g_{YM}^2 N.
\end{align*}
\]

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 \( \alpha' \) 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 \( \alpha' \) corrections no longer has the Einstein-Hilbert form, and correspondingly the Bekenstein-Hawking prescription no longer agrees with the free energy computed as \( \beta I \) 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 $e^{-I_{\text{SUGRA}}}$ 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 $I_{\text{SUGRA}}$ is the one that dominates the path integral.

When there are two or more solutions competing to minimize $I_{\text{SUGRA}}$, 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

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

Explicitly, the metric is

$$ds^2 = f dt^2 + \frac{1}{f} dr^2 + r^2 d\Omega_3^2,$$

$$f = 1 + \frac{r^2}{R^2} - \frac{\mu}{r^2}. \quad (74)$$

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

$$\beta = \frac{2\pi R^2 r_+}{2r_+^2 + R^2}. \quad (75)$$

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 $\mathcal{X}_2$. 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, $\text{Tr}(-1)^F e^{-\beta H}$, 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, $\text{Tr}e^{-\beta H}$, 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

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

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 $F_g(T)$ induced by the presence of the external charge $q$, $\langle W(C) \rangle \sim \exp (-F_g(T)/T)$. In a confining phase $F_g(T)$ is infinite and therefore $\langle W(C) \rangle = 0$. In the deconfined phase $F_g(T)$ 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 $\psi$ with period $2\pi$, which contributes $i\psi$ to the string worldsheet action. The string theory partition function includes now an integral over all values of $\psi$, making $\langle W(C) \rangle = 0$ on $S^3$. In contrast, on $\mathbb{R}^3$ one integrates over the local fluctuations of $\psi$ 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


2 (1978) 421.

236.

[88] M. Flato and C. Fronsdal, “Quantum field theory of singletons: the

Particles, Conformal Group and de Sitter Universe,” Phys. Rev. D23

[90] H. Nicolai and E. Sezgin, “Singleton representations of $OSp(N,4)$,”

[91] M. Gunaydin and N. Marcus, “The unitary supermultiplet of $N = 8$
conformal superalgebra involving fields of spin $\leq 2$,” Class. Quant.

(1986) 412.

field theory on AdS$_5$, IIB three-branes and duality,” Class. Quant.

[94] S. Ferrara and C. Fronsdal, “Gauge fields and singletons of AdS$_{2p+1}$,”

[95] S. Ferrara and C. Fronsdal, “Gauge fields as composite boundary exci-

[96] V. G. Kac, “Representations of classical Lie superalgebras,”. Proceed-
ings, Differential Geometrical Methods In Mathematical Physics.II.,

[97] V. K. Dobrev and V. B. Petkova, “Group theoretical approach to ex-
tended conformal supersymmetry: function space realizations and in-

[98] V. K. Dobrev and V. B. Petkova, “All positive energy unitary irre-
ducible representations of extended conformal supersymmetry,” Phys.


